(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0019843
(43) 공개일자 2023년02월09일

(51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) *C12N 15/11* (2006.01)
C12N 15/113 (2010.01) *C12N 9/12* (2006.01)
C12N 9/22 (2006.01)
(52) CPC특허분류
C12N 15/102 (2013.01)
C12N 15/111 (2013.01)
(21) 출원번호 10-2022-7042608
(22) 출원일자(국제) 2021년05월07일
심사청구일자 없음
(85) 번역문제출일자 2022년12월05일
(86) 국제출원번호 PCT/US2021/031439
(87) 국제공개번호 WO 2021/226558
국제공개일자 2021년11월11일
(30) 우선권주장
63/022,397 2020년05월08일 미국(US)
63/116,785 2020년11월20일 미국(US)

(71) 출원인
더 브로드 인스티튜트, 인코퍼레이티드
미국 02142 매사 추세츠주 캠브리지 메인 스트리트 415
프레지던트 앤드 펠로우즈 오브 하바드 칼리지
미합중국, 메사추세츠 02138, 캠브리지, 퀸시스트리트17
(72) 발명자
류, 데이빗, 알.
미국 02142 매사추세츠 캠브리지 메인 스트리트 415 더 브로드 인스티튜트, 인코퍼레이티드 내
안잘론, 앤드루, 비토
미국 02142 매사추세츠 캠브리지 메인 스트리트 415 더 브로드 인스티튜트, 인코퍼레이티드 내
(뒷면에 계속)
(74) 대리인
양영준, 이상남

전체 청구항 수 : 총 232 항

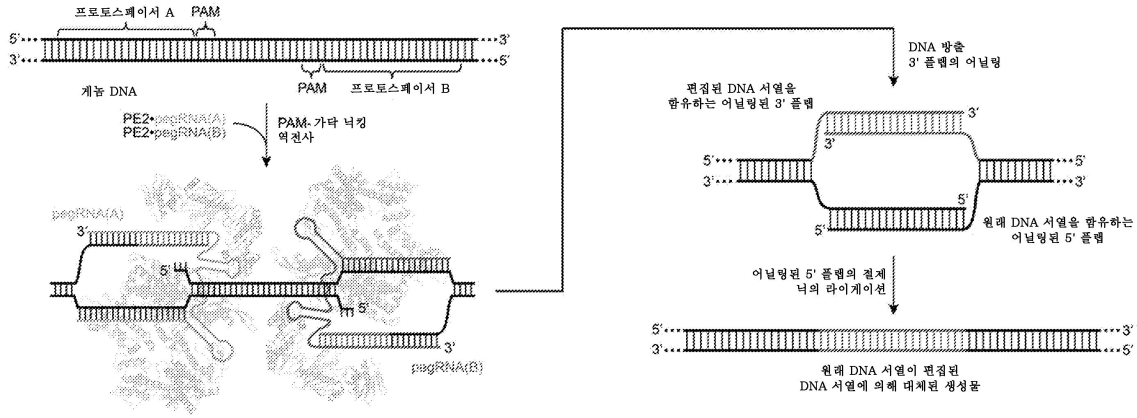
(54) 발명의 명칭 **표적 이중 가닥 뉴클레오티드 서열의 두 가닥의 동시 편집을 위한 방법 및 조성물**

(57) 요 약

본 개시내용은 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템, 조성물 및 방법을 제공한다. 일부 측면에서, 시스템은 제1 및 제2 프라임 편집제 복합체를 포함하고, 여기서 각각의 제1 및 제2 프라임 편집제 복합체는 (1) (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 (ii) RNA-의존성

(뒷면에 계속)

대 표 도



DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 프라임 편집제; 및 (2) 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함하는 pegRNA를 포함하고, 여기서 DNA 합성 주형은 목적하는 DNA 서열 또는 그의 상보체를 코딩하고, 여기서 목적하는 DNA 서열 및 그의 상보체는 편집된 부분을 포함하는 듀플렉스를 형성하고, 이는 편집될 표적 부위 내로 통합된다. 일부 측면에서, 시스템은 각각 프라임 편집제 및 PEgRNA를 포함하는 제1, 제2, 제3 및 제4 프라임 편집제 복합체를 포함한다. 또한, 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하는 방법이 본원에 제공된다. 추가로, 제약 조성물, 폴리뉴클레오티드, 벡터, 세포 및 키트가 본원에 제공된다.

(52) CPC특허분류
C12N 15/113 (2013.01)
C12N 9/1276 (2013.01)
C12N 9/22 (2013.01)
C12Y 207/07049 (2013.01)
C12Y 301/00 (2013.01)
C12N 2310/20 (2017.05)
C12N 2310/3519 (2013.01)

(72) 발명자
레비, 조너선, 마
미국 02142 매사추세츠 캠브리지 메인 스트리트 415 더 브로드 인스티튜트, 인코퍼레이티드 내
가오, 신
미국 02142 매사추세츠 캠브리지 메인 스트리트 415 더 브로드 인스티튜트, 인코퍼레이티드 내
포드래키, 크리스토퍼, 제이.
미국 02142 매사추세츠 캠브리지 메인 스트리트 415 더 브로드 인스티튜트, 인코퍼레이티드 내

명 세 서

청구범위

청구항 1

편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템으로서, 여기서 시스템은 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체를 포함하고, 여기서 각각의 제1 및 제2 프라임 편집제 복합체는 (1) (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 (ii) RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 프라임 편집제; 및 (2) 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함하는 pegRNA를 포함하고, 여기서 제1 프라임 편집제 복합체의 pegRNA의 DNA 합성 주형은 제1 단일 가닥 DNA 서열을 코딩하고, 제2 프라임 편집제 복합체의 pegRNA의 DNA 합성 주형은 제2 단일 가닥 DNA 서열을 코딩하고, 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고, 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 편집될 표적 부위에서의 DNA 서열과 비교하여 편집된 부분을 포함하는 듀플렉스를 형성하고, 이는 편집될 표적 부위 내로 통합되는 것인 시스템.

청구항 2

제1항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 시스템.

청구항 3

제1항 또는 제2항에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 시스템.

청구항 4

제1항 또는 제2항에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 상이한 프라임 편집제를 포함하는 것인 시스템.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, napDNAbp가 Cas9 도메인 또는 그의 변이체인 시스템.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, Cas9 닉카제 도메인 또는 그의 변이체인 시스템.

청구항 7

제1항 내지 제4항 중 어느 한 항에 있어서, napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, CasX, CasY 및 아르고노트로 이루어진 군으로부터 선택된 것인 시스템.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, napDNAbp가 닉카제 활성을 갖는 것인 시스템.

청구항 9

제1항 내지 제4항 중 어느 한 항에 있어서, napDNAbp가 서열식별번호: 18-88 중 어느 하나의 아미노산 서열 또는 서열식별번호: 18-88 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드가 리버스 트랜스크립타제인 시스템.

청구항 11

제1항 내지 제10항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드가 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159 및 700-766 중 어느 하나의 아미노산 서열 또는 서열식별번호: 89-100, 105-122, 128-129 및 132 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서, 프라임 편집제가 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 시스템.

청구항 13

제12항에 있어서, 링커가 서열식별번호: 166-177 중 어느 하나의 아미노산 서열 또는 서열식별번호: 166-177 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 14

제12항 또는 제13항에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 시스템.

청구항 15

제1항 내지 제14항 중 어느 한 항에 있어서, pegRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 시스템.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, 제1 프라임 편집제 복합체의 pegRNA의 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 시스템.

청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, 제2 프라임 편집제 복합체의 pegRNA의 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 시스템.

청구항 18

제17항에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 pegRNA의 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

청구항 19

제18항에 있어서, 제1 및 제2 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 시스템.

청구항 20

제19항에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍,

적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 21

제19항에 있어서, 인접 영역이 편집된 부분을 포함하는 듀플렉스에 의해 대체되는 것인 시스템.

청구항 22

제1항 내지 제21항 중 어느 한 항에 있어서, DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적

어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 23

제1항 내지 제22항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 시스템.

청구항 24

제1항 내지 제22항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 시스템.

청구항 25

제1항 내지 제24항 중 어느 한 항에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 시스템.

청구항 26

제25항에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이이고, 목적하는 DNA 서열 및 그의 상보체의 길이 이하인 시스템.

청구항 27

제25항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역이 편집된 부분을 포함하는 듀플렉스를 형성하는 것인 시스템.

청구항 28

제1항 내지 제27항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 시스템.

청구항 29

제21항에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 시스템.

청구항 30

편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템으로서, 상기 시스템은 하기를 포함하고:

a) 하기를 포함하는 제1 프라임 편집제 복합체:

i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

ii. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

b) 하기를 포함하는 제2 프라임 편집제 복합체:

iii. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

iv. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

여기서 제1 PEgRNA는 제1 단일 가닥 DNA 서열을 코딩하는 제1 DNA 합성 주형을 포함하고, 제2 PEgRNA는 제2 단일 가닥 DNA 서열을 코딩하는 제2 DNA 합성 주형을 포함하고, 여기서 이중 가닥 DNA 서열의 제1 가닥 및 제2 가닥은 서로 상보적인 것인 시스템.

청구항 31

제30항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 시스템.

청구항 32

제30항에 있어서, 제1 및 제2 단일 가닥 DNA 서열이 각각 다른 것에 대한 상보성 영역을 포함하는 것인 시스템.

청구항 33

제30항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열이 편집된 부분을 포함하는 듀플렉스를 형성하고, 이는 편집될 표적 부위 내로 통합되는 는 것인 시스템.

청구항 34

제30항에 있어서, 제1 단일 가닥 DNA 서열 및/또는 제2 단일 가닥 서열이 표적 부위에서의 DNA 서열과 비교하여 서열 상동성을 포함하지 않는 것인 시스템.

청구항 35

제30항에 있어서, 제1 PEgRNA가 제1 스페이서 서열, 제1 gRNA 코어 및 제1 프라이머 결합 부위를 추가로 포함하고, 제2 PEgRNA가 제2 스페이서 서열, 제2 gRNA 코어 및 제2 프라이머 결합 부위를 추가로 포함하는 것인 시스템.

청구항 36

제30항 내지 제35항 중 어느 한 항에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 시스템.

청구항 37

제30항 내지 제36항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드가 동일한 것인 시스템.

청구항 38

제30항 내지 제37항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9 도메인 또는 그의 변이체인 시스템.

청구항 39

제30항 내지 제38항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, Cas9 닉카제 도메인 또는 그의 변이체인 시스템.

청구항 40

제30항 내지 제39항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, CasX, CasY 및 아르고노트로 이루어진 군으로부터 선택된 것인 시스템.

청구항 41

제30항 내지 제40항 중 어느 한 항에 있어서, napDNAbp가 닉카제 활성을 갖는 것인 시스템.

청구항 42

제30항 내지 제41항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 서열식별번호: 18-88 중 어느 하나의 아미노산 서열 또는 서열식별번호: 18-88 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 43

제30항 내지 제42항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 리버스 트랜스크립타제인 시스템.

청구항 44

제30항 내지 제43항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159 및 700-766 중 어느 하나의 아미노산 서열 또는 서열식별번호: 89-100, 105-122, 128-129 및 132 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 45

제30항 내지 제44항 중 어느 한 항에 있어서, 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나; 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 시스템.

청구항 46

제45항에 있어서, 링커가 서열식별번호: 166-177 중 어느 하나의 아미노산 서열 또는 서열식별번호: 166-177 중

어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 47

제45항에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 시스템.

청구항 48

제30항 내지 제47항 중 어느 한 항에 있어서, 제1 및/또는 제2 pegRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 시스템.

청구항 49

제30항 내지 제48항 중 어느 한 항에 있어서, 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 시스템.

청구항 50

제35항에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 시스템.

청구항 51

제35항에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 제1 및/또는 제2 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

청구항 52

제30항에 있어서, 제1 및 제2 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 시스템.

청구항 53

제52항에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어

도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 54

제52항에 있어서, 인접 영역이 편집된 부분을 포함하는 듀플렉스에 의해 대체되는 것인 시스템.

청구항 55

제30항 내지 제54항 중 어느 한 항에 있어서, 제1 및/또는 제2 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 56

제30항 내지 제55항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 시스템.

청구항 57

제30항 내지 제56항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 시스템.

청구항 58

제30항 내지 제57항 중 어느 한 항에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 시스템.

청구항 59

제30항 내지 제58항 중 어느 한 항에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 60

제30항 내지 제59항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

청구항 61

제30항 내지 제60항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 시스템.

청구항 62

제54항에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 시스템.

청구항 63

제1항 내지 제62항 중 어느 한 항의 시스템을 코딩하는 폴리뉴클레오티드.

청구항 64

제1항 내지 제62항 중 어느 한 항의 제1 및 제2 프라임 편집제를 코딩하는 폴리뉴클레오티드.

청구항 65

제1항 내지 제62항 중 어느 한 항의 제1 및/또는 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

청구항 66

제30항 내지 제62항 중 어느 한 항의 제1 및/또는 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

청구항 67

제1항 내지 제29항 중 어느 한 항의 제1 및 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

청구항 68

제30항 내지 제62항 중 어느 한 항의 제1 및 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

청구항 69

제63항 내지 제68항 중 어느 한 항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 70

제63항 내지 제68항 중 어느 한 항의 폴리뉴클레오티드 또는 제69항의 벡터를 포함하는 세포.

청구항 71

제63항 내지 제68항 중 어느 한 항의 폴리뉴클레오티드, 제69항의 벡터 또는 제70항의 세포 및 제약 부형제를 포함하는 제약 조성물.

청구항 72

제63항 내지 제68항 중 어느 한 항의 폴리뉴클레오티드 또는 제69항의 벡터 및 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하는 것에 대한 지침서를 포함하는 키트.

청구항 73

표적 부위에서 이중 가닥 DNA 서열의 제1 및 제2 상보적 가닥을 동시에 편집하는 방법으로서, 상기 방법은 이중 가닥 DNA 서열을 프라임 편집제 복합체의 쌍과 접촉시키는 것을 포함하고, 상기 쌍은 하기를 포함하고:

a. 하기를 포함하는 제1 프라임 편집제 복합체:

i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

ii. 표적 부위의 상류의 게놈 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

b. 하기를 포함하는 제2 프라임 편집제 복합체:

i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

여기서 제1 프라임 편집제 복합체는 제1 결합 부위에 상보적인 서열에서 제1 닉을 유발하고, 후속하여 제1 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제1 단일 가닥 DNA 서열의 중합을 유발하고;

여기서 제2 프라임 편집제 복합체는 제2 결합 부위에 상보적인 서열에서 제2 닉을 유발하고, 후속하여 제2 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제2 단일 가닥 DNA 서열의 중합을 유발하고;

여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이고, 편집을 포함하는 듀플렉스를 형성하고;

여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하는 것인 방법.

청구항 74

제73항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 방법.

청구항 75

제73항에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 방법.

청구항 76

제73항에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 상이한 프라임 편집제를 포함하는 것인 방법.

청구항 77

제73항 내지 제76항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9 도메인 또는 그의 변이체인 방법.

청구항 78

제73항 내지 제77항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, Cas9 닉카제 도메인 또는 그의 변이체인 방법.

청구항 79

제73항 내지 제78항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, CasX, CasY 및 아르고노트로 이루어진 군으로부터 선택된 것인 방법.

청구항 80

제73항 내지 제79항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 닉카제 활성을 갖는 것인 방법.

청구항 81

제73항 내지 제80항 중 어느 한 항에 있어서, 제1 및/또는 제2 napDNAbp가 서열식별번호: 18-88 중 어느 하나의 아미노산 서열 또는 서열식별번호: 18-88 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

청구항 82

제73항 내지 제81항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 리버스 트랜스크립타제인 방법.

청구항 83

제73항 내지 제82항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159 및 700-766 중 어느 하나의 아미노산 서열 또는 서열식별번호: 89-100, 105-122, 128-129 및 132 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

청구항 84

제73항 내지 제83항 중 어느 한 항에 있어서, 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나, 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 방법.

청구항 85

제84항에 있어서, 링커가 서열식별번호: 166-177 중 어느 하나의 아미노산 서열 또는 서열식별번호: 166-177 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

청구항 86

제84항 또는 제85항에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 방법.

청구항 87

제73항 내지 제86항 중 어느 한 항에 있어서, 제1 및/또는 제2 pegRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 방법.

청구항 88

제73항 내지 제87항 중 어느 한 항에 있어서, 제1 및/또는 제2 PEgRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 방법.

청구항 89

제73항 내지 제88항 중 어느 한 항에 있어서, 제1 프라임 편집제 복합체의 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 방법.

청구항 90

제73항 내지 제89항 중 어느 한 항에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 방법.

청구항 91

제90항에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 제1 및 제2 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 방법.

청구항 92

제91항에 있어서, 제1 및 제2 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 방법.

청구항 93

제92항에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍,

적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

청구항 94

제92항 또는 제93항에 있어서, 인접 영역이 편집된 부분을 포함하는 듀플렉스에 의해 대체되는 것인 방법.

청구항 95

제73항 내지 제94항 중 어느 한 항에 있어서, 제1 및/또는 제2 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개

뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

청구항 96

제73항 내지 제95항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 방법.

청구항 97

제73항 내지 제96항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 방법.

청구항 98

제73항 내지 제97항 중 어느 한 항에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 방법.

청구항 99

제73항 내지 제98항 중 어느 한 항에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이이고, 목적하는 DNA 서열 및 그의 상보체의 길이 이하인 방법.

청구항 100

제73항 내지 제99항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 방법.

청구항 101

제73항 내지 제100항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 방법.

청구항 102

제94항에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 방법.

청구항 103

제73항 내지 제102항 중 어느 한 항에 있어서, 편집을 포함하는 듀플렉스가 1개 이상의 레콤비나제 부위를 포함하는 것인 방법.

청구항 104

제103항에 있어서, 편집될 표적 부위에서 DNA 서열의 레콤비나제-매개된 편집을 촉진하기 위해 레콤비나제를 제공하는 것을 추가로 포함하는 방법.

청구항 105

이중 프라임 편집에 사용하기 위한 pegRNA의 쌍으로서, 여기서 쌍은

a. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA)이며, 여기서 제1 PEgRNA는 제1 스페이서 서열, 제1 gRNA 코어, 제1 DNA 합성 주형 및 제1 프라이머 결합 부위를 포함하고, 여기서 제1 DNA 합성 주형은 편집된 부분을 포함하는 제1 단일 가닥 DNA 서열을 코딩하는 것인 제1 PEgRNA;

b. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA)이며, 여기서 제2 PEgRNA는 제2 스페이서 서열, 제2 gRNA 코어, 제2 DNA 합성 주형 및 제2 프라이머 결합 부위를 포함하고, 여기서 제2 DNA 합성 주형은 편집된 부분을 포함하는 제2 단일 가닥 DNA 서열을 코딩하고, 여기서 제2 단일 가닥 DNA 서열은 제1 단일 가닥 DNA 서열에 대한 상보성 영역을 포함하는 것인 제2 PEgRNA

를 포함하는 것인 pegRNA의 쌍.

청구항 106

제105항에 있어서, 제1 PEgRNA가 하기 구조:

5'-[제1 gRNA 스페이서 서열]-[제1 gRNA 코어]-[제1 DNA 합성 주형]-[제1 프라이머 결합 부위]-3'를 포함하고,

제2 PEgRNA가 하기 구조:

5'-[제2 gRNA 스페이서 서열]-[제2 gRNA 코어]-[제2 DNA 합성 주형]-[제2 프라이머 결합 부위]-3'를 포함하는 것인 pegRNA의 쌍.

청구항 107

제105항 또는 제106항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열이 듀플렉스를 형성하고, 이는 편집될 표적 부위에서 이중 가닥 DNA 서열 내로 통합되는 것인 pegRNA의 쌍.

청구항 108

제107항에 있어서, 듀플렉스의 통합이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 pegRNA의 쌍.

청구항 109

제108항에 있어서, 삽입이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 pegRNA의 쌍.

청구항 110

제108항에 있어서, 결실이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72

개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 pegRNA의 쌍.

청구항 111

제108항에 있어서, 대체가 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 pegRNA의 쌍.

청구항 112

제105항 내지 제111항 중 어느 한 항의 pegRNA의 쌍을 코딩하는 폴리뉴클레오티드.

청구항 113

세포에서 pegRNA의 쌍을 발현시키기 위한 제112항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 114

제113항의 벡터를 포함하는 세포.

청구항 115

제105항 내지 제111항 중 어느 한 항의 pegRNA의 쌍, 제113항의 벡터 또는 제114항의 세포 및 제약 부형제를 포함하는 제약 조성물.

청구항 116

프라임 편집제 복합체의 쌍으로서, 상기 쌍은

a. 하기를 포함하는 제1 프라임 편집제 복합체:

i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

ii. 표적 부위의 상류의 게놈 DNA 서열의 가닥 상의 제1 표적 서열에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

b. 하기를 포함하는 제2 프라임 편집제 복합체:

i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제2 표적 서열에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA)

를 포함하는 것인 프라임 편집제 복합체의 쌍.

청구항 117

제116항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 프라임 편집제 복합체의 쌍.

청구항 118

제116항 또는 제117항의 프라임 편집제 복합체의 쌍을 코딩하는 폴리뉴클레오티드.

청구항 119

세포에서 프라임 편집제 복합체의 쌍을 발현시키기 위한 제118항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 120

제119항의 벡터를 포함하는 세포.

청구항 121

제116항 또는 제117항의 프라임 편집제 복합체의 쌍, 제119항의 벡터 또는 제120항의 세포 및 제약 부형제를 포함하는 제약 조성물.

청구항 122

제1항 내지 제62항 및 제73항 내지 제104항 중 어느 한 항에 있어서, 편집된 부분을 포함하는 듀플렉스의 통합이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 시스템 또는 방법.

청구항 123

제122항에 있어서, 삽입이 펩티드 태그인 시스템 또는 방법.

청구항 124

제123항에 있어서, 펩티드 태그가 폴리-히스티딘 (예를 들어, HHHHHH) (서열식별번호: 252-262), FLAG (예를 들어, DYKDDDDK) (서열식별번호: 2), V5 (예를 들어, GKPIPNPLLGLDST) (서열식별번호: 3), GCN4, HA (예를 들어, YPYDVPDYA) (서열식별번호: 5), Myc (예를 들어, EQKLISEED) (서열식별번호: 6) 또는 GST인 시스템 또는 방법.

청구항 125

제123항에 있어서, 펩티드 태그가 서열식별번호: 1-6, 245-249, 252-262, 264-273, 275-276, 281, 278-288 및 622로 이루어진 군으로부터 선택된 아미노산 서열을 갖는 것인 시스템 또는 방법.

청구항 126

제122항에 있어서, 삽입이 이뮤노에피토프 태그인 시스템 또는 방법.

청구항 127

제126항에 있어서, 이뮤노에피토프 태그가 파상풍 톡소이드 (서열식별번호: 396); 디프테리아 독소 돌연변이체 CRM197 (서열식별번호: 630); 볼거리 이뮤노에피토프 1 (서열식별번호: 400); 볼거리 이뮤노에피토프 2 (서열식별번호: 402); 볼거리 이뮤노에피토프 3 (서열식별번호: 404); 풍진 바이러스 (서열식별번호: 406); 헤마글루티닌 (서열식별번호: 408); 뉴라미니다제 (서열식별번호: 410); TAP1 (서열식별번호: 412); TAP2 (서열식별번호: 414); 부류 I HLA에 대한 헤마글루티닌 에피토프 (서열식별번호: 416); 부류 I HLA에 대한 뉴라미니다제 에피토프 (서열식별번호: 418); 부류 II HLA에 대한 헤마글루티닌 에피토프 (서열식별번호: 420); 부류 II HLA에 대한 뉴라미니다제 에피토프 (서열식별번호: 422); 헤마글루티닌 에피토프 H5N1-결합된 부류 I 및 부류 II HLA (서열식별번호: 424); 및 뉴라미니다제 에피토프 H5N1-결합된 부류 I 및 부류 II HLA (서열식별번호: 426)로 이루어진 군으로부터 선택된 것인 시스템 또는 방법.

청구항 128

제122항에 있어서, 삽입이 이량체화 도메인인 시스템 또는 방법.

청구항 129

제128항에 있어서, 이량체화 도메인이 서열식별번호: 488의 FKBP12, 서열식별번호: 489의 FKBP12-F36V 또는 서열식별번호: 490 또는 493-494의 시클로필린의 소분자 결합 도메인인 시스템 또는 방법.

청구항 130

편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템으로서, 시스템은 하기를 포함하고:

a) 하기를 포함하는 제1 프라임 편집제 복합체:

i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

ii. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

b) 하기를 포함하는 제2 프라임 편집제 복합체:

i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

ii. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

c) 하기를 포함하는 제3 프라임 편집제 복합체:

i. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

ii. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 제3 프라임 편

집 가이드 RNA (제3 PEgRNA);

d) 하기를 포함하는 제4 프라임 편집제 복합체:

i. 제4 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

ii. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

여기서 제1 PEgRNA는 제1 단일 가닥 DNA 서열을 코딩하는 제1 DNA 합성 주형을 포함하고, 제2 PEgRNA는 제2 단일 가닥 DNA 서열을 코딩하는 제2 DNA 합성 주형을 포함하고, 제3 PEgRNA는 제3 단일 가닥 DNA 서열을 코딩하는 제3 DNA 합성 주형을 포함하고, 제4 PEgRNA는 제4 단일 가닥 DNA 서열을 코딩하는 제4 DNA 합성 주형을 포함하고;

여기서 제1 및 제3 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고; 여기서 제2 및 제4 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하는 것인 시스템.

청구항 131

1개 이상의 이중 가닥 DNA 서열을 편집하기 위한 시스템으로서, 시스템은 하기를 포함하고:

a) 하기를 포함하는 제1 프라임 편집제 복합체:

i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

ii. 편집될 제1 표적 부위에서 제1 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

b) 하기를 포함하는 제2 프라임 편집제 복합체:

i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

ii. 편집될 제1 표적 부위에서 제1 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

c) 하기를 포함하는 제3 프라임 편집제 복합체:

i. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

ii. 편집될 제2 표적 부위에서 제2 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA);

d) 하기를 포함하는 제4 프라임 편집제 복합체:

i. 제4 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

ii. 편집될 제2 표적 부위에서 제2 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

여기서 제1 PEgRNA는 제1 단일 가닥 DNA 서열을 코딩하는 제1 DNA 합성 주형을 포함하고, 제2 PEgRNA는 제2 단일 가닥 DNA 서열을 코딩하는 제2 DNA 합성 주형을 포함하고, 제3 PEgRNA는 제3 단일 가닥 DNA 서열을 코딩하는 제3 DNA 합성 주형을 포함하고, 제4 PEgRNA는 제4 단일 가닥 DNA 서열을 코딩하는 제4 DNA 합성 주형을 포함하고;

여기서 제1 및 제3 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고; 여기서 제2 및 제4 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하는 것인 시스템.

청구항 132

제131항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및/또는 제4 프라임 편집제 복합체의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 시스템.

청구항 133

제130항 내지 제132항 중 어느 한 항에 있어서,

(i) 제1 PEgRNA가 제1 스페이서 서열, 제1 gRNA 코어 및 제1 프라이머 결합 부위를 추가로 포함하고;

(ii) 제2 PEgRNA가 제2 스페이서 서열, 제2 gRNA 코어 및 제2 프라이머 결합 부위를 추가로 포함하고;

(iii) 제3 PEgRNA가 제3 스페이서 서열, 제3 gRNA 코어 및 제3 프라이머 결합 부위를 추가로 포함하고;

(iv) 제4 PEgRNA가 제4 스페이서 서열, 제4 gRNA 코어, 제4 DNA 합성 주형 및 제4 프라이머 결합 부위를 포함하는 것인 시스템.

청구항 134

제131항에 있어서, 제1과 제3 단일 가닥 DNA 서열의 상보성 영역 및/또는 제2와 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

청구항 135

제134항에 있어서, 제1 스페이서 서열이 편집될 제1 표적 부위의 상류의 제1 이중 가닥 DNA 서열의 제1 결합 부위에 결합하고, 제2 스페이서 서열이 편집될 제1 표적 부위의 하류의 제1 이중 가닥 DNA 서열 상의 제2 결합 부위에 결합하는 것인 시스템.

청구항 136

제134항 또는 제135항에 있어서, 제3 스페이서 서열이 편집될 제2 표적 부위의 상류의 제2 이중 가닥 DNA 서열의 제1 결합 부위에 결합하고, 제4 스페이서 서열이 편집될 제2 표적 부위의 하류의 제2 이중 가닥 DNA 서열 상의 제2 결합 부위에 결합하는 것인 시스템.

청구항 137

제134항 내지 제136항 중 어느 한 항에 있어서, 제1 및 제2 결합 부위가 제1 이중 가닥 DNA의 제1 인접 영역의 2개의 단부를 규정하는 것인 시스템.

청구항 138

제134항 내지 제137항 중 어느 한 항에 있어서, 제3 및 제4 결합 부위가 제2 이중 가닥 DNA의 제2 인접 영역의 2개의 단부를 규정하는 것인 시스템.

청구항 139

제134항에 있어서, 제1 및 제2 결합 부위 사이 또는 제3 및 제4 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드,

적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 140

제134항 내지 제139항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역 및/또는 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

청구항 141

제140항에 있어서, 듀플렉스가 제1 이중 가닥 DNA 및/또는 제2 이중 가닥 DNA에 혼입되는 것인 시스템.

청구항 142

제134항 내지 제141항 중 어느 한 항에 있어서, 제1 이중 가닥 DNA 내로 제1 인접 영역의 제2 인접 영역에 의한 대체를 발생시키는 시스템.

청구항 143

제134항 내지 제142항 중 어느 한 항에 있어서, 제1 이중 가닥 DNA와 제2 이중 가닥 DNA 사이의 제1 인접 영역 및 제2 인접 영역의 교환을 발생시키는 시스템.

청구항 144

제134항 내지 제143항 중 어느 한 항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및 제4 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 시스템.

청구항 145

제134항 내지 제144항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1, 제2, 제3 및 제4 폴리펩티드가 동일한 것인 시스템.

청구항 146

제134항 내지 제145항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9 도메인 또는 그의 변이체인 시스템.

청구항 147

제134항 내지 제146항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인 또는 Cas9 닉카제 도메인 또는 그의 변이체인 시스템.

청구항 148

제134항 내지 제147항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c 및 아르고노트로 이루어진 군으로부터 선택된 것인 시스템.

청구항 149

제134항 내지 제148항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 닉카제 활성을 갖는 것인 시스템.

청구항 150

제134항 내지 제149항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 서열식별번호: 2-65 중 어느 하나의 아미노산 서열 또는 서열식별번호: 2-65 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 151

제134항 내지 제150항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 리버스 트랜스크립타제인 시스템.

청구항 152

제134항 내지 제151항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나의 아미노산 서열 또는 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 153

제134항 내지 제152항 중 어느 한 항에 있어서,

(i) 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

(ii) 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

(iii) 제3 프라임 편집제가 제3 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제3 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

(iv) 제4 프라임 편집제가 제4 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제4 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 시스템.

청구항 154

제153항에 있어서, 링커가 서열식별번호: 119-128 중 어느 하나의 아미노산 서열 또는 서열식별번호: 119-128 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

청구항 155

제153항 또는 제154항에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 시스템.

청구항 156

제134항 내지 제155항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 PEgRNA가 서열식별번호: 192-203 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 192-203 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 시스템.

청구항 157

제134항 내지 제156항 중 어느 한 항에 있어서, 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥

DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 시스템.

청구항 158

제134항 내지 제157항 중 어느 한 항에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 시스템.

청구항 159

제134항 내지 제158항 중 어느 한 항에 있어서, 제3 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 것인 시스템.

청구항 160

제134항 내지 제159항 중 어느 한 항에 있어서, 제4 프라임 편집제 복합체의 제4 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 것인 시스템.

청구항 161

제160항에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 제1 및/또는 제2 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

청구항 162

제161항에 있어서, PAM 서열의 존재 하에서의 제3 및 제4 프라임 편집제 복합체의 스페이서 서열의 결합이 각각의 제3 및 제4 프라임 편집제 복합체의 제3 및/또는 제4 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

청구항 163

제133항에 있어서, 제1 및 제3 결합 부위 또는 제2 및 제4 결합 부위가 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 시스템.

청구항 164

제163항에 있어서, 제1 및 제3 결합 부위 사이 또는 제2 및 제4 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드,

적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 165

제130항 내지 제164항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열의 상보성 영역 및/또는 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

청구항 166

제164항에 있어서, 인접 영역이 듀플렉스에 의해 대체되는 것인 시스템.

청구항 167

제166항에 있어서, 듀플렉스가 표적 부위에서 이중 가닥 DNA에 혼입되는 것인 시스템.

청구항 168

제122항 내지 제167항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

청구항 169

제122항 내지 제168항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제

3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 시스템.

청구항 170

제122항 내지 제169항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 시스템.

청구항 171

제122항 내지 제170항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 단일 가닥 DNA 서열의 상보성 영역이 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 시스템.

청구항 172

제122항 내지 제171항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 단일 가닥 DNA 서열의 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이이고, 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열, 및 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열의 길이 이하인 시스템.

청구항 173

제130항 내지 제172항 중 어느 한 항에 있어서, 제1, 제2, 제3 및 제4 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 시스템.

청구항 174

제173항에 있어서, 편집된 부분을 포함하는 듀플렉스의 혼입이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발

생시키는 것인 시스템.

청구항 175

제130항 내지 제174항 중 어느 한 항의 제1, 제2, 제3 및 제4 프라임 편집제를 코딩하는 폴리뉴클레오티드.

청구항 176

제130항 내지 제175항 중 어느 한 항의 제1, 제2, 제3 및 제4 PEgRNA를 코딩하는 폴리뉴클레오티드.

청구항 177

제175항 또는 제176항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 178

제175항 또는 제176항의 폴리뉴클레오티드 또는 제177항의 벡터를 포함하는 세포.

청구항 179

제175항 또는 제176항의 폴리뉴클레오티드, 제177항의 벡터 또는 제178항의 세포 및 제약 부형제를 포함하는 제약 조성물.

청구항 180

제175항 또는 제176항의 폴리뉴클레오티드 또는 제177항의 벡터 및 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하는 것에 대한 지침서를 포함하는 키트.

청구항 181

표적 부위에서 이중 가닥 DNA 서열의 제1 및 제2 상보적 가닥을 동시에 편집하는 방법으로서, 방법은 이중 가닥 DNA 서열을 프라임 편집제 복합체의 쌍과 접촉시키는 것을 포함하고, 쌍은 하기를 포함하고:

(a) 하기를 포함하는 제1 프라임 편집제 복합체:

i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

ii. 표적 부위의 상류의 게놈 DNA 서열의 제1 가닥 상의 제1 표적 서열에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

(b) 하기를 포함하는 제2 프라임 편집제 복합체:

i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

ii. 표적 부위의 상류의 게놈 DNA 서열의 제2 가닥 상의 제2 표적 서열에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

(c) 하기를 포함하는 제3 프라임 편집제 복합체:

i. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

ii. 표적 부위의 하류의 게놈 DNA 서열의 제1 가닥 상의 제3 표적 서열에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA);

(d) 하기를 포함하는 제4 프라임 편집제 복합체:

i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제4 표적 서열에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

여기서 제1 프라임 편집제 복합체는 제1 표적 서열에서 제1 닉을 유발하고, 후속하여 제1 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제1 단일 가닥 DNA 서열의 중합을 유발하고;

여기서 제2 프라임 편집제 복합체는 제2 표적 서열에서 제2 닉을 유발하고, 후속하여 제2 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제2 단일 가닥 DNA 서열의 중합을 유발하고;

여기서 제3 프라임 편집제 복합체는 제3 표적 서열에서 제3 닉을 유발하고, 후속하여 제3 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제3 단일 가닥 DNA 서열의 중합을 유발하고;

여기서 제4 프라임 편집제 복합체는 제4 표적 서열에서 제4 닉을 유발하고, 후속하여 제4 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제4 단일 가닥 DNA 서열의 중합을 유발하고;

여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이며 듀플렉스를 형성하고, 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하고;

여기서 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이며 듀플렉스를 형성하고, 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하는 것인 방법.

청구항 182

제181항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및/또는 제4 프라임 편집제 복합체의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 방법.

청구항 183

제181항 또는 제182항에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및 제4 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 방법.

청구항 184

제181항 내지 제183항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 프라임 편집제 복합체가 상이한 프라임 편집제를 포함하는 것인 방법.

청구항 185

제181항 내지 제184항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9 도메인 또는 그의 변이체인 방법.

청구항 186

제181항 내지 제185항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인 또는 Cas9 닉카제 도메인 또는 그의 변이체인 방법.

청구항 187

제181항 내지 제184항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c로 이루어진 군으로부터 선택된 것인 방법.

청구항 188

제181항 내지 제187항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 닉카제 활성을 갖는 것인 방법.

청구항 189

제181항 내지 제188항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 서열식별번호: 2-65 중 어느 하나의 아미노산 서열 또는 서열식별번호: 2-65 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

청구항 190

제181항 내지 제189항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 리버스 트랜스크립타제인 방법.

청구항 191

제181항 내지 제190항 중 어느 한 항에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나의 아미노산 서열 또는 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

청구항 192

제181항 내지 제191항 중 어느 한 항에 있어서,

(i) 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

(ii) 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

(iii) 제3 프라임 편집제가 제3 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제3 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

(iv) 제4 프라임 편집제가 제4 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제4 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 방법.

청구항 193

제192항에 있어서, 링커가 서열식별번호: 119-128 중 어느 하나의 아미노산 서열 또는 서열식별번호: 119-128 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

청구항 194

제192항 또는 제193항에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 방법.

청구항 195

제192항 내지 제194항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 PEgRNA가 서열식별번호: 192-203 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 192-203 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 방법.

청구항 196

제192항 내지 제195항 중 어느 한 항에 있어서, 제1 프라임 편집제 복합체의 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 방법.

청구항 197

제192항 내지 제196항 중 어느 한 항에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 방법.

청구항 198

제192항 내지 제197항 중 어느 한 항에 있어서, 제3 프라임 편집제 복합체의 제3 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 것인 방법.

청구항 199

제192항 내지 제198항 중 어느 한 항에 있어서, 제4 프라임 편집제 복합체의 제4 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 것인 방법.

청구항 200

제199항에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 제1 및 제2 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 방법.

청구항 201

제200항에 있어서, PAM 서열의 존재 하에서의 제3 및 제4 프라임 편집제 복합체의 제3 및 제4 스페이서 서열의 결합이 각각의 제3 및 제4 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 방법.

청구항 202

제200항 또는 제201항에 있어서, 제1 및 제2 결합 부위 및/또는 제3 및 제4 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 방법.

청구항 203

제202항에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역 및/또는 제3 및 제4 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

청구항 204

제202항 또는 제203항에 있어서, 인접 영역이 듀플렉스에 의해 대체되는 것인 방법.

청구항 205

제181항 내지 제204항 중 어느 한 항에 있어서, 제1, 제2, 제3 및/또는 제4 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

청구항 206

제181항 내지 제205항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 방법.

청구항 207

제181항 내지 제206항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 방법.

청구항 208

제181항 내지 제207항 중 어느 한 항에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 방법.

청구항 209

제181항 내지 제208항 중 어느 한 항에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티

드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

청구항 210

제188항 내지 제209항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 방법.

청구항 211

제188항 내지 제210항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 방법.

청구항 212

제204항에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 방법.

청구항 213

제181항 내지 제212항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 표적 DNA 서열의 단부 상에 있고, 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 동일한 표적 DNA 서열의 대향 단부 상에 있는 것인 방법.

청구항 214

제213항에 있어서, 표적 DNA 서열이 역전된 것인 방법.

청구항 215

제181항 내지 제214항 중 어느 한 항에 있어서, 원형 DNA 공여자를 제공하는 것을 추가로 포함하는 방법.

청구항 216

제215항에 있어서, 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열이 표적 DNA 서열의 대향 단부 상에 있고, 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 원형 DNA 공여자 상에 있는 것인 방법.

청구항 217

제216항에 있어서, 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 원형 DNA 공여자의 부분이 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 표적 서열을 대체하는 것인 방법.

청구항 218

제185항 내지 제217항 중 어느 한 항에 있어서, 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열이 제1 핵산 분자 상에 있고, 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 제2 핵산 분자 상에 있는 것인 방법.

청구항 219

제218항에 있어서, 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 제1 핵산 분자의 부분이 제2 핵산 분자 내로 혼입되고, 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 제2 핵산 분자의 부분이 제1 핵산 분자 내로 혼입되는 것인 방법.

청구항 220

제218항 또는 제219항에 있어서, 제1 핵산 분자가 제1 염색체이고, 제2 핵산 분자가 제2 염색체인 방법.

청구항 221

다중-플랩 프라임 편집에 사용하기 위한 PEgRNA의 세트로서, 여기서 PEgRNA는

(a) 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA)이며, 여기서 제1 PEgRNA는 제1 스페이서 서열, 제1 gRNA 코어, 제1 DNA 합성 주형 및 제1 프라이머 결합 부위를 포함하고, 여기서 제1 DNA 합성 주형은 제1 단일 가닥 DNA 서열을 코딩하는 것인 제1 PEgRNA;

(b) 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA)이며, 여기서 제2 PEgRNA는 제2 스페이서 서열, 제2 gRNA 코어, 제2 DNA 합성 주형 및 제2 프라이머 결합 부위를 포함하고, 여기서 제2 DNA 합성 주형은 편집된 부분을 포함하는 제2 단일 가닥 DNA 서열을 코딩하고, 여기서 제2 단일 가닥 DNA 서열은 제1 단일 가닥 DNA 서열에 상보적인 것인 제2 PEgRNA;

(c) 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA)이며, 여기서 제3 PEgRNA는 제3 스페이서 서열, 제3 gRNA 코어, 제3 DNA 합성 주형 및 제3 프라이머 결합 부위를 포함하고, 여기서 제3 DNA 합성 주형은 제3 단일 가닥 DNA 서열을 코딩하는 것인 제3 PEgRNA;

(d) 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA)이며, 여기서 제4 PEgRNA는 제4 스페이서 서열, 제4 gRNA 코어, 제4 DNA 합성 주형 및 제4 프라이머 결합 부위를 포함하고, 여기서 제4 DNA 합성 주형은 제4 단일 가닥 DNA 서열을 코딩하고, 여기서 제4 단일 가닥 DNA 서열은 제3 단일 가닥 DNA 서열에 대한 상보성 영역을 포함하는 것인 제4 PEgRNA

를 포함하는 것인 PEgRNA의 세트.

청구항 222

제221항에 있어서,

제1 PEgRNA가 하기 구조:

5'-[제1 gRNA 스페이서 서열]-[제1 gRNA 코어]-[제1 DNA 합성 주형]-[제1 프라이머 결합 부위]-3'를 포함하고;

제2 PEgRNA가 하기 구조:

5'-[제2 gRNA 스페이서 서열]-[제2 gRNA 코어]-[제2 DNA 합성 주형]-[제2 프라이머 결합 부위]-3'를 포함하고;

제3 PEgRNA가 하기 구조:

5'-[제3 gRNA 스페이서 서열]-[제3 gRNA 코어]-[제3 DNA 합성 주형]-[제3 프라이머 결합 부위]-3'를 포함하고;

제4 PEgRNA가 하기 구조:

5'-[제4 gRNA 스페이서 서열]-[제4 gRNA 코어]-[제4 DNA 합성 주형]-[제4 프라이머 결합 부위]-3'를 포함하는 것인 PEgRNA의 세트.

청구항 223

제221항 또는 제222항에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열이 듀플렉스를 형성하고, 이는 편집될 표적 부위에서 이중 가닥 DNA 서열 내로 통합되는 것인 PEgRNA의 세트.

청구항 224

제221항 내지 제223항 중 어느 한 항에 있어서, 듀플렉스의 통합이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 PEgRNA의 세트.

청구항 225

제224항에 있어서, 삽입이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 PEgRNA의 세트.

청구항 226

제224항에 있어서, 결실이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 PEgRNA의 세트.

청구항 227

제224항에 있어서, 대체가 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80

개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 PEgRNA의 세트.

청구항 228

제221항 내지 제227항 중 어느 한 항의 PEgRNA의 세트를 코딩하는 폴리뉴클레오티드.

청구항 229

세포에서 복수의 PEgRNA를 발현시키기 위한 제228항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 230

제229항의 벡터를 포함하는 세포.

청구항 231

제221항 내지 제227항 중 어느 한 항의 복수의 PEgRNA, 제229항의 벡터 또는 제230항의 세포 및 제약 부형제를 포함하는 제약 조성물.

청구항 232

세포 내 게놈을 변형시키는 방법으로서,

(i) 게놈을 이중-플랩 프라임 편집 시스템과 접촉시키는 단계이며, 상기 시스템은 표적 부위의 제1 가닥 상의 제1 닉 부위에 제1 3' 핵산 플랩을 설치할 수 있는 제1 프라임 편집제/pegRNA 복합체 및 표적 부위의 제2 가닥 상의 제2 닉 부위에 제2 3' 핵산 플랩을 설치할 수 있는 제2 프라임 편집제/pegRNA 복합체를 포함하고,

여기서 제1 3' 핵산 플랩 및 제2 3' 핵산 플랩은 듀플렉스를 형성하는 역 상보체 서열이고,

여기서 듀플렉스는 제1 레콤비나제 부위를 포함하고,

여기서 세포는 제1 및 제2 닉 부위 사이에 위치한 내인성 듀플렉스 대신에 표적 부위에 제1 레콤비나제 부위를 포함하는 듀플렉스를 설치하는 것인 단계;

(ii) 제1 레콤비나제 부위를 제2 레콤비나제 부위를 포함하는 공여자 핵산과 접촉시키는 단계; 및

(iii) 게놈을 제1 및 제2 레콤비나제 부위 사이의 재조합이 가능한 레콤비나제와 접촉시킴으로써, 게놈 내 표적 부위에서 공여자 핵산을 통합시키는 단계

를 포함하는 방법.

발명의 설명

기 술 분 야

[0001] 정부 지원

[0002] 본 발명은 미국 국립 보건원에 의해 수여된 승인 번호 U01AI142756, RM1HG009490, R01EB022376 및 R35GM118062 하의 정부 지원으로 이루어졌다. 정부는 본 발명에 특정 권리를 갖는다.

[0003] 관련 출원 및 참조로 포함

[0004] 본 출원은 2020년 5월 8일에 출원된 미국 가출원 번호 63/022,397 및 2020년 11월 20일에 출원된 미국 가출원 번호 63/116,785를 우선권 주장하며, 이들 각각의 전체 내용은 그 전문이 본원에 참조로 포함된다.

[0005] 이 미국 가출원은 또한 하기 출원, 즉 2019년 3월 19일에 출원된 미국 가출원 번호 62/820,813 (대리인 문서 번호 B1195.70074US00), 2019년 6월 7일에 출원된 미국 가출원 번호 62/858,958 (대리인 문서 번호 B1195.70074US01), 2019년 8월 21일에 출원된 미국 가출원 번호 62/889,996 (대리인 문서 번호 B1195.70074US02), 2019년 8월 21일에 출원된 미국 가출원 번호 62/922,654 (대리인 문서 번호 B1195.70083US00), 2019년 10월 10일에 출원된 미국 가출원 번호 62/913,553 (대리인 문서 번호 B1195.70074US03), 2019년 10월 10일에 출원된 미국 가출원 번호 62/973,558 (대리인 문서 번호 B1195.70083US01), 2019년 11월 5일에 출원된 미국 가출원 번호 62/931,195 (대리인 문서 번호 B1195.70074US04), 2019년 12월 5일에 출원된 미국 가출원 번호 62/944,231 (대리인 문서 번호 B1195.70074US05), 2019년 12월 5일에 출원된 미국 가출원 번호 62/974,537 (대리인 문서 번호 B1195.70083US02), 2020년 3월 17일에 출원된 미국 가출원 번호 62/991,069 (대리인 문서 번호 B1195.70074US06), 및 2020년 3월 17일에 출원된 미국 가출원 번호 (63/100,548) (대리인 문서 번호 B1195.70083US03)를 참조하고 참조로 포함시킨다. 추가로, 이 미국 가출원은 각각 2020년 3월 19일에 출원된 국제 PCT 출원 번호 PCT/US20/23721; PCT/US20/23730; PCT/US20/23713; PCT/US20/23712; PCT/US20/23727; PCT/US20/23724; PCT/US20/23725; PCT/US20/23728; PCT/US20/23732; PCT/US20/23723; PCT/US20/23553; 및 PCT/US20/23583을 참조하고 참조로 포함시킨다.

배 경 기 술

[0006] 병원성 단일 뉴클레오티드 돌연변이는 일부 추정에 따르면 유전적 성분이 존재하는 인간 질환의 대략 50%에 기여한다[7]. 불행하게도, 이들 유전 장애를 갖는 환자에 대한 치료 옵션은 수십년의 유전자 요법 탐구[8]에도 불구하고 극히 제한적인 것으로 남아있다. 아마도 이러한 치료 과제에 대한 가장 경제적인 해결책은 환자 게놈 내의 단일 뉴클레오티드 돌연변이의 직접적 교정이며, 이는 질환의 근본 원인을 다룰 것이고, 지속적인 이익을 제공할 가능성이 있다. 이러한 전략은 이전에 생각할 수 없었지만, CRISPR/Cas 시스템[9]의 출현에 의해 야기된 게놈 편집 능력에서의 최근의 개선은 이제 이러한 치료 접근법을 도달할 수 있는 범위 내로 가져왔다. 표적 DNA 서열에 상보적인 ~20개의 뉴클레오티드를 함유하는 가이드 RNA (gRNA) 서열의 간단한 설계에 의해, 거의 모든 고려가능한 게놈 부위가 CRISPR 연관 (Cas) 뉴클레아제[1,2]에 의해 특이적으로 접근될 수 있다. 지금까지, 여러 단량체 박테리아 Cas 뉴클레아제 시스템이 확인되었고, 게놈 편집 적용[10]을 위해 적합화되었다. 이러한 Cas 뉴클레아제의 자연적 다양성은, 증가하는 조작된 변이체[11-14]의 집합물과 함께, 새로운 게놈 편집 기술을 개발하기 위한 비옥한 기반을 제공한다.

[0007] CRISPR을 사용한 유전자 파괴는 이제 성숙한 기술이지만, 인간 게놈 내의 단일 염기 쌍의 정밀 편집은 주요 과제로 남아있다[3]. 상동성 지정 복구 (HDR)는 목적하는 편집을 코딩하는 공여자 DNA 복구 주형을 사용하여 이중 가닥 파괴 (DSB)의 부위에서 DNA 서열을 삽입, 교정 또는 교환하기 위해 인간 세포 및 다른 유기체에서 오랫동안 사용되어 왔다[15]. 그러나, 전통적인 HDR은 대부분의 인간 세포 유형, 특히 비-분열 세포에서 매우 낮은 효율을 갖고, 경쟁 비-상동 단부 연결 (NHEJ)은 삽입-결실 (indel) 부산물을 우세하게 유도한다[16]. 다른 문제는 DSB의 생성에 관한 것이며, 이는 표적 유전자좌에서 큰 염색체 재배열 및 결실을 일으킬 수 있거나[17], 또는 p53 축을 활성화시켜 성장 정지 및 아폽토시스를 유도할 수 있다[18,19].

[0008] 이들 HDR의 결점을 다루는 여러 접근법이 탐구되어 왔다. 예를 들어, 올리고뉴클레오티드 공여자에 의한 단일 가닥 DNA 파괴 (닉)의 복구는 indel 형성을 감소시키는 것으로 밝혀졌지만, 목적하는 복구 생성물의 수율은 여전히 낮다[20]. 다른 전략은 소분자 및 생물학적 시약을 사용하여 NHEJ보다는 HDR 쪽으로 복구를 편향시키려 시도한다[21-23]. 그러나, 이들 방법의 유효성은 세포-유형 의존성일 가능성이 있고, 정상 세포 상태의 교란은 바람직하지 않은 예측불가능한 효과를 유도할 수 있다.

[0009] 최근에, 본 발명자들은 데이비드 리우(David Liu) 교수 등의 주도로, DSB를 생성하지 않으면서 또는 HDR에 의존하지 않으면서 표적 뉴클레오티드를 편집하는 기술로서 염기 편집을 개발하였다[4-6,24-27]. Cas-융합된 데아미나제 효소에 의한 DNA 염기의 직접적 변형은 매우 높은 효율로 짧은 표적 윈도우 (~5-7개 염기)에서, C·G에서 T·A로의 또는 A·T에서 G·C로의 염기 쌍 전환을 가능하게 한다. 그 결과, 염기 편집제는 과학계에 의해 신속하게

채택되었다. 그러나, 하기 인자는 정밀 게놈 편집에 대한 그의 일반성을 제한한다: (1) 표적 윈도우 내의 비-표적 C 또는 A 염기의 "방관자 편집"이 관찰됨; (2) 표적 뉴클레오티드 생성물 혼합물이 관찰됨; (3) 표적 염기는 PAM 서열의 15±2개 뉴클레오티드 상류에 위치해야 함; 및 (5) 작은 삽입 및 결실 돌연변이의 복구는 가능하지 않음.

[0010] 따라서, 임의의 목적하는 단일 뉴클레오티드 변화를 유연하게 도입할 수 있고/거나 염기 쌍 삽입 또는 결실 (예를 들어, 적어도 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100개 또는 그 초과의 염기 쌍 삽입 또는 결실)을 설치할 수 있고/거나 표적 부위에서의 뉴클레오티드 서열을 높은 특이성 및 효율로 변경 또는 변형시킬 수 있는 프로그램가능한 편집제의 개발은 CRISPR에 기초한 게놈 편집 기술의 범주 및 치료 잠재력을 실질적으로 확장시킬 것이다.

발명의 내용

[0011] 본 발명은 "다중-플랩 프라임 편집" (예를 들어, "이중-플랩 프라임 편집" 및 "사중-플랩 프라임 편집" 포함)으로 불리는 게놈 편집을 위한 새로운 플랫폼을 기재하고, 문헌 [Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149-157 (2019)] (이는 본원에 참조로 포함됨)에서 본 발명자들에 의해 기재된 바와 같은 "프라임 편집" 또는 "전형적 프라임 편집"의 혁신적인 진보를 나타낸다. 전형적 프라임 편집은 다양한 실시양태에서 동일한 가닥 상의 표적 핵산 내로 통합되는 단일 3' 플랩을 닉 부위에서 중합시키는 반면, 본원에 기재된 다중-플랩 프라임 편집 시스템은 다양한 실시양태에서 상이한 가닥들 상에 3' 플랩의 쌍 또는 다중 쌍을 생성하는 별개의 구축물, 시스템 및 방법론을 수반하며, 상기 쌍이 목적하는 편집을 포함하는 듀플렉스를 형성하고, 예를 들어 게놈 내 특이적 유전자좌 또는 편집 부위에서 표적 핵산 분자 내로 혼입된다. 다양한 측면에서, 3' 플랩의 쌍 또는 다중 쌍은 본원에 기재된 프라임 편집제에 의해 생성되었을 때 서로 어닐링하는 역 상보적 서열을 포함하기 때문에 듀플렉스를 형성한다. 듀플렉스는 인접한 닉 부위 사이에 위치한 내인성 듀플렉스 서열을 자연적으로 대체하는 세포-구동 메카니즘에 의해 표적 부위 내로 혼입된다. 특정 실시양태에서, 새로운 듀플렉스 서열은 1개 이상의 위치 (예를 들어, 인접한 게놈 유전자좌 또는 2개의 상이한 염색체 위치)에 도입될 수 있고, 1개 이상의 관심 서열, 예를 들어 단백질-코딩 서열, 펩티드-코딩 서열 또는 RNA-코딩 서열을 포함할 수 있다. 한 실시양태에서, 다중-플랩 프라임 편집 시스템에 의해 설치된 새로운 듀플렉스 서열은 레콤비나제 부위, 예를 들어 Bxb1 레콤비나제 attB (38 bp) 및/또는 attP (50 bp) 부위, 또는 Hin 레콤비나제, Gin 레콤비나제, Tn3 레콤비나제, β-6 레콤비나제, CinH 레콤비나제, ParA 레콤비나제, γδ 레콤비나제, ΦC31 레콤비나제, TP901 레콤비나제, TG1 레콤비나제, ϕBT1 레콤비나제, R4 레콤비나제, ϕRV1 레콤비나제, ϕFC1 레콤비나제, MR11 레콤비나제, A118 레콤비나제, U153 레콤비나제 및 gp29 레콤비나제, Cre 레콤비나제, FLP 레콤비나제, R 레콤비나제, 람다 레콤비나제, HK101 레콤비나제, HK022 레콤비나제 및 pSAM2 레콤비나제에 의해 인식되는 레콤비나제 부위를 포함할 수 있다.

[0012] 본 발명자들은 최근에 오류-유발 이중 가닥 DNA 파괴를 요구하지 않으면서 게놈 DNA 서열의 삽입, 결실 또는 대체를 가능하게 하는 프라임 편집을 개발하였다. 프라임 편집은 Cas9의 표적 게놈 부위로의 지시 및 목적하는 편집을 설치하기 위한 정보의 코딩 둘 다를 수행하는 조작된 프라임 편집 가이드 RNA (pegRNA)와 쌍형성된 조작된 Cas9 닉카제-리버스 트랜스크립타제 융합 단백질 (PE1 또는 PE2)을 사용한다. 프라임 편집은 다단계 편집 과정을 통해 진행된다: 1) Cas9 도메인은 pegRNA의 스페이서 서열에 의해 특정된 표적 게놈 DNA 부위에 결합하여 닉킹하고; 2) 리버스 트랜스크립타제 도메인은 닉킹된 게놈 DNA를 프라이머로서 사용하여, 역전사를 위한 주형으로서 pegRNA 상의 조작된 연장부를 사용하여 편집된 DNA 가닥의 합성을 개시하고-이는 편집된 DNA 서열을 함유하는 단일 가닥 3' 플랩을 생성하고; 3) 세포 DNA 복구는 5' 플랩 종의 변위에 의해 3' 플랩 중간체를 분해하고, 이는 편집된 3' 플랩에 의한 침입, 원래 DNA 서열을 함유하는 5' 플랩의 절제 및 새로운 3' 플랩의 라이게이션을 통해 발생하여 편집된 DNA 가닥을 혼입시킴으로써, 1개의 편집된 가닥 및 1개의 비편집된 가닥의 헤테로듀플렉스를 형성하고; 4) 세포 DNA 복구는 복구를 위한 주형으로서 편집된 가닥을 사용하여 헤테로듀플렉스 내의 비편집된 가닥을 대체함으로써, 편집 과정을 완료한다.

[0013] 목적하는 편집의 효율적인 혼입은 새로 합성된 3' 플랩이 게놈 DNA 부위에 상동인 서열의 부분을 함유하는 것을 요구한다. 이러한 상동성은 편집된 3' 플랩이 DNA 듀플렉스 내로의 혼입을 위해 내인성 DNA 가닥 (상응하는 5' 플랩)과 경쟁하는 것을 가능하게 한다. 편집된 3' 플랩은 내인성 5' 플랩보다 더 적은 서열 상동성을 함유할 것이기 때문에, 경쟁은 5' 플랩 가닥에 유리할 것으로 예상된다. 따라서, 프라임 편집의 효율에 있어서 잠재적 제한 인자는 내인성 DNA의 3' 플랩의 침입 및 후속하여 5' 플랩 가닥의 변위 및 대체의 효율일 수 있다. 더욱이, 성공적인 3' 플랩 침입 및 5' 플랩의 제거는 이중 가닥 DNA 게놈의 한 가닥 상에만 편집을 혼입시킨다. 편

집의 영구적 설치는 주형으로서 편집된 가닥을 사용하여 비편집된 상보적 DNA 가닥을 대체하기 위한 세포 DNA 복구를 요구한다. 세포는 2차 sgRNA (PE3 시스템)를 사용하여 편집에 인접한 비편집된 가닥에 닉을 도입함으로써 편집된 가닥에 비해 비편집된 가닥의 대체를 유리하게 할 수 있지만 (상기 단계 4), 이 과정은 여전히 DNA 복구의 제2 단계에 의존한다. 이들 DNA 복구 단계는 긴 5' 및 3' 플랩 중간체의 평형을 요구하거나 긴 비-상동 영역, 예컨대 긴 삽입 또는 긴 결실을 함유하는 편집에 특히 비효율적일 수 있다. 프라임 편집의 추가의 개발은 관련 기술분야를 진보시킬 것이다.

[0014] 다양한 측면에서, 본 명세서는 두 DNA 가닥을 동시에 편집함으로써 플랩 평형 및 비-편집된 상보적 게놈 DNA 가닥 내로의 편집의 후속 혼입과 연관된 도전과제를 다루는 다중-플랩 프라임 편집 시스템 (예를 들어, 이중 프라임 편집 시스템 및 사중 프라임 편집 시스템 포함)을 기재한다. 예를 들어 이중-플랩 프라임 편집 시스템에서, 2개의 pegRNA는 게놈 부위의 대향 가닥을 표적화하고 편집된 DNA 서열을 함유하는 2개의 상보적 3' 플랩의 합성을 지시하는 데 사용된다 (도 91). 전형적 프라임 편집과는 달리, 상보적인 편집된 가닥이 대신 혼성화에 이용가능하기 때문에, 편집된 DNA 가닥의 쌍 (3' 플랩)이 내인성 게놈 DNA에서 5' 플랩과 직접 경쟁하는 것이 요구되지 않는다. 듀플렉스의 두 가닥이 편집된 DNA로서 합성되기 때문에, 이중-플랩 프라임 편집 시스템은 전형적 프라임 편집에 의해 요구되는 비-편집된 상보적 DNA 가닥의 대체에 대한 필요를 제거한다. 대신, 세포 DNA 복구 기구는 단지 쌍형성된 5' 플랩 (원래 게놈 DNA)만을 절제하고, 쌍형성된 3' 플랩 (편집된 DNA)을 유전자좌 내로 라이게이션시킬 필요가 있다. 따라서, 또한 새로 합성된 DNA 가닥에 게놈 DNA에 상동인 서열을 포함할 필요가 없으므로, 새로운 가닥의 선택적 혼성화를 허용하고 최소 게놈 상동성을 함유하는 편집을 용이하게 한다. DNA의 두 가닥을 절단하는 프라임 편집제의 뉴클레아제-활성 버전은 또한 원래 DNA 서열의 제거를 가속화하는 데 사용될 수 있다. 4개의 pegRNA를 사용하는 사중-플랩 프라임 편집 시스템은 유사한 이점을 제공한다.

[0015] 전형적 프라임 편집과 같이, 다중-플랩 프라임 편집은 핵산 프로그램가능한 DNA 결합 단백질 ("napDNAbp") (이는 폴리머라제와 회합하여 (즉, 융합 단백질의 형태로 또는 달리 napDNAbp와 트랜스로 제공됨) 작동함)을 사용하여 특정된 DNA 부위 내로 새로운 유전자 정보를 직접 기록하는 다재다능한 정확한 게놈 편집 방법이며, 여기서 프라임 편집 시스템은 표적 부위를 특정하고 또한 가이드 RNA 상에 (예를 들어, 5' 또는 3' 단부에 또는 가이드 RNA의 내부 부분에) 조작된 연장부 (DNA 또는 RNA)에 의해 대체 DNA 가닥의 형태로 목적하는 편집의 합성의 주형이 되는 프라임 편집 (PE) 가이드 RNA ("PEgRNA")로 프로그램화된다. 목적하는 편집 (예를 들어, 단일 핵염기 치환)을 함유하는 대체 가닥은 (목적하는 편집을 포함하는 것을 제외하고는) 편집될 표적 부위의 내인성 가닥과 동일한 서열을 공유한다. DNA 복구 및/또는 복제 기구를 통해, 표적 부위의 내인성 가닥은 목적하는 편집을 함유하는 새로 합성된 대체 가닥에 의해 대체된다. 일부 경우에, 프라임 편집은 "검색-및-대체" 게놈 편집 기술로서 생각될 수 있으며, 이는 프라임 편집제가, 본원에 기재된 바와 같이, 편집될 목적하는 표적 부위를 검색하고 위치를 찾아낼 뿐만 아니라, 동시에, 상응하는 표적 부위 내인성 DNA 가닥 대신에 설치된 목적하는 편집을 함유하는 대체 가닥을 코딩하기 때문이다.

[0016] 본 개시내용의 다중-플랩 프라임 편집제는, 부분적으로, 표적-프라이밍된 역전사 (TPRT) 또는 "프라임 편집"의 메카니즘이 (예를 들어, 도 1a-1f의 다양한 실시양태에 도시된 바와 같이) 높은 효율 및 유전적 유연성으로 정밀 CRISPR/Cas-기반 게놈 편집을 수행하기 위해 활용 또는 적합화될 수 있다는 발견에 관한 것이다. TPRT는 이동성 DNA 요소, 예컨대 포유동물 비-LTR 레트로트랜스포손 및 박테리아 그룹 II 인트론에 의해 자연적으로 사용된다[28,29]. 본 발명자들은 본원에서 Cas 단백질-리버스 트랜스크립타제 융합체 또는 관련 시스템을 사용하여 가이드 RNA에 의해 특이적 DNA 서열을 표적화하고, 표적 부위에서 단일 가닥 닉을 생성하고, 닉킹된 DNA를 가이드 RNA와 통합된 조작된 리버스 트랜스크립타제 주형의 역전사를 위한 프라이머로서 사용하였다. 그러나, 상기 개념이 DNA 폴리머라제 성분으로서 리버스 트랜스크립타제를 사용하는 프라임 편집제로 시작되지만, 본원에 기재된 다중-플랩 프라임 편집제는 리버스 트랜스크립타제로 제한되는 것이 아니라, 실질적으로 임의의 DNA 폴리머라제의 사용을 포함할 수 있다. 실제로, 본 출원은 전반에 걸쳐 "리버스 트랜스크립타제"를 사용한 다중-플랩 프라임 편집제를 언급할 수 있지만, 리버스 트랜스크립타제는 다중-플랩 프라임 편집으로 작동할 수 있는 DNA 폴리머라제의 단지 한 유형인 것으로 본원에 제시된다. 따라서, 본 명세서가 "리버스 트랜스크립타제"를 언급할 때마다, 관련 기술분야의 통상의 기술자는 리버스 트랜스크립타제 대신에 임의의 적합한 DNA 폴리머라제가 사용될 수 있다는 것을 인지하여야 한다. 따라서, 한 측면에서, 다중-플랩 프라임 편집제는 Cas9 (또는 등가의 napDNAbp)를 포함할 수 있으며, 이는 그것을 표적 DNA 내의 프로토스페이서 서열의 상보체에 어닐링하는 스페이서 서열을 함유하는 특수화된 가이드 RNA (즉, PEgRNA)와 회합시킴으로써 DNA 서열을 표적화하도록 프로그램화된다. 특수화된 가이드 RNA는 또한 표적 부위에서 상응하는 내인성 DNA 가닥을 대체하는 데 사용되는 목적하는 유전자 변경을 함유하는 DNA의 대체 가닥을 코딩하는 연장부의 형태로 새로운 유전자 정보를 함유한다. PEgRNA

로부터 표적 DNA로 정보를 전달하기 위해, 다중-플랩 프라임 편집의 메카니즘은 DNA의 한 가닥 내의 표적 부위를 닉킹하여 3'-히드록실 기를 노출시키는 것을 수반한다. 이어서, 노출된 3'-히드록실 기를 사용하여 표적 부위 내로 직접 PEgRNA 상의 편집-코딩 연장부의 DNA 중합을 프라이밍할 수 있다. 다양한 실시양태에서, 편집을 함유하는 대체 가닥의 중합을 위한 주형을 제공하는 연장부는 RNA 또는 DNA로부터 형성될 수 있다. RNA 연장부의 경우에, 프라임 편집제의 폴리머라제는 RNA-의존성 DNA 폴리머라제 (예컨대, 리버스 트랜스크립타제)일 수 있다. DNA 연장부의 경우에, 프라임 편집제의 폴리머라제는 DNA-의존성 DNA 폴리머라제일 수 있다.

[0017] 전형적 프라임 편집에서, 본원에 개시된 프라임 편집제에 의해 형성된 새로 합성된 가닥 (즉, 목적하는 편집을 함유하는 대체 DNA 가닥)은 목적하는 뉴클레오티드 변화 (예를 들어, 단일 뉴클레오티드 변화, 결실 또는 삽입 또는 그의 조합)의 포함을 제외하고는 게놈 표적 서열에 상동일 것이다 (즉, 그와 동일한 서열을 가질 것임). DNA의 새로 합성된 (또는 대체) 가닥은 또한 단일 가닥 DNA 플랩으로서 지칭될 수 있으며, 이는 상보적 상동 내인성 DNA 가닥과의 혼성화에 대해 경쟁함으로써 상응하는 내인성 가닥을 대체할 것이다. 특정 실시양태에서, 시스템은 오류-유발 리버스 트랜스크립타제 효소 (예를 들어, Cas9 도메인과의 융합 단백질로서 제공되거나 또는 Cas9 도메인에 대해 트랜스로 제공됨)의 사용과 조합될 수 있다. 오류-유발 리버스 트랜스크립타제 효소는 단일 가닥 DNA 플랩의 합성 동안 변경을 도입할 수 있다. 따라서, 특정 실시양태에서, 오류-유발 리버스 트랜스크립타제를 이용하여 표적 DNA에 뉴클레오티드 변화를 도입할 수 있다. 시스템과 함께 사용되는 오류-유발 리버스 트랜스크립타제에 따라, 변화는 무작위 또는 비-무작위일 수 있다.

[0018] 전형적 프라임 편집에서, 혼성화된 중간체 (내인성 DNA 가닥에 혼성화된 리버스 트랜스크립타제에 의해 합성된 단일 가닥 DNA 플랩을 포함함)의 분해는 생성된 내인성 DNA의 대체 플랩의 제거 (예를 들어, 5' 단부 DNA 플랩 엔도뉴클레아제인 FEN1을 사용함), 합성된 단일 가닥 DNA 플랩의 표적 DNA에 대한 라이게이션, 및 세포 DNA 복구 및/또는 복제 과정의 결과로서 목적하는 뉴클레오티드 변화의 동화를 포함할 수 있다. 주형화된 DNA 합성은 삽입 및 결실을 포함한 임의의 뉴클레오티드의 변형을 위한 단일 뉴클레오티드 정밀도를 제공하기 때문에, 이러한 접근법의 범주는 매우 넓고, 기초 과학 및 요법에서 무수한 적용에 사용될 수 있을 것으로 예측된다.

[0019] 일부 측면에서, 본 명세서는 각각 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 DNA 폴리머라제를 포함하는 프라임 편집제의 쌍을 제공한다. 일부 실시양태에서, 각각의 프라임 편집제는 연장된 가이드 RNA의 존재 하에 표적-프라이밍된 역전사에 의해 게놈 편집을 수행할 수 있다.

[0020] 일부 측면에서, 본 명세서는 각각 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 DNA 폴리머라제를 포함하며, 여기서 DNA 폴리머라제는 napDNAbp와 트랜스로 제공되는 것인 프라임 편집제의 쌍을 제공한다. 다양한 실시양태에서, 각각의 프라임 편집제는 연장된 가이드 RNA의 존재 하에 표적-프라이밍된 역전사에 의해 게놈 편집을 수행할 수 있다.

[0021] 일부 측면에서, 본 명세서는 각각 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 리버스 트랜스크립타제를 포함하는 프라임 편집제의 쌍을 제공한다. 다양한 실시양태에서, 각각의 프라임 편집제는 연장된 가이드 RNA의 존재 하에 표적-프라이밍된 역전사에 의해 게놈 편집을 수행할 수 있다.

[0022] 일부 측면에서, 본 명세서는 각각 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 리버스 트랜스크립타제를 포함하며, 여기서 리버스 트랜스크립타제는 napDNAbp와 트랜스로 제공되는 것인 프라임 편집제의 쌍을 제공한다. 다양한 실시양태에서, 각각의 프라임 편집제는 연장된 가이드 RNA의 존재 하에 표적-프라이밍된 역전사에 의해 게놈 편집을 수행할 수 있다.

[0023] 특정 실시양태에서, napDNAbp는 닉카제 활성을 갖는다. napDNAbp는 또한 Cas9 단백질 또는 그의 기능적 등가물, 예컨대 뉴클레아제 활성 Cas9, 뉴클레아제 불활성 Cas9 (dCas9) 또는 Cas9 닉카제 (nCas9)일 수 있다.

[0024] 특정 실시양태에서, napDNAbp는 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c 및 아르고노트로 이루어진 군으로부터 선택되고, 임의로 닉카제 활성을 갖는다.

[0025] 다른 실시양태에서, 이중 프라임 편집제의 각각의 프라임 편집제는 연장된 가이드 RNA와 복합체화된 경우에 표적 DNA 서열에 결합할 수 있다.

[0026] 또 다른 실시양태에서, 표적 DNA 서열은 표적 가닥 및 상보적 비-표적 가닥을 포함한다.

[0027] 다른 실시양태에서, 연장된 가이드 RNA에 복합체화된 프라임 편집제의 결합은 R-루프를 형성한다. R-루프는 (i) 연장된 가이드 RNA 및 표적 가닥을 포함하는 RNA-DNA 하이브리드 및 (ii) 상보적 비-표적 가닥을 포함할 수 있다.

[0028] 또 다른 실시양태에서, 상보적 비-표적 가닥은 닉킹되어 유리 3' 단부를 갖는 리버스 트랜스크립타제 프라이밍 서열을 형성한다.

[0029] 다양한 실시양태에서, 연장된 가이드 RNA는 (a) 가이드 RNA 및 (b) 가이드 RNA의 5' 또는 3' 단부에 또는 가이드 RNA 내의 분자내 위치에 RNA 연장부를 포함한다. RNA 연장부는 (i) 목적하는 뉴클레오티드 변화를 포함하는 역전사 주형 서열, (ii) 역전사 프라이머 결합 부위 및 (iii) 임의로 링커 서열을 포함할 수 있다. 다양한 실시양태에서, 역전사 주형 서열은 닉 부위에 인접한 내인성 DNA 서열에 상보적인 단일 가닥 DNA 플랩을 코딩할 수 있으며, 여기서 단일 가닥 DNA 플랩은 목적하는 뉴클레오티드 변화를 포함한다.

[0030] 다양한 실시양태에서, RNA 연장부는 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드 또는 적어도 25개 뉴클레오티드 길이이다.

[0031] 또 다른 실시양태에서, 단일 가닥 DNA 플랩은 닉 부위에 인접한 내인성 DNA 서열에 혼성화함으로써, 목적하는 뉴클레오티드 변화를 설치할 수 있다. 또 다른 실시양태에서, 단일 가닥 DNA 플랩은 닉 부위에 인접하고 유리 5' 단부를 갖는 내인성 DNA 서열을 대체한다. 특정 실시양태에서, 대체된 5' 단부를 갖는 내인성 DNA는 세포에 의해 절제된다.

[0032] 다양한 실시양태에서, 단일 가닥 DNA 플랩의 세포 복구는 목적하는 뉴클레오티드 변화의 설치를 유발함으로써, 목적하는 생성물을 형성한다.

[0033] 다양한 다른 실시양태에서, 목적하는 뉴클레오티드 변화는 PAM 서열의 약 -4 내지 +10인 편집 윈도우에 설치된다.

[0034] 또 다른 실시양태에서, 목적하는 뉴클레오티드 변화는 닉 부위의 약 -5 내지 +5 또는 닉 부위의 약 -10 내지 +10 또는 닉 부위의 약 -20 내지 +20 또는 닉 부위의 약 -30 내지 +30 또는 닉 부위의 약 -40 내지 +40 또는 닉 부위의 약 -50 내지 +50 또는 닉 부위의 약 -60 내지 +60 또는 닉 부위의 약 -70 내지 +70 또는 닉 부위의 약 -80 내지 +80 또는 닉 부위의 약 -90 내지 +90 또는 닉 부위의 약 -100 내지 +100 또는 닉 부위의 약 -200 내지 +200인 편집 윈도우에 설치된다.

[0035] 다양한 실시양태에서, 이중 프라임 편집제의 napDNAbp는 각각 서열식별번호(SEQ ID NO): 18의 아미노산 서열을 포함한다. 다양한 다른 실시양태에서, napDNAbp는 서열식별번호: 26-39, 42-61, 75-76, 126, 130, 137, 141, 147, 153, 157, 445, 460, 467 및 482-487 (Cas9); 서열식별번호: 77-86 (CP-Cas9); 서열식별번호: 18-25 및 87-88 (SpCas9); 및 서열식별번호: 62-72 (Cas12) 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함한다.

[0036] 다른 실시양태에서, 개시된 프라임 편집제 및/또는 이중 프라임 편집제의 조성물의 리버스 트랜스크립타제는 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159, 235, 454, 471, 516, 662, 700-716, 739-742 및 766의 아미노산 서열 중 어느 하나를 포함할 수 있다. 또 다른 실시양태에서, 리버스 트랜스크립타제는 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159, 235, 454, 471, 516, 662, 700-716, 739-742 및 766 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함할 수 있다. 이들 서열은, 예를 들어 레트로바이러스 또는 레트로트랜스포손으로부터의 자연 발생 리버스 트랜스크립타제 서열일 수 있고, 서열은 재조합일 수 있다.

[0037] 다양한 다른 실시양태에서, 본원에 개시된 이중 프라임 편집제의 프라임 편집제는 다양한 구조적 구성을 포함할 수 있다. 예를 들어, 프라임 편집제가 융합 단백질로서 제공되는 실시양태에서, 각각의 이중 프라임 편집제 융합 단백질은 구조 NH_2-[napDNAbp]-[리버스 트랜스크립타제]-COOH; 또는 NH_2-[리버스 트랜스크립타제]-[napDNAbp]-COOH를 포함할 수 있으며, 여기서 각각의 경우의 "]-["는 임의적인 링커 서열의 존재를 나타낸다.

[0038] 다양한 실시양태에서, 링커 서열은 서열식별번호: 127, 165-176, 446, 453 및 767-769의 아미노산 서열 또는 서열식별번호: 127, 165-176, 446, 453 및 767-769의 링커 아미노산 서열 중 어느 하나에 대해 적어도 80%, 85% 또는 90% 또는 95% 또는 99% 동일한 아미노산 서열을 포함한다.

[0039] 다양한 실시양태에서, 표적 DNA 내로 혼입되는 목적하는 뉴클레오티드 변화는 단일 뉴클레오티드 변화 (예를 들

어, 전이 또는 전환), 1개 이상의 뉴클레오티드의 삽입 또는 1개 이상의 뉴클레오티드의 결실일 수 있다.

[0040] 특정 경우에, 삽입은 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 200개, 적어도 300개, 적어도 400개 또는 적어도 500개 뉴클레오티드 길이이다.

[0041] 특정의 다른 경우에, 결실은 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 200개, 적어도 300개, 적어도 400개 또는 적어도 500개 뉴클레오티드 길이이다.

[0042] 또 다른 측면에서, 본 개시내용은 가이드 RNA 및 적어도 1개의 RNA 연장부를 포함하는 연장된 가이드 RNA를 제공한다. RNA 연장부는 가이드 RNA의 3' 단부에 위치할 수 있다. 다른 실시양태에서, RNA 연장부는 가이드 RNA의 5'에 위치할 수 있다. 또 다른 실시양태에서, RNA 연장부는 가이드 RNA 내의 분자내 위치에 위치할 수 있지만, 그러나 연장된 부분의 분자내 위치하는 것이 프로토스페이서의 기능을 방해하지 않는 것이 바람직하다.

[0043] 다양한 실시양태에서, 연장된 가이드 RNA는 napDNAbp에 결합할 수 있고 napDNAbp를 표적 DNA 서열로 지시할 수 있다. 표적 DNA 서열은 표적 가닥 및 상보적 비-표적 가닥을 포함할 수 있으며, 여기서 가이드 RNA는 표적 가닥에 혼성화하여 RNA-DNA 하이브리드 및 R-루프를 형성한다.

[0044] 연장된 가이드 RNA의 다양한 실시양태에서, 적어도 1개의 RNA 연장부는 역전사 주형 서열을 포함할 수 있다. 다양한 다른 실시양태에서, RNA 연장부는 역전사 프라이머 결합 부위를 추가로 포함할 수 있다. 추가 실시양태에서, RNA 연장부는 RNA 연장부를 가이드 RNA에 연결하는 링커 또는 스페이서 서열을 포함할 수 있다.

[0045] 다양한 실시양태에서, RNA 연장부는 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 150개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드 또는 적어도 500개 뉴클레오티드 길이일 수 있다.

[0046] 다른 실시양태에서, 역전사 주형 서열은 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드 또는 적어도 500개 뉴클레오티드 길이이다.

[0047] 또 다른 실시양태에서, 역전사 프라이머 결합 부위 서열은 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드 또는 적어도 500개 뉴클레오

티드 길이이다.

[0048] 다른 실시양태에서, 임의적인 링커 또는 스페이서 서열은 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드 또는 적어도 500개 뉴클레오티드 길이이다.

[0049] 연장된 가이드 RNA의 다양한 실시양태에서, 역전사 주형 서열은 닉 부위에 인접한 내인성 DNA 서열에 상보적인 단일 가닥 DNA 플랩을 코딩할 수 있으며, 여기서 단일 가닥 DNA 플랩은 목적하는 뉴클레오티드 변화를 포함한다. 단일 가닥 DNA 플랩은 닉 부위에서 내인성 단일 가닥 DNA를 대체할 수 있다. 닉 부위에서 대체된 내인성 단일 가닥 DNA는 5' 단부를 가질 수 있고, 내인성 플랩을 형성할 수 있으며, 이는 세포에 의해 절제될 수 있다. 다양한 실시양태에서, 5' 단부 내인성 플랩의 절제는 생성물 형성을 구동하는 것을 도울 수 있으며, 이는 5' 단부 내인성 플랩의 제거가 단일 가닥 3' DNA 플랩의 상응하는 상보적 DNA 가닥에 대한 혼성화 및 단일 가닥 3' DNA 플랩에 의해 운반된 목적하는 뉴클레오티드 변화의 표적 DNA 내로의 혼입 또는 동화를 촉진하기 때문이다.

[0050] 연장된 가이드 RNA의 다양한 실시양태에서, 단일 가닥 DNA 플랩의 세포 복구는 목적하는 뉴클레오티드 변화의 설치를 유발함으로써, 목적하는 생성물을 형성한다.

[0051] 특정 실시양태에서, PEgRNA는 서열식별번호: 101-104, 181-183, 223-234, 237-244, 277, 324-330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 394, 429-442, 499-505, 641-649, 678-692, 735-736, 757-761, 776-777, 2997-3103, 3113-3121, 3305-3455, 3479-3493, 3522-3540, 3549-3556, 3628-3698, 3755-3810, 3874, 3890-3901, 3905-3911, 3913-3929 및 3972-3989의 뉴클레오티드 서열 또는 서열식별번호: 101-104, 181-183, 223-234, 237-244, 277, 324-330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 394, 429-442, 499-505, 641-649, 678-692, 735-736, 757-761, 776-777, 2997-3103, 3113-3121, 3305-3455, 3479-3493, 3522-3540, 3549-3556, 3628-3698, 3755-3810, 3874, 3890-3901, 3905-3911, 3913-3929 및 3972-3989 중 어느 하나와 적어도 85% 또는 적어도 90% 또는 적어도 95% 또는 적어도 98% 또는 적어도 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함한다.

[0052] 본 발명의 또 다른 측면에서, 본 명세서는 본원에 기재된 프라임 편집제 및 상기 기재된 임의의 연장된 가이드 RNA를 포함하는 복합체를 제공한다.

[0053] 본 발명의 또 다른 측면에서, 본 명세서는 napDNAbp 및 연장된 가이드 RNA를 포함하는 복합체를 제공한다. napDNAbp는 Cas9 닉카제일 수 있거나, 또는 서열식별번호: 42-57 (Cas9 닉카제) 및 65 (AsCas12a 닉카제)의 아미노산 서열 또는 서열식별번호: 42-57 (Cas9 닉카제) 및 65 (AsCas12a 닉카제) 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열일 수 있다.

[0054] 복합체를 수반하는 다양한 실시양태에서, 연장된 가이드 RNA는 napDNAbp를 표적 DNA 서열로 지시할 수 있다. 다양한 실시양태에서, 리버스 트랜스크립타제는 트랜스로 제공될 수 있으며, 즉 복합체 자체와 상이한 공급원으로부터 제공될 수 있다. 예를 들어, 리버스 트랜스크립타제는 복합체를 갖는 동일한 세포에 리버스 트랜스크립타제를 별개로 코딩하는 별개의 벡터를 도입함으로써 제공될 수 있다.

[0055] 또 다른 측면에서, 본 개시내용은 제1 및 제2 프라임 편집제 복합체를 포함하며, 각각의 복합체는 프라임 편집제 및 프라임 편집 가이드 RNA (PEgRNA)를 포함하는 것인 시스템을 제공한다. 일부 실시양태에서, 각각의 프라임 편집제는 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하고, 각각의 PEgRNA는 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 특정 실시양태에서, 각각의 DNA 합성 주형은 편집된 부분을 포함하는 단일 가닥 DNA 서열을 코딩한다. 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대해 상보적일 수 있고, 편집될 표적 부위 내로 통합되는 듀플렉스를 형성할 수 있다. 일부 실시양태에서, 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대한 상보성 영역을 포함

할 수 있다. 특정 실시양태에서, 코딩된 2개의 단일 가닥 DNA 서열은 적어도 2 bp, 적어도 3 bp, 적어도 4 bp, 적어도 5 bp, 적어도 10 bp, 적어도 20 bp, 적어도 30 bp, 적어도 40 bp, 적어도 50 bp, 적어도 100 bp, 적어도 200 bp, 적어도 300 bp, 적어도 400 bp, 적어도 500 bp, 적어도 600 bp, 적어도 700 bp, 적어도 800 bp, 적어도 900 bp 또는 적어도 1000 bp 길이인 서로에 대한 상보성 영역을 포함할 수 있다. 일부 실시양태에서, 프라임 편집제는 융합 단백질로서 제공된다. 특정 실시양태에서, 프라임 편집제의 성분 (즉, napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드)은 트랜스로 제공된다.

[0056] 또 다른 측면에서, 본 개시내용은 제1, 제2, 제3 및 제4 프라임 편집제 복합체를 포함하며, 각각의 복합체는 프라임 편집제 및 프라임 편집 가이드 RNA (PEgRNA)를 포함하는 것인 시스템을 제공한다. 일부 실시양태에서, 각각의 프라임 편집제는 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하고, 각각의 PEgRNA는 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 특정 실시양태에서, 각각의 DNA 합성 주형은 편집된 부분을 포함하는 단일 가닥 DNA 서열을 코딩한다. 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대해 상보적일 수 있고, 편집될 표적 부위 내로 통합되는 듀플렉스를 형성할 수 있다. 일부 실시양태에서, 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대한 상보성 영역을 포함할 수 있다. 특정 실시양태에서, 코딩된 2개의 단일 가닥 DNA 서열은 적어도 2 bp, 적어도 3 bp, 적어도 4 bp, 적어도 5 bp, 적어도 10 bp, 적어도 20 bp, 적어도 30 bp, 적어도 40 bp, 적어도 50 bp, 적어도 100 bp, 적어도 200 bp, 적어도 300 bp, 적어도 400 bp, 적어도 500 bp, 적어도 600 bp, 적어도 700 bp, 적어도 800 bp, 적어도 900 bp 또는 적어도 1000 bp 길이인 서로에 대한 상보성 영역을 포함할 수 있다. 일부 실시양태에서, 프라임 편집제는 융합 단백질로서 제공된다. 특정 실시양태에서, 프라임 편집제의 성분 (즉, napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드)은 트랜스로 제공된다.

[0057] 일부 실시양태에서, 각각의 napDNAbp는 Cas9 도메인 또는 그의 변이체이다. 일부 실시양태에서, 각각의 napDNAbp는 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인 또는 Cas9 닉카제 도메인 또는 그의 변이체이다. 특정 실시양태에서, 각각의 napDNAbp는 독립적으로 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c 및 아르고노트로 이루어진 군으로부터 선택되고, 임의로 닉카제 활성을 갖는다. 다양한 실시양태에서, 각각의 napDNAbp는 서열식별번호: 2-65 중 어느 하나의 아미노산 서열 또는 서열식별번호: 2-65 중 어느 하나에 대해 적어도 80%, 85%, 90%, 95% 또는 99% 동일한 아미노산 서열을 포함한다.

[0058] 일부 실시양태에서, RNA-의존성 DNA 폴리머라제 활성을 포함하는 폴리펩티드는 리버스 트랜스크립타제이다. 특정 실시양태에서, RNA-의존성 DNA 폴리머라제 활성을 포함하는 폴리펩티드는 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나의 아미노산 서열 또는 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함한다.

[0059] 일부 실시양태에서, 각각의 프라임 편집제는 napDNAbp와 리버스 트랜스크립타제를 연결하는 링커를 포함할 수 있다. 특정 실시양태에서, 링커는 서열식별번호: 119-128 중 어느 하나의 아미노산 서열 또는 서열식별번호: 119-128 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함한다. 각각의 링커는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이일 수 있다.

[0060] 일부 실시양태에서, 각각의 PEgRNA는 독립적으로 서열식별번호: 192-203 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 192-203 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함할 수 있다.

[0061] 다양한 실시양태에서, 각각의 PEgRNA의 각각의 스페이서 서열은 편집될 표적 부위에 인접한 이중 가닥 DNA 서열의 특이적 결합 부위에 결합할 수 있다. 일부 실시양태에서, 하나의 프라임 편집제 복합체의 스페이서 서열이 이중 가닥 DNA 서열의 한 가닥에 결합하는 것 및 또 다른 프라임 편집제 복합체의 스페이서 서열이 이중 가닥 DNA 서열의 대향 가닥에 결합하는 것은 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 두 DNA 가닥의 닉킹을 발생시킨다.

[0062] 또 다른 측면에서, 본 개시내용은 폴리뉴클레오티드를 제공한다. 일부 실시양태에서, 폴리뉴클레오티드는 본원에 기재된 임의의 복합체를 코딩할 수 있다. 특정 실시양태에서, 폴리뉴클레오티드는 본원에 기재된 임의의 PEgRNA를 코딩할 수 있다.

[0063] 또 다른 측면에서, 본 명세서는 폴리뉴클레오티드를 제공한다. 특정 실시양태에서, 폴리뉴클레오티드는 본원에

개시된 임의의 프라임 편집제를 코딩할 수 있다. 특정의 다른 실시양태에서, 폴리뉴클레오티드는 본원에 개시된 임의의 napDNAbp를 코딩할 수 있다. 추가 실시양태에서, 폴리뉴클레오티드는 본원에 개시된 임의의 리버스 트랜스크립타제를 코딩할 수 있다. 또 다른 실시양태에서, 폴리뉴클레오티드는 본원에 개시된 임의의 연장된 가이드 RNA, 임의의 역전사 주형 서열 또는 임의의 역전사 프라이머 부위 또는 임의의 임의적인 링커 서열을 코딩할 수 있다.

[0064] 또 다른 측면에서, 본 명세서는 본원에 기재된 폴리뉴클레오티드를 포함하는 벡터를 제공한다. 따라서, 특정 실시양태에서, 벡터는 napDNAbp 및 리버스 트랜스크립타제를 포함하는 프라임 편집제 (즉, 융합 단백질로서 또는 트랜스로 발현됨)를 코딩하기 위한 폴리뉴클레오티드를 포함한다. 특정 실시양태에서, 벡터는 본원에 기재된 임의의 복합체를 코딩하기 위한 폴리뉴클레오티드를 포함한다. 다른 실시양태에서, 벡터는 napDNAbp 및 리버스 트랜스크립타제를 개별적으로 코딩하는 폴리뉴클레오티드를 포함한다. 또 다른 실시양태에서, 벡터는 연장된 가이드 RNA를 코딩하는 폴리뉴클레오티드를 포함할 수 있다. 다양한 실시양태에서, 벡터는 napDNAbp, 리버스 트랜스크립타제 및 동일한 또는 별개의 벡터 상의 연장된 가이드 RNA를 코딩하는 1개 이상의 폴리뉴클레오티드를 포함할 수 있다. 일부 실시양태에서, 벡터는 본원에 기재된 임의의 pegRNA를 코딩하기 위한 폴리뉴클레오티드를 포함한다.

[0065] 또 다른 측면에서, 본 명세서는 본원에 기재된 바와 같은 프라임 편집제 및 연장된 가이드 RNA를 포함하는 세포를 제공한다. 세포는 프라임 편집제, napDNAbp, 리버스 트랜스크립타제 및 연장된 가이드 RNA를 포함하는 벡터로 형질전환될 수 있다. 이들 유전 요소는 동일한 벡터 상에 또는 상이한 벡터 상에 포함될 수 있다. 일부 실시양태에서, 세포는 본원에 기재된 임의의 시스템 또는 복합체를 포함한다. 세포는 본원에 개시된 임의의 시스템, 복합체 및/또는 pegRNA를 코딩하는 폴리뉴클레오티드 또는 본원에 개시된 임의의 시스템, 복합체 또는 pegRNA를 코딩하는 폴리뉴클레오티드를 포함하는 벡터로 형질전환될 수 있다.

[0066] 또 다른 측면에서, 본 명세서는 제약 조성물을 제공한다. 특정 실시양태에서, 제약 조성물은 napDNAbp, 프라임 편집제, 리버스 트랜스크립타제 및 연장된 가이드 RNA 중 1종 이상을 포함한다. 특정 실시양태에서, 제약 조성물은 본원에 기재된 임의의 시스템 및/또는 복합체를 포함한다. 특정 실시양태에서, 제약 조성물은 본원에 기재된 임의의 프라임 편집제, 시스템 또는 복합체 및 제약상 허용되는 부형제를 포함한다. 다른 실시양태에서, 제약 조성물은 본원에 기재된 임의의 연장 가이드 RNA 및 제약상 허용되는 부형제를 포함한다. 또 다른 실시양태에서, 제약 조성물은 본원에 기재된 임의의 프라임 편집제와 조합된 본원에 기재된 임의의 연장 가이드 RNA 및 제약상 허용되는 부형제를 포함한다. 또 다른 실시양태에서, 제약 조성물은 napDNAbp, 프라임 편집제, 리버스 트랜스크립타제 및 연장된 가이드 RNA 중 1종 이상을 코딩하는 임의의 폴리뉴클레오티드 서열 또는 본원에 개시된 임의의 벡터를 포함한다. 또 다른 실시양태에서, 본원에 개시된 다양한 성분은 1종 이상의 제약 조성물 내로 분리될 수 있다. 예를 들어, 제1 제약 조성물은 프라임 편집제 또는 napDNAbp를 포함할 수 있고, 제2 제약 조성물은 리버스 트랜스크립타제를 포함할 수 있고, 제3 제약 조성물은 연장된 가이드 RNA를 포함할 수 있다.

[0067] 추가 측면에서, 본 개시내용은 키트를 제공한다. 한 실시양태에서, 키트는 프라임 편집제, napDNAbp, 리버스 트랜스크립타제 및 연장된 가이드 RNA를 포함한 1종 이상의 성분을 코딩하는 1종 이상의 폴리뉴클레오티드를 포함한다. 키트는 또한 벡터, 세포 및 본원에 개시된 임의의 프라임 편집제, napDNAbp 또는 리버스 트랜스크립타제를 포함한 폴리펩티드의 단리된 제제를 포함할 수 있다.

[0068] 또 다른 측면에서, 본 개시내용은 표적 부위에서 이중 가닥 DNA 서열의 두 상보적 가닥을 동시에 편집하기 위해 본원에 기재된 임의의 시스템을 사용하는 방법을 포함한, 개시된 물질의 조성물을 사용하는 방법을 제공한다. 일부 실시양태에서, 방법은 이중 가닥 DNA 서열을 본원에 개시된 임의의 시스템과 접촉시키는 것을 포함한다.

[0069] 한 측면에서, 본 개시내용은 표적 부위에서 이중 가닥 DNA 서열을 제1 및 제2 프라임 편집제 복합체와 접촉시키는 것을 포함하며, 각각의 복합체는 프라임 편집제 및 프라임 편집 가이드 RNA (PEgRNA)를 포함하는 것인 방법을 제공한다. 일부 실시양태에서, 각각의 프라임 편집제는 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하고, 각각의 PEgRNA는 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 일부 실시양태에서, 각각의 프라임 편집제는 융합 단백질로서 제공된다. 일부 실시양태에서, 프라임 편집제의 성분은 트랜스로 제공된다. 특정 실시양태에서, 각각의 DNA 합성 주형은 편집된 부분을 포함하는 단일 가닥 DNA 서열을 코딩한다. 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대해 상보적일 수 있고, 편집될 표적 부위 내로 통합되는 듀플렉스를 형성할 수 있다. 프라임 편집제 복합체의 다양한 요소는 본원에 개시된 시스템의 임의의 실시양태를 포함할 수 있다.

[0070] 또 다른 측면에서, 본 개시내용은 표적 부위에서 이중 가닥 DNA 서열을 제1, 제2, 제3 및 제4 프라임 편집제 복합체와 접촉시키는 것을 포함하며, 각각의 복합체는 프라임 편집제 및 프라임 편집 가이드 RNA (PEgRNA)를 포함하는 것인 방법을 제공한다. 일부 실시양태에서, 각각의 프라임 편집제는 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하고, 각각의 PEgRNA는 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 일부 실시양태에서, 각각의 프라임 편집제는 융합 단백질로서 제공된다. 일부 실시양태에서, 프라임 편집제의 성분은 트랜스로 제공된다. 특정 실시양태에서, 각각의 DNA 합성 주형은 단일 가닥 DNA 서열을 코딩한다. 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대해 상보적일 수 있고, 편집될 표적 부위 내로 통합되는 듀플렉스를 형성할 수 있다. 프라임 편집제 복합체의 다양한 요소는 본원에 개시된 시스템의 임의의 실시양태를 포함할 수 있다.

[0071] 일부 실시양태에서, 본원에 제공된 방법은 표적 DNA 서열의 역위를 허용한다. 일부 실시양태에서, 제1 DNA 합성 주형에 의해 코딩된 제1 단일 가닥 DNA 서열 및 제2 DNA 합성 주형에 의해 코딩된 제2 단일 가닥 DNA 서열은 표적 DNA 서열의 대향 단부 상에 있고, 제3 DNA 합성 주형에 의해 코딩된 제3 단일 가닥 DNA 서열 및 제4 DNA 합성 주형에 의해 코딩된 제4 단일 가닥 DNA 서열은 동일한 표적 DNA 서열의 대향 단부 상에 있다.

[0072] 일부 실시양태에서, 본원에 제공된 방법은 원형 DNA 공여자를 제공하는 것을 추가로 포함한다. 특정 실시양태에서, 제1 DNA 합성 주형에 의해 코딩된 제1 단일 가닥 DNA 서열 및 제3 DNA 합성 주형에 의해 코딩된 제3 단일 가닥 DNA 서열은 표적 DNA 서열의 대향 단부 상에 있고, 제2 DNA 합성 주형에 의해 코딩된 제2 단일 가닥 DNA 서열 및 제4 DNA 합성 주형에 의해 코딩된 제4 단일 가닥 DNA 서열은 원형 DNA 공여자 상에 있다. 일부 실시양태에서, 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 원형 DNA 공여자의 부분은 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 표적 DNA 서열을 대체한다.

[0073] 일부 실시양태에서, 본원에 제공된 방법은 제1 핵산 분자 (예를 들어, 제1 염색체)로부터 제2 핵산 분자 (예를 들어, 제2 염색체)로의 표적 DNA 서열의 전위를 허용한다. 일부 실시양태에서, 제1 DNA 합성 주형에 의해 코딩된 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 합성 주형에 의해 코딩된 제3 단일 가닥 DNA 서열은 제1 핵산 분자 상에 있고, 제2 DNA 합성 주형에 의해 코딩된 제2 단일 가닥 DNA 서열 및 제4 DNA 합성 주형에 의해 코딩된 제4 단일 가닥 DNA 서열은 제2 핵산 분자 상에 있다. 특정 실시양태에서, 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 제1 핵산 분자의 부분은 제2 핵산 분자 내로 혼입된다. 특정 실시양태에서, 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 제2 핵산 분자의 부분은 제1 핵산 분자 내로 혼입된다.

[0074] 또 다른 측면에서, 본 개시내용은 다중-플랩 프라임 편집에 사용하기 위한 PEgRNA의 쌍을 제공한다. 일부 실시양태에서, 쌍은 제1 PEgRNA 및 제2 PEgRNA를 포함하고, 각각의 PEgRNA는 독립적으로 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 특정 실시양태에서, 각각의 DNA 합성 주형은 단일 가닥 DNA 서열을 코딩한다. 다양한 실시양태에서, 다중-플랩 프라임 편집제는 별개의 프라임 편집제를 표적 부위의 어느 한 측에 표적화하는 PEgRNA의 쌍과 관련하여 사용되며, 여기서 PEgRNA의 쌍은 각각 서로의 역 상보체인 핵산 서열을 포함하는 3' 핵산 플랩을 코딩한다. 다양한 실시양태에서, 역 상보체 서열을 포함하는 3' 플랩은 서로 어닐링하여, 목적하는 편집 또는 PEgRNA를 코딩하는 핵산 서열을 포함하는 듀플렉스를 형성할 수 있다. 이어서, 듀플렉스는 인접한 닉 부위 사이에 위치한 상응하는 내인성 듀플렉스의 대체에 의해 표적 부위 내로 통합된다.

[0075] 또 다른 측면에서, 본 개시내용은 다중-플랩 프라임 편집에 사용하기 위한 복수의 PEgRNA를 제공한다. 일부 실시양태에서, 복수는 제1, 제2, 제3 및 제4 PEgRNA를 포함한다. 일부 실시양태에서, 각각의 4개의 PEgRNA는 독립적으로 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 특정 실시양태에서, 각각의 DNA 합성 주형은 단일 가닥 DNA 서열을 코딩한다. 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대해 상보적일 수 있다.

[0076] 다양한 측면에서, 본 개시내용은 본원에 기재된 임의의 쌍 또는 복수의 PEgRNA를 코딩하는 폴리뉴클레오티드를 제공한다. 특정 측면에서, 본 개시내용은 본원에 기재된 임의의 쌍 또는 복수의 PEgRNA를 코딩하는 폴리뉴클레오티드를 코딩하는 벡터를 제공한다. 또 다른 측면에서, 본 개시내용은 본원에 기재된 임의의 쌍 또는 복수의 PEgRNA를 코딩하는 폴리뉴클레오티드를 코딩하는 벡터를 포함하는 세포를 제공한다. 다른 측면에서, 본 개시내용은 본원에 기재된 임의의 쌍 또는 복수의 PEgRNA, 본원에 기재된 임의의 쌍 또는 복수의 PEgRNA를 코딩하는 벡터, 또는 본원에 기재된 임의의 쌍 또는 복수의 PEgRNA를 코딩하는 벡터를 포함하는 세포를 포함하는 제약 조성물을 제공한다. 특정 실시양태에서, 제약 조성물은 제약 부형제를 포함한다.

[0077] 한 실시양태에서, 방법은 이중 가닥 DNA 서열에 목적하는 뉴클레오티드 변화를 설치하는 방법에 관한 것이다. 방법은 먼저 이중 가닥 DNA 서열을 프라임 편집제 및 연장된 가이드 RNA를 포함하는 복합체와 접촉시키는 것을

포함하며, 여기서 프라임 편집제는 napDNAbp 및 리버스 트랜스크립타제를 포함하고, 여기서 연장된 가이드 RNA는 목적하는 뉴클레오티드 변화를 포함하는 역전사 주형 서열을 포함한다. 일부 실시양태에서, 각각의 프라임 편집제는 융합 단백질로서 제공된다. 일부 실시양태에서, 프라임 편집제의 성분은 트랜스로 제공된다. 이어서, 방법은 비-표적 가닥 상의 이중 가닥 DNA 서열을 닉킹함으로써 3' 단부를 갖는 유리 단일 가닥 DNA를 생성하는 것을 포함한다. 이어서, 방법은 유리 단일 가닥 DNA의 3' 단부를 역전사 주형 서열에 혼성화시킴으로써 리버스 트랜스크립타제 도메인을 프라이밍하는 것을 포함한다. 이어서, 방법은 3' 단부로부터 DNA의 가닥을 중합시킴으로써 목적하는 뉴클레오티드 변화를 포함하는 단일 가닥 DNA 플랩을 생성하는 것을 포함한다. 이어서, 방법은 절단 부위에 인접한 내인성 DNA 가닥을 단일 가닥 DNA 플랩으로 대체함으로써, 이중 가닥 DNA 서열에 목적하는 뉴클레오티드 변화를 설치하는 것을 포함한다.

[0078] 다른 실시양태에서, 본 개시내용은 DNA 분자를 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 napDNAbp를 표적 유전자좌로 표적화하는 가이드 RNA와 접촉시키는 것을 포함하며, 여기서 가이드 RNA는 적어도 1개의 목적하는 뉴클레오티드 변화를 포함하는 리버스 트랜스크립타제 (RT) 주형 서열을 포함하는 것인, 표적 유전자좌에서 DNA 분자의 뉴클레오티드 서열에 1개 이상의 변화를 도입하는 방법을 제공한다. 이어서, 방법은 표적 유전자좌에서 DNA 가닥에 노출된 3' 단부를 형성한 다음, 노출된 3' 단부를 RT 주형 서열에 혼성화시켜 역전사를 프라이밍하는 것을 포함한다. 이어서, RT 주형 서열에 기초하여 적어도 1개의 목적하는 뉴클레오티드 변화를 포함하는 단일 가닥 DNA 플랩이 리버스 트랜스크립타제에 의해 합성되거나 중합된다. 마지막으로, 적어도 1개의 목적하는 뉴클레오티드 변화가 상응하는 내인성 DNA 내로 혼입되어, 표적 유전자좌에서 DNA 분자의 뉴클레오티드 서열에 1개 이상의 변화를 도입한다.

[0079] 또 다른 실시양태에서, 본 개시내용은 (a) 표적 유전자좌에서 DNA 분자를 (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 리버스 트랜스크립타제를 포함하는 프라임 편집제 및 (ii) 목적하는 뉴클레오티드 변화를 포함하는 RT 주형을 포함하는 가이드 RNA와 접촉시키는 단계; (b) RT 주형의 표적-프라이밍된 역전사를 수행하여 목적하는 뉴클레오티드 변화를 포함하는 단일 가닥 DNA를 생성하는 단계; 및 (c) 목적하는 뉴클레오티드 변화를 DNA 복구 및/또는 복제 과정을 통해 표적 유전자좌에서 DNA 분자 내로 혼입시키는 단계를 포함하는, 표적-프라이밍된 역전사에 의해 표적 유전자좌에서 DNA 분자의 뉴클레오티드 서열에 1개 이상의 변화를 도입하는 방법을 제공한다.

[0080] 특정 실시양태에서, 내인성 DNA 가닥을 대체하는 단계는 (i) 단일 가닥 DNA 플랩을 절단 부위에 인접한 내인성 DNA 가닥에 혼성화시켜 서열 미스매치를 생성시키는 단계; (ii) 내인성 DNA 가닥을 절제하는 단계; 및 (iii) 미스매치를 복구시켜 DNA의 두 가닥에 목적하는 뉴클레오티드 변화를 포함하는 목적하는 생성물을 형성하는 단계를 포함한다.

[0081] 다양한 실시양태에서, 목적하는 뉴클레오티드 변화는 단일 뉴클레오티드 치환 (예를 들어 전이 또는 전환 변화), 결실 또는 삽입일 수 있다. 예를 들어, 목적하는 뉴클레오티드 변화는 (1) G에서 T로의 치환, (2) G에서 A로의 치환, (3) G에서 C로의 치환, (4) T에서 G로의 치환, (5) T에서 A로의 치환, (6) T에서 C로의 치환, (7) C에서 G로의 치환, (8) C에서 T로의 치환, (9) C에서 A로의 치환, (10) A에서 T로의 치환, (11) A에서 G로의 치환 또는 (12) A에서 C로의 치환일 수 있다.

[0082] 다른 실시양태에서, 목적하는 뉴클레오티드 변화는 (1) G:C 염기쌍을 T:A 염기쌍으로, (2) G:C 염기쌍을 A:T 염기쌍으로, (3) G:C 염기쌍을 C:G 염기쌍으로, (4) T:A 염기쌍을 G:C 염기쌍으로, (5) T:A 염기쌍을 A:T 염기쌍으로, (6) T:A 염기쌍을 C:G 염기쌍으로, (7) C:G 염기쌍을 G:C 염기쌍으로, (8) C:G 염기쌍을 T:A 염기쌍으로, (9) C:G 염기쌍을 A:T 염기쌍으로, (10) A:T 염기쌍을 T:A 염기쌍으로, (11) A:T 염기쌍을 G:C 염기쌍으로 또는 (12) A:T 염기쌍을 C:G 염기쌍으로 전환시킬 수 있다.

[0083] 또 다른 실시양태에서, 방법은 삽입인 목적하는 뉴클레오티드 변화를 도입한다. 특정 경우에, 삽입은 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 200개, 적어도 300개, 적어도 400개 또는 적어도 500개 뉴클레오티드 길이이다.

[0084] 다른 실시양태에서, 방법은 결실인 목적하는 뉴클레오티드 변화를 도입한다. 특정의 다른 경우에, 결실은 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개,

적어도 18개, 적어도 19개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 200개, 적어도 300개, 적어도 400개 또는 적어도 500개 뉴클레오티드 길이이다.

[0085] 다양한 실시양태에서, 목적하는 뉴클레오티드 변화는 질환-연관 유전자를 교정한다. 질환-연관 유전자는 아데노신 데아미나제 (ADA) 결핍; 알파-1 항트립신 결핍; 낭성 섬유증; 뒤시엔느 근육 이영양증; 갈락토스혈증; 혈색소증; 헌팅톤병; 메이플 시럽뇨 질환; 마르팡 증후군; 제1형 신경섬유종증; 선천성 손발톱비대증; 페닐케톤뇨; 중증 복합 면역결핍; 겸상 적혈구 질환; 스미스-렘리-오피츠 증후군; 및 테이-삭스병으로 이루어진 군으로부터 선택된 단일유전자 장애와 연관될 수 있다. 다른 실시양태에서, 질환-연관 유전자는 심장 질환; 고혈압; 알츠하이머병; 관절염; 당뇨병; 암; 및 비만으로 이루어진 군으로부터 선택된 다유전자 장애와 연관될 수 있다.

[0086] 본원에 개시된 방법은 뉴클레아제 기능상실 Cas9 (dCas9), Cas9 닉카제 (nCas9) 또는 뉴클레아제 활성 Cas9인 napDNAbp를 갖는 융합 단백질을 포함할 수 있다. 다른 실시양태에서, napDNAbp 및 리버스 트랜스크립타제는 단일 융합 단백질로서 코딩되기 보다는, 별개의 구축물로 제공될 수 있다. 따라서, 일부 실시양태에서, 리버스 트랜스크립타제는 napDNAbp에 대해 트랜스로 (융합 단백질로 보다는) 제공될 수 있다.

[0087] 방법을 수반하는 다양한 실시양태에서, napDNAbp는 서열식별번호: 26-61, 75-76, 126, 130, 137, 141, 147, 153, 157, 445, 460, 467 및 482-487 (Cas9); (SpCas9); 서열식별번호: 77-86 (CP-Cas9); 서열식별번호: 18-25 및 87-88 (SpCas9); 및 서열식별번호: 62-72 (Cas12)의 아미노산 서열을 포함할 수 있다. napDNAbp는 또한 서열식별번호: 26-61, 75-76, 126, 130, 137, 141, 147, 153, 157, 445, 460, 467 및 482-487 (Cas9); (SpCas9); 서열식별번호: 77-86 (CP-Cas9); 서열식별번호: 18-25 및 87-88 (SpCas9); 및 서열식별번호: 62-72 (Cas12) 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함할 수 있다.

[0088] 방법을 수반하는 다양한 실시양태에서, 리버스 트랜스크립타제는 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159, 235, 454, 471, 516, 662, 700-716, 739-742 및 766의 아미노산 서열 중 어느 하나를 포함할 수 있다. 리버스 트랜스크립타제는 또한 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159, 235, 454, 471, 516, 662, 700-716, 739-742 및 766 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함할 수 있다.

[0089] 방법은 서열식별번호: 101-104, 181-183, 223-234, 237-244, 277, 324-330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 394, 429-442, 499-505, 641-649, 678-692, 735-736, 757-761, 776-777, 2997-3103, 3113-3121, 3305-3455, 3479-3493, 3522-3540, 3549-3556, 3628-3698, 3755-3810, 3874, 3890-3901, 3905-3911, 3913-3929 및 3972-3989의 뉴클레오티드 서열 또는 그와 적어도 80% 또는 적어도 85% 또는 적어도 90% 또는 적어도 95% 또는 적어도 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 PEgRNA의 사용을 수반할 수 있다. 방법은 3' 단부에 RNA 연장부를 포함하는 연장된 가이드 RNA의 사용을 포함할 수 있으며, 여기서 RNA 연장부는 역전사 주형 서열을 포함한다.

[0090] 방법은 5' 단부에 RNA 연장부를 포함하는 연장된 가이드 RNA의 사용을 포함할 수 있으며, 여기서 RNA 연장부는 역전사 주형 서열을 포함한다.

[0091] 방법은 가이드 RNA 내의 분자내 위치에 RNA 연장부를 포함하는 연장된 가이드 RNA의 사용을 포함할 수 있으며, 여기서 RNA 연장부는 역전사 주형 서열을 포함한다.

[0092] 방법은 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 200개, 적어도 300개, 적어도 400개 또는 적어도 500개 뉴클레오티드 길이인 1개 이상의 RNA 연장부를 갖는 연장된 가이드 RNA의 사용을 포함할 수 있다.

[0093] 상기 개념 및 하기 논의된 추가의 개념은, 본 개시내용이 이와 관련하여 제한되지 않기 때문에, 임의의 적합한 조합으로 배열될 수 있는 것으로 인지되어야 한다. 추가로, 본 개시내용의 다른 이점 및 신규 특색은 첨부 도면과 함께 고려되는 경우에 다양한 비제한적 실시양태의 하기 상세한 설명으로부터 분명해질 것이다.

도면의 간단한 설명

[0094] 하기 도면은 본 명세서의 일부를 형성하고, 본 개시내용의 특정 측면을 추가로 입증하기 위해 포함되며, 이는 이들 도면 중 하나 이상을 본원에 제시된 구체적 실시양태의 상세한 설명과 함께 조합하여 참조함으로써 보다 잘 이해될 수 있다.

도 1a는 연장된 가이드 RNA 분자와 복합체를 이룬 Cas9 단백질에 융합된 리버스 트랜스크립타제를 포함하는 융합 단백질을 사용하여 단일 뉴클레오티드 변화 및/또는 삽입 및/또는 결실을 DNA 분자 (예를 들어, 게놈) 내로 도입하기 위한 예시적인 과정의 개략도를 제공한다. 이러한 실시양태에서, 가이드 RNA는 3' 단부에서 연장되어 리버스 트랜스크립타제 주형 서열을 포함한다. 개략도는 가이드 RNA (gRNA)와 복합체를 이룬 Cas9 닉카제에 융합된 리버스 트랜스크립타제 (RT)가 DNA 표적 부위에 결합하고 표적 뉴클레오티드에 인접한 PAM-함유 DNA 가닥을 닉킹하는 방법을 보여준다. RT 효소는 gRNA로부터의 DNA 합성을 위한 프라이머로서 닉킹된 DNA를 사용하며, 이는 목적하는 편집을 코딩하는 새로운 DNA 가닥의 합성을 위한 주형으로서 사용된다. 제시된 편집 과정은 표적-프라이밍된 역전사 편집 (TRT 편집) 또는 동등하게, "프라임 편집"으로 지칭될 수 있다.

도 1b는 도 1a에서와 동일한 표현을 제공하되, 예외로 프라임 편집제 복합체가 보다 일반적으로 [napDNAbp]-[P]:PEgRNA 또는 [P]-[napDNAbp]:PEgRNA로 나타내어지며, 여기서 "P"는 임의의 폴리머라제 (예를 들어, 리버스 트랜스크립타제)를 지칭하고, "napDNAbp"는 핵산 프로그램가능한 DNA 결합 단백질 (예를 들어, SpCas9)을 지칭하고, "PEgRNA"는 프라임 편집 가이드 RNA를 지칭하고, "]-["는 임의적인 링커를 지칭한다. 다른 곳, 예를 들어 도 3a-3g에 기재된 바와 같이, PEgRNA는 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 5' 연장 아암을 포함한다. 제시되지 않았지만, PEgRNA의 연장 아암 (즉, 프라이머 결합 부위 및 DNA 합성 주형을 포함함)은 DNA 또는 RNA일 수 있는 것으로 고려된다. 이 구성에서 고려되는 특정한 폴리머라제는 DNA 합성 주형의 성질에 좌우될 것이다. 예를 들어, DNA 합성 주형이 RNA이면, 폴리머라제는 RNA-의존성 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)이다. DNA 합성 주형이 DNA이면, 폴리머라제는 DNA-의존성 DNA 폴리머라제일 수 있다.

도 1c는 연장된 가이드 RNA 분자와 복합체를 이룬 Cas9 단백질에 융합된 리버스 트랜스크립타제를 포함하는 융합 단백질을 사용하여 단일 뉴클레오티드 변화 및/또는 삽입 및/또는 결실을 DNA 분자 (예를 들어, 게놈) 내로 도입하기 위한 예시적인 과정의 개략도를 제공한다. 이러한 실시양태에서, 가이드 RNA는 5' 단부에서 연장되어 리버스 트랜스크립타제 주형 서열을 포함한다. 개략도는 가이드 RNA (gRNA)와 복합체를 이룬 Cas9 닉카제에 융합된 리버스 트랜스크립타제 (RT)가 DNA 표적 부위에 결합하고 표적 뉴클레오티드에 인접한 PAM-함유 DNA 가닥을 닉킹하는 방법을 보여준다. RT 효소는 gRNA로부터의 DNA 합성을 위한 프라이머로서 닉킹된 DNA를 사용하며, 이는 목적하는 편집을 코딩하는 새로운 DNA 가닥의 합성을 위한 주형으로서 사용된다. 제시된 편집 과정은 표적-프라이밍된 역전사 편집 (TRT 편집) 또는 동등하게, "프라임 편집"으로 지칭될 수 있다.

도 1d는 도 1c에서와 동일한 표현을 제공하되, 예외로 프라임 편집제 복합체가 보다 일반적으로 [napDNAbp]-[P]:PEgRNA 또는 [P]-[napDNAbp]:PEgRNA로 나타내어지며, 여기서 "P"는 임의의 폴리머라제 (예를 들어, 리버스 트랜스크립타제)를 지칭하고, "napDNAbp"는 핵산 프로그램가능한 DNA 결합 단백질 (예를 들어, SpCas9)을 지칭하고, "PEgRNA"는 프라임 편집 가이드 RNA를 지칭하고, "]-["는 임의적인 링커를 지칭한다. 다른 곳, 예를 들어 도 3a-3g에 기재된 바와 같이, PEgRNA는 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 3' 연장 아암을 포함한다. 제시되지 않았지만, PEgRNA의 연장 아암 (즉, 프라이머 결합 부위 및 DNA 합성 주형을 포함함)은 DNA 또는 RNA일 수 있는 것으로 고려된다. 이 구성에서 고려되는 특정한 폴리머라제는 DNA 합성 주형의 성질에 좌우될 것이다. 예를 들어, DNA 합성 주형이 RNA이면, 폴리머라제는 RNA-의존성 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)이다. DNA 합성 주형이 DNA이면, 폴리머라제는 DNA-의존성 DNA 폴리머라제일 수 있다. 다양한 실시양태에서, PEgRNA는 DNA-기반 DNA 합성 주형이 혼입되도록 조작 또는 합성될 수 있다.

도 1e는 합성된 DNA 단일 가닥 (목적하는 뉴클레오티드 변화를 포함함)이 분해되어 목적하는 뉴클레오티드 변화가 DNA 내로 혼입되도록 하는 방법의 예시적인 과정을 도시하는 개략도이다. 제시된 바와 같이, 편집된 가닥 (또는 "돌연변이유발 가닥")의 합성 후에, 내인성 가닥과의 평형, 내인성 가닥의 플랩 절단 및 라이게이션은 미스매칭된 DNA 듀플렉스의 분해 후에 내인성 DNA 복구 및/또는 복제 과정의 작용을 통해 DNA 편집의 혼입을 유도한다.

도 1f는 "대향 가닥 닉킹"이 도 1e의 분해 방법 내로 혼입되어 복귀 생성물에 비해 목적하는 생성물의 형성을 구동하는 것을 도울 수 있음을 보여주는 개략도이다. 대향 가닥 닉킹에서, 제2 Cas9/gRNA 복합체를 사용하여, 초기 닉킹된 가닥으로부터의 대향 가닥 상에 제2 닉을 도입한다. 이는 내인성 세포 DNA 복구 및/또는 복제 과정을 유도하여 비편집된 가닥 (즉, 제2 닉 부위를 함유하는 가닥)을 우선적으로 대체한다.

도 1g는 연장된 가이드 RNA와 복합체화된 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)을 사용하여 단일 뉴

클레오티드 변화 및/또는 삽입 및/또는 결실을 표적 유전자좌의 DNA 분자 (예를 들어, 게놈) 내로 도입하기 위한 예시적인 과정의 또 다른 개략도를 제공한다. 이 과정은 프라임 편집의 실시양태로 지칭될 수 있다. 연장된 가이드 RNA는 가이드 RNA의 3' 또는 5' 단부에 또는 가이드 RNA 내의 분자내 위치에 연장부를 포함한다. 단계 (a)에서, napDNAbp/gRNA 복합체는 DNA 분자와 접촉하고, gRNA는 napDNAbp가 표적 유전자좌에 결합하도록 가이드한다. 단계 (b)에서, 표적 유전자좌의 DNA 가닥 (R-루프 가닥 또는 PAM-함유 가닥 또는 비-표적 DNA 가닥 또는 프로토스페이서 가닥) 중 하나에 닉을 도입하여 (예를 들어, 뉴클레아제 또는 화학적 작용제에 의해), 표적 유전자좌의 가닥 중 하나에 이용가능한 3' 단부를 생성한다. 특정 실시양태에서, 닉은 R-루프 가닥에 상응하는 DNA의 가닥, 즉 가이드 RNA 서열에 혼성화되지 않는 가닥에서 생성된다. 단계 (c)에서, 3' 단부 DNA 가닥은 역전사를 프라이밍하기 위해 가이드 RNA의 연장된 부분과 상호작용한다. 특정 실시양태에서, 3' 단부 DNA 가닥은 가이드 RNA의 연장된 부분 상의 특이적 RT 프라이밍 서열에 혼성화한다. 단계 (d)에서, 프라이밍된 부위의 3' 단부로부터 가이드 RNA의 3' 단부를 향해 DNA의 단일 가닥을 합성하는 리버스 트랜스크립타제가 도입된다. 이는 목적하는 뉴클레오티드 변화 (예를 들어, 단일 염기 변화, 삽입 또는 결실 또는 그의 조합)를 포함하는 단일 가닥 DNA 플랩을 형성한다. 단계 (e)에서, napDNAbp 및 가이드 RNA가 방출된다. 단계 (f) 및 (g)는 목적하는 뉴클레오티드 변화가 표적 유전자좌 내로 혼입되도록 하는 단일 가닥 DNA 플랩의 분해에 관한 것이다. 이러한 과정은 3' 단일 가닥 DNA 플랩이 다른 가닥 상의 상보적 서열에 침입하고 혼성화하였을 때 형성되는 상응하는 5' 내인성 DNA 플랩을 제거함으로써, 목적하는 생성물 형성을 향해 구동될 수 있다. 과정은 또한 도 1f에 예시된 바와 같이 제2 가닥 닉킹에 의해 생성물 형성을 향해 구동될 수 있다. 이 과정은 하기 유전자 변화: 전환, 전이, 결실 및 삽입 중 적어도 하나 이상을 도입할 수 있다.

도 1h는 본원에 기재된 프라임 편집 과정으로 가능한 유전자 변화의 유형을 도시하는 개략도이다. 프라임 편집에 의해 달성가능한 뉴클레오티드 변화의 유형은 결실 (짧은 및 긴 결실 포함), 단일-뉴클레오티드 변화 (전이 및 전환 포함), 역위 및 삽입 (짧은 및 긴 결실 포함)을 포함한다.

도 1i는 PE3b (PE3b = PE2 프라임 편집제 융합 단백질 + PEgRNA + 제2 가닥 닉킹 가이드 RNA)에 의해 예시된 시간적 제2 가닥 닉킹을 도시하는 개략도이다. 시간적 제2 가닥 닉킹은 목적하는 편집된 생성물의 형성을 용이하게 하기 위한 제2 가닥 닉킹의 변형이다. "시간적"이라는 용어는 목적하는 편집이 편집 가닥에 설치된 후에만 비편집 가닥에 대한 제2 가닥 닉이 발생한다는 사실을 지칭한다. 이는 이중 가닥 DNA 파괴로 이어지는 두 가닥 상의 공동 닉을 피한다.

도 1j-1k는 napDNAbp (예를 들어, SpCas9 닉카제)를 임의의 프로그램가능한 뉴클레아제 도메인, 예컨대 아연 핑거 뉴클레아제 (ZFN) 또는 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)로 대체하는, 본원에서 고려되는 프라임 편집의 변형을 도시한다. 따라서, 적합한 뉴클레아제가 반드시 핵산 표적화 분자 (예컨대 가이드 RNA)에 의해 "프로그램화"될 필요는 없고, 오히려 DNA-결합 도메인, 예컨대 특히 뉴클레아제의 특이성을 규정함으로써 프로그램화될 수 있는 것으로 고려된다. napDNAbp 모이어티를 사용한 프라임 편집에서와 같이, 이러한 대안적 프로그램가능한 뉴클레아제는 표적 DNA의 한 가닥만이 절단되도록 변형되는 것이 바람직하다. 다시 말해서, 프로그램가능한 뉴클레아제는 바람직하게는 닉카제로서 기능해야 한다. 프로그램가능한 뉴클레아제 (예를 들어, ZFN 또는 TALEN)가 선택되면, 추가의 기능성을 시스템 내로 조작하여 그가 프라임 편집-유사 메카니즘에 따라 작동하도록 할 수 있다. 예를 들어, 프로그램가능한 뉴클레아제는 RNA 또는 DNA 연장 아암을 그에 (예를 들어, 화학적 링커를 통해) 커플링시킴으로써 변형될 수 있으며, 여기서 연장 아암은 프라이머 결합 부위 (PBS) 및 DNA 합성 주형을 포함한다. 프로그램가능한 뉴클레아제는 또한 폴리머라제에 (예를 들어, 화학적 또는 아미노산 링커를 통해) 커플링될 수 있으며, 이의 성질은 연장 아암이 DNA인지 또는 RNA인지에 좌우될 것이다. RNA 연장 아암의 경우에, 폴리머라제는 RNA-의존성 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)일 수 있다. DNA 연장 아암의 경우에, 폴리머라제는 DNA-의존성 DNA 폴리머라제 (예를 들어, Pol I, Pol II 또는 Pol III를 포함한 원핵 폴리머라제, 또는 Pol a, Pol b, Pol g, Pol d, Pol e 또는 Pol z를 포함한 진핵 폴리머라제)일 수 있다. 시스템은 또한 프로그램가능한 뉴클레아제에 대한 융합체로서 부가되거나 또는 트랜스로 부가되어 반응을 전체로서 용이하게 하기 위한 다른 기능성을 포함할 수 있다 (예를 들어, (a) 절단 부위에서 DNA를 풀어내어 3' 단부를 갖는 절단 가닥을 프라이머로서 이용가능하게 만드는 헬리카제, (b) 절단 가닥 상의 내인성 가닥을 제거하도록 도와 합성된 가닥으로 내인성 가닥을 대체하는 쪽으로 반응을 구동하는 플랩 엔도뉴클레아제 (예를 들어, FEN1), 또는 (c) 대향 가닥 상에 제2 부위 닉을 생성하여, 비-편집된 가닥의 선호되는 세포 복구를 통해 합성 복구의 통합을 구동하는 것을 도울 수 있는 nCas9:gRNA 복합체). napDNAbp에 의한 프라임 편집과 유사한 방식으로, 그 외의 다른 프로그램가능한 뉴클레아제를 갖는 이러한 복합체를 사용하여 관심 편집을 보유하는 DNA의 새로 합성된 대체 가닥을 합성한 다음, DNA의 표적 부위 내로 영구적으로 설치할 수 있다.

도 1l은, 한 실시양태에서, 프라임 편집에 의해 편집될 수 있는 표적 DNA의 해부학적 특색을 도시한다. 표적 DNA는 "비-표적 가닥" 및 "표적 가닥"을 포함한다. 표적-가닥은 PAM 부위 (이 경우에, 정규 SpCas9-기반 프라임 편집제에 의해 인식되는 NGG)를 인식하는 프라임 편집제 복합체의 PEgRNA의 스페이서에 어닐링되는 가닥이다. 표적 가닥은 또한 "비-PAM 가닥" 또는 "비-편집 가닥"으로 지칭될 수 있다. 대조적으로, 비-표적 가닥 (즉, 프로토스페이서 및 NGG의 PAM 서열을 함유하는 가닥)은 "PAM-가닥" 또는 "편집 가닥"으로 지칭될 수 있다. 다양한 실시양태에서, PE 복합체의 닉 부위는 PAM-가닥 상의 프로토스페이서 내에 존재할 것이다 (예를 들어, SpCas9-기반 PE에 의한 경우). 닉의 위치는 PE를 형성하는 특정한 Cas9의 특징일 것이다. 예를 들어, SpCas9-기반 PE에 의한 경우에, 포스포디에스테르 결합 내의 닉 부위는 염기 3 (PAM 서열의 위치 1에 대해 "-3" 위치)과 4 (PAM 서열의 위치 1에 대해 "-4" 위치) 사이이다. 프로토스페이서 내의 닉 부위는 유리 3' 히드록실 기를 형성하고, 이는 하기 도면에서 보이는 바와 같이, PEgRNA의 연장 아암의 프라이머 결합 부위와 복합체화되고, PEgRNA의 연장 아암의 DNA 합성 주형에 의한 DNA 코드의 단일 가닥의 중합을 시작하는 기질을 제공한다. 이 중합 반응은 5'에서 3' 방향으로 PE 융합 단백질의 폴리머라제 (예를 들어, 리버스 트랜스크립타제)에 의해 촉매된다. 중합은 gRNA 코어에 도달하기 전에 종결되어 (예를 들어, 중합 종결 신호, 또는 PE의 중합 활성을 종결시키는 기능을 하는 2차 구조의 포함에 의해), 닉킹된 PAM 가닥의 원래 3' 히드록실 기로부터 연장된 단일 가닥 DNA 플랩을 생성한다. DNA 합성 주형은 PAM 가닥 상의 닉 부위에 바로 이어지는 DNA의 내인성 5'-단부 단일 가닥에 상동인 단일 가닥 DNA를 코딩하고, 목적하는 뉴클레오티드 변화 (예를 들어, 단일 염기 치환, 삽입, 결실, 역위)를 포함한다. 목적하는 편집의 위치는 PAM 가닥 상의 닉 부위의 하류에 하기 임의의 위치에 있을 수 있으며, 이는 (닉 부위의 하류 위치에 대해) 위치 +1, +2, +3, +4 (PAM 부위의 시작), +5 (PAM 부위의 위치 2), +6 (PAM 부위의 위치 3), +7, +8, +9, +10, +11, +12, +13, +14, +15, +16, +17, +18, +19, +20, +21, +22, +23, +24, +25, +26, +27, +28, +29, +30, +31, +32, +33, +34, +35, +36, +37, +38, +39, +40, +41, +42, +43, +44, +45, +46, +47, +48, +49, +50, +51, +52, +53, +54, +55, +56, +57, +58, +59, +60, +61, +62, +63, +64, +65, +66, +67, +68, +69, +70, +71, +72, +73, +74, +75, +76, +77, +78, +79, +80, +81, +82, +83, +84, +85, +86, +87, +88, +89, +90, +91, +92, +93, +94, +95, +96, +97, +98, +99, +100, +101, +102, +103, +104, +105, +106, +107, +108, +109, +110, +111, +112, +113, +114, +115, +116, +117, +118, +119, +120, + 121, +122, +123, +124, +125, +126, +127, +128, +129, +130, +131, +132, +133, +134, +135, +136, +137, +138, +139, +140, +141, +142, +143, +144, +145, +146, +147, +148, +149 또는 +150 또는 그 초과를 포함할 수 있다. 3' 단부 단일 가닥 DNA (관심 편집을 함유함)가 내인성 5' 단부 단일 가닥 DNA를 대체하면, DNA 복구 및 복제 과정은 PAM 가닥 상의 편집 부위의 영구적 설치, 및 이어서 편집 부위에 존재할 비-PAM 가닥 상의 미스매치의 교정을 유발할 것이다. 이러한 방식으로, 편집은 표적 DNA 부위 상의 DNA 두 가닥으로 연장될 것이다. "편집된 가닥" 및 "비-편집된" 가닥에 대한 언급은 단지 PE 메카니즘에 수반되는 DNA의 가닥을 서술하도록 의도되는 것으로 인지될 것이다. "편집된 가닥"은 닉 부위의 바로 하류의 5' 단부 단일 가닥 DNA를 목적하는 편집을 함유하는 합성된 3' 단부 단일 가닥 DNA로 대체함으로써 처음 편집되는 가닥이다. "비-편집된" 가닥은 편집된 가닥과의 가닥 쌍이지만, 이는 또한 그 자체가 복구 및/또는 복제를 통해 편집되어 편집된 가닥, 특히 관심 편집에 상보적이 된다.

도 1m은 표적 DNA, 프라임 편집제 복합체의 해부학적 특색 및 PEgRNA와 표적 DNA 사이의 상호작용을 보여주는 프라임 편집의 메카니즘을 도시한다. 먼저, 폴리머라제 (예를 들어, 리버스 트랜스크립타제) 및 napDNAbp (예를 들어, SpCas9 닉카제, 예를 들어 HNH 뉴클레아제 도메인 내의 탈활성화 돌연변이 (예를 들어, H840A) 또는 RuvC 뉴클레아제 도메인 내의 탈활성화 돌연변이 (D10A)를 갖는 SpCas9)를 갖는 융합 단백질을 포함하는 프라임 편집제는 편집될 표적 DNA를 갖는 DNA 및 PEgRNA와 복합체화된다. PEgRNA는 스페이서, gRNA 코어 (일명 gRNA 스캐폴드 또는 gRNA 백본) (이는 napDNAbp에 결합함), 및 연장 아암을 포함한다. 연장 아암은 3' 단부, 5' 단부 또는 PEgRNA 분자 내의 어딘가에 있을 수 있다. 제시된 바와 같이, 연장 아암은 PEgRNA의 3' 단부에 있다. 연장 아암은 3'에서 5' 방향으로 프라이머 결합 부위 및 DNA 합성 주형 (관심 편집 및 PAM 가닥 상의 닉 부위에 바로 이어지는 5' 단부 단일 가닥 DNA와 상동인 상동성의 영역 (즉, 상동성 아암) 둘 다를 포함함)을 포함한다. 제시된 바와 같이, 닉이 도입되어 닉 부위의 바로 상류에 유리 3' 히드록실 기를 생성하면, PAM 가닥 상의 닉 부위의 바로 상류에 있는 영역은 "프라이머 결합 부위"로 지칭되는 연장 아암의 3' 단부에서 상보적 서열에 어닐링되어, 이용가능한 3' 히드록실 단부를 갖는 짧은 이중 가닥 영역을 생성하고, 이는 프라임 편집제 복합체의 폴리머라제에 대한 기질을 형성한다. 이어서, 폴리머라제 (예를 들어, 리버스 트랜스크립타제)는 3' 히드록실 단부로부터 연장 아암의 단부까지 DNA의 가닥으로서 중합시킨다. 단일 가닥 DNA의 서열은 새로운 DNA를 합성하기 위해 폴리머라제에 의해 "판독"되는 연장 아암의 부분 (즉, 프라이머 결합 부위 제외)인 DNA 합성 주형에 의해 코딩된다. 이러한 중합은 초기 닉 부위의 원래 3' 히드록실 단부의 서열을 효과적으로 연장시킨다. DNA 합

성 주형은 목적하는 편집뿐만 아니라 PAM 가닥 상의 닉 부위의 바로 하류의 DNA의 내인성 단일 가닥에 상동인 영역을 포함하는 DNA의 단일 가닥을 코딩한다. 다음으로, DNA의 코딩된 3' 단부 단일 가닥 (즉, 3' 단일 가닥 DNA 플랩)은 PAM 가닥 상의 닉 부위의 바로 하류의 DNA의 상응하는 상동 내인성 5'-단부 단일 가닥을 대체하여, 5'-단부 단일 가닥 DNA 플랩을 갖는 DNA 중간체를 형성하고, 이는 세포에 의해 (예를 들어, 플랩 엔도뉴클레아제에 의해) 제거된다. 내인성 5'-단부 단일 가닥 DNA 플랩의 상보체에 어닐링하는 3'-단부 단일 가닥 DNA 플랩은 5' DNA 플랩이 제거된 후에 내인성 가닥에 라이게이션된다. 이제 어닐링되고 라이게이션된 3' 단부 단일 가닥 DNA 플랩 내의 목적하는 편집은 상보체 가닥과의 미스매치를 형성하며, 이는 DNA 복구 및/또는 복제 라운드를 겪음으로써, 두 가닥 상에 목적하는 편집을 영구적으로 설치한다.

도 2는 본원에 기재된 프라임 편집제에 사용될 수 있는 3종의 Cas 복합체 (SpCas9, SaCas9 및 LbCas12a) 및 그의 PAM, gRNA 및 DNA 절단 특색을 보여준다. 도면은 SpCas9, SaCas9 및 LbCas12a를 포함하는 복합체에 대한 설계를 보여준다.

도 3a-3f는 조작된 5' 프라임 편집제 gRNA (도 3a), 3' 프라임 편집제 gRNA (도 3b) 및 분자내 연장부 (도 3c)에 대한 설계를 보여준다. 연장된 가이드 RNA (또는 연장된 gRNA)는 또한 본원에서 PEgRNA 또는 "프라임 편집 가이드 RNA"로 지칭될 수 있다. 도 3d 및 도 3e는 각각 3' 및 5' 프라임 편집제 gRNA (PEgRNA)의 추가의 실시양태를 제공한다. 도 3f는 3' 단부 프라임 편집제 가이드 RNA와 표적 DNA 서열 사이의 상호작용을 예시한다. 도 3a-3c의 실시양태는 3', 5' 및 분자내 버전의 연장된 부분에서 역전사 주형 서열 (즉 또는 보다 넓게는 나타낸 바와 같은 DNA 합성 주형으로 지칭되며, 이는 RT가 프라임 편집제의 맥락에서 사용될 수 있는 유일한 1가지 유형의 폴리머라제이기 때문임), 프라이머 결합 부위 및 임의적인 링커 서열의 예시적인 배열, 뿐만 아니라 스페이서 및 코어 영역의 일반적 배열을 도시한다. 개시된 프라임 편집 과정은 이들 연장된 가이드 RNA의 구성으로 제한되지는 않는다. 도 3d의 실시양태는 본원에서 고려된 예시적인 PEgRNA의 구조를 제공한다. PEgRNA는 5'에서 3' 방향으로 정렬된 3종의 주요 성분 요소, 즉: 스페이서, gRNA 코어 및 3' 단부에서의 연장 아암을 포함한다. 연장 아암은 5'에서 3' 방향으로 하기 구조적 요소, 즉: 프라이머 결합 부위 (A), 편집 주형 (B) 및 상동성 아암 (C)으로 추가로 나뉠 수 있다. 추가로, PEgRNA는 임의적인 3' 단부 변형제 영역 (e1) 및 임의적인 5' 단부 변형제 영역 (e2)을 포함할 수 있다. 또한 추가로, PEgRNA는 PEgRNA의 3' 단부에 전사 종결 신호를 포함할 수 있다 (도시되지 않음). 이들 구조적 요소는 본원에 추가로 정의된다. PEgRNA의 구조의 도시는 제한적인 것으로 의도되지 않고, 요소의 배열에서의 변형을 포괄한다. 예를 들어, 임의적인 서열 변형제 (e1) 및 (e2)는 제시된 임의의 다른 영역 내에 또는 사이에 위치할 수 있고, 3' 및 5' 단부에 위치하는 것으로 제한되지는 않는다. PEgRNA는, 특정 실시양태에서, 2차 RNA 구조, 예컨대 비제한적으로 헤어핀, 스템/루프, 토루프, RNA-결합 단백질 동원 도메인 (예를 들어, MS2cp 단백질을 동원하고 이에 결합하는 MS2 압타머)을 포함할 수 있다. 예를 들어, 이러한 2차 구조는 스페이서, gRNA 코어 또는 연장 아암 내에, 특히 e1 및/또는 e2 변형제 영역 내에 위치할 수 있다. 2차 RNA 구조에 추가로, PEgRNA는 (예를 들어, e1 및/또는 e2 변형제 영역 내에) 화학적 링커 또는 폴리(N) 링커 또는 테일을 포함할 수 있으며, 여기서 "N"은 임의의 핵염기일 수 있다. 일부 실시양태에서 (예를 들어, 도 72(c)에 제시된 바와 같이), 화학적 링커는 sgRNA 스캐폴드 또는 코어의 역전사를 방지하는 기능을 할 수 있다. 추가로, 특정 실시양태에서 (예를 들어, 도 72(c) 참조), 연장 아암 (3)은 RNA 또는 DNA로 구성될 수 있고/거나, 1개 이상의 핵염기 유사체 (예를 들어, 온도 탄력성과 같은 기능성을 부가할 수 있음)를 포함할 수 있다. 또한 추가로, 연장 아암 (3)의 배향은 자연적인 5'에서 3' 방향일 수 있거나 또는 3'에서 5' 방향의 반대 배향으로 합성될 수 있다 (전체 PEgRNA 분자의 배향에 대해). 또한, 관련 기술 분야의 통상의 기술자는 napDNAbp와의 융합체로서 구현되거나 또는 목적하는 편집을 포함하는 목적하는 주형-코딩된 3' 단일 가닥 DNA 플랩을 합성하기 위한 별개의 모이어티로서 트랜스로 제공되어 구현될 수 있는 프라임 편집에 사용하기 위해, 연장 아암의 핵산 물질 (즉, DNA 또는 RNA)의 성질에 따라 적절한 DNA 폴리머라제를 선택할 수 있을 것임을 주목한다. 예를 들어, 연장 아암이 RNA이면, DNA 폴리머라제는 리버스 트랜스크립타제 또는 임의의 다른 적합한 RNA-의존성 DNA 폴리머라제일 수 있다. 그러나, 연장 아암이 DNA이면, DNA 폴리머라제는 DNA-의존성 DNA 폴리머라제일 수 있다. 다양한 실시양태에서, DNA 폴리머라제의 제공은, 예를 들어 RNA-단백질 동원 도메인 (예를 들어, PEgRNA 상에 (예를 들어, e1 또는 e2 영역에 또는 다른 곳에) 설치된 MS2 헤어핀 및 DNA 폴리머라제에 융합된 MS2cp 단백질, 이에 의해 DNA 폴리머라제를 PEgRNA로 공동-국재화함)의 사용을 통해 트랜스로 이루어질 수 있다. 추가로, 프라이머 결합 부위는 일반적으로 목적하는 편집을 포함하는 생성된 3' 단일 가닥 DNA 플랩을 코딩하기 위해 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)에 의해 사용되는 주형의 일부를 형성하지 않는다는 것을 주목한다. 따라서, "DNA 합성 주형"의 명칭은 편집을 함유하는 목적하는 3' 단일 가닥 DNA 플랩 및 프라임 편집 DNA 합성의 3' 단일 가닥 DNA 가닥 생성물에 의해 대체되는 5' 내인성 단일 가닥 DNA 플랩에 대한 상동성의 영역을 코딩하기 위해 DNA 폴리머라제에 의해 주형으로서 사용되는 연

장 아암 (3)의 영역 또는 부분을 지칭한다. 일부 실시양태에서, DNA 합성 주형은 "편집 주형" 및 "상동성 아암", 또는 예를 들어 편집 주형 전후의 1개 이상의 상동성 아암을 포함한다. 편집 주형은 작게는 단일 뉴클레오티드 치환일 수 있거나, 또는 DNA의 삽입 또는 역위일 수 있다. 추가로, 편집 주형은 또한 결실을 포함할 수 있으며, 이는 목적하는 결실을 함유하는 상동성 아암을 코딩함으로써 조작될 수 있다. 다른 실시양태에서, DNA 합성 주형은 또한 e2 영역 또는 그의 한 부분을 포함할 수 있다. 예를 들어, e2 영역이 DNA 폴리머라제 활성의 종결을 유발하는 2차 구조를 포함한다면, e2 영역의 임의의 부분이 DNA로 실제로 코딩되기 전에 DNA 폴리머라제 기능이 종결될 가능성이 있다. e2 영역의 일부 또는 심지어 전부가 DNA로 코딩되는 것이 또한 가능하다. e2의 얼마나 많은 부분이 주형으로서 실제로 사용되는지는 그의 구성 및 그 구성이 DNA 폴리머라제 기능을 방해하는지 여부에 좌우될 것이다.

도 3e의 실시양태는 본원에서 고려되는 또 다른 PEgRNA의 구조를 제공한다. PEgRNA는 5'에서 3' 방향으로 정렬된 3종의 주요 성분 요소, 즉: 스페이서, gRNA 코어 및 3' 단부에서의 연장 아암을 포함한다. 연장 아암은 5'에서 3' 방향으로 하기 구조적 요소, 즉: 프라이머 결합 부위 (A), 편집 주형 (B) 및 상동성 아암 (C)으로 추가로 나뉠 수 있다. 추가로, PEgRNA는 임의적인 3' 단부 변형제 영역 (e1) 및 임의적인 5' 단부 변형제 영역 (e2)을 포함할 수 있다. 또한 추가로, PEgRNA는 PEgRNA의 3' 단부에 전사 종결 신호를 포함할 수 있다 (도시되지 않음). 이들 구조적 요소는 본원에 추가로 정의된다. PEgRNA의 구조의 도시는 제한적인 것으로 의도되지 않고, 요소의 배열에서의 변형을 포괄한다. 예를 들어, 임의적인 서열 변형제 (e1) 및 (e2)는 제시된 임의의 다른 영역 내에 또는 사이에 위치할 수 있고, 3' 및 5' 단부에 위치하는 것으로 제한되지는 않는다. PEgRNA는, 특정 실시양태에서, 2차 RNA 구조, 예컨대 비제한적으로 헤어핀, 스템/루프, 토 루프, RNA-결합 단백질 동원 도메인 (예를 들어, MS2cp 단백질을 동원하고 이에 결합하는 MS2 압타머)을 포함할 수 있다. 이들 2차 구조는 PEgRNA 분자 내 어느 곳에든 위치할 수 있다. 예를 들어, 이러한 2차 구조는 스페이서, gRNA 코어 또는 연장 아암 내에, 특히 e1 및/또는 e2 변형제 영역 내에 위치할 수 있다. 2차 RNA 구조에 추가로, PEgRNA는 (예를 들어, e1 및/또는 e2 변형제 영역 내에) 화학적 링커 또는 폴리(N) 링커 또는 테일을 포함할 수 있으며, 여기서 "N"은 임의의 핵염기일 수 있다. 일부 실시양태에서 (예를 들어, 도 72(c)에 제시된 바와 같이), 화학적 링커는 sgRNA 스캐폴드 또는 코어의 역전사를 방지하는 기능을 할 수 있다. 추가로, 특정 실시양태에서 (예를 들어, 도 72(c) 참조), 연장 아암 (3)은 RNA 또는 DNA로 구성될 수 있고/거나, 1개 이상의 핵염기 유사체 (예를 들어, 온도 탄력성과 같은 기능성을 부가할 수 있음)를 포함할 수 있다. 또한 추가로, 연장 아암 (3)의 배향은 자연적인 5'에서 3' 방향일 수 있거나 또는 3'에서 5' 방향의 반대 배향으로 합성될 수 있다 (전체 PEgRNA 분자의 배향에 대해). 또한, 관련 기술분야의 통상의 기술자는 napDNAbp와의 융합체로서 구현되거나 또는 목적하는 편집을 포함하는 목적하는 주형-코딩된 3' 단일 가닥 DNA 플랩을 합성하기 위한 별개의 모이어티로서 트랜스로 제공되어 구현될 수 있는 프라임 편집에 사용하기 위해, 연장 아암의 핵산 물질 (즉, DNA 또는 RNA)의 성질에 따라 적절한 DNA 폴리머라제를 선택할 수 있을 것임을 주목한다. 예를 들어, 연장 아암이 RNA이면, DNA 폴리머라제는 리버스 트랜스크립타제 또는 임의의 다른 적합한 RNA-의존성 DNA 폴리머라제일 수 있다. 그러나, 연장 아암이 DNA이면, DNA 폴리머라제는 DNA-의존성 DNA 폴리머라제일 수 있다. 다양한 실시양태에서, DNA 폴리머라제의 제공은, 예를 들어 RNA-단백질 동원 도메인 (예를 들어, PEgRNA 상에 (예를 들어, e1 또는 e2 영역에 또는 다른 곳에) 설치된 MS2 헤어핀 및 DNA 폴리머라제에 융합된 MS2cp 단백질, 이에 의해 DNA 폴리머라제를 PEgRNA로 공동-국재화함)의 사용을 통해 트랜스로 이루어질 수 있다. 추가로, 프라이머 결합 부위는 일반적으로 목적하는 편집을 포함하는 생성된 3' 단일 가닥 DNA 플랩을 코딩하기 위해 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)에 의해 사용되는 주형의 일부를 형성하지 않는다는 것을 주목한다. 따라서, "DNA 합성 주형"의 명칭은 편집을 함유하는 목적하는 3' 단일 가닥 DNA 플랩 및 프라임 편집 DNA 합성의 3' 단일 가닥 DNA 가닥 생성물에 의해 대체되는 5' 내인성 단일 가닥 DNA 플랩에 대한 상동성의 영역을 코딩하기 위해 DNA 폴리머라제에 의해 주형으로서 사용되는 연장 아암 (3)의 영역 또는 부분을 지칭한다. 일부 실시양태에서, DNA 합성 주형은 "편집 주형" 및 "상동성 아암", 또는 예를 들어 편집 주형 전후의 1개 이상의 상동성 아암을 포함한다. 편집 주형은 작게는 단일 뉴클레오티드 치환일 수 있거나, 또는 DNA의 삽입 또는 역위일 수 있다. 추가로, 편집 주형은 또한 결실을 포함할 수 있으며, 이는 목적하는 결실을 함유하는 상동성 아암을 코딩함으로써 조작될 수 있다. 다른 실시양태에서, DNA 합성 주형은 또한 e2 영역 또는 그의 한 부분을 포함할 수 있다. 예를 들어, e2 영역이 DNA 폴리머라제 활성의 종결을 유발하는 2차 구조를 포함한다면, e2 영역의 임의의 부분이 DNA로 실제로 코딩되기 전에 DNA 폴리머라제 기능이 종결될 가능성이 있다. e2 영역의 일부 또는 심지어 전부가 DNA로 코딩되는 것이 또한 가능하다. e2의 얼마나 많은 부분이 주형으로서 실제로 사용되는지는 그의 구성 및 그 구성이 DNA 폴리머라제 기능을 방해하는지 여부에 좌우될 것이다.

도 3f의 개략도는 전형적인 PEgRNA와 이중 가닥 DNA의 표적 부위의 상호작용 및 관심 유전자 변화를 함유하는

3' 단일 가닥 DNA 플랩의 동반 생성을 도시한다. 이중 가닥 DNA는 3'에서 5' 배향의 상단 가닥 (즉, 표적 가닥) 및 5'에서 3' 방향의 하단 가닥 (즉, PAM 가닥 또는 비-표적 가닥)으로 제시된다. 상단 가닥은 "프로토스페이서"의 상보체 및 PAM 서열의 상보체를 포함하며, PEgRNA의 스페이서에 의해 표적화되고 이에 어닐링되는 가닥이기 때문에 "표적 가닥"으로 지칭된다. 상보적 하단 가닥은 PAM 서열 (예를 들어, NGG) 및 프로토스페이서를 함유하기 때문에 "비-표적 가닥" 또는 "PAM 가닥" 또는 "프로토스페이서 가닥"으로 지칭된다. 제시되지는 않았지만, 도시된 PEgRNA는 프라임 편집제 융합 단백질의 Cas9 또는 등가 도메인과 복합체화될 것이다. 개략도에 제시된 바와 같이, PEgRNA의 스페이서는 표적 가닥 상의 프로토스페이서의 상보적 영역에 어닐링된다. 이러한 상호작용은 스페이서 RNA와 프로토스페이서 DNA의 상보체 사이에 DNA/RNA 하이브리드로서 형성되고, 프로토스페이서에 R 루프의 형성을 유도한다. 본원의 다른 곳에 교시된 바와 같이, Cas9 단백질 (제시되지 않음)은 이어서 제시된 바와 같은 비-표적 가닥에 닉을 유도한다. 이어서, 이는 *z*에 따라 프라이머 결합 부위에서 PEgRNA의 3' 단부와 상호작용하는 닉 부위의 바로 상류의 3' ssDNA 플랩 영역의 형성을 유도한다. ssDNA 플랩의 3' 단부 (즉, 리버스 트랜스크립타제 프라이머 서열)는 PEgRNA 상의 프라이머 결합 부위 (A)에 어닐링함으로써, 리버스 트랜스크립타제를 프라이밍한다. 이어서, 리버스 트랜스크립타제 (예를 들어, Cas9 구축물에 부착된 융합 단백질로서 시스로 제공되거나 또는 트랜스로 제공됨)는 DNA 합성 주형 (편집 주형 (B) 및 상동성 아암 (C) 포함)에 의해 코딩되는 DNA의 단일 가닥을 중합시킨다. 중합은 연장 아암의 5' 단부 쪽으로 계속된다. ssDNA의 중합된 가닥은, 다른 곳에 기재된 바와 같이 (예를 들어, 도 1g에 제시된 바와 같이), 내인성 DNA를 침입하고, 상응하는 내인성 가닥 (이는 내인성 DNA의 5' 단부 DNA 플랩으로서 제거됨)을 대체하며, 자연 발생 DNA 복구/복제 라운드를 통해 목적하는 뉴클레오티드 편집 (단일 뉴클레오티드 염기 쌍 변화, 결실, 삽입 (전체 유전자 포함))을 설치하는 ssDNA 3' 단부 플랩을 형성한다.

도 3g는 본원에서 고려되는 프라임 편집의 또 다른 실시양태를 도시한다. 특히, 상단 개략도는 링커에 의해 연결된 napDNAbp (예를 들어, SpCas9) 및 폴리머라제 (예를 들어, 리버스 트랜스크립타제)의 융합 단백질을 포함하는 프라임 편집제 (PE)의 한 실시양태를 도시한다. PE는 PEgRNA의 gRNA 코어에 결합함으로써 PEgRNA와 복합체를 형성한다. 제시된 실시양태에서, PEgRNA에는 3' 단부에서 시작하여 프라이머 결합 부위 (PBS)에 이어서 DNA 합성 주형을 포함하는 3' 연장 아암이 장착되어 있다. 하단 개략도는 "트랜스 프라임 편집제 (tPE)"로 지칭되는 프라임 편집제의 변이체를 도시한다. 이러한 실시양태에서, DNA 합성 주형 및 PBS는 PEgRNA로부터 탈커플링되고, RNA-단백질 동원 도메인 (예를 들어, MS2 헤어핀)을 포함하는 트랜스 프라임 편집제 RNA 주형 ("tPERT")으로 지칭되는 별개의 분자 상에 제시된다. PE 그 자체는 RNA-단백질 동원 도메인을 특이적으로 인식하고 이에 결합하는 단백질인 rPERT 동원 단백질 ("RP")에 대한 융합체를 포함하도록 추가로 변형된다. RNA-단백질 동원 도메인이 MS2 헤어핀인 예에서, 상응하는 rPERT 동원 단백질은 MS2 태그부착 시스템의 MS2cp일 수 있다. MS2 태그부착 시스템은 MS2 박테리오파지 코트 단백질 ("MCP" 또는 "MS2cp")과 파지의 게놈에 존재하는 스템-루프 또는 헤어핀 구조, 즉 "MS2 헤어핀" 또는 "MS2 압타머"의 천연 상호작용에 기초한다. 트랜스 프라임 편집의 경우에, 도 3h에 도시된 예에 제시된 바와 같이, RP-PE:gRNA 복합체는 적절한 RNA-단백질 동원 도메인을 갖는 tPERT를 "동원"하여 PE:gRNA 복합체와 공동-국재화함으로써, 프라임 편집에 사용하기 위한 PBS 및 DNA 합성 주형을 트랜스로 제공한다.

도 3h는 트랜스 프라임 편집의 과정을 도시한다. 이러한 실시양태에서, 트랜스 프라임 편집제는 MS2cp 단백질 (즉, MS2 압타머를 인식하고 이에 결합하는 동원 단백질의 유형)에 융합되고 sgRNA (즉, PEgRNA와 대조적으로 표준 가이드 RNA)와 복합체화된 "PE2" 프라임 편집제 (즉, Cas9(H840A) 및 변이체 MMLV RT의 융합체)를 포함한다. 트랜스 프라임 편집제는 표적 DNA에 결합하고, 비표적 가닥을 닉킹한다. MS2cp 단백질은 tPERT 분자 상의 RNA-단백질 동원 도메인과의 특이적 상호작용을 통해 tPERT를 트랜스로 동원한다. tPERT는 트랜스 프라임 편집제와 공동-국재화됨으로써, 3' 단부를 갖고 DNA 합성 주형에 의해 코딩되는 목적하는 유전자 정보를 함유하는 단일 가닥 DNA 플랩을 합성하기 위해 리버스 트랜스크립타제 폴리머라제에 의해 사용하기 위한 PBS 및 DNA 합성 주형 기능을 트랜스로 제공한다.

도 4a-4e는 시험관내 TPRT 검정 (즉, 프라임 편집 검정)을 나타낸다. 도 4a는 RT 효소에 의한 형광 표지된 DNA 기질 gRNA 주형화된 연장의 개략도 (PAGE)이다. 도 4b는 사전-닉킹된 기질, dCas9 및 합성 주형 길이가 상이한 5'-연장된 gRNA를 사용한 TPRT (즉, 프라임 편집)를 보여준다. 도 4c는 Cas9의 부재 하에 사전-닉킹된 DNA 기질과의 RT 반응을 보여준다. 도 4d는 Cas9(H840A) 및 5'-연장된 gRNA를 사용한 전체 dsDNA 기질에 대한 TPRT (즉, 프라임 편집)를 보여준다. 도 4e는 사전-닉킹된 및 전체 dsDNA 기질을 사용한 3'-연장된 gRNA 주형을 보여준다. 모든 반응은 M-MLV RT에 의한다.

도 5는 합성 주형 길이가 다양한 5'-연장된 gRNA를 사용한 시험관내 검증을 보여준다. 형광 표지된 (Cy5) DNA

표적을 기질로서 사용하고, 이 실험 세트에서 사전-닉킹하였다. 이들 실험에 사용된 Cas9는 촉매 기능상실 Cas9 (dCas9)이고, 사용된 RT는 슈퍼스크립트 III이며, 이는 몰로니-뮤린 백혈병 바이러스 (M-MLV)로부터 유래된 상업용 RT이다. dCas9:gRNA 복합체를 정제된 성분으로부터 형성하였다. 이어서, 형광 표지된 DNA 기질을 dNTP 및 RT 효소와 함께 첨가하였다. 37℃에서 1시간의 인큐베이션 후에, 반응 생성물을 변성 우레아-폴리아크릴아미드 겔 전기영동 (PAGE)에 의해 분석하였다. 겔 영상은 원래 DNA 가닥이 역전사 주형의 길이와 일치하는 길이로 연장된 것을 보여준다.

도 6은 합성 주형 길이가 다양한 5'-연장된 gRNA를 사용한 시험관내 검증을 보여주며, 이는 도 5에 제시된 것과 밀접하게 유사하다. 그러나, DNA 기질은 이 실험 세트에서 사전-닉킹되지 않는다. 이들 실험에 사용된 Cas9는 Cas9 닉카제 (SpyCas9 H840A 돌연변이체)이고, 사용된 RT는 슈퍼스크립트 III이며, 이는 몰로니-뮤린 백혈병 바이러스 (M-MLV)로부터 유래된 상업용 RT이다. 반응 생성물을 변성 우레아-폴리아크릴아미드 겔 전기영동 (PAGE)에 의해 분석하였다. 겔에 제시된 바와 같이, 닉카제는 표준 gRNA가 사용되는 경우에 DNA 가닥을 효율적으로 절단한다 (gRNA_0, 레인 3).

도 7은 3' 연장이 DNA 합성을 지지하고 Cas9 닉카제 활성에 유의하게 영향을 미치지 않는다는 것을 나타낸다. 사전-닉킹된 기질 (흑색 화살표)은 dCas9 또는 Cas9 닉카제가 사용될 때 RT 생성물로 거의-정량적으로 전환된다 (레인 4 및 5). RT 생성물 (적색 화살표)로의 50% 초과의 전환이 전체 기질 (레인 3)에서 관찰된다. Cas9 닉카제 (SpyCas9 H840A 돌연변이체), 촉매 기능상실 Cas9 (dCas9) 및 슈퍼스크립트 III (몰로니-뮤린 백혈병 바이러스 (M-MLV)로부터 유래된 상업용 RT)이 사용된다.

도 8은 RT 반응이 시스의 gRNA (동일한 복합체에 결합됨)와 우선적으로 발생하는지 여부를 결정하는 데 사용된 이중 색상 실험을 나타낸다. 5'-연장된 및 3'-연장된 gRNA에 대해 2개의 별개의 실험을 수행하였다. 생성물을 PAGE에 의해 분석하였다. 생성물 비는 (Cy3시스/Cy3트랜스) / (Cy5트랜스/Cy5시스)로서 계산하였다.

도 9a-9d는 플랩 모델 기질을 나타낸다. 도 9a는 플랩-지정 돌연변이유발을 위한 이중-FP 리포터를 보여준다. 도 9b는 HEK 세포에서의 정지 코돈 복구를 보여준다. 도 9c는 플랩 복구 후 서열분석된 효모 클론을 보여준다. 도 9d는 인간 세포에서의 상이한 플랩 특색의 시험을 보여준다.

도 10은 플라스미드 기질에 대한 프라임 편집을 나타낸다. 효모 (에스. 세레비지아에(*S. cerevisiae*)) 발현을 위해 이중-형광 리포터 플라스미드를 구축하였다. 효모에서의 이러한 구축물의 발현은 GFP만을 생성한다. 시험관내 프라임 편집 반응은 점 돌연변이를 도입하고, 모 플라스미드 또는 시험관내 Cas9(H840A) 닉킹된 플라스미드를 효모 내로 형질전환시킨다. 콜로니가 형광 영상화에 의해 가시화된다. 효모 이중-FP 플라스미드 형질전환체가 제시된다. 모 플라스미드 또는 시험관내 Cas9(H840A) 닉킹된 플라스미드를 형질전환시키는 것은 녹색 GFP 발현 콜로니만을 생성한다. 5'-연장된 또는 3'-연장된 gRNA와의 프라임 편집 반응은 녹색과 황색 콜로니의 혼합을 생성한다. 후자는 GFP 및 mCherry 둘 다를 발현한다. 더 많은 황색 콜로니가 3'-연장된 gRNA에서 관찰된다. 정지 코돈을 함유하지 않는 양성 대조군이 또한 제시된다.

도 11은 도 10에서의 실험과 유사한 플라스미드 기질에 대한 프라임 편집을 보여주지만, 정지 코돈에 점 돌연변이를 설치하는 대신에, 프라임 편집은 프레임시프트 돌연변이를 복구하고 하류 mCherry의 합성을 가능하게 하는 단일 뉴클레오티드 삽입 (좌측) 또는 결실 (우측)을 설치한다. 실험 둘 다는 3' 연장된 gRNA를 사용하였다.

도 12는 생어 서열분석에 의해 특징화된, 플라스미드 기질에 대한 프라임 편집의 편집 생성물을 보여준다. TRT 형질전환으로부터의 개별 콜로니를 선택하고, 생어 서열분석에 의해 분석하였다. 정확한 편집은 선택 콜로니를 서열분석함으로써 관찰되었다. 녹색 콜로니는 원래 DNA 서열을 갖는 플라스미드를 함유한 반면, 황색 콜로니는 프라임 편집 gRNA에 의해 설계된 정확한 돌연변이를 함유하였다. 다른 점 돌연변이 또는 indel은 관찰되지 않았다.

도 13은 새로운 프라임 편집 기술에 대한 잠재적 범주가 제시되고 데아미나제-매개된 염기 편집제 기술과 비교된다는 것을 보여준다.

도 14는 인간 세포에서의 편집의 개략도를 보여준다.

도 15는 gRNA에서의 프라이머 결합 부위의 연장을 나타낸다.

도 16은 인접한 표적화를 위한 말단절단된 gRNA를 보여준다.

도 17a-17c는 인간 배아 신장 (HEK) 세포에서의 성분의 형질감염 후 표적 뉴클레오티드에서의 % T에서 A로의 전환을 나타내는 그래프이다. 도 17a는 Cas9(H840A) 닉카제에 대한 야생형 MLV 리버스 트랜스크립타제의 N-말단

융합체 (32개-아미노산 링커)를 사용한 결과를 제시하는 데이터를 보여준다. 도 17b는 도 17a와 유사하지만, RT 효소의 C-말단 융합체에 대한 것이다. 도 17c는 도 17a와 유사하지만, MLV RT와 Cas9 사이의 링커가 32개 아미노산 대신 60개 아미노산 길이이다.

도 18은 고처리량 앰플리콘 서열분석에 의한 HEK3 부위에서의 고순도 T에서 A로의 편집을 보여준다. 서열분석 분석의 결과는 편집된 세포의 가장 풍부한 유전자형을 디스플레이한다.

도 19는 indel 비율 (오렌지색 막대)과 함께 표적 뉴클레오티드에서의 편집 효율 (청색 막대)을 보여준다. WT는 야생형 MLV RT 효소를 지칭한다. 돌연변이체 효소 (M1 내지 M4)는 우측에 열거된 돌연변이를 함유한다. 편집 비율을 게놈 DNA 앰플리콘의 고처리량 서열분석에 의해 정량화하였다.

도 20은 단일 가닥 닉이 표적 뉴클레오티드에 근접하여 상보적 DNA 가닥에 도입된 경우의 표적 뉴클레오티드의 편집 효율을 보여준다. 표적 뉴클레오티드로부터 다양한 거리에서의 닉킹을 시험하였다 (삼각형). 표적 염기쌍에서의 편집 효율 (청색 막대)이 indel 형성 비율 (오렌지색 막대)과 함께 제시된다. "부재" 예는 상보적 가닥 닉킹 가이드 RNA를 함유하지 않는다. 편집 비율을 게놈 DNA 앰플리콘의 고처리량 서열분석에 의해 정량화하였다.

도 21은 목적하는 T에서 A로의 전환 돌연변이 및 다른 주요 게놈 편집 부산물의 일반적 부재를 보여주는 프로세싱된 고처리량 서열분석 데이터를 나타낸다.

도 22는 연장된 가이드 RNA와 복합체화된 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)을 사용하여 표적 유전자좌 상에서 오류-유발 리버스 트랜스크립타제에 의해 표적화된 돌연변이유발을 수행하기 위한, 즉 오류-유발 RT에 의해 프라임 편집을 수행하기 위한 예시적인 과정의 개략도를 제공한다. 이 과정은 표적화된 돌연변이유발을 위한 프라임 편집의 실시양태로 지칭될 수 있다. 연장된 가이드 RNA는 가이드 RNA의 3' 또는 5' 단부에 또는 가이드 RNA 내의 분자내 위치에 연장부를 포함한다. 단계 (a)에서, napDNAbp/gRNA 복합체는 DNA 분자와 접촉하고, gRNA는 napDNAbp가 돌연변이유발될 표적 유전자좌에 결합하도록 가이드한다. 단계 (b)에서, 표적 유전자좌의 DNA의 가닥 중 하나에 닉을 도입하여 (예를 들어, 뉴클레아제 또는 화학적 작용제에 의해), 표적 유전자좌의 가닥 중 하나에 이용가능한 3' 단부를 생성한다. 특정 실시양태에서, 닉은 R-루프 가닥에 상응하는 DNA의 가닥, 즉 가이드 RNA 서열에 혼성화되지 않는 가닥에서 생성된다. 단계 (c)에서, 3' 단부 DNA 가닥은 역전사를 프라이밍하기 위해 가이드 RNA의 연장된 부분과 상호작용한다. 특정 실시양태에서, 3' 단부 DNA 가닥은 가이드 RNA의 연장된 부분 상의 특이적 RT 프라이밍 서열에 혼성화한다. 단계 (d)에서, DNA의 돌연변이유발된 단일 가닥을 프라이밍된 부위의 3' 단부로부터 가이드 RNA의 3' 단부를 향해 합성하는 오류-유발 리버스 트랜스크립타제가 도입된다. 예시적인 돌연변이는 별표 "*"로 표시된다. 이는 목적하는 돌연변이유발된 영역을 포함하는 단일 가닥 DNA 플랩을 형성한다. 단계 (e)에서, napDNAbp 및 가이드 RNA가 방출된다. 단계 (f) 및 (g)는 목적하는 돌연변이유발된 영역이 표적 유전자좌 내로 혼입되도록 하는 단일 가닥 DNA 플랩 (돌연변이유발된 영역을 포함함)의 분해에 관한 것이다. 이러한 과정은 3' 단일 가닥 DNA 플랩이 다른 가닥 상의 상보적 서열에 침입하고 혼성화하였을 때 형성되는 상응하는 5' 내인성 DNA 플랩을 제거함으로써, 목적하는 생성물 형성을 향해 구동될 수 있다. 과정은 또한 도 1f에 예시된 바와 같이 제2 가닥 닉킹에 의해 생성물 형성을 향해 구동될 수 있다. 내인성 DNA 복구 및/또는 복제 과정 후에, 돌연변이유발된 영역은 DNA 유전자좌의 DNA의 두 가닥 내로 혼입된다.

도 23은 트리뉴클레오티드 반복 서열을 축소시키기 위한 gRNA 설계 및 TPRT 게놈 편집 (즉, 프라임 편집)에 의한 트리뉴클레오티드 반복부 축소의 개략도이다. 트리뉴클레오티드 반복부 확장은 헌팅톤병, 유약 X 증후군 및 프리드라이히 운동실조를 포함한 다수의 인간 질환과 연관된다. 가장 흔한 트리뉴클레오티드 반복부는 CAG 트리플렛을 함유하지만, GAA 트리플렛 (프리드라이히 운동실조) 및 CGG 트리플렛 (유약 X 증후군)이 또한 발생한다. 확장에 대한 소인이 유전되거나 또는 이미 확장된 모 대립유전자를 획득하는 것은 질환을 획득할 가능성을 증가시킨다. 병원성 트리뉴클레오티드 반복부 확장은 프라임 편집을 사용하여 가설적으로 교정될 수 있다. 반복 영역의 상류 영역은 RNA-가이드된 뉴클레아제에 의해 닉킹되고, 이어서 건강한 수의 반복부 (이는 특정한 유전자 및 질환에 의존함)를 함유하는 새로운 DNA 가닥의 합성을 프라이밍하는 데 사용될 수 있다. 반복 서열 후에, 반복부의 다른 단부 (적색 가닥)에 인접한 서열의 정체에 매칭되는 짧은 상동성 스트레치가 부가된다. 새로 합성된 가닥의 침입 및 내인성 DNA의 새로 합성된 플랩으로의 후속 대체는 축소된 반복 대립유전자를 유도한다.

도 24는 프라임 편집에 의한 정확한 10개-뉴클레오티드 결실을 보여주는 개략도이다. HEK3 유전자좌를 표적화하는 가이드 RNA를 닉 부위 뒤 10개-뉴클레오티드 결실을 코딩하는 역전사 주형으로 설계하였다. 형질감염된

HEK 세포에서의 편집 효율을 앰플리콘 서열분석을 사용하여 평가하였다.

도 25는 내인성 게놈 유전자좌에서의 펩티드 태그부착 유전자에 대한 gRNA 설계 및 TPRT 게놈 편집 (즉, 프라임 편집)에 의한 펩티드 태그부착을 보여주는 개략도이다. FlAsH 및 ReAsH 태그부착 시스템은 다음 2가지 부분을 포함한다: (1) 형광단-이중비소 프로브 및 (2) 서열 FLNCCPGCCMEP (서열식별번호: 1)에 의해 예시되는 테트라시스테인 모티프를 함유하는 유전자 코딩된 펩티드. 세포 내에서 발현되는 경우에, 테트라시스테인 모티프를 함유하는 단백질은 형광단-비소 프로브로 형광 표지될 수 있다 (참고문헌: [J. Am. Chem. Soc., 2002, 124 (21), pp 6063-6076. DOI: 10.1021/ja017687n] 참조). "소르태깅(sortagging)" 시스템은 표지된 펩티드 프로브를 적합한 펩티드 기질을 함유하는 단백질에 공유 접합시키는 박테리아 소르타제 효소를 사용한다 (참고문헌: [Nat. Chem. Biol. 2007 Nov;3(11):707-8. DOI: 10.1038/nchembio.2007.31] 참조). FLAG-태그 (DYKDDDDK (서열식별번호: 2)), V5-태그 (GKPIPNPLLGLDST (서열식별번호: 3)), GCN4-태그 (EELLSKNYHLENEVARLKK (서열식별번호: 4)), HA-태그 (YPYDVPDYA (서열식별번호: 5)) 및 Myc-태그 (EQKLISEEDL (서열식별번호: 6))는 면역검정을 위한 에피토프 태그로서 통상적으로 사용된다. 파이-클램프는 펜타플루오로-방향족 기질로 표지될 수 있는 펩티드 서열 (FCPF (서열식별번호: 622))을 코딩한다 (참고문헌: [Nat. Chem. 2016 Feb;8(2):120-8. doi: 10.1038/nchem.2413]).

도 26a는 His_6-태그 및 FLAG-태그의 게놈 DNA 내로의 정확한 설치를 보여준다. HEK3 유전자좌를 표적화하는 가이드 RNA를 18-nt His-태그 삽입 또는 24-nt FLAG-태그 삽입을 코딩하는 역전사 주형으로 설계하였다. 형질감염된 HEK 세포에서의 편집 효율을 앰플리콘 서열분석을 사용하여 평가하였다. FLAG-태그의 전체 24-nt 서열은 보이는 프레임의 외부에 있다는 것을 주목한다 (서열분석으로 전체 및 정확한 삽입을 확인함). 도 26b는 (a) 단백질을 가용성 또는 불용성으로 만드는 것, (b) 단백질의 세포 국재화를 변화시키거나 추적하는 것, (c) 단백질의 반감기를 연장시키는 것, (d) 단백질 정제를 용이하게 하는 것, 및 (e) 단백질의 검출을 용이하게 하는 것을 포함하는, 단백질/펩티드 태그부착을 수반하는 다양한 적용을 약술하는 개략도를 보여준다.

도 27은 프리온 질환의 진행을 방지하거나 중지시키는 보호 돌연변이를 PRNP에 설치하는 것에 의한 프라임 편집의 개관을 보여준다. PEgRNA 서열은 좌측 (즉, sgRNA 스캐폴드의 5')에서 서열식별번호: 810의 잔기 번호 1-20 및 우측 (즉, sgRNA 스캐폴드의 3')에서 서열식별번호: 810의 잔기 번호 21-43에 상응한다.

도 28a는 RNA 모티프를 코딩하는 서열의 PE-기반 삽입의 개략도이다. 도 28b는 잠재적으로 삽입될 수 있는 일부 예시 모티프 및 그의 기능의 목록 (총체적은 아님)이다.

도 29a는 프라임 편집제의 도시이다. 도 29b는 PE에 의해 지시된 게놈, 플라스미드 또는 바이러스 DNA에 대한 가능한 변형을 보여준다. 도 29c는 펩티드 루프의 라이브러리를 PEgRNA의 라이브러리를 통해 특정된 단백질 (이 경우에 GFP) 내로 삽입하기 위한 예시 계략도를 보여준다. 도 29d는 상이한 PEgRNA를 사용한 가능한 프로그램가능한 코돈 결실 또는 단백질의 N- 또는 C-말단의 말단절단의 예를 보여준다. 결실은 프레임시프트 돌연변이의 최소 생성과 함께 발생하는 것으로 예측될 것이다.

도 30은 연속 진화 시스템, 예컨대 PACE에서의 코돈의 반복적 삽입에 대한 가능한 개략도를 보여준다.

도 31은 gRNA 코어, 표적화된 유전자의 서열과 매칭되는 ~20nt 스페이서, 면역원성 에피토프 뉴클레오티드 서열을 갖는 역전사 주형 및 표적화된 유전자의 서열과 매칭되는 프라이머 결합 부위를 보여주는 조작된 gRNA의 예시이다.

도 32는 내인성 또는 외래 게놈 DNA 내로 공지된 면역원성 에피토프를 삽입하기 위한 수단으로서 프라임 편집을 사용하여 상응하는 단백질의 변형을 생성하는 것을 보여주는 개략도이다.

도 33은 프라이머 결합 서열 삽입을 위한 PEgRNA 설계, 및 오프-타겟 편집을 결정하기 위해 프라임 편집을 사용한 게놈 DNA 내로의 프라이머 결합 삽입을 보여주는 개략도이다. 이러한 실시양태에서, 프라임 편집은 살아있는 세포, 조직 또는 동물 모델 내부에서 수행된다. 제1 단계로서, 적절한 PEgRNA를 설계한다. 상단 개략도는 이러한 측면에서 사용될 수 있는 예시적인 PEgRNA를 보여준다. PEgRNA 내의 스페이서 ("프로토스페이서"로 표지됨)는 게놈 표적의 가닥 중 하나에 상보적이다. PE:PEgRNA 복합체 (즉, PE 복합체)는 코딩된 프라이머 결합 서열 및 절단 부위의 바로 하류의 영역 (적색)에 상보적인 상동성 영역 (PEgRNA의 상동성 아암에 의해 코딩됨)을 함유하는 단일 가닥 3' 단부 플랩을 닉 부위에 설치한다. 플랩 침입 및 DNA 복구/복제 과정을 통해, 합성된 가닥은 DNA 내로 혼입되어 프라이머 결합 부위를 설치한다. 이러한 과정은 목적하는 게놈 표적에서 발생할 수 있지만, 또한 오프-타겟 방식으로 PEgRNA와 상호작용할 수 있는 다른 게놈 부위에서도 발생할 수 있다 (즉, PEgRNA는 의도된 게놈 부위가 아닌 다른 게놈 부위에 대한 스페이서 영역의 상보성으로 인해 PE 복합체를 다른

오프-타겟 부위로 가이드함). 따라서, 프라이머 결합 서열을 목적하는 게놈 표적뿐만 아니라 게놈 내 다른 곳의 오프-타겟 게놈 부위에 설치할 수도 있다. 의도된 게놈 표적 부위 및 오프-타겟 게놈 부위 둘 다에서의 이들 프라이머 결합 부위의 삽입을 검출하기 위해, 게놈 DNA (PE-후)를 단리하고, 단편화하고, 어댑터 뉴클레오티드 (적색으로 제시됨)에 라이게이션할 수 있다. 이어서, 프라이머 결합 부위가 PE에 의해 삽입된 온-타겟 및 오프-타겟 게놈 DNA 영역을 증폭시키기 위해 어댑터 및 삽입된 프라이머 결합 서열에 어닐링하는 PCR 올리고뉴클레오티드를 사용하여 PCR을 수행할 수 있다. 이어서, 고처리량 서열분석을 수행하고, 서열 정렬하여 온-타겟 부위 또는 오프-타겟 부위에서 PE-삽입된 프라이머 결합 서열의 삽입 지점을 확인할 수 있다.

도 34는 PE에 의한 유전자의 정확한 삽입을 보여주는 개략도이다.

도 35a는 천연 인슐린 신호전달 경로를 보여주는 개략도이다. 도 35b는 FK1012에 의해 제어되는 FKBP12-태그부착된 인슐린 수용체 활성화를 보여주는 개략도이다.

도 36은 소분자 단량체를 보여준다. 참조: 범핑된 FK506 모방체 (2)[107]

도 37a-37b는 소분자 이량체를 보여준다. 참조: FK1012 4[95,96]; FK1012 5[108]; FK1012 6[107]; AP1903 7[107]; 시클로스포린 A 이량체 8[98]; FK506-시클로스포린 A 이량체 (FkCsA) 9[100].

도 38a-38f는 시험관내 및 효모 세포에서의 프라임 편집 및 실행가능성 연구의 개관을 제공한다. 도 38a는 유형별로 분류된, 클린바(ClinVar) (2019년 7월 액세스)에서의 75,122종의 공지된 병원성 인간 유전자 변이체를 보여준다. 도 38b는 프라임 편집 복합체가 조작된 리버스 트랜스크립타제 도메인에 융합되고 프라임 편집 가이드 RNA (PEgRNA)와 복합체화된, RNA-가이드된 DNA-닉킹 도메인, 예컨대 Cas9 닉카제를 함유하는 프라임 편집제 (PE) 단백질로 이루어진다는 것을 보여준다. PE:PEgRNA 복합체는 표적 DNA 부위에 결합하고, 표적 부위의 프로토스페이서 인접 모티프 (PAM) 전 또는 후의 광범위한 DNA 위치에서 매우 다양한 정확한 DNA 편집을 가능하게 한다. 도 38c는 DNA 표적 결합시에, PE:PEgRNA 복합체가 PAM-함유 DNA 가닥을 닉킹하는 것을 보여준다. 생성된 유리 3' 단부는 PEgRNA의 프라이머-결합 부위에 혼성화한다. 리버스 트랜스크립타제 도메인은 PEgRNA의 RT 주형을 사용하여 프라이머 연장을 촉매함으로써, 목적하는 편집을 함유하는 새로 합성된 DNA 가닥 (3' 플랩)을 생성한다. 편집된 3' 플랩과 원래 DNA를 함유하는 비편집된 5' 플랩 사이의 평형, 이어서 세포 5' 플랩 절단 및 라이게이션, 및 DNA 복구 또는 복제로 헤테로듀플렉스 DNA를 분해하는 것은, 안정하게 편집된 DNA를 생성한다. 도 38d는 5'-Cy5 표지된 PAM 가닥을 함유하는 사전-닉킹된 dsDNA 기질, dCas9 및 상업용 M-MLV RT 변이체 (RT, 슈퍼스크립트 III)를 사용한 시험관내 5'-연장된 PEgRNA 프라이머 연장 검정을 보여준다. dCas9를 다양한 길이의 RT 주형을 함유하는 PEgRNA와 복합체화한 다음, 표시된 성분과 함께 DNA 기질에 첨가하였다. 반응물을 37℃에서 1시간 동안 인큐베이션한 다음, 변성 우레아 PAGE에 의해 분석하고, Cy5 형광에 대해 가시화하였다. 도 38e는 dCas9 또는 Cas9 H840A 닉카제와 사전-복합체화된 3'-연장된 PEgRNA 및 사전-닉킹된 또는 비-닉킹된 5'-Cy5-표지된 dsDNA 기질을 사용하여 도 38d에서와 같이 수행된 프라이머 연장 검정을 보여준다. 도 38f는 시험관내에서 PEgRNA, Cas9 닉카제 및 RT로 편집된 GFP-mCherry 융합 리포터 플라스미드로 형질전환된 효모 콜로니를 보여준다. GFP와 mCherry 사이의 넌센스 또는 프레임시프트 돌연변이를 함유하는 플라스미드를, 전환 돌연변이, 1-bp 삽입 또는 1-bp 결실을 통해 mCherry 번역을 회복시키는 5'-연장된 또는 3'-연장된 PEgRNA로 편집하였다. GFP 및 mCherry 이중-양성 세포 (황색)는 성공적인 편집을 반영한다.

도 39a-39d는 PE1 및 PE2에 의한 인간 세포에서의 게놈 DNA의 프라임 편집을 조여준다. 도 39a는 PEgRNA가 스페이서 서열, sgRNA 스캐폴드, 및 프라이머-결합 부위 (녹색) 및 역전사 (RT) 주형 (자주색) (이는 편집된 염기(들) (적색)를 함유함)을 함유하는 3' 연장부를 함유한다는 것을 보여준다. 프라이머-결합 부위는 닉킹 부위의 바로 상류에서 PAM-함유 DNA 가닥에 혼성화한다. RT 주형은 코딩된 편집을 제외하고는 닉의 하류 DNA 서열에 상동이다. 도 39b는 야생형 M-MLV 리버스 트랜스크립타제에 융합된 Cas9 H840A 닉카제 (PE1) 및 다양한 프라이머-결합 부위 길이의 PEgRNA를 사용한 HEK293T 세포 내 HEK3 부위에서의 T·A에서 A·T로의 전환 편집의 설치를 보여준다. 도 39c는 PE2에서의 조작된 오중돌연변이체 M-MLV 리버스 트랜스크립타제 (D200N, L603W, T306K, W313F, T330P)의 사용이 HEK293T 세포 내 5개 게놈 부위에서의 프라임 편집 전환 효율 및 HEK3에서의 작은 삽입 및 작은 결실 편집을 실질적으로 개선시킨다는 것을 보여준다. 도 39d는 HEK293T 세포 내 5개 게놈 부위에서의 다양한 RT 주형 길이에 의한 PE2 편집 효율의 비교이다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 40a-40c는 PE3 및 PE3b 시스템이 비-편집된 가닥을 닉킹하여 프라임 편집 효율을 증가시킨다는 것을 보여준다. 도 40a는 PE3에 의한 프라임 편집의 개관이다. 편집된 가닥의 초기 합성 후에, DNA 복구는 편집을 함유하

는 새로 합성된 가닥을 제거하거나 (3' 플랩 절제) 또는 원래 게놈 DNA 가닥을 제거할 것이다 (5' 플랩 절제). 5' 플랩 절제는 1개의 편집된 가닥 및 1개의 비-편집된 가닥을 함유하는 DNA 헤테로듀플렉스를 남긴다. 미스매치 복구 기구 또는 DNA 복제는 헤테로듀플렉스를 분해하여 편집된 또는 비-편집된 생성물을 제공할 수 있다. 비-편집된 가닥을 닉킹하는 것은 그 가닥의 복구를 선호하여, 목적하는 편집을 함유하는 안정한 듀플렉스 DNA의 우선적인 생성을 유발한다. 도 40b는 PE3-매개된 프라임 편집 효율 및 indel 형성에 대한 상보적 가닥 닉킹의 효과를 보여준다. "부재"는 상보적 가닥을 닉킹하지 않는 PE2 대조군을 지칭한다. 도 40c는 PE2 (상보적 가닥 닉 부재), PE3 (일반적 상보적 가닥 닉) 및 PE3b (편집-특이적 상보적 가닥 닉)를 사용한 편집 효율의 비교이다. 모든 편집 수율은 분류 없이 모든 처리된 세포 중에서 의도된 편집을 함유하고 indel을 함유하지 않는 총 서열분석 판독물의 백분율을 반영한다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 41a-41k는 HEK293T 세포 내 7개의 내인성 인간 게놈 유전자좌에서 PE3에 의한 표적화된 삽입, 결실 및 모든 12가지 유형의 점 돌연변이를 보여준다. 도 41a는 10-nt RT 주형을 사용한 HEK3 부위의 위치 +1 내지 +8 (PEgRNA-유도된 닉의 위치를 위치 +1과 -1 사이로서 카운팅하여)의 모든 12가지 유형의 단일-뉴클레오티드 전이 및 전환 편집을 보여주는 그래프이다. 도 41b는 34-nt RT 주형을 사용한 HEK3 부위에서의 긴-범위 PE3 전환 편집을 보여주는 그래프이다. 도 41c-41h는 (도 41c) RNF2, (도 41d) FANCF, (도 41e) EMX1, (도 41f) RUNX1, (도 41g) VEGFA 및 (도 41h) DNMT1에 대한 프라임 편집 윈도우 내의 다양한 위치에서의 모든 12가지 유형의 전이 및 전환 편집을 보여주는 그래프이다. 도 41i는 7개의 내인성 게놈 유전자좌에서의 PE3에 의한 표적화된 1- 및 3-bp 삽입 및 1- 및 3-bp 결실을 보여주는 그래프이다. 도 41j는 HEK3 표적 부위에서의 5 내지 80 bp의 표적화된 정확한 결실을 보여주는 그래프이다. 도 41k는 3개의 내인성 게놈 유전자좌에서의 삽입 및 결실, 삽입 및 점 돌연변이, 결실 및 점 돌연변이, 및 이중 점 돌연변이의 조합 편집을 보여주는 그래프이다. 모든 편집 수율은 분류 없이 모든 처리된 세포 중에서 의도된 편집을 함유하고 indel을 함유하지 않는 총 서열분석 판독물의 백분율을 반영한다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 42a-42h는 프라임 편집 및 염기 편집, 및 공지된 Cas9 오프-타겟 부위에서의 Cas9 및 PE3에 의한 오프-타겟 편집의 비교를 보여준다. 도 42a는 HEK293T 세포 내 내인성 HEK3, FANCF 및 EMX1 부위에서의 PE2, PE3, BE2max 및 BE4max에 대한 동일한 표적 뉴클레오티드에서의 총 C·G에서 T·A로의 편집 효율을 보여준다. 도 42b는 도 42a에서의 처리로부터의 indel 빈도를 보여준다. 도 42c는 HEK3, FANCF 및 EMX1에서의 PE2, PE3, BE2max 및 BE4max에 대한 정확한 C·G에서 T·A로의 편집 (방관자 편집 또는 indel 부재)의 편집 효율을 보여준다. EMX1의 경우, 3개의 표적화된 뉴클레오티드에서의 C·G에서 T·A로의 전환의 모든 가능한 조합의 정확한 PE 조합 편집이 또한 제시된다. 도 42d는 HEK3 및 FANCF에서의 PE2, PE3, ABEdmax 및 ABEmax에 대한 총 A·T에서 G·C로의 편집 효율을 보여준다. 도 42e는 HEK3 및 FANCF에 대한 방관자 편집 또는 indel이 없는 정확한 A·T에서 G·C로의 편집 효율을 보여준다. 도 42f는 도 42d에서의 처리로부터의 indel 빈도를 보여준다. 도 42g는 4개의 온-타겟 및 16개의 공지된 오프-타겟 부위에서의 Cas9 뉴클레아제에 대한 HEK293T 세포에서의 평균 삼중 편집 효율 (indel을 갖는 서열분석 판독물 백분율)을 보여준다. 검사된 16개의 오프-타겟 부위는 4개의 온-타겟 부위 각각에 대해 이전에 보고된 상위 4개의 오프-타겟 부위[118,159]였다. 각각의 온-타겟 부위에 대해, Cas9를 sgRNA 또는 동일한 프로토스페이서를 인식하는 각각 4개의 PEgRNA와 쌍형성시켰다. 도 42h는 (도 42g)에서의 각각의 PEgRNA와 쌍형성된 PE2 또는 PE3에 대한 HEK293T 세포에서의 평균 삼중 온-타겟 및 오프-타겟 편집 효율 및 indel 효율 (하기 괄호 안)을 보여준다. 온-타겟 편집 수율은 분류 없이 모든 처리된 세포 중에서 의도된 편집을 함유하고 indel을 함유하지 않는 총 서열분석 판독물의 백분율을 반영한다. 오프-타겟 편집 수율은 프라임 편집과 일치하는 오프-타겟 유전자좌 변형을 반영한다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 43a-43i는 다양한 인간 세포주 및 1차 마우스 피질 뉴런에서의 프라임 편집, 병원성 전환, 삽입 또는 결실 돌연변이의 설치 및 교정, 및 프라임 편집과 HDR의 비교를 보여준다. 도 43a는 HEK293T 세포에서의 HBB 내 병원성 E6V 돌연변이의 설치 (T·A에서 A·T로의 전환을 통함) 및 교정 (A·T에서 T·A로의 전환을 통함)을 보여주는 그래프이다. 야생형 HBB 또는 PEgRNA PAM을 파괴하는 침묵 돌연변이를 함유하는 HBB로의 교정이 제시된다. 도 43b는 HEK293T 세포에서의 병원성 HEXA 1278+TATC 대립유전자의 설치 (4-bp 삽입을 통함) 및 교정 (4-bp 결실을 통함)을 보여주는 그래프이다. 야생형 HEXA 또는 PEgRNA PAM을 파괴하는 침묵 돌연변이를 함유하는 HEXA로의 교정이 제시된다. 도 43c는 G·C에서 T·A로의 전환을 통한 HEK293T 세포에서의 PRNP 내 보호 G127V 변이체의 설치를 보여주는 그래프이다. 도 43d는 K562 (백혈병성 골수 세포), U2OS (골육종 세포) 및 HeLa (자궁경부암 세포)를 포함한 다른 인간 세포주에서의 프라임 편집을 보여주는 그래프이다. 도 43e는 이중

분할-인테인 PE3 렌티바이러스 시스템을 사용한 마우스 1차 피질 뉴런의 DNMT1 내 G·C에서 T·A로의 전환 돌연변이의 설치를 보여주는 그래프이며, 여기서 N-말단 절반은 N-인테인에 및 P2A 자기-절단 펩티드를 통해 GFP-KASH에 융합된 Cas9 (1-573)이고, C-말단 절반은 PE2의 나머지에 융합된 C-인테인이다. PE2 절반은 성숙 뉴런에 고도로 특이적인 인간 시냅신 프로모터로부터 발현된다. 분류 값은 GFP-양성 핵으로부터의 편집 또는 indel을 반영하는 반면, 비분류 값은 모든 핵으로부터의 것이다. 도 43f는 HEK293T 세포 내 내인성 게놈 유전자좌에서의 PE3과 Cas9-매개된 HDR의 편집 효율의 비교이다. 도 43g는 K562, U2OS 및 HeLa 세포 내 내인성 게놈 유전자좌에서의 PE3과 Cas9-매개된 HDR의 편집 효율의 비교이다. 도 43h는 HEK293T, K562, U2OS 및 HeLa 세포에서의 PE3과 Cas9-매개된 HDR의 indel 부산물 생성의 비교이다. 도 43i는 PE3에 의한 HEK293T 세포에서의 His6 태그 (18 bp), FLAG 에피토프 태그 (24 bp) 또는 연장된 LoxP 부위 (44 bp)의 표적화된 삽입을 보여준다. 모든 편집 수율은 모든 처리된 세포 중 의도된 편집을 함유하고 indel을 함유하지 않는 총 서열분석 판독물의 백분율을 반영한다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 44a-44g는 형광 표지된 DNA 기질을 사용한 시험관내 프라임 편집 검증 연구를 보여준다. 도 44a는 dCas9, 5'-연장된 PEgRNA 및 5'-Cy5-표지된 DNA 기질을 사용한 전기영동 이동성 변화 검정을 보여준다. PEgRNA 1 내지 5는 스페이서와 PBS 사이의 15-nt 링커 서열 (PEgRNA 1의 경우 링커 A, PEgRNA 2 내지 5의 경우 링커 B), 5-nt PBS 서열 및 7 nt (PEgRNA 1 및 2), 8 nt (PEgRNA 3), 15 nt (PEgRNA 4) 및 22 nt (PEgRNA 5)의 RT 주형을 함유한다. PEgRNA는 도 44e 및 44f에 사용된 것이고; 전체 서열은 표 2A-2C에 열거된다. 도 44b는 5'-연장된 및 3'-연장된 PEgRNA를 사용한 Cas9 H840A의 시험관내 닉킹 검정을 보여준다. 도 44c는 5'-연장된 및 3'-연장된 PEgRNA를 사용한 HEK293T 세포 내 HEK3에서의 Cas9-매개된 indel 형성을 보여준다. 도 44d는 프라임 편집 시험관내 생화학적 검정의 개관을 보여준다. 5'-Cy5-표지된 사전-닉킹된 및 비-닉킹된 dsDNA 기질을 시험하였다. sgRNA, 5'-연장된 PEgRNA 또는 3'-연장된 PEgRNA를 dCas9 또는 Cas9 H840A 닉카제와 사전-복합체화한 다음, dsDNA 기질, M-MLV RT 및 dNTP와 합하였다. 반응을 37℃에서 1시간 동안 진행시킨 후, 변성 우레아 PAGE에 의해 분리하고, Cy5 형광에 의해 가시화하였다. 도 44e는 5'-연장된 PEgRNA, 사전-닉킹된 DNA 기질 및 dCas9를 사용한 프라이머 연장 반응이 RT 생성물로의 유의한 전환을 유도한다는 것을 보여준다. 도 44f는 도 44b에서와 같은 5'-연장된 PEgRNA를 비-닉킹된 DNA 기질 및 Cas9 H840A 닉카제와 함께 사용한 프라이머 연장 반응을 보여준다. 생성물 수율은 사전-닉킹된 기질과 비교하여 크게 감소된다. 도 44g는 3'-PEgRNA를 사용한 시험관내 프라이머 연장 반응이 변성 우레아 PAGE에 의해 단일 겉보기 생성물을 생성한다는 것을 보여준다. RT 생성물 밴드를 절제하고, 겔로부터 용리시킨 다음, dGTP 또는 dATP를 사용한 말단 트랜스퍼라제 (TdT)에 의한 단독중합체 테일링에 적용하였다. 테일링된 생성물을 폴리-T 또는 폴리-C 프라이머에 의해 연장시키고, 생성된 DNA를 서열분석하였다. 생어 트레이스는 gRNA 스캐폴드로부터 유래된 3개의 뉴클레오티드가 역전사되었다는 것을 나타낸다 (DNA 생성물에 최종 3' 뉴클레오티드로서 부가됨). 포유동물 세포 프라임 편집 실험에서, PEgRNA 스캐폴드 삽입은 시험관내에서보다 훨씬 더 드물며 (도 56a-56d), 이는 잠재적으로 테더링된 리버스 트랜스크립타제가 Cas9-결합된 가이드 RNA 스캐폴드에 접근할 수 없는 것 및/또는 PEgRNA 스캐폴드 서열을 함유하는 3' 플랩의 미스매칭된 3' 단부의 세포 절제에 기인한다.

도 45a-45g는 시험관내 프라임 편집 반응으로부터의 3' DNA 플랩의 효모 내 세포 복구를 보여준다. 도 45a는 이중 형광 단백질 리포터 플라스미드가 인-프레임 정지 코돈, +1 프레임시프트 또는 -1 프레임시프트를 코딩하는 표적 부위에 의해 분리된 GFP 및 mCherry 오픈 리딩 프레임을 함유한다는 것을 보여준다. 프라임 편집 반응을 Cas9 H840A 닉카제, PEgRNA, dNTP 및 M-MLV 리버스 트랜스크립타제를 사용하여 시험관내에서 수행한 다음, 효모 내로 형질전환시켰다. 비편집된 플라스미드를 함유하는 콜로니는 GFP를 생산하지만, mCherry는 생산하지 않는다. 편집된 플라스미드를 함유하는 효모 콜로니는 GFP 및 mCherry 둘 다를 융합 단백질로서 생산한다. 도 45b는 GFP와 mCherry 사이에 정지 코돈을 함유하는 리포터 플라스미드 (비편집된 음성 대조군, 상단) 또는 GFP와 mCherry 사이에 정지 코돈 또는 프레임시프트를 함유하지 않는 리포터 플라스미드 (사전-편집된 양성 대조군, 하단)로 형질전환된 효모 콜로니에 대한 GFP 및 mCherry 형광의 오버레이를 보여준다. 도 45c-45f는 시험관내 프라임 편집 반응 생성물로 형질전환된 효모 콜로니로부터의 mCherry 및 GFP 형광의 가시화를 보여준다. 도 45c는 도 45d에 제시된 바와 같은 3'-연장된 PEgRNA 또는 5'-연장된 PEgRNA를 사용한 T·A에서 A·T로의 전환을 통한 정지 코돈 교정을 보여준다. 도 45e는 3'-연장된 PEgRNA를 사용한 1-bp 결실을 통한 +1 프레임시프트 교정을 보여준다. 도 45f는 3'-연장된 PEgRNA를 사용한 1-bp 삽입을 통한 -1 프레임시프트 교정을 보여준다. 도 45g는 도 45b에서의 GFP-단독 콜로니 및 도 45c에서의 GFP 및 mCherry 이중-양성 콜로니로부터 단리된 플라스미드로부터의 생어 DNA 서열분석 트레이스를 보여준다.

도 46a-46f는 PE1을 사용한 정확한 편집 대 indel 생성을 보여준다. 도 46a는 10-nt RT 주형 및 8-17 nt 범위의 PBS 서열을 함유하는 PEgRNA를 사용한 HEK3의 +1 위치에서의 PE1에 의한 T·A에서 A·T로의 전환 편집 효율

및 indel 생성을 보여준다. 도 46b는 13-nt RT 주형 및 9-17 nt 범위의 PBS 서열을 함유하는 PEgRNA를 사용한 EMX1의 +5 위치에서의 PE1에 의한 G·C에서 T·A로의 전환 편집 효율 및 indel 생성을 보여준다. 도 46c는 17-nt RT 주형 및 8-17 nt 범위의 PBS 서열을 함유하는 PEgRNA를 사용한 FANCF의 +5 위치에서의 PE1에 의한 G·C에서 T·A로의 전환 편집 효율 및 indel 생성을 보여준다. 도 46d는 11-nt RT 주형 및 9-17 nt 범위의 PBS 서열을 함유하는 PEgRNA를 사용한 RNF2의 +1 위치에서의 PE1에 의한 C·G에서 A·T로의 전환 편집 효율 및 indel 생성을 보여준다. 도 46e는 13-nt RT 주형 및 7-15 nt 범위의 PBS 서열을 함유하는 PEgRNA를 사용한 HEK4의 +2 위치에서의 PE1에 의한 G·C에서 T·A로의 전환 편집 효율 및 indel 생성을 보여준다. 도 46f는 13-nt PBS 및 10-nt RT 주형을 사용한 HEK3 부위에서의 PE1-매개된 +1 T 결실, +1 A 삽입 및 +1 CTT 삽입을 보여준다. PEgRNA의 서열은 도 39c에서 사용된 것이다 (표 3A-3R 참조). 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 47a-47s는 프라임 편집을 위한 M-MLV RT 변이체의 평가를 보여준다. 도 47a는 이 도면에 사용된 프라임 편집제 변이체에 대한 약어를 보여준다. 도 47b는 HEK3 유전자좌에서 PE1에 의한 표적화된 삽입 및 결실 편집을 보여준다. 도 47c-47h는 M-MLV RT 변이체를 함유하는 18종의 프라임 편집제 구축물을 도 47c에 제시된 바와 같은 HEK3에서의 +2 G·C에서 C·G로의 전환 편집, 도 47d에 제시된 바와 같은 HEK3에서의 24-bp FLAG 삽입, 도 47e에 제시된 바와 같은 RNF2에서의 +1 C·G에서 A·T로의 전환 편집, 도 47f에 제시된 바와 같은 EMX1에서의 +1 G·C에서 C·G로의 전환 편집, 도 47g에 제시된 바와 같은 HBB에서의 +2 T·A에서 A·T로의 전환 편집 및 도 47h에 제시된 바와 같은 FANCF에서의 +1 G·C에서 C·G로의 전환 편집을 설치하는 그의 능력에 대해 비교한 것을 보여준다. 도 47i-47n은 M-MLV 변이체를 함유하는 4종의 프라임 편집제 구축물을 독립적 실험의 제2 라운드에서 도 47c-47h에 제시된 편집을 설치하는 그의 능력에 대해 비교한 것을 보여준다. 도 47o-47s는 5개의 게놈 유전자좌에서 다양한 PBS 길이를 사용한 PE2 편집 효율을 보여준다. 도 47o는 HEK3에서의 +1 T·A에서 A·T로의 변이를 보여준다. 도 47p는 EMX1에서의 +5 G·C에서 T·A로의 변이를 보여준다. 도 47q는 FANCF에서의 +5 G·C에서 T·A로의 변이를 보여준다. 도 47r은 RNF2에서의 +1 C·G에서 A·T로의 변이를 보여준다. 도 47s는 HEK4에서의 +2 G·C에서 T·A로의 변이를 보여준다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 48a-48c는 PEgRNA PBS 및 RT 주형 서열의 설계 특색을 보여준다. 도 48a는 RT 주형 길이의 함수로서 HEK293T 세포 내 VEGFA에서의 PE2-매개된 +5 G·C에서 T·A로의 전환 편집 효율 (청색 선)을 보여준다. Indel (회색 선)이 비교를 위해 플롯팅되어 있다. 그래프 아래의 서열은 PEgRNA에 의한 합성을 위해 주형화된 마지막 뉴클레오티드를 보여준다. G 뉴클레오티드 (PEgRNA에서 C에 의해 주형화됨)가 강조되고; C에서 끝나는 RT 주형은 프라임 편집 효율을 최대화하기 위해 PEgRNA 설계 동안 피해야 한다. 도 48b는 도 48a에서와 같은 DNMT1에 대한 +5 G·C에서 T·A로의 전환 편집 및 indel을 보여준다. 도 48c는 도 48a에서와 같은 RUNX1에 대한 +5 G·C에서 T·A로의 전환 편집 및 indel을 보여준다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 49a-49b는 세포 생존율에 대한 PE2, PE2 R110S K103L, Cas9 H840A 닉카제 및 dCas9의 효과를 보여준다. HEK293T 세포를 HEK3-표적화 PEgRNA 플라스미드와 함께 PE2, PE2 R110S K103L, Cas9 H840A 닉카제 또는 dCas9를 코딩하는 플라스미드로 형질감염시켰다. 세포 생존율을 셀타이터-글로 2.0 검정 (프로메가(Promega))을 사용하여 3일 동안 형질감염 후 24시간마다 측정하였다. 도 49a는 형질감염 후 1, 2 또는 3일에 발광에 의해 측정된 생존율을 보여준다. 값 및 오차 막대는 각각 기술적 삼중으로 수행된 3회의 독립적인 생물학적 반복실험의 평균 및 s.e.m.을 반영한다. 도 49b는 +5 G에서 A로의 편집을 코딩하는 HEK3-표적화 PEgRNA 플라스미드와 함께, PE2, PE2 R110S K103L, Cas9 H840A 닉카제 또는 dCas9에 대한 퍼센트 편집 및 indel을 보여준다. 편집 효율은 도 49a에서 생존율을 검정하는 데 사용된 세포와 함께 처리된 세포로부터 형질감염 후 제3일에 측정하였다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 50a-50b는 다양한 PEgRNA에 의한 PE3-매개된 HBB E6V 교정 및 HEXA 1278+TATC 교정을 보여준다. 도 50a는 PE3에 의한 HEK293T 세포에서의 HBB E6V 대립유전자의 교정에 대한 14개 PEgRNA의 스크린을 보여준다. 평가된 모든 PEgRNA는 HBB E6V 대립유전자를 임의의 침묵 PAM 돌연변이의 도입 없이 다시 야생형 HBB로 전환시킨다. 도 50b는 PE3 또는 PE3b에 의한 HEK293T 세포에서의 HEXA 1278+TATC 대립유전자의 교정에 대한 41개 PEgRNA의 스크린을 보여준다. 이들 HEXA 표지된 PEgRNA는 PAM을 파괴하고 침묵 돌연변이를 남기는 이동된 4-bp 결실에 의해 병원성 대립유전자를 교정한다. 이들 HEXA 표지된 PEgRNA는 병원성 대립유전자를 다시 야생형으로 교정한다. "b"로 끝나는 항목은 PEgRNA와 조합하여 편집-특이적 닉킹 sgRNA를 사용한다 (PE3b 시스템). 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 51a-51g는 인간 세포주에서의 PE3 활성, 및 PE3과 Cas9-개시 HDR의 비교를 나타낸다. 도 51a에 제시된 바와 같은 HEK293T 세포, 도 51b에 제시된 바와 같은 K562 세포, 도 51c에 제시된 바와 같은 U2OS 세포 및 도 51d에 제시된 바와 같은 HeLa 세포에서 PE3 및 Cas9-개시된 HDR에 대한 정확한 편집 (indel 부재) 생성 효율 및 indel 빈도. 각각의 괄호친 편집 비교는 PE3 및 Cas9-개시된 HDR에 의해 동일한 편집을 설치한다. 비-표적화 대조군은 비-표적 유전자좌를 표적화하는 PEgRNA 및 PE3이다. 도 51e는 야생형 Cas9 HDR 실험과 비교하여 비-표적화 PEgRNA+PE3 및 dCas9+sgRNA를 사용한 대조군 실험을 보여주며, 이는 겉보기 HDR 효율을 인위적으로 상승시키는 공통 오염물인 ssDNA 공여자 HDR 주형이 도 51a-51d에서의 HDR 측정치에 기여하지 않는다는 것을 확인시켜 준다. 도 51f-51g는 PE3 또는 Cas9-개시된 HDR에 의한 편집 후 K562 세포로부터 단리된 게놈 DNA 샘플로부터의 예시 HEK3 부위 대립유전자 표를 보여준다. 대립유전자를 일루미나 MiSeq 상에서 서열분석하고, CRISPResso2[178]로 분석하였다. 이 영역으로부터의 참조 HEK3 서열은 상단에 있다. 비-표적화 PEgRNA 음성 대조군, PE3을 사용한 HEK3에서의 +1 CTT 삽입 및 Cas9-개시된 HDR을 사용한 HEK3에서의 +1 CTT 삽입에 대한 대립유전자 표가 제시된다. 대립유전자 빈도 및 상응하는 일루미나 서열분석 판독물 카운트가 각각의 대립유전자에 대해 제시된다. 빈도 ≥0.20%로 관찰된 모든 대립유전자가 제시된다. 값 및 오차 막대는 3개의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 52a-52d는 클린바 데이터베이스에서의 병원성 삽입, 중복, 결실 및 indel의 길이별 분포를 보여준다. 클린바 변이체 요약을 2019년 7월 15일에 NCBI로부터 다운로드하였다. 보고된 삽입, 결실 및 중복의 길이는 참조 및 대안적 대립유전자, 변이체 출발 및 정지 위치, 또는 변이체 명칭에서의 적절한 확인 정보를 사용하여 계산하였다. 상기 정보 중 어느 것도 보고하지 않은 변이체는 분석에서 제외하였다. 보고된 indel (참조 게놈에 비해 삽입 및 결실 둘 다를 포함하는 단일 변이체)의 길이는 참조와 대안적 대립유전자 사이의 최상의 쌍별 정렬에서 미스매치 또는 갭의 수를 결정함으로써 계산하였다.

도 53a-53e는 GFP-양성 세포 분류에 대한 FACS 게이팅 예를 보여준다. 하기는 HEXA 1278+TATC 및 HBB E6V HEK293T 세포주를 생성하는 데 사용된 분류 전략을 약술하는 원래 배치 분석 파일의 예이다. 셀 소터 소프트웨어 v. 3.0.5를 사용하여 소니 LE-MA900 세포측정기 상에서 영상 데이터를 생성하였다. 그래픽 1은 GFP를 발현하지 않는 세포에 대한 게이팅 플롯을 보여준다. 그래픽 2는 HBB E6V HEK293T 세포주를 단리하는 데 사용된 P2A-GFP-발현 세포의 예시 분류를 보여준다. HEK293T 세포를 초기에 FSC-A/BSC-A (게이트 A)를 사용하여 집단에 대해 게이팅한 다음, FSC-A/FSC-H (게이트 B)를 사용하여 싱글렛에 대해 분류하였다. 살아있는 세포를 DAPI-음성 세포의 게이팅에 의해 분류하였다 (게이트 C). 음성-대조군 세포의 경우보다 높은 GFP 형광 수준을 갖는 세포를 형광색소로서 EGFP를 사용하여 분류하였다 (게이트 D). 도 53a는 HEK293T 세포 (GFP-음성)를 보여준다. 도 53b는 PE2-P2A-GFP를 발현하는 세포에 대한 FACS 게이팅의 대표적인 플롯을 보여준다. 도 53c는 HEXA 1278+TATC 동형접합체 HEK293T 세포에 대한 유전자형을 보여준다. 도 53d-53e는 HBB E6V 동형접합체 HEK293T 세포주에 대한 대립유전자 표를 보여준다.

도 54는 PEgRNA 클로닝 절차를 요약한 개략도이다.

도 55a-55g는 PEgRNA 설계의 개략도이다. 도 55a는 도메인이 표지된 (좌측) 및 도메인이 게놈 부위에서 nCas9에 결합된 (우측) PEgRNA의 간단한 다이어그램을 보여준다. 도 55b는 활성을 증가시킬 것으로 예상되는 PEgRNA에 대한 다양한 유형의 변형을 보여준다. 도 55c는 프로모터 선택 및 5', 3' 프로세싱 및 종결을 통해 보다 긴 RNA의 전사를 증가시키는 PEgRNA에 대한 변형을 보여준다. 도 55d는 스캐폴드 변형의 예인 P1 시스템의 연장을 보여준다. 도 55e는 주형 영역 내에 또는 PEgRNA 내의 다른 곳에 합성 변형의 혼입이 활성을 증가시킬 수 있다는 것을 보여준다. 도 55f는 주형 내 설계된 최소 2차 구조의 혼입이 보다 길고 보다 억제성인 2차 구조의 형성을 방지할 수 있다는 것을 보여준다. 도 55g는 PEgRNA의 3' 단부에서 RNA 요소에 의해 앵커링된 제2 주형 서열을 갖는 분할 PEgRNA를 보여준다 (좌측). PEgRNA의 5' 또는 3' 단부에의 요소의 혼입은 RT 결합을 증진시킬 수 있다.

도 56a-56d는 표적 유전자좌 내로의 PEgRNA 스캐폴드 서열의 혼입을 보여준다. HTS 데이터를 도 60a-60b에 기재된 바와 같이 PEgRNA 스캐폴드 서열 삽입에 대해 분석하였다. 도 56a는 EMX1 유전자좌에 대한 분석을 보여준다. RT 주형에 인접한 삽입 내에 1개 이상의 PEgRNA 스캐폴드 서열 뉴클레오티드를 함유하는 총 서열분석 판독물의 % (좌측); 명시된 길이의 PEgRNA 스캐폴드 서열 삽입을 함유하는 총 서열분석 판독물의 백분율 (중간); 및 X 축 상에 명시된 길이를 포함하여 그까지의 PEgRNA 삽입의 누적 총 백분율이 제시된다. 도 56b는 FANCF를 제외하고는 도 56a와 동일한 것을 보여준다. 도 56c는 HEK3을 제외하고는 도 56a에서와 동일한 것을 보여준다. 도 56d는 RNF2에 대한 것을 제외하고는 도 56a와 동일한 것을 보여준다. 값 및 오차 막대는 3개의 독립적인 생

물학적 반복실험의 평균 및 s.d.를 반영한다.

도 57a-57i는 트랜스크립톰-전반 RNA 존재비에 대한 PE2, PE2-dRT 및 Cas9 H840A 닉카제의 효과를 보여준다. PE2, PE2-dRT 또는 Cas9 H840A 닉카제 및 PRNP-표적화 또는 HEXA-표적화 PEgRNA를 발현하는 HEK293T 세포로부터 단리된, 리보솜 RNA가 고갈된, 세포 RNA의 분석. 14,410개의 유전자 및 14,368개의 유전자에 상응하는 RNA가 각각 PRNP 및 HEXA 샘플에서 검출되었다. 도 57a-57f는 (도 57a) PRNP-표적화 PEgRNA를 사용한 PE2 대 PE2-dRT, (도 57b) PRNP-표적화 PEgRNA를 사용한 PE2 대 Cas9 H840A, (도 57c) PRNP-표적화 PEgRNA를 사용한 PE2-dRT 대 Cas9 H840A, (도 57d) HEXA-표적화 PEgRNA를 사용한 PE2 대 PE2-dRT, (도 57e) HEXA-표적화 PEgRNA를 사용한 PE2 대 Cas9 H840A, (도 57f) HEXA-표적화 PEgRNA를 사용한 PE2-dRT 대 Cas9 H840A를 비교하는, 각각의 RNA에 대한 전사체 존재비의 log2-배수 변화 대 -log10 FDR-조정된 p-값을 디스플레이하는 볼케이노 플롯을 보여준다. 적색 점은 통계적으로 유의한 상대 존재비의 ≥2배 변화를 보여주는 유전자를 나타낸다 (FDR-조정된 $p < 0.05$). 도 57g-57i는 (도 57g) PE2 대 PE2-dRT, (도 57h) PE2 대 Cas9 H840A, 및 (도 57i) PE2-dRT 대 Cas9 H840A에 대한 PRNP 및 HEXA 샘플을 비교하는 상향조절 및 하향조절 전사체 (≥2배 변화)의 벤 다이어그램이다.

도 58a-58b는 뉴런 핵 분류에 대한 대표적인 FACS 게이팅을 보여준다. 핵을 다이사이클 루비 신호, FSC/SSC 비, SSC-폭/SSC-높이 비 및 GFP/다이사이클 비에 기초하여 순차적으로 게이팅하였다.

도 59a-59g는 골든 게이트 조립에 의해 3'-연장된 PEgRNA를 포유동물 U6 발현 벡터 내로 클로닝하기 위한 프로토콜을 보여준다. 도 59a는 클로닝 개관을 보여준다. 도 59b는 '단계 1: pU6-PEgRNA-GG-벡터 플라스미드를 소화시킨다 (성분 1)'를 보여준다. 도 59c는 '단계 2 및 3: 올리고뉴클레오티드 부분을 정렬하고 어닐링한다 (성분 2, 3 및 4)'을 보여준다. 도 59d는 '단계 2.b.ii.: sgRNA 스캐폴드 인산화 (인산화된 올리고뉴클레오티드를 구입한 경우 불필요)'를 보여준다. 도 59e는 '단계 4: PEgRNA 조립'을 보여준다. 도 59f는 '단계 5 및 6: 조립된 플라스미드의 형질전환'을 보여준다. 도 59g는 PEgRNA 클로닝 프로토콜을 요약한 다이어그램을 보여준다.

도 60a-60b는 PEgRNA 스캐폴드 통합을 정량화하기 위한 파이썬(Python) 스크립트를 보여준다. 맞춤형 파이썬 스크립트를 생성하여 표적 게놈 유전자좌에서의 PEgRNA 삽입을 특징화하고 정량화하였다. 스크립트는 참조 서열 (가이드 RNA 스캐폴드 서열)로부터 취한 증가하는 길이의 텍스트 스트링을 fastq 파일 내의 서열분석 판독물에 반복적으로 매칭시키고, 검색 질의에 매칭되는 서열분석 판독물의 수를 카운팅한다. 각각의 연속적인 텍스트 스트링은 가이드 RNA 스캐폴드 서열의 추가의 뉴클레오티드에 상응한다. 정확한 길이 적분 및 명시된 길이까지의 누적 적분을 이러한 방식으로 계산하였다. 참조 서열의 시작부에, sgRNA의 짧은 절편의 정렬 및 정확한 카운팅을 보장하기 위해 리버스 트랜스크립타제에 의해 합성된 새로운 DNA 가닥의 3' 단부의 5 내지 6개의 염기가 포함된다.

도 61은 SaCas9(N580A)-MMLV RT HEK3+6 C>A에 대해 명시된 편집을 갖는 총 서열분석 판독물의 퍼센트를 보여주는 그래프이다. 정확한 편집뿐만 아니라 indel에 대한 값이 제시된다.

도 62a-62b는 프라임 편집에 의한 정확한 위치에서의 목적하는 편집의 효율적인 설치를 위한 프로토스페이서의 중요성을 보여준다. 도 62a는 다양한 HEK3 유전자좌에 대해 표적 T·A 염기 쌍이 A·T로 전환된 총 서열분석 판독물의 퍼센트를 보여주는 그래프이다. 도 62b는 이를 보여주는 서열 분석이다.

도 63은 PAM 편집에서 SpCas9 PAM 변이체를 보여주는 그래프이다 (N=3). 표적화된 PAM 편집을 갖는 총 서열분석 판독물의 퍼센트가 SpCas9(H840A)-VRQR-MMLV RT (여기서 NGA > NTA) 및 SpCas9(H840A)-VRER-MMLV RT (여기서 NGCG > NTCG)에 대해 제시된다. 사용된 PEgRNA 프라이머 결합 부위 (PBS) 길이, RT 주형 (RT) 길이 및 PE 시스템이 열거된다.

도 64a-64f는 PE를 사용한 게놈 내로의 다양한 부위-특이적 레콤비나제 (SSR) 표적의 도입을 보여주는 개략도를 도시한다. 도 64a는 프라임 편집제에 의한 레콤비나제 표적 서열의 삽입의 일반적 개략도를 제공한다. 도 64b는 어떻게 PE에 의해 삽입된 단일 SSR 표적이 DNA 공여자 주형의 게놈 통합을 위한 부위로서 사용될 수 있는지를 보여준다. 도 64c는 어떻게 SSR 표적 부위의 탠덤 삽입이 게놈의 부분을 결실시키는데 사용될 수 있는지를 보여준다. 도 64d는 어떻게 SSR 표적 부위의 탠덤 삽입이 게놈의 부분을 역전시키는데 사용될 수 있는지를 보여준다. 도 64e는 어떻게 2개의 원위 염색체 영역에서의 2개의 SSR 표적 부위의 삽입이 염색체 전위를 유발할 수 있는지를 보여준다. 도 64f는 어떻게 게놈 내 2개의 상이한 SSR 표적 부위의 삽입이 DNA 공여자 주형으로부터의 카세트를 교환하는 데 사용될 수 있는지를 보여준다.

도 65는 1) 인간 세포 게놈 내 SSR 표적 부위의 PE-매개된 합성 및 2) GFP 발현 마커를 포함하는 DNA 공여자 주

형을 통합하기 위한 상기 SSR 표적 부위의 사용을 보여준다. 성공적으로 통합되면, GFP는 세포가 형광을 발하도록 한다.

도 66은 각각의 프라임 편집제 절반 단백질의 단부 또는 시작부에 위치한 분할-인테인 절반의 자기-스플라이싱 작용을 통해 전체 프라임 편집제로서 재생하는 2개의 PE 절반 단백질로서 제공되는 프라임 편집제의 한 실시양태를 도시한다.

도 67a-67b는 폴리펩티드 서열로부터의 인테인 제거 및 N-말단과 C-말단의 엑스테인 서열 사이의 펩티드 결합의 재형성의 메카니즘을 도시한다. 도 67a는 각각 인테인 서열의 절반을 함유하는 2개의 절반 단백질의 일반적 메카니즘을 도시하며, 이들 단백질은 세포 내에서 접촉하는 경우에 완전-기능적 인테인을 생성하고, 이는 이어서 자기-스플라이싱 및 절제를 겪는다. 절제 과정은 N-말단 단백질 절반 (또는 "N 엑스테인")과 C-말단 단백질 절반 (또는 "C 엑스테인") 사이의 펩티드 결합의 형성을 유발하여 N 엑스테인 및 C 엑스테인 부분을 포함하는 전체 단일 폴리펩티드를 형성한다. 다양한 실시양태에서, N 엑스테인은 분할 프라임 편집제의 융합 단백질의 N-말단 절반에 상응할 수 있고, C 엑스테인은 분할 프라임 편집제의 C-말단 절반에 상응할 수 있다. (b)는 인테인 절제 및 N 엑스테인 절반 (적색 절반)과 C 엑스테인 절반 (청색 절반)을 연결하는 펩티드 결합의 재형성의 화학적 메카니즘을 보여준다. 분할 인테인 (즉, 분할 인테인 구성에서 N 인테인 및 C 인테인)의 절제는 또한 트랜스로 제공되는 2개의 개별 성분의 스플라이싱 작용을 수반하기 때문에 "트랜스 스플라이싱"으로 지칭될 수 있다.

도 68a는 링커에서의 SpPE (서열식별번호: 3875, 3876)의 두 분할 인테인 절반의 전달이 HEK293T 세포 내로 공동-형질감염될 경우에 3개의 시험 유전자좌에서 활성을 유지한다는 것을 입증한다.

도 68b는 SaPE2의 두 분할 인테인 절반 (예를 들어, 서열식별번호: 443 및 서열식별번호: 450)의 전달이 HEK293T 세포 내로 공동-형질감염될 경우에 전장 SaPE2 (서열식별번호: 134)의 활성을 재현한다는 것을 입증한다. 따옴표로 표시된 잔기는 인테인 트랜스 스플라이싱 반응에 중요한 SaCas9 내 아미노산 741-743 (C-말단 엑스테인의 제1 잔기)의 서열이다. 'SMP'는 본 발명자들이 또한 'CFN' 컨센서스 스플라이싱 서열로 돌연변이시킨 천연 잔기이다. 컨센서스 서열은 프라임 편집 백분율에 의해 측정시 최고 재구성을 생성하는 것으로 제시된다.

도 68c는 다양한 개시된 PE 리보핵단백질 복합체 (고농도의 PE2, 고농도의 PE3 및 저농도의 PE3)가 이러한 방식으로 전달될 수 있다는 것을 보여주는 데이터를 제공한다.

도 69는 PANCE에서의 PE 유효성을 결정하기 위한 박테리오파지 플라크 검정을 보여준다. 플라크 (어두운 원형)는 이. 콜라이를 성공적으로 감염시킬 수 있는 파지를 나타낸다. L-람노스 농도의 증가는 PE 발현의 증가 및 플라크 형성의 증가를 유발한다. 플라크의 서열분석은 PE-설치된 게놈 편집의 존재를 밝혀내었다.

도 70a-70i는 프라임 편집을 위해 PEgRNA 및 닉킹-sgRNA를 설계하기 위한 단계별 지침의 예시로서 편집된 표적 서열의 예를 제공한다. 도 70a: 단계 1. 표적 서열 및 편집을 정의한다. 목적하는 편집 (점 돌연변이, 삽입, 결실 또는 그의 조합)의 위치 주위에 집중된 표적 DNA 영역 (~200bp)의 서열을 검색한다. 도 70b: 단계 2. 표적 PAM의 위치를 찾아낸다. 편집 위치에 근접한 PAM을 확인한다. 두 가닥에서 PAM을 확실히 찾아낸다. 편집 위치에 근접한 PAM이 바람직하지만, 편집 위치로부터 ≥ 30 nt에 닉을 배치하는 PAM 및 프로토스페이서를 사용하여 편집을 설치하는 것이 가능하다. 도 70c: 단계 3. 닉 부위를 위치시킨다. 고려되는 각각의 PAM에 대해, 상응하는 닉 부위를 확인한다. Sp Cas9 H840A 닉카제의 경우, 절단은 PAM-함유 가닥에서 NGG PAM에 대해 5'의 세번째 및 네번째 염기 사이에서 일어난다. 모든 편집된 뉴클레오티드는 닉 부위의 3'에 존재해야 하고, 따라서 적절한 PAM은 PAM-함유 가닥 상의 표적 편집에 대해 5'에 닉을 배치해야 한다. 하기 제시된 예에서, 2종의 가능한 PAM이 존재한다. 단순성을 위해, 나머지 단계는 PAM 1만을 사용한 PEgRNA의 설계를 입증할 것이다. 도 70d: 단계 4. 스페이서 서열을 설계한다. Sp Cas9의 프로토스페이서는 PAM-함유 가닥 상의 NGG PAM에 대해 5'의 20개 뉴클레오티드에 상응한다. 효율적인 Pol III 전사 개시는 G가 제1 전사된 뉴클레오티드일 것을 요구한다. 프로토스페이서의 제1 뉴클레오티드가 G인 경우에, PEgRNA에 대한 스페이서 서열은 단순히 프로토스페이서 서열이다. 프로토스페이서의 제1 뉴클레오티드가 G가 아닌 경우에, PEgRNA의 스페이서 서열은 G에 이어서 프로토스페이서 서열이 후속한다. 도 70e: 단계 5. 프라이머 결합 부위 (PBS)를 설계한다. 출발 대립유전자 서열을 사용하여, PAM-함유 가닥 상의 DNA 프라이머를 확인한다. DNA 프라이머의 3' 단부는 닉 부위의 바로 상류에 있는 뉴클레오티드이다 (즉, Sp Cas9의 경우 NGG PAM에 대해 5'의 네번째 염기). PE2 및 PE3과 함께 사용하기 위한 일반적 설계 원리로서, DNA 프라이머에 대해 12 내지 13개 뉴클레오티드의 상보성을 함유하는 PEgRNA 프라이머 결합 부위 (PBS)가 ~40-60% GC 함량을 함유하는 서열에 대해 사용될 수 있다. 낮은 GC 함량을 갖는 서열의 경우에, 보다 긴 (14- 내지 15-nt) PBS가 시험되어야 한다. 보다 높은 GC 함량을 갖는 서열의 경우에,

보다 짧은 (8- 내지 11-nt) PBS가 시험되어야 한다. 최적의 PBS 서열은 GC 함량에 상관없이 실험적으로 결정되어야 한다. 길이-p PBS 서열을 설계하기 위해, 출발 대립유전자 서열을 사용하여 PAM-함유 가닥에서 닉 부위의 5'의 첫번째 p 뉴클레오티드의 역 상보체를 취한다. 도 70f: 단계 6. RT 주형을 설계한다. RT 주형은 설계된 편집 및 편집에 인접한 서열에 대한 상동성을 코딩한다. 최적의 RT 주형 길이는 표적 부위에 기초하여 달라진다. 짧은-범위 편집 (위치 +1 내지 +6)의 경우, 짧은 (9 내지 12 nt), 중간 (13 내지 16 nt), 및 긴 (17 내지 20 nt) RT 주형을 시험하는 것이 권장된다. 긴-범위 편집 (위치 +7 및 그 초과)의 경우, 충분한 3' DNA 플랩 상동성을 허용하기 위해 편집 위치를 지나 적어도 5개의 nt (바람직하게는 10개 이상의 nt) 연장된 RT 주형을 사용하는 것이 권장된다. 긴-범위 편집의 경우, 여러 RT 주형은 기능적 설계를 확인하기 위해 스크리닝되어야 한다. 보다 큰 삽입 및 결실 (≥5 nt)을 위해, RT 주형 내로의 보다 큰 3' 상동성 (~20 nt 이상)의 혼입이 권장된다. 편집 효율은 전형적으로 RT 주형이 역전사된 DNA 생성물에서 마지막 뉴클레오티드로서 G의 합성 (PEgRNA의 RT 주형에서의 C에 상응함)을 코딩하는 경우에 손상된다. 많은 RT 주형이 효율적인 프라임 편집을 지지하기 때문에, RT 주형을 설계할 때 최종 합성된 뉴클레오티드로서 G를 회피하는 것이 권장된다. 길이-r RT 주형 서열을 설계하기 위해, 목적하는 대립유전자 서열을 사용하고, 원래 PAM을 함유하는 가닥에서 닉 부위의 3'의 첫번째 r 뉴클레오티드의 역 상보체를 취한다. SNP 편집과 비교하여, 동일한 길이의 RT 주형을 사용한 삽입 또는 결실 편집은 동일한 상동성을 함유하지 않을 것임을 주목한다. 도 70g: 단계 7. 전체 PEgRNA 서열을 조립한다. PEgRNA 성분을 하기 순서 (5'에서 3')로: 스페이서, 스캐폴드, RT 주형 및 PBS로 연결한다. 도 70h: 단계 8. PE3에 대한 닉킹-sgRNA의 설계. 편집의 상류 및 하류의 비-편집된 가닥 상에서 PAM을 확인한다. 최적의 닉킹 위치는 고도로 유전자좌-의존적이고, 실험적으로 결정되어야 한다. 일반적으로, PEgRNA-유도된 닉에 걸친 위치에 대해 5'의 40 내지 90개 뉴클레오티드에 배치된 닉은 보다 높은 편집 수율 및 보다 적은 indel을 유도한다. 닉킹 sgRNA는 출발 대립유전자에서 20-nt 프로토스페이서와 매칭되는 스페이서 서열을 가지며, 여기서 프로토스페이서가 G로 시작되지 않는 경우 5'-G가 부가된다. 도 70i: 단계 9. PE3b 닉킹-sgRNA의 설계. PAM이 상보적 가닥에 존재하고 그의 상응하는 프로토스페이서가 편집을 위해 표적화된 서열과 중첩되는 경우에, 이러한 편집은 PE3b 시스템에 대한 후보일 수 있다. PE3b 시스템에서, 닉킹-sgRNA의 스페이서 서열은 출발 대립유전자가 아닌 목적하는 편집된 대립유전자의 서열과 매칭된다. PE3b 시스템은 편집된 뉴클레오티드(들)가 닉킹-sgRNA 프로토스페이서의 시드 영역 (PAM에 인접한 ~10개의 nt) 내에 속할 때 효율적으로 작동한다. 이는 편집된 가닥의 설치 후까지 상보적 가닥의 닉킹을 방지하여, 표적 DNA에의 결합에 대한 PEgRNA와 sgRNA 사이의 경쟁을 방지한다. PE3b는 또한 두 가닥 상의 동시 닉의 생성을 피하고, 따라서 높은 편집 효율을 유지하면서 indel 형성을 유의하게 감소시킨다. PE3b sgRNA는 필요한 경우에 5' G를 부가하여 목적하는 대립유전자에서 20-nt 프로토스페이서와 매칭되는 스페이서 서열을 가져야 한다.

도 71a는 3' 단부에서 "UUU"로 종결되고 토루프 요소를 함유하지 않는 SpCas9 PEgRNA 분자의 뉴클레오티드 서열을 보여준다 (상단). 도면의 하단 부분은 동일한 SpCas9 PEgRNA 분자를 도시하지만, "UUU" 3' 단부 바로 앞에 서열 5'-"GAAANNNNN"-3'가 삽입된 토루프 요소를 함유하도록 추가로 변형된다. "N"은 임의의 핵염기일 수 있다.

도 71b는 HEK 세포 또는 EMX 세포에서의 프라임 편집의 효율이 PEgRNA 함유 토루프 요소를 사용하여 증가되는 반면, indel 형성의 퍼센트는 크게 변화되지 않는다는 것을 입증한다.

도 72a-72c는 프라임 편집에 사용될 수 있는 대안적 PEgRNA 구성을 도시한다. 도 72a는 프라임 편집의 PE2:PEgRNA 실시양태를 도시한다. 이러한 실시양태는 PEgRNA (도 1a-1i 및/또는 도 3a-3e에 또한 기재된 바와 같음)와 복합체화된 PE2 (Cas9 및 리버스 트랜스크립타제를 포함하는 융합 단백질)를 포함한다. 이러한 실시양태에서, 역전사를 위한 주형은 sgRNA 상의 3' 연장 아암 내로 혼입되어 PEgRNA를 만들고, DNA 폴리머라제 효소는 Cas9에 직접 융합된 리버스 트랜스크립타제 (RT)이다. 도 72b는 MS2cp-PE2:sgRNA + tPERT 실시양태를 도시한다. 이러한 실시양태는 MS2 박테리오파지 코트 단백질 (MS2cp)에 추가로 융합되어 MS2cp-PE2 융합 단백질을 형성하는 PE2 융합체 (Cas9 + 리버스 트랜스크립타제)를 포함한다. 프라임 편집을 달성하기 위해, MS2cp-PE2 융합 단백질은 복합체를 DNA 내의 특이적 표적 부위에 표적화하는 sgRNA와 복합체화된다. 이어서, 실시양태는 PEgRNA 대신에 별개의 분자 상에 프라이머 결합 부위 (PBS) 및 DNA 합성 주형을 제공함으로써 작동하는 트랜스 프라임 편집 RNA 주형 ("tPERT"), 즉 MS2 압타머 (스템 루프)가 또한 장착된 tPERT의 도입을 포함한다. MS2cp 단백질은 분자의 MS2 압타머에 결합함으로써 tPERT를 동원한다. 도 72c는 핵산 분자의 화학적 합성에 대한 공지된 방법을 통해 달성될 수 있는 PEgRNA에 대한 대안적 설계를 도시한다. 예를 들어, 화학적 합성은 프라임 편집에 사용하기 위한 하이브리드 RNA/DNA PEgRNA 분자를 합성하는 데 사용될 수 있으며, 여기서 하이브리드 PEgRNA의 연장 아암은 RNA 대신에 DNA이다. 이러한 실시양태에서, DNA-의존성 DNA 폴리머라제가 리버스 트랜스크립타제 대신에 사용되어 프라임 편집에 의해 형성된 목적하는 유전자 변화를 포함하는 3' DNA 플랩을 합성할

수 있다. 또 다른 실시양태에서, 연장 아암은 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)가 주형으로서 sgRNA 스캐폴드 또는 백본을 사용하는 것을 방지하는 화학적 링커를 포함하도록 합성될 수 있다. 또 다른 실시양태에서, 연장 아암은 PEgRNA 분자의 전체 배향에 대해 역배향을 갖는 DNA 합성 주형을 포함할 수 있다. 예를 들어, 5'에서 3' 배향이고 연장부가 sgRNA 스캐폴드의 3' 단부에 부착된 PEgRNA에 대해 제시된 바와 같이, DNA 합성 주형은 반대 방향, 즉 3'에서 5' 방향으로 배향된다. 이러한 실시양태는 gRNA의 3' 단부에 위치한 연장 아암을 갖는 PEgRNA 실시양태에 유리할 수 있다. 연장 아암의 배향을 역전시킴으로써, 폴리머라제 (예를 들어, 리버스 트랜스크립타제)에 의한 DNA 합성은 그것이 연장 아암의 새로 배향된 5'에 도달하면 종결될 것이고, 따라서 주형으로서 gRNA 코어를 사용할 위험이 없을 것이다.

도 73은 tPERT 및 MS2 동원 시스템 (일명 MS2 태그부착 기술)을 사용한 프라임 편집을 나타낸다. 프라임 편집제 단백질 (PE2)을 표적 유전자좌로 표적화하는 sgRNA는 프라이머 결합 부위 (13-nt 또는 17-nt PBS), His6 태그 삽입 및 상동성 아암을 코딩하는 RT 주형, 및 MS2 압타머 (tPERT 분자의 5' 또는 3' 단부에 위치함)를 함유하는 tPERT와 조합되어 발현된다. 프라임 편집제 단백질 (PE2) 또는 PE2의 N-말단에 대한 MS2cp의 융합체를 사용하였다. 이전에 개발된 PE3 시스템에서와 같이, 상보적-가닥 닉킹 sgRNA의 존재 또는 부재 하에 편집을 수행하였다 (x-축에서 각각 표지 "PE2+닉" 또는 "PE2"로 지정됨). 이는 또한 본원에서 "제2 가닥 닉킹"으로 지칭되고 정의된다.

도 74는 트랜스의 리버스 트랜스크립타제의 MS2 압타머 발현 및 MS2 압타머 시스템에 의한 그의 동원을 나타낸다. PEgRNA는 2개의 sgRNA 스캐폴드 헤어핀 중 하나 내로 삽입된 MS2 RNA 압타머를 함유한다. 야생형 M-MLV 리버스 트랜스크립타제는 MS2 코트 단백질 (MCP)에 대한 N-말단 또는 C-말단 융합체로서 발현된다. 편집은 HEK293T 세포 내 HEK3 부위에서 이루어진다.

도 75는 다양한 세포주에서의 상이한 표적 부위에 걸친 PE2, PE2-말단절단, PE3 및 PE3-말단절단의 효율 (즉, "명시된 편집 또는 indel을 갖는 총 서열분석 판독물의 %)을 비교하는 막대 그래프를 제공한다. 데이터는 말단절단된 RT 변이체를 포함하는 프라임 편집제가 대략 비-말단절단된 RT 단백질을 포함하는 프라임 편집제만큼 효율적이었음을 보여준다.

도 76은 인테인-분할 프라임 편집제의 편집 효율을 나타낸다. HEK239T 세포를 전장 PE2 또는 인테인-분할 PE2, PEgRNA 및 닉킹 가이드 RNA를 코딩하는 플라스미드로 형질감염시켰다. 컨센서스 서열 (C 말단 엑스테인의 대부분의 아미노-말단 잔기)이 표시된다. 2개의 부위에서의 편집 퍼센트가 제시된다: HEK3 +1 CTT 삽입 및 PRNP +6 G에서 T로. n=3 독립적 형질감염을 반복한다.

도 77은 인테인-분할 프라임 편집제의 편집 효율을 나타낸다. P0 마우스에게 ICV 주사에 의해 SpPE3 절반당 5E10vg 및 소량 1E10의 핵-국재화된 GFP:KASH의 전달시 벌크 피질 및 GFP+ 하위집단에서의 표적화된 심층 서열분석에 의해 편집을 평가하였다. 편집제 및 GFP를 EFS 프로모터와 함께 AAV9에 패키징하였다. 마우스를 주사 3주 후에 수거하고, GFP+ 핵을 유동 세포측정법에 의해 단리하였다. 개별 데이터 포인트가 제시되며, 조건당 1-2마리의 마우스를 분석한 것이다.

도 78은 인테인-분할 프라임 편집제의 편집 효율을 나타낸다. 구체적으로, 본 도면은 AAV 분할-SpPE3 구축물을 도시한 것이다. SpPE3-N 및 SpPE3-C를 개별적으로 발현하는 AAV 입자에 의한 공동-형질도입은 PE3 활성을 재현한다. N-말단 게놈은 닉킹 sgRNA를 발현하는 U6-sgRNA 카세트를 함유하고, C-말단 게놈은 PEgRNA를 발현하는 U6-PEgRNA 카세트를 함유한다는 것을 주목한다.

도 79는 특정의 최적화된 링커의 편집 효율을 보여준다. 특히, 데이터는 전이, 전환, 삽입 및 결실 편집을 위한 대표적인 PEgRNA에 대해 HEK3, EMX1, FANCF, RNF2 유전자좌에서, 현재 링커를 갖는 PE2 구축물 (PE2 - 백색 박스로 나타냄)의 편집 효율을 표시된 바와 같은 서열로 대체된 링커를 갖는 다양한 버전과 비교한 것을 보여준다. 대체 링커는 "1x SGGS" (서열식별번호: 174), "2x SGGS" (서열식별번호: 446), "3x SGGS" (서열식별번호: 3889), "1x XTEN" (서열식별번호: 171), "링커 부재", "1x Gly", "1x Pro", "1x EAAAK" (서열식별번호: 3968), "2x EAAAK" (서열식별번호: 3969) 및 "3x EAAAK" (서열식별번호: 3970)로 지칭된다. 편집 효율은 PE2의 "대조군" 편집 효율 대비 막대 그래프 포맷으로 측정된다. PE2의 링커는 SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (서열식별번호: 127)이다. 모든 편집을 PE3 시스템 (즉 이는 PE2 편집 구축물 플러스 최적의 2차 sgRNA 닉킹 가이드의 첨가를 지칭함)의 맥락에서 행하였다.

도 80. PE2 대비 평균 배수 효능을 취하면 제시된 그래프가 산출되며, 이는 1x XTEN (서열식별번호: 171) 링커 서열의 사용이 편집 효율을 평균 1.14배 개선시킨다는 것을 나타낸다 (n=15).

도 81은 상이한 프로모터로부터의 PEgRNA의 전사 수준을 도시한다.

도 82는 비변형된 PEgRNA 대비 PEgRNA 구조 상의 상이한 유형의 변형이 편집 효율에 대해 미치는 영향을 도시한다.

도 83은 HEK3 유전자의 표적화된 편집, 구체적으로 닉 부위에 대해 위치 +1에서의 10 nt 삽입의 삽입을 표적화하고, PE3을 사용하는 PE 실험을 도시한다.

도 84a는 스페이서, gRNA 코어 및 연장 아암 (RT 주형 + 프라이머 결합 부위)을 갖는 예시적인 PEgRNA를 도시하며, 이는 PEgRNA의 3' 단부에서 UCU 링커를 통해 커플링된 tRNA 분자로 변형된다. tRNA는 다양한 전사후 변형을 포함한다. 그러나, 상기 변형은 요구되지 않는다.

도 84b는 PEgRNA 구조를 변형시키는데 사용될 수 있는 tRNA의 구조를 도시한다. P1은 길이가 가변적일 수 있다. P1은 PEgRNA-tRNA 융합체의 RNAseP 프로세싱을 방지하는 것을 돕기 위해 연장될 수 있다.

도 85는 FANCF 유전자의 표적화된 편집, 구체적으로 닉 부위에 대해 위치 +5에서 G에서 T로의 전환을 특이적으로 표적화하고, PE3 구축물을 사용하는 PE 실험을 도시한다.

도 86은 HEK3 유전자의 표적화된 편집, 구체적으로 닉 부위에 대해 위치 +1에서의 71 nt FLAG 태그 삽입의 삽입을 표적화하고, PE3 구축물을 사용하는 PE 실험을 도시한다.

도 87은 pegRNA가 1412Adel을 설치한 N2A 세포에서의 스크린으로부터의 결과를 프라이머 결합 부위 (PBS) 길이 및 리버스 트랜스크립타제 (RT) 주형 길이에 관한 세부사항과 함께 제시한다. (indel의 존재 및 부재 하에 제시됨).

도 88은 pegRNA가 1412Adel을 설치한 N2A 세포에서의 스크린으로부터의 결과를 프라이머 결합 부위 (PBS) 길이 및 리버스 트랜스크립타제 (RT) 주형 길이에 관한 세부사항과 함께 제시한다. (indel의 존재 및 부재 하에 제시됨).

도 89는 pegRNA 및 닉킹 gRNA에 대한 편집제의 농도를 변화시키면서, 건강한 HSC에서 β-글로빈 유전자 내의 프록시 유전자좌 및 HEK3에서의 편집의 결과를 도시한다.

도 90은 이중-플랩 프라임 편집의 한 실시양태의 개략도를 제공한다. DNA 표적 서열은 2개의 프라임 편집 복합체 (pegRNA-A 및 pegRNA-B에 의해 가이드됨)에 의해 작용한다. 2개의 pegRNA는 이중 나선의 대향 가닥을 표적화한다. 각각의 프라임 편집제 (PE2·pegRNA)는 단일 DNA 가닥을 닉킹한 다음, 주형으로서 pegRNA를 사용하여 3' DNA 플랩을 합성한다. 2개의 프라임 편집제 복합체의 작용은 DNA의 대향 가닥 상에 2개의 3' 플랩을 함유하는 중간체의 생산을 유발한다. 2개의 3' 플랩은 그들의 3' 단부에서 서로에 대해 상보성이다. 3' 플랩의 3' 단부의 어닐링은 이중-듀플렉스 구조의 형성을 유발하고, 여기서 하나의 듀플렉스는 새로운 DNA 서열 (적색)을 함유하는 쌍형성된 3' 플랩으로 이루어지고, 다른 듀플렉스는 원래 DNA 서열 (흑색)을 함유하는 쌍형성된 5' 플랩으로 이루어진다. 개재되는 원래의 DNA 듀플렉스 (흑색 쌍형성된 5' 플랩)의 절제는 원래 DNA 서열을 대체한 목적하는 새로운 DNA 서열 (적색)을 함유하는 이중-닉킹된 DNA 종을 생성한다. 두 닉의 라이게이션은 편집 과정을 완료한다.

도 91a-91b는 실시예 7로부터의 결과를 제공한다. 도 91a는 이중 프라임 편집제를 사용하여 HEK293T 세포 내 HEK3 부위에서 90-bp 서열을 새로운 22-bp 서열로 대체한 Crispresso2 산출 대립유전자 표 (목적하는 생성물에 대해 정렬됨)를 보여준다. 목적하는 생성물은 서열분석 판독물의 80% 초과를 차지한다. 비교를 위해 참조 출발 대립유전자가 서열분석된 대립유전자 위에 제시된다. 도 91b는 도 91a에 제시된 서열 대체를 달성하는 데 사용된 pegRNA의 서열을 보여준다. pegRNA1 및 pegRNA2는 DNA 이중 나선의 상이한 가닥을 표적화하고, 도 90에 도시된 바와 같이 5' 대체된 닉을 생성한다.

도 92는 이중-플랩 프라임 편집을 위한 pegRNA 설계의 설계 실시양태를 보여준다. 2개의 pegRNA가 이중-플랩 프라임 편집에 사용되며, 이는 도면에 pegRNA A 및 pegRNA B로서 제시된다. 각각의 pegRNA는 프라임 편집 복합체를 표적 DNA 부위로 가이드하는 스페이서 서열 (암청색)을 함유한다. 2개의 pegRNA는 DNA 이중 나선의 대향 가닥을 표적화한다. 다른 pegRNA와 같이, 이중-플랩 프라임 편집 pegRNA는 닉킹된 게놈 DNA 가닥에 어닐링하여 역전사를 개시하는 프라이머 결합 서열 (PBS, 녹색) 및 리버스 트랜스크립타제 효소에 의한 새로운 DNA의 합성의 주형이 되는 역전사 주형 (RT 주형, 담청색)을 갖는 3' 연장부를 함유한다. 내인성 5' 플랩과 경쟁하기 위해 새로 합성된 편집된 3' 플랩을 필요로 하는, 전형적 프라임 편집에 사용된 pegRNA와 달리, RT 주형 내의 표적 부위에 대한 상동성을 코딩할 필요는 없다. 대신, 2개의 합성된 3' 플랩의 3' 단부는 서로에 대한 상보성만

을 함유할 필요가 있다 (즉, 3' 플랩의 3' 단부는 서로의 역 상보체 서열임). 이러한 상보성은 2개의 3' 플랩이 어닐링되고 목적하는 편집된 DNA 서열의 형성을 촉진하도록 한다.

도 93a-93b는 이중-플랩 프라임 편집에 의해 Bxb1 attB 및 attP 부위를 설치하기 위해 이중 프라임 편집을 사용한 실시예 7의 결과를 보여준다. 도 93a는 HEK3에서의 38-bp Bxb1 attB 부위의 설치를 보여준다. 6개의 pegRNA를 구축하였으며, 3개는 (+) 가닥을 표적화하고 (A1, A2 및 A3), 3개는 (-) 가닥을 표적화하였다 (B1, B2 및 B3). 이들은 RT 주형에서 코딩되는 attB 서열의 양이 상이하여, 2개의 플랩 사이에 상이한 수의 상보적 뉴클레오티드를 생성한다. pegRNA의 3x3 매트릭스를 HEK293T 세포 내 표적 게놈 위치에서의 attB 서열의 설치에 대해 평가하였다. 도 93a 및 도 93b는 HEK3에서의 50-bp Bxb1 attP 부위의 설치를 보여준다. 6개의 pegRNA를 구축하였으며, 3개는 (+) 가닥을 표적화하고 (A1, A2 및 A3), 3개는 (-) 가닥을 표적화하였다 (B1, B2 및 B3). 이들은 RT 주형에서 코딩되는 attP 서열의 양이 상이하여, 2개의 플랩 사이에 상이한 수의 상보적 뉴클레오티드를 생성한다. pegRNA의 3x3 매트릭스를 HEK293T 세포 내 표적 게놈 위치에서의 attP 서열의 설치에 대해 평가하였다. 도 93a 및 도 93b에서의 편집 둘 다에 대해, attB 또는 attP 부위의 설치는 2개의 닉 부위 사이에 위치한 90-bp의 게놈 DNA 서열의 공동 결실과 함께 발생한다.

도 94는 이중-플랩 프라임 편집에 의한 인간 세이프-하버 유전자좌에서의 Bxb1 attB 및 attP 부위의 설치의 결과를 보여준다. HEK293T 세포 내 AAVS1 유전자좌에서의 Bxb1 attP 부위 (좌측) 또는 CCR5 유전자좌에서의 Bxb1 attB 부위 (우측)의 설치. 정확한 편집은 청색 (AAVS1) 및 적색 (CCR5)으로 제시되고, 한편 indel 부산물은 회색으로 제시된다.

도 95는 사중-플랩 프라임 편집을 사용한 게놈 서열 역위의 개략도를 제공한다. 게놈 DNA의 영역이 역위를 위해 표적화된다 (녹색 및 오렌지색 절편). 4개의 pegRNA가 PE2 프라임 편집제에 의해 세포에 전달된다. 한 쌍의 pegRNA은 단일 게놈 DNA 가닥을 표적화하고, 2개의 상보적 DNA 플랩 (A 및 A', 청색)의 합성의 주형이 되고, 한편 제2 쌍은 다른 게놈 DNA 가닥을 표적화하고, 직교 DNA 서열을 갖는 2개의 상보적 DNA 플랩 (B 및 B', 분홍색)의 합성의 주형이 된다. 상보적 플랩은 어닐링되어 3' 오버행 듀플렉스를 형성한다. 5' 오버행 듀플렉스는 내인성 세포 복구 효소에 의해 절제된다. 닉은 라이게이션되어 역전된 DNA 서열 (녹색 및 오렌지색 절편) 및 역위 접합부에서의 pegRNA-주형화된 서열 (청색 및 분홍색 절편)을 함유하는 생성물 대립유전자를 생산한다.

도 96은 AAVS1 역위 접합부의 앰플리콘 서열분석의 결과를 보여준다. 사중-플랩 프라이밍된 편집 전략을 사용한 HEK293T 세포 내 AAVS1 유전자좌에서의 2.7-kb 역위의 CRISPResso2 분석 결과. 예상된 역위 접합부의 PCR 증폭 및 서열분석은 역위 접합부에 삽입된 Bxb1 attP 또는 attB 서열을 갖는 목적하는 생성물을 보여주었다.

도 97a-97b는 사중-플랩 프라임 편집을 사용한 게놈 내로의 원형 DNA 플라스미드의 표적화된 통합의 결과를 보여준다. (도 97a) 게놈 DNA의 영역 및 플라스미드 DNA의 영역이 사중 플랩 프라임 편집 통합을 위해 표적화된다. 4개의 pegRNA가 PE2 프라임 편집제에 의해 세포에 전달된다. 2개의 pegRNA는 상보적 서열의 주형이 되며, 하나는 단일 게놈 DNA 가닥을 표적화하고, 다른 하나는 단일 플라스미드 DNA 가닥을 표적화한다 (청색 플랩을 생성함). 다른 2개의 pegRNA는 처음 2개의 pegRNA의 것으로부터의 대향 게놈 DNA 및 플라스미드 DNA 가닥을 표적화하고, 이들은 제1 쌍에 직교하는 2개의 상보적 DNA 플랩 (분홍색)의 합성의 주형이 된다. 상보적 플랩은 어닐링되어 3' 오버행 듀플렉스를 형성한다. 5' 오버행 듀플렉스는 내인성 세포 복구 효소에 의해 절제된다. 닉은 라이게이션되어 통합된 플라스미드 DNA 서열 (녹색 및 오렌지색 절편) 및 통합 접합부에서의 pegRNA-주형화된 서열 (청색 및 분홍색 절편)을 함유하는 생성물 대립유전자를 생산한다. (도 97b) pegRNA-주형화된 attP 서열에 의해 가교된 플라스미드 백본 및 게놈 DNA 서열을 보여주는, 예상 접합부의 앰플리콘 서열분석의 CRISPResso2 분석.

도 98a-98b는 사중-플랩 프라임 편집을 사용한 표적화된 염색체 전위의 결과를 보여준다. (도 98a) pegRNA는 상이한 염색체 상의 2개의 영역을 표적화한다. 상보적 3' DNA 플랩은 2개의 염색체 서열을 가교시키고, 전위의 배향을 지시한다. (도 98b) HEK239T 세포 내 MYC 및 TIMM44 유전자좌 사이의 표적화된 전위. 염색체 8 상의 MYC 유전자좌와 염색체 19 상의 TIMM44 유전자좌 사이의 전위로부터의 예상 접합부의 앰플리콘 서열분석으로부터의 CRISPResso2 분석 결과. 대다수의 서열분석 판독물은 목적하는 대립유전자 서열에 상응한다.

도 99는 IDS 유전자좌에서 이중-플랩 프라임 편집에 의한 Bxb1 attB 및 attP 부위의 설치를 보여준다. HEK293T 세포를 PE2 및 상이한 쌍의 pegRNA (예를 들어, 제1 칼럼에서 pegRNA A1_a 및 pegRNA B2_a (attP 부위를 정방향으로 설치하기 위한 주형을 가짐))로 형질감염시켰다. 효율을 HTS에 의해 측정하였다. 이 데이터는 이중-플랩 편집이 관심 서열을 최대 ~80%의 효율로 IDS 유전자좌에 성공적으로 삽입할 수 있다는 것을 보여준다.

도 100a-100b는 이중-플랩-매개된 중복을 기재한다. 도 100a는 293T 세포 내 AAVS1 유전자좌에서의 이중-플랩-매개된 중복을 보여주는 개략도이다. 도 100b는 AAVS1에서 유전자 서열의 중복을 유도하기 위해 PE2와 함께 이중-플랩 pegRNA를 사용한 결과를 보여준다.

도 101은 MYC-CCR5의 다중-플랩 유도된 새로운 전위를 보여준다. MYC-CCR5 전위를 사중-플랩 pegRNA 및 PE2에 의해 유도하였다. MYC-CCR5 전위 사건을 사중-pegRNA에 의해 유도하였다. 4개의 상이한 세트의 pegRNA를 HEK293T 세포에서 시험하였다. 유도된 chr8과 chr3 사이의 전위 접합부 생성물을 접합부 프라이머에 의해 증폭시켰다. 예상 접합부 대립유전자에 대해 정렬된 판독물의 %가 그래프에 제시된다. 결과는 사중-플랩이 MYC 및 CCR5 유전자의 전위를 매개할 수 있으며, 생성물 순도는 접합부 1에서 거의 100% 및 접합부 2에서 ~50%라는 것을 보여준다. 대표적인 대립유전자 플롯은 접합부 1에서 예상 대립유전자 서열에 대해 정렬된 서열을 보여준다.

도 102a-102b는 다른 인간 세포주에서의 이중-플랩 및 다중-플랩 편집을 보여준다. 도 102a는 4종의 상이한 인간 세포주에서의 이중-플랩 편집을 보여준다. 3종의 상이한 게놈 유전자좌 (IDS, MYC 및 TIMM44)를 편집하기 위해 HEK293T 및 HeLa 세포를 이중 pegRNA 및 PE2로 형질감염시켰다. U2OS 및 K562 세포를 동일한 성분으로 뉴클레오펙션시켰다. 이중-플랩은 모든 4종의 인간 세포에 걸쳐, 특히 HEK293T 및 K562 세포에서 표적화된 유전자좌에서 강건한 편집을 보여주었다. 이중-플랩 편집을 가능하게 하는 세포 메카니즘은 많은 인간 세포 유형에 걸쳐 보존된다. 도 102b는 HeLa 세포 내 AAVS1에서의 다중-플랩 (사중-pegRNA) 지시된 2.7 kb 역위를 보여준다.

도 103a-103b는 CCR5 유전자좌에서 HTS에 의한 역위 효율 측정치의 결과를 보여준다. 예상 역위 편집 대립유전자의 백분율을 HTS에 의해 측정하였다. PE2의 4개의 사중-pegRNA 세트를 각각 HEK293T 세포에 형질감염시켰다. 도 103a는 HEK293T 세포 내 CCR5 유전자좌에서의 이중-플랩 매개된 서열 중복 (~100nt)을 보여준다. 편집 효율은 HTS 300-사이클 쌍-단부 서열 분석을 통해 ~1.5%를 달성한다. 도 103b는 HeLa 세포 내 CCR5 유전자좌에서의 사중-플랩-매개된 서열 역위 (~95-117nt)를 보여준다. 편집 효율은 HTS 300-사이클 쌍-단부 서열 분석을 통해 ~1.2%를 달성한다. 이러한 결과는 다중-플랩이 정확하게 CCR5 유전자좌에서의 중복 및 역위를 성공적으로 매개할 수 있다는 것을 보여준다. 편집 특이성은 표적화된 서열이 중복되는 경우에 높다 (indel의 백분율 < 2%).

도 104a-104e는 이중-플랩 편집을 위한 표적화된 세포 복구 경로를 보여준다. 도 104a-104d에서, HEK293T 세포를 pegRNA 및 PE2와 함께 Exo1, Fen1, 적색 형광 단백질 (대조군), DNA2, Mlh1 neg 및 P53 억제제를 발현하는 플라스미드로 형질감염시켰다. 편집 효율을 HTS에 의해 측정하였다. 편집 효율을 후보와 RFP 대조군 사이에 비교하였다. 도 104e에서, HEK293T 세포를 각각의 표적 유전자좌에 대해 siRNA 플라스미드 및 pegRNA 및 PE2로 형질감염시켰다. 비-표적화 siRNA (siNT)를 대조군으로서 사용하였다. 편집 효율을 HTS에 의해 측정하였다. 편집 효율을 각각의 표적 유전자좌에서의 각각의 siRNA 녹다운과 siNT ctrl 사이에 비교하였다. HEK293T 세포를 이중-pegRNA, PE2 및 각각 Exo1, Fen1, 적색 형광 단백질 (ctrl), DNA2, Mlh1 neg 및 P53 억제제를 발현하는 플라스미드로 형질감염시켰다. 편집 효율을 HTS에 의해 측정하였다. 편집 효율을 후보와 RFP ctrl 사이에 비교하였다. 양측, 대응표본 스튜던트 t-검정을 사용하여 각각의 처리군과 RFP 대조군 사이의 통계적 차이를 측정하였다 ($P < 0.05$, *; $P < 0.01$, **; $P < 0.001$, ***). FEN1의 과다발현은 모든 4종의 표적화된 유전자좌 (MYC, TIMM44, IDS, CCR5)에서 이중 플랩 편집 효율을 개선시킨다.

도 105a-105b는 AAVS1 유전자좌에서의 이중-플랩-매개된 서열 중복을 보여준다. 도 105a는 AAVS1 유전자좌에서의 이중-플랩-매개된 서열 중복의 개략적 다이어그램을 보여준다. 도 105b는 2개의 고유한 3' 플랩 구조를 생성하는 이중 pegRNA를 사용함으로써, 293T 세포 내 AAVS1 유전자좌에서 ~300bp 서열 중복이 유도되었다는 것을 보여준다. 예상 대립유전자를 특이적 프라이머로 증폭시키고, HTS에 적용하였다. ~94% 판독물을 중복을 갖는 예상 대립유전자에 대해 정렬하였다. 중복 생성물은 비처리된 샘플에서는 관찰되지 않는다.

도 106은 사중-플랩 프라임 편집에 의한 표적화된 IDS 게놈 서열 역위를 보여준다. 헌터 증후군 환자의 ~13%는 IDS 유전자 서열의 역위를 갖는 것으로 나타났다 (Bondeson et al., Human Molecular Genetics, 1995). 사중-플랩 프라임 편집을 적용하여 HEK293T 세포에서 ~40kb IDS 게놈 서열의 이러한 병원성 역위를 유도하였다. 6세트의 사중 pegRNA를, HEK293T 세포를 pegRNA 및 PE2로 형질감염시킴으로써 시험하였다. 프라이머를 사용하여 접합부 "ab" 및 접합부 "cd"에서 역전된 서열을 특이적으로 증폭시켰다. 예상된 역전된 대립유전자 서열의 ~95%가 IDS_QF1의 두 접합부에서 관찰되었다. 다른 세트의 pegRNA는 또한 두 접합부에서 높은 백분율의 예상 대립유전자 서열을 생성한다. 역전된 접합부 생성물은 비처리된 샘플에서는 관찰되지 않았다.

도 107은 PegRNA 3' 모티프 변형이 IDS 유전자좌에서 이중-플랩 편집 효율을 개선시킨다는 것을 보여준다. 이

중 플랩 편집 효율을 추가로 개선시키기 위해, 유사매듭형 evoPreQ1 모티프를 도입하여 pegRNA 3' 단부를 보호하였다. 비변형된 및 evoPreQ1-변형된 이중 pegRNA에 의해 생성된 편집 효율을 비교하면, 표적화된 IDS 유전자좌에서 변형된 pegRNA에 의한 편집 효율의 전체 증가가 존재한다. 이중-플랩 편집 효율의 개선은 최대 5.3배까지 도달할 수 있다.

도 108a-108c는 트윈PE 및 트윈PE-매개된 서열 대체의 개관을 보여준다. 도 108a는 트윈PE 시스템이 DNA의 대향 가닥 상에 2개의 프로토스페이서 서열을 함유하는 게놈 DNA 서열을 표적화한다는 것을 보여준다. PE2·pegRNA 복합체는 각각의 프로토스페이서를 표적화하고, 단일 가닥 닉을 생성하고, 목적하는 삽입 서열을 함유하는 pegRNA-코딩된 주형을 역전사시킨다. 3' DNA 플랩의 합성 및 방출 후, 편집된 DNA 서열을 함유하는 어닐링된 3' 플랩 및 원래 DNA 서열을 함유하는 어닐링된 5' 플랩을 보유하는 가설적 중간체가 존재한다. 5' 플랩에 함유된 원래 DNA 서열을 절제하고, 이어서 3' 플랩을 상응하는 절제 부위에 라이게이션하여, 목적하는 편집된 생성물을 생성한다. 도 108b는 HEK 부위 3에서의 90-bp 서열을 38-bp Bxb1 attB 서열로 트윈PE-매개된 대체한 예를 보여준다. 도 108c는 다양한 길이의 삽입 서열의 주형이 되는 pegRNA를 사용한, 도 108b에 제시된 바와 같은 38-bp Bxb1 attB 부위 또는 HEK 부위 3에서의 50-bp Bxb1 attP 부위의 설치에 대한 HEK293T 세포 내 트윈PE의 평가를 보여준다. 값 및 오차 막대는 3회의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 109a-109e는 인간 세포에서 트윈PE에 의한 표적화된 서열 삽입, 결실 및 재코딩을 보여준다. 도 109a는 HEK293T 세포 내 HEK 부위 3에서의 PE3 (12 bp, 36 bp, 108 bp 또는 321 bp) 또는 트윈PE (108 bp)에 의한 FKBP 코딩 서열 단편의 삽입을 보여준다. 도 109b는 트윈PE를 사용한 HEK293T 세포 내 PAH에서의 엑손 4 및 7 내의 서열의 재코딩을 보여준다. 엑손 4 내의 64-bp 표적 서열을 24, 36 또는 59 bp의 중첩 플랩을 사용하여 편집하거나, 엑손 7 내의 46-bp 표적 서열을 22 또는 42 bp의 중첩 플랩을 사용하여 편집하거나, 또는 엑손 7 내의 64-bp 서열을 24 또는 47 bp의 중첩 플랩을 사용하여 편집하였다. 표준 pegRNA 또는 3' evoPreQ1 모티프를 함유하는 epegRNA를 사용하여 편집 활성을 비교하였다. 도 109c는 표적화된 결실을 수행하기 위해 조사된 3가지 별개의 이중-플랩 결실 전략의 개략적 다이어그램이다. "베이직-앵커(Basic-Anchor; BA)" 트윈PE 전략은 하나의 닉 부위의 3'의 임의적 위치에서 시작하여 다른 닉 부위에서 끝나는 유연한 결실을 가능하게 한다. "하이브리드-앵커(Hybrid-Anchor; HA)" 트윈PE 전략은 2개의 닉 부위 사이의 임의적으로 선택된 위치에서 서열의 유연한 결실을 가능하게 한다. 본원에서 시험된 "프라임델(PrimeDel; PD)" 전략은 하나의 닉 부위에서 시작하여 또 다른 닉 부위에서 끝나는 서열의 결실을 가능하게 한다. 도 109d는 동일한 프로토스페이서 쌍을 표적화하는 BA-트윈PE, HA-트윈PE 또는 PD 전략을 사용한 HEK293T 세포 내 HEK 부위 3에서의 서열의 결실을 보여준다. 표준 pegRNA 또는 3' evoPreQ1 모티프를 함유하는 epegRNA를 사용하여 편집 활성을 비교하였다. 도 109e는 BA-트윈PE, PD, 쌍형성된 Cas9 뉴클레아제 또는 트윈PE-매개된 attB 서열 대체를 사용한 HEK293T 세포 내 DMD 유전자좌에서의 엑손 51 서열의 결실을 보여준다. 값 및 오차 막대는 3회의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다. DMD 엑손 51 스킵핑 실험에서, 적어도 2회의 독립적인 생물학적 반복실험을 수행하였다.

도 110a-110e는 인간 세포에서 트윈PE 및 Bxb1 레콤비나제에 의한 DNA 카고의 부위-특이적 게놈 통합을 보여준다. 도 110a는 HEK293T 세포 내 AAVS1 유전자좌에서의 Bxb1 attP 서열의 설치에 대한 트윈PE pegRNA 쌍의 스크리닝을 보여준다. 도 110b는 HEK293T 세포 내 CCR5 유전자좌에서의 Bxb1 attB 서열의 설치에 대한 트윈PE pegRNA 쌍의 스크리닝을 보여준다. 도 110c는 CCR5 (4개의 최좌측 막대) 또는 AAVS1 (3개의 최우측 막대)을 표적화하는 트윈PE pegRNA 쌍을 사용한 5.6-kb DNA 공여자의 단일 형질감염 녹-인(knock-in)을 보여준다. 트윈PE pegRNA는 CCR5에 attB를 또는 AAVS1에 attP를 설치한다. 이어서, Bxb1은 상응하는 부착 부위를 보유하는 공여자를 게놈 부착 부위 내로 통합시킨다. 도 110d는 531/584 트윈PE pegRNA 쌍을 사용한 CCR5에서의 단일 형질감염 녹-인의 최적화를 보여준다. 주형화된 편집의 정체 (attB 대 attP), 중심 디뉴클레오티드의 정체 (야생형 GT 대 직교 돌연변이체 GA) 및 플랩 사이의 중첩의 길이를 달리하여 최고 녹-인 효율을 지지하는 조합을 확인하였다. 도 110e는 HEK293T 및 Huh7 세포주에서 ALB의 인트론 1 내의 Bxb1 attB 서열의 삽입을 보여준다. 도 110f는 HEK293T 및 Huh7 세포주에서 CCR5 및 ALB에서의 단일 형질감염 녹-인 효율의 비교를 보여준다.

도 111a-111e는 인간 세포에서 트윈PE 및 Bxb1 레콤비나제를 사용한 부위-특이적 게놈 서열 역위를 보여준다. 도 111a는 병원성 39-kb 역위를 유도하는 IDS 및 IDS2에서의 재조합 핫 스폿 및 IDS 역위 돌연변이를 설치 또는 교정하기 위한 조합된 트윈PE-Bxb1 전략의 개략적 다이어그램이다. 도 111b는 특이적 DNA 표적을 갖는 IDS 및 IDS2 유전자좌에서의 attP 또는 attB 재조합 부위 삽입의 설치에 대한 IDS 및 IDS2에서의 pegRNA 쌍의 스크린을 보여준다. 도 111c는 순차적 DNA 형질감염을 사용한 attP 또는 attB 삽입의 DNA 서열 분석을 보여준다. 도 111d는 역전된 접합부 1 및 접합부 2 (순차적 형질감염)에서의 역위 생성물 순도를 보여주며, 이는 2개의 접합

부에서의 성공적인 역위를 나타낸다. 도 111e는 접합부 (순차적 형질감염 및 "원-포트" RNA 뉴클레오펙션)에서의 역위 효율의 정량화를 보여준다.

도 112는 트윈PE를 통한 HEK293T 세포 내 PAH에서의 엑손 10, 11 및 12 내의 서열의 재코딩을 보여준다. 엑손 10 내의 64-bp 표적 서열을 28 bp의 중첩 플랩을 사용하여 편집하거나, 엑손 11 내의 61-bp 및 55-bp 표적 서열을 25 bp의 중첩 플랩을 사용하여 편집하거나, 또는 엑손 12 내의 68-bp 및 58-bp 서열을 각각 27 및 24 bp의 중첩 플랩을 사용하여 편집하였다. 값 및 오차 막대는 3회의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 113은 BxBI 플라스미드 및 attP-함유 공여자 DNA 플라스미드에 의한 동형접합 attB 부위 삽입을 함유하는 HEK293T 클론 세포주의 형질감염을 보여준다. 녹-인 효율은 ddPCR에 의해 측정 시 표적 부위에서 12-17%이다.

도 114는 트윈PE 및 BxBI-매개된 원-포트 녹-인 후에 attL 및 attR 재조합 생성물을 함유하는 예상 접합부 서열의 HTS 측정치를 보여준다. 생성물 순도는 71-95% 범위이다. 값 및 오차 막대는 3회의 독립적인 생물학적 반복실험의 평균 및 s.d.를 반영한다.

도 115a는 이중 pegRNA에서의 감소된 플랩 중첩 길이에 의한 트윈PE 매개된 attB 삽입 효율을 보여준다.

도 115b는 아가로스 겔 상에 제시된 공여자 DNA와 pegRNA 플라스미드 사이의 재조합을 포획하기 위한 특이적 프라이머에 의해 증폭된 PCR 생성물을 보여준다. 공여자 DNA와 pegRNA 플라스미드 사이의 재조합은 보다 작은 플랩 중첩으로 감소되었다.

도 116은 IDS 역위 효율을 정량화하기 위해 개발된 PCR 전략의 개략적 다이어그램이다.

발명을 실시하기 위한 구체적인 내용

[0095] 정의

[0096] 달리 정의되지 않는 한, 본원에 사용된 모든 기술 과학 용어는 본 발명이 속하는 기술분야의 통상의 기술자에 의해 통상적으로 이해되는 의미를 갖는다. 하기 참고문헌은 통상의 기술자에게 본 발명에 사용된 많은 용어의 일반적 정의를 제공한다: [Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); 및 Hale & Marham, The Harper Collins Dictionary of Biology (1991)]. 본원에 사용된 하기 용어는 달리 명시되지 않는 한 이들에게 부여된 의미를 갖는다.

[0097] 안티센스 가닥

[0098] 유전학에서, 이중 가닥 DNA 내의 절편의 "안티센스" 가닥은 주형 가닥이고, 이는 3'에서 5' 배향으로 진행하는 것으로 간주된다. 대조적으로, "센스" 가닥은 5'에서 3'로 진행하는 이중 가닥 DNA 내의 절편이며, 이는 3'에서 5'로 진행하는 DNA의 안티센스 가닥 또는 주형 가닥에 상보적이다. 단백질을 코딩하는 DNA 절편의 경우에, 센스 가닥은 mRNA와 동일한 서열을 갖는 DNA의 가닥이며, 이는 전사 동안 안티센스 가닥을 그의 주형으로서 취하고, 결국 (전형적으로, 항상은 아님) 단백질로의 번역을 겪는다. 따라서, 안티센스 가닥은 나중에 단백질로 번역되는 RNA를 담당하는 반면, 센스 가닥은 mRNA의 것과 거의 동일한 구성을 갖는다. dsDNA의 각각의 절편에 대해, 판독하는 방향에 따라 센스 및 안티센스의 2개의 세트가 존재할 가능성이 있다는 것 (센스 및 안티센스는 시각이 상대적이기 때문에)을 주목한다. 궁극적으로, dsDNA의 한 절편 중 어느 가닥이 센스 또는 안티센스로 지칭되는지를 좌우하는 것은 유전자 생성물 또는 mRNA이다.

[0099] 이중특이적 리간드

[0100] 본원에 사용된 용어 "이중특이적 리간드" 또는 "이중특이적 모이어티"는 2종의 상이한 리간드-결합 도메인에 결합하는 리간드를 지칭한다. 특정 실시양태에서, 리간드는 소분자 화합물 또는 펩티드 또는 폴리펩티드이다. 다른 실시양태에서, 리간드-결합 도메인은 펩티드 태그로서 단백질 상에 설치될 수 있는 "이량체화 도메인"이다. 다양한 실시양태에서, 각각 동일하거나 상이한 이량체화 도메인을 포함하는 2종의 단백질은 이중특이적 리간드에 대한 각각의 이량체화 도메인의 결합을 통해 이량체화되도록 유도될 수 있다. 본원에 사용된 "이중특이적 리간드"는 동등하게 "이량체화의 화학적 유도제" 또는 "CID"를 지칭할 수 있다.

[0101] Cas9

[0102] 용어 "Cas9" 또는 "Cas9 뉴클레아제"는 Cas9 도메인 또는 그의 단편 (예를 들어, Cas9의 활성 또는 불활성 DNA 절단 도메인 및/또는 Cas9의 gRNA 결합 도메인을 포함하는 단백질)을 포함하는 RNA-가이드 뉴클레아제를 지칭한다. 본원에 사용된 "Cas9 도메인"은 Cas9의 활성 또는 불활성 절단 도메인 및/또는 Cas9의 gRNA 결합 도메인을 포함하는 단백질 단편이다. "Cas9 단백질"은 전장 Cas9 단백질이다. Cas9 뉴클레아제는 또한 때때로 casn1 뉴클레아제 또는 CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부)-연관 뉴클레아제로 지칭된다. CRISPR은 이동성 유전 요소 (바이러스, 전위가능 요소 및 접합 플라스미드)에 대한 보호를 제공하는 적응 면역계이다. CRISPR 클러스터는 스페이서, 이전의 이동성 요소에 상보적인 서열 및 표적 침입 핵산을 함유한다. CRISPR 클러스터는 CRISPR RNA (crRNA)로 전사 및 프로세싱된다. 유형 II CRISPR 시스템에서, 프리-crRNA의 정확한 프로세싱은 트랜스-코딩된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 도메인을 요구한다. tracrRNA는 프리-crRNA의 리보뉴클레아제 3-보조 프로세싱을 위한 가이드로서의 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 원형 dsDNA 표적을 핵산내부분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 핵산내부분해적으로 절단된 다음, 핵산외부분해적으로 3'-5' 트리밍된다. 사실상, DNA-결합 및 절단은 전형적으로 단백질 및 둘 다의 RNA를 요구한다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 간단히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 측면을 단일 RNA 종 내로 혼입하도록 조작될 수 있다. 예를 들어, 문헌 [Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)]을 참조하며, 이의 전체 내용은 본원에 참조로 포함된다. Cas9는 CRISPR 반복 서열 내의 짧은 모티프 (PAM 또는 프로토스페이서 인접 모티프)를 인식하여 자기 대 비-자기 구별을 돕는다. Cas9 뉴클레아제 서열 및 구조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 ["Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); 및 "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)] 참조, 이들 각각의 전체 내용은 본원에 참조로 포함됨). Cas9 오르토로그는 에스. 피오게네스(*S. pyogenes*) 및 에스. 써모필루스(*S. thermophilus*)를 포함하나 이에 제한되지는 않는 다양한 종에서 기재되어 있다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이고, 이러한 Cas9 뉴클레아제 및 서열은 그의 전체 내용이 본원에 참조로 포함되는 문헌 [Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737]에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다. 일부 실시양태에서, Cas9 뉴클레아제는 DNA 절단 도메인을 부분적으로 손상 또는 불활성화시키는 1개 이상의 돌연변이를 포함한다.

[0103] 뉴클레아제-불활성화된 Cas9 도메인은 상호교환가능하게 "dCas9" 단백질 (뉴클레아제-"기능상실" Cas9의 경우)로 지칭될 수 있다. 불활성 DNA 절단 도메인을 갖는 Cas9 도메인 (또는 그의 단편)을 생성하는 방법은 공지되어 있다 (예를 들어, 문헌 [Jinek et al., Science. 337:816-821(2012); Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) Cell. 28;152(5):1173-83] 참조, 이들 각각의 전체 내용은 본원에 참조로 포함됨). 예를 들어, Cas9의 DNA 절단 도메인은 2개의 서브도메인, 즉 HNH 뉴클레아제 서브도메인 및 RuvC1 서브도메인을 포함하는 것으로 공지되어 있다. HNH 서브도메인은 gRNA에 상보적인 가닥을 절단하는 반면, RuvC1 서브도메인은 비-상보적 가닥을 절단한다. 이들 서브도메인 내의 돌연변이는 Cas9의 뉴클레아제 활성을 침묵시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 에스. 피오게네스 Cas9의 뉴클레아제 활성을 완전히 불활성화시킨다 (Jinek et al., Science. 337:816-821(2012); Qi et al., Cell. 28;152(5):1173-83 (2013)). 일부 실시양태에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시양태에서, 단백질은 다음 2개의 Cas9 도메인 중 하나를 포함한다: (1) Cas9의 gRNA 결합 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 실시양태에서, Cas9 또는 그의 단편을 포함하는 단백질은 "Cas9 변이체"로 지칭된다. Cas9 변이체는 Cas9 또는 그의 단편에 대한 상동성을 공유한다. 예를 들어, Cas9 변이체는 야생형 Cas9 (예를 들어 서열식별번호: 18의 SpCas9)에 대해 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 96% 동일하거나, 적어도 약 97% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나, 적어도 약 99.8% 동일하거나 또는 적어도 약 99.9% 동일하다. 일부 실시양태에서, Cas9 변이체는 야생형 Cas9 (예를

들어 서열식별번호: 18의 SpCas9)와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 또는 그 초과의 아미노산 변화를 가질 수 있다. 일부 실시양태에서, Cas9 변이체는 서열식별번호: 18 Cas9의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함하며, 이러한 단편은 야생형 Cas9 (예를 들어, 서열식별번호: 18의 SpCas9)의 상응하는 단편에 대해 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 96% 동일하거나, 적어도 약 97% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나 또는 적어도 약 99.9% 동일하다. 일부 실시양태에서, 단편은 상응하는 야생형 Cas9 (예를 들어, 서열식별번호: 18의 SpCas9)의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 동일, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5%이다.

[0104] cDNA

[0105] 용어 "cDNA"는 RNA 주형으로부터 카피된 DNA의 가닥을 지칭한다. cDNA는 RNA 주형에 상보적이다.

[0106] 원형 순열체

[0107] 본원에 사용된 용어 "원형 순열체"는 단백질의 아미노산 서열에 나타나는 아미노산의 순서의 변화를 수반하는 단백질의 구조적 구성의 변화인 원형 순열을 포함하는 단백질 또는 폴리펩티드 (예를 들어, Cas9)를 지칭한다. 다시 말해서, 원형 순열체는 야생형 대응물과 비교하여 N- 및 C-말단이 변경된 단백질이며, 예를 들어 단백질의 야생형 C-말단 절반이 새로운 N-말단 절반이 된다. 원형 순열 (또는 CP)은 본질적으로, 그의 서열을 상이한 위치에서 분할하여 새로운 인접한 N- 및 C-말단을 생성하면서 동시에 그의 N- 및 C-말단을 종종 펩티드 링커와 연결하는, 단백질의 1차 서열의 위상 재배열이다. 그 결과는, 상이한 연결성을 갖지만 종종 동일한 전체 유사한 3차원 (3D) 형상을 가질 수 있고, 가능하게는 감소된 단백질분해 감수성, 개선된 촉매 활성, 변경된 기질 또는 리간드 결합 및/또는 개선된 열안정성을 포함한 개선된 또는 변경된 특징을 포함할 수 있는 단백질 구조이다. 원형 순열체 단백질은 자연에서 발생할 수 있다 (예를 들어, 콘카나발린 A 및 렉틴). 추가로, 원형 순열은 번역후 변형의 결과로서 발생할 수 있거나 또는 재조합 기술을 사용하여 조작될 수 있다.

[0108] 원형 순열 Cas9

[0109] 용어 "원형 순열 Cas9"는 그의 N- 및 C-말단이 국소로 재배열된, 원형 순열체로서 발생한 임의의 Cas9 단백질 또는 그의 변이체를 지칭한다. 이러한 원형 순열 Cas9 단백질 ("CP-Cas9") 또는 그의 변이체는 가이드 RNA (gRNA)와 복합체화되는 경우 DNA에 결합하는 능력을 보유한다. 문헌 [Oakes et al., "Protein Engineering of Cas9 for enhanced function," Methods Enzymol, 2014, 546:491-511 and Oakes et al., "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," Cell, January 10, 2019, 176: 254-267]을 참조하며, 이들 각각은 본원에 참조로 포함된다. 본 개시내용은 임의의 이전에 공지된 CP-Cas9를 고려하거나 또는 생성된 원형 순열 단백질이 가이드 RNA (gRNA)와 복합체화되는 경우 DNA에 결합하는 능력을 보유하는 한, 새로운 CP-Cas9의 사용을 고려한다. 예시적인 CP-Cas9 단백질은 서열식별번호: 77-86이다.

[0110] CRISPR

[0111] CRISPR은 원핵생물을 침입한 바이러스에 의한 이전 감염의 조각(snippets)을 나타내는 박테리아 및 고세균에서의 DNA 서열 (즉, CRISPR 클러스터)의 패밀리이다. DNA의 조각은 유사한 바이러스에 의한 후속 공격으로부터 DNA를 검출하고 파괴하기 위해 원핵 세포에 의해 사용되고, CRISPR-연관 단백질 (Cas9 및 그의 상동체 포함) 및 CRISPR-연관 RNA의 어레이와 함께 원핵 면역 방어 시스템을 효과적으로 구성한다. 사실상, CRISPR 클러스터는 CRISPR RNA (crRNA)로 전사 및 프로세싱된다. 특정 유형의 CRISPR 시스템 (예를 들어, 유형 II CRISPR 시스템)에서, 프리-crRNA의 정확한 프로세싱은 트랜스-코딩된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질을 요구한다. tracrRNA는 프리-crRNA의 리보뉴클레아제 3-보조 프로세싱을 위한 가이드로서의 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 RNA에 상보적인 선형 또는 원형 dsDNA 표적을 핵산내부분해적으로 절단한다. 구체적으로, crRNA에 상보적이지 않은 표적 가닥은 먼저 핵산내부분해적으로 절단된 다음, 핵산외부분해적으로 3'-5' 트리밍된다. 사실상, DNA-결합 및 절단은 전형적으로 단백질 및 둘 다의 RNA를 요구한다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 간단히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 측면을 단일 RNA 종 - 가이드 RNA 내로 혼입하도록 조작될 수 있다. 예를 들어, 문헌 [Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)]을 참조하며, 이의 전체 내용

은 본원에 참조로 포함된다. Cas9는 CRISPR 반복 서열 내의 짧은 모티프 (PAM 또는 프로토스페이서 인접 모티프)를 인식하여 자기 대 비-자기 구별을 돕는다. CRISPR 생물학, 뿐만 아니라 Cas9 뉴클레아제 서열 및 구조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 ["Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); 및 "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)] 참조, 이들 각각의 전체 내용은 본원에 참조로 포함됨). Cas9 오르토로그는 에스. 피오게네스 및 에스. 써모필루스를 포함하나 이에 제한되지는 않는 다양한 종에서 기재되어 있다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이고, 이러한 Cas9 뉴클레아제 및 서열은 그의 전체 내용이 본원에 참조로 포함되는 문헌 [Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737]에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다.

[0112] 특정 유형의 CRISPR 시스템 (예를 들어, 유형 II CRISPR 시스템)에서, 프리-crRNA의 정확한 프로세싱은 트랜스-코딩된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질을 요구한다. tracrRNA는 프리-crRNA의 리보뉴클레아제 3-보조 프로세싱을 위한 가이드로서의 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 RNA에 상보적인 선형 또는 원형 핵산 표적을 핵산내부분해적으로 절단한다. 구체적으로, crRNA에 상보적이지 않은 표적 가닥은 먼저 핵산내부분해적으로 절단된 다음, 핵산외부분해적으로 3'-5' 트리밍된다. 사실상, DNA-결합 및 절단은 전형적으로 단백질 및 둘 다의 RNA를 요구한다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 간단히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 실시양태를 단일 RNA 종-가이드 RNA 내로 혼입하도록 조작될 수 있다.

[0113] 일반적으로, "CRISPR 시스템"은 Cas 유전자, tracr (트랜스-활성화 CRISPR) 서열 (예를 들어 tracrRNA 또는 활성 부분 tracrRNA), tracr 메이트 서열 (내인성 CRISPR 시스템의 맥락에서 "직접 반복부" 및 tracrRNA-프로세싱된 부분 직접 반복부를 포괄함), 가이드 서열 (내인성 CRISPR 시스템의 맥락에서 "스페이서"로도 지칭됨), 또는 CRISPR 유전자좌로부터의 다른 서열 및 전사체를 코딩하는 서열을 포함한, CRISPR-연관 ("Cas") 유전자의 발현에 수반되거나 또는 그의 활성을 지시하는 전사체 및 다른 요소를 집합적으로 지칭한다. 시스템의 tracrRNA는 가이드 RNA 상에 존재하는 tracr 메이트 서열에 (완전히 또는 부분적으로) 상보적이다.

[0114] DNA 합성 주형

[0115] 본원에 사용된 용어 "DNA 합성 주형"은 목적하는 편집을 함유하는 3' 단일 가닥 DNA 플랩을 코딩하기 위해 프라임 편집제의 폴리머라제에 의해 주형 가닥으로서 이용되고, 이어서 프라임 편집의 메카니즘을 통해 표적 부위에서 DNA의 상응하는 내인성 가닥을 대체하는 PEgRNA의 연장 아암의 영역 또는 부분을 지칭한다. 다양한 실시양태에서, DNA 합성 주형은 도 3a (5' 연장 아암을 포함하는 PEgRNA의 맥락에서), 도 3b (3' 연장 아암을 포함하는 PEgRNA의 맥락에서), 도 3c (내부 연장 아암의 맥락에서), 도 3d (3' 연장 아암의 맥락에서), 및 도 3e (5' 연장 아암의 맥락에서)에 제시된다. DNA 합성 주형을 포함한 연장 아암은 DNA 또는 RNA로 구성될 수 있다. RNA의 경우에, 프라임 편집제의 폴리머라제는 RNA-의존성 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)일 수 있다. DNA의 경우에, 프라임 편집제의 폴리머라제는 DNA-의존성 DNA 폴리머라제일 수 있다. 다양한 실시양태에서 (예를 들어, 도 3d-3e에 도시된 바와 같이), DNA 합성 주형 (4)은 "편집 주형" 및 "상동성 아암", 및 임의적인 5' 단부 변형제 영역 e2의 모두 또는 부분을 포함할 수 있다. 즉, e2 영역의 성질 (예를 들어, 헤어핀, 토루프 또는 스템/루프 2차 구조를 포함하는지 여부)에 따라, 폴리머라제는 또한 e2 영역 중 어느 것도 코딩하지 않을 수 있거나, 일부 또는 모두를 코딩할 수 있다. 상기 또 다른 방식으로, 3' 연장 아암의 경우에, DNA 합성 주형 (3)은 프라이머 결합 부위 (PBS)의 5' 단부로부터, 폴리머라제 (예를 들어, 리버스 트랜스크립타제)에 의한 DNA의 단일 가닥의 합성을 위한 주형으로서 작동할 수 있는 gRNA 코어의 3' 단부까지 걸쳐 있는 연장 아암 (3)의 부분을 포함할 수 있다. 5' 연장 아암의 경우에, DNA 합성 주형 (3)은 PEgRNA 분자의 5' 단부로부터 편집 주형의 3' 단부까지 걸쳐 있는 연장 아암 (3)의 부분을 포함할 수 있다. 바람직하게는, DNA 합성 주형은 3' 연장 아암 또는 5' 연장 아암을 갖는 PEgRNA의 프라이머 결합 부위 (PBS)를 배제한다. 본원에 기재된 특정 실시양태 (예를 들어, 도 71a)는 편집 주형 및 상동성 아암을 포함하는 "RT 주형", 즉 DNA 합성 동안 주형

으로서 실제로 사용되는 PEgRNA 연장 아암의 서열을 지칭한다. 용어 "RT 주형"은 용어 "DNA 합성 주형"과 동등하다.

[0116] 트랜스 프라임 편집 (예를 들어, 도 3g 및 도 3h)의 경우에, 프라이머 결합 부위 (PBS) 및 DNA 합성 주형은 트랜스 프라임 편집제 RNA 주형 (tPERT)으로 지칭되는 개별 분자 내로 조작될 수 있다.

[0117] 이량체화 도메인

[0118] 용어 "이량체화 도메인"은 이중특이적 리간드의 결합 모이어티에 결합하는 리간드-결합 도메인을 지칭한다. "제1" 이량체화 도메인은 이중특이적 리간드의 제1 결합 모이어티에 결합하고, "제2" 이량체화 도메인은 동일한 이중특이적 리간드의 제2 결합 모이어티에 결합한다. 제1 이량체화 도메인이 (예를 들어, 본원에 논의된 바와 같은 PE를 통해) 제1 단백질에 융합되고, 제2 이량체화 도메인이 (예를 들어, 본원에 논의된 바와 같은 PE를 통해) 제2 단백질에 융합되는 경우에, 제1 및 제2 단백질은 이중특이적 리간드의 존재 하에 이량체화되며, 여기서 이중특이적 리간드는 제1 이량체화 도메인에 결합하는 적어도 하나의 모이어티 및 제2 이량체화 도메인에 결합하는 적어도 또 다른 모이어티를 갖는다.

[0119] 하류

[0120] 본원에 사용된 용어 "상류" 및 "하류"는 5'에서 3' 방향으로 배향되는 핵산 분자 (단일 가닥이든 이중 가닥이든)에 위치한 적어도 2개의 요소의 선형 위치를 정의하는 상대성 용어이다. 특히, 제1 요소는 핵산 분자에서 제2 요소의 상류에 있으며, 여기서 제1 요소는 제2 요소에 대해 5'인 어딘가에 위치한다. 예를 들어, SNP가 닉 부위의 5' 측면 상에 있는 경우에 SNP는 Cas9-유도된 닉 부위의 상류에 있다. 반대로, 제1 요소는 핵산 분자에서 제2 요소의 하류에 있으며, 여기서 제1 요소는 제2 요소에 대해 3'인 어딘가에 위치한다. 예를 들어, SNP가 닉 부위의 3' 측면 상에 있는 경우에 SNP는 Cas9-유도된 닉 부위의 하류에 있다. 핵산 분자는 DNA (이중 또는 단일 가닥), RNA (이중 또는 단일 가닥), 또는 DNA와 RNA의 하이브리드일 수 있다. 분석은 단일 가닥 핵산 분자 및 이중 가닥 분자에 대해 동일하며, 이는 용어 상류 및 하류가 이중 가닥 분자의 어느 가닥이 고려되는지를 선택할 필요가 있다는 것을 제외하고는 핵산 분자의 단일 가닥만을 언급하기 때문이다. 종종, 적어도 2개의 요소의 위치적 상대성을 결정하는 데 사용될 수 있는 이중 가닥 DNA의 가닥은 "센스" 또는 "코딩" 가닥이다. 유전학에서, "센스" 가닥은 5'에서 3'로 진행하는 이중 가닥 DNA 내의 절편이며, 이는 3'에서 5'로 진행하는 DNA의 안티센스 가닥 또는 주형 가닥에 상보적이다. 따라서, 예로서, SNP 핵염기는 SNP 핵염기가 센스 또는 코딩 가닥 상의 프로모터의 3' 측면 상에 존재하는 경우에 게놈 DNA (이중 가닥임) 내의 프로모터 서열의 "하류"에 있다.

[0121] 편집 주형

[0122] 용어 "편집 주형"은 폴리머라제, 예를 들어 DNA-의존성 DNA 폴리머라제, RNA-의존성 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)에 의해 합성된 단일 가닥 3' DNA 플랩에서 목적하는 편집을 코딩하는 연장 아암의 부분을 지칭한다. 본원에 기재된 특정 실시양태 (예를 들어, 도 71a)는 편집 주형 및 상동성 아암 둘 다를 함께 지칭하는 "RT 주형", 즉 DNA 합성 동안 주형으로서 실제로 사용되는 PEgRNA 연장 아암의 서열을 지칭한다. 용어 "RT 편집 주형"은 또한 용어 "DNA 합성 주형"과 동등하지만, 여기서 RT 편집 주형은 리버스 트랜스크립타제인 폴리머라제를 갖는 프라임 편집제의 사용을 반영하고, 여기서 DNA 합성 주형은 임의의 폴리머라제를 갖는 프라임 편집제의 사용을 보다 광범위하게 반영한다.

[0123] 유효량

[0124] 본원에 사용된 용어 "유효량"은 목적하는 생물학적 반응을 도출하기에 충분한 생물학적 활성제의 양을 지칭한다. 예를 들어, 일부 실시양태에서, 프라임 편집제 (PE)의 유효량은 표적 부위 뉴클레오티드 서열, 예를 들어 게놈을 편집하기에 충분한 편집제의 양을 지칭할 수 있다. 일부 실시양태에서, 본원에 제공된 프라임 편집제 (PE), 예를 들어 닉카제 Cas9 도메인 및 리버스 트랜스크립타제를 포함하는 융합 단백질의 유효량은 융합 단백질에 의해 특이적으로 결합되고 편집되는 표적 부위의 편집을 유도하기에 충분한 융합 단백질의 양을 지칭할 수 있다. 통상의 기술자에 의해 인지될 바와 같이, 작용제, 예를 들어 융합 단백질, 뉴클레아제, 하이브리드 단백질, 단백질 이량체, 단백질 (또는 단백질 이량체) 및 폴리뉴클레오티드의 복합체, 또는 폴리뉴클레오티드의 유효량은 다양한 인자, 예를 들어 목적하는 생물학적 반응, 예를 들어 편집될 특정 대립유전자, 게놈 또는 표적 부위, 표적화될 세포 또는 조직, 및 사용될 작용제에 따라 달라질 수 있다.

[0125] 오류-유발 리버스 트랜스크립타제

[0126] 본원에 사용된 용어 "오류-유발" 리버스 트랜스크립타제 (또는 보다 넓게는, 임의의 폴리머라제)는 자연 발생하거나 또는 야생형 M-MLV 리버스 트랜스크립타제의 오류율보다 더 적은 오류율을 갖는 또 다른 리버스 트랜스크립타제 (예를 들어, 야생형 M-MLV 리버스 트랜스크립타제)로부터 유래된 리버스 트랜스크립타제 (또는 보다 넓게는, 임의의 폴리머라제)를 지칭한다. 야생형 M-MLV 리버스 트랜스크립타제의 오류율은 15,000 (더 높음) 내지 27,000개 (더 낮음)에서 1개의 오류 범위인 것으로 보고되어 있다. 15,000개 중 1개의 오류율은 6.7×10^{-5}의 오류율에 상응한다. 27,000개 중 1개의 오류율은 3.7×10^{-5}의 오류율에 상응한다. 본원에 참조로 포함되는 문헌 [Boutabout et al. (2001) "DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1," Nucleic Acids Res 29(11):2217-2222]을 참조한다. 따라서, 본 출원의 목적상, 용어 "오류 유발"은 15,000개 핵염기 혼입에서 1개의 오류 초과 (6.7×10^{-5} 이상), 예를 들어 14,000개 핵염기에서 1개의 오류 (7.14×10^{-5} 이상), 13,000개 핵염기 이하에서 1개의 오류 (7.7×10^{-5} 이상), 12,000개 핵염기 이하에서 1개의 오류 (7.7×10^{-5} 이상), 11,000개 핵염기 이하에서 1개의 오류 (9.1×10^{-5} 이상), 10,000개 핵염기 이하에서 1개의 오류 (1×10^{-4} 또는 0.0001 이상), 9,000개 핵염기 이하에서 1개의 오류 (0.00011 이상), 8,000개 핵염기 이하에서 1개의 오류 (0.00013 이상), 7,000개 핵염기 이하에서 1개의 오류 (0.00014 이상), 6,000개 핵염기 이하에서 1개의 오류 (0.00016 이상), 5,000개 핵염기 이하에서 1개의 오류 (0.0002 이상), 4,000개 핵염기 이하에서 1개의 오류 (0.00025 이상), 3,000개 핵염기 이하에서 1개의 오류 (0.00033 이상), 2,000개 핵염기 이하에서 1개의 오류 (0.00050 이상), 또는 1,000개 핵염기 이하에서 1개의 오류 (0.001 이상), 또는 500개 핵염기 이하에서 1개의 오류 (0.002 이상), 또는 250개 핵염기 이하에서 1개의 오류 (0.004 이상)인 오류율을 갖는 RT를 지칭한다.

[0127] 엑스테인

[0128] 본원에 사용된 용어 "엑스테인"은 인테인이 플랭킹되어 있는 폴리펩티드 서열을 지칭하고, 단백질 스플라이싱 과정 동안 또 다른 엑스테인에 라이게이션되어 성숙한 스플라이싱된 단백질을 형성한다. 전형적으로, 인테인은 인테인이 그 자신의 절제를 촉매할 때 함께 라이게이션되는 2개의 엑스테인 서열이 플랭킹되어 있다. 따라서, 엑스테인은 mRNA에서 발견되는 엑손과 유사한 단백질이다. 예를 들어, 인테인을 포함하는 폴리펩티드는 구조 엑스테인(N) - 인테인 - 엑스테인(C)의 것일 수 있다. 인테인의 절제 및 2개의 엑스테인의 스플라이싱 후, 생성된 구조는 엑스테인(N) - 엑스테인(C) 및 유리 인테인이다. 다양한 구성에서, 엑스테인은 각각 분할-인테인에 융합된 별개의 단백질 (예를 들어, Cas9 또는 PE 융합 단백질의 절반)일 수 있으며, 여기서 분할 인테인의 절제는 엑스테인 서열들이 함께 스플라이싱되도록 유발한다.

[0129] 연장 아암

[0130] 용어 "연장 아암"은 리버스 트랜스크립타제에 대한 프라이머 결합 부위 및 편집 주형을 포함한, 여러 기능을 제공하는 PEgRNA의 뉴클레오티드 서열 성분을 지칭한다. 일부 실시양태, 예를 들어 도 3d에서, 연장 아암은 가이드 RNA의 3' 단부에 위치한다. 다른 실시양태, 예를 들어 도 3e에서, 연장 아암은 가이드 RNA의 5' 단부에 위치한다. 일부 실시양태에서, 연장 아암은 또한 상동성 아암을 포함한다. 다양한 실시양태에서, 연장 아암은 하기 성분을 5'에서 3' 방향으로 포함한다: 상동성 아암, 편집 주형 및 프라이머 결합 부위. 리버스 트랜스크립타제의 중합 활성이 5'에서 3' 방향으로 존재하기 때문에, 상동성 아암, 편집 주형 및 프라이머 결합 부위의 바람직한 배열은 5'에서 3' 방향으로 존재하여, 리버스 트랜스크립타제가 어닐링된 프라이머 서열에 의해 프라이밍되면, 상보적 주형 가닥으로서 편집 주형을 사용하여 DNA의 단일 가닥을 중합시킨다. 추가의 세부사항, 예컨대 연장 아암의 길이는 본원의 다른 곳에 기재되어 있다.

[0131] 연장 아암은 또한, 예를 들어 도 3g (상단)에 제시된 바와 같이, 일반적으로 2개의 영역: 프라이머 결합 부위 (PBS) 및 DNA 합성 주형을 포함하는 것으로 기재될 수 있다. 프라이머 결합 부위는 프라임 편집제 복합체에 의해 닉킹될 때 표적 부위의 내인성 DNA 가닥으로부터 형성된 프라이머 서열에 결합하여, 내인성 닉킹된 가닥 상의 3' 단부를 노출시킨다. 본원에 설명된 바와 같이, PEgRNA의 연장 아암 상의 프라이머 결합 부위에 대한 프라이머 서열의 결합은 노출된 3' 단부 (즉, 프라이머 서열의 3')를 갖는 듀플렉스 영역을 생성하고, 이는 이어서 폴리머라제가 DNA 합성 주형의 길이를 따라 노출된 3' 단부로부터 DNA의 단일 가닥을 중합시키기 시작하기 위한 기질을 제공한다. 단일 가닥 DNA 생성물의 서열은 DNA 합성 주형의 상보체이다. 중합은 중합이 종결될 때까지 DNA 합성 주형 (또는 연장 아암)의 5'를 향해 계속된다. 따라서, DNA 합성 주형은 프라임 편집제 복합체의 폴리머라제에 의해 단일 가닥 DNA 생성물 (즉, 목적하는 유전자 편집 정보를 함유하는 3' 단일 가닥 DNA

플랩)로 코딩되고 궁극적으로 PE-유도된 닉 부위의 바로 하류에 위치하는 표적 부위의 상응하는 내인성 DNA 가닥을 대체하는 연장 아암의 부분을 나타낸다. 이론에 얽매이지는 않지만, DNA 합성 주형의 중합은 종결 사건까지 연장 아암의 5' 단부를 향해 계속된다. 중합은 (a) PEgRNA의 5' 말단에 도달하는 것 (예를 들어, DNA 폴리머라제가 단순히 주형으로부터 진행하는 5' 연장 아암의 경우), (b) 통과할 수 없는 RNA 2차 구조 (예를 들어, 헤어핀 또는 스템/루프)에 도달하는 것, 또는 (c) 복제 종결 신호, 예를 들어 폴리머라제를 차단 또는 억제하는 특이적 뉴클레오티드 서열, 또는 핵산 위상 신호, 예컨대 슈퍼코일드 DNA 또는 RNA에 도달하는 것을 포함하나 이에 제한되지는 않는 다양한 방식으로 종결될 수 있다.

[0132] 플랩 엔도뉴클레아제 (예를 들어, FEN1)

[0133] 본원에 사용된 용어 "플랩 엔도뉴클레아제"는 5' 단일 가닥 DNA 플랩의 제거를 촉매하는 효소를 지칭한다. 이들은 DNA 복제를 비롯한 세포 과정 동안 형성된 5' 플랩의 제거를 프로세싱하는 자연 발생 효소이다. 본원에 기재된 프라임 편집 방법은 프라임 편집 동안 표적 부위에 형성된 내인성 DNA의 5' 플랩을 제거하기 위해 내인성으로 공급된 플랩 엔도뉴클레아제 또는 트랜스로 제공된 것을 이용할 수 있다. 플랩 엔도뉴클레아제는 관련 기술분야에 공지되어 있고, 문헌 [Patel et al., "Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends," Nucleic Acids Research, 2012, 40(10): 4507-4519, Tsutakawa et al., "Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily," Cell, 2011, 145(2): 198-211, 및 Balakrishnan et al., "Flap Endonuclease 1," Annu Rev Biochem, 2013, Vol 82: 119-138] (이들 각각은 본원에 참조로 포함됨)에 기재된 바와 같이 발견될 수 있다. 예시적인 플랩 엔도뉴클레아제는 하기 아미노산 서열로 나타내어질 수 있는 FEN1이다:

[0134]

설명	서열	SEQ ID NO:
FEN1 야생형	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVDL CILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDPN KYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPNE EELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 7

[0135] 기능적 등가물

[0136] 용어 "기능적 등가물"은 제1 생체분자와 기능상 등가이지만 구조상 반드시 등가인 것은 아닌 제2 생체분자를 지칭한다. 예를 들어, "Cas9 등가물"은 Cas9와 동일하거나 실질적으로 동일한 기능을 갖지만 반드시 동일한 아미노산 서열을 갖는 것은 아닌 단백질을 지칭한다. 본 개시내용의 문맥에서, 본 명세서는 전반에 걸쳐 "단백질 X 또는 그의 기능적 등가물"을 언급한다. 이러한 문맥에서, 단백질 X의 "기능적 등가물"은 등가의 기능을 보유하는 단백질 X의 임의의 상동체, 파라로그, 단편, 자연 발생, 조작된, 돌연변이된 또는 합성 버전을 포괄한다.

[0137] 융합 단백질

[0138] 본원에 사용된 용어 "융합 단백질"은 적어도 2종의 상이한 단백질로부터의 단백질 도메인을 포함하는 하이브리드 폴리펩티드를 지칭한다. 1종의 단백질은 융합 단백질의 아미노-말단 (N-말단) 부분에 또는 카르복시-말단 (C-말단) 단백질에 위치하며, 따라서 각각 "아미노-말단 융합 단백질" 또는 "카르복시-말단 융합 단백질"을 형성할 수 있다. 단백질은 상이한 도메인, 예를 들어 핵산 결합 도메인 (예를 들어, 표적 부위에 대한 단백질의 결합을 지시하는 Cas9의 gRNA 결합 도메인) 및 핵산 절단 도메인 또는 핵산 편집 단백질의 촉매 도메인을 포함할 수 있다. 또 다른 예는 리버스 트랜스크립타제에 대한 Cas9 또는 그의 등가물을 포함한다. 본원에 제공된 임의의 단백질은 관련 기술분야에 공지된 임의의 방법에 의해 생산될 수 있다. 예를 들어, 본원에 제공된 단백질은 재조합 단백질 발현 및 정제를 통해 생산될 수 있으며, 이는 펩티드 링커를 포함하는 융합 단백질에 특히 적합하다. 재조합 단백질 발현 및 정제를 위한 방법은 널리 공지되어 있고, 문헌 [Green and Sambrook, Molecular Cloning: A Laboratory Manual (4^{th} ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 기재된 것을 포함하며, 이의 전체 내용은 본원에 참조로 포함된다.

[0139] 관심 유전자 (GOI)

[0140] 용어 "관심 유전자" 또는 "GOI"는 관심 생체분자 (예를 들어, 단백질 또는 RNA 분자)를 코딩하는 유전자를 지칭한다. 관심 단백질은 임의의 세포내 단백질, 막 단백질 또는 세포외 단백질, 예를 들어 핵 단백질, 전사 인자, 핵 막 수송체, 세포내 소기관 연관 단백질, 막 수용체, 촉매 단백질 및 효소, 치료 단백질, 막 단백질, 막 수송 단백질, 신호 전달 단백질 또는 면역학적 단백질 (예를 들어, IgG 또는 다른 항체 단백질) 등을 포함할 수 있다. 관심 유전자는 또한 메신저 RNA (mRNA), 전달 RNA (tRNA), 리보솜 RNA (rRNA), 소형 핵 RNA (snRNA), 안티센스 RNA, 가이드 RNA, 마이크로RNA (miRNA), 소형 간섭 RNA (siRNA) 및 세포-유리 RNA (cfRNA)를 포함하나 이에 제한되지는 않는 RNA 분자를 코딩할 수 있다.

[0141] 가이드 RNA ("gRNA")

[0142] 본원에 사용된 용어 "가이드 RNA"는 대부분 통상적으로 CRISPR-Cas9의 Cas 단백질과 회합되고 Cas9와 회합되어, Cas9 단백질을 가이드 RNA의 프로토스페이서 서열에 대한 상보성을 포함하는 DNA 분자 내의 특이적 서열로 지시하는 가이드 핵산의 특정한 유형이다. 그러나, 이 용어는 또한 자연 발생이든 비-자연 발생이든 (예를 들어, 조작된 것이든 재조합이든), Cas9 등가물, 상동체, 오르토로그 또는 파라로그와 회합하고, 달리 Cas9 등가물을 특이적 표적 뉴클레오티드 서열에 국재화하도록 프로그램화하는 등가의 가이드 핵산 분자를 포괄한다. Cas9 등가물은 Cpf1 (유형-V CRISPR-Cas 시스템), C2c1 (유형 V CRISPR-Cas 시스템), C2c2 (유형 VI CRISPR-Cas 시스템) 및 C2c3 (유형 V CRISPR-Cas 시스템)을 포함한 임의의 유형의 CRISPR 시스템 (예를 들어, 유형 II, V, VI)으로부터의 다른 napDNAbp를 포함할 수 있다. 추가의 Cas-등가물은 문헌 [Makarova et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector," Science 2016; 353(6299)]에 기재되어 있으며, 이의 내용은 본원에 참조로 포함된다. 가이드 RNA의 예시적인 서열 및 구조가 본원에 제공된다. 추가로, 적절한 가이드 RNA 서열을 설계하는 방법이 본원에 제공된다. 본원에 사용된 "가이드 RNA"는 또한 본원에 개시된 프라임 편집 방법 및 조성물에 대해 발명된 "프라임 편집 가이드 RNA" (또는 "PEgRNA")로 지칭되는 변형된 형태의 가이드 RNA와 대조되도록 "전통적인 가이드 RNA"로 지칭될 수 있다.

[0143] 가이드 RNA 또는 PEgRNA는 하기를 포함하나 이에 제한되지는 않는 다양한 구조적 요소를 포함할 수 있다:

[0144] 스페이서 서열 - 표적 DNA 내의 프로토스페이서에 결합하는 가이드 RNA 또는 PEgRNA 내의 서열 (약 20개 nt 길이를 가짐).

[0145] gRNA 코어 (또는 gRNA 스캐폴드 또는 백본 서열) - Cas9 결합을 담당하는 gRNA 내의 서열을 지칭하며, 이는 Cas9를 표적 DNA로 가이드하는 데 사용되는 20 bp 스페이서/표적화 서열을 포함하지 않음.

[0146] 연장 아암 - 프라이머 결합 부위 및 폴리머라제 (예를 들어, 리버스 트랜스크립타제)를 통해 관심 유전자 변화를 함유하는 단일 가닥 DNA 플랩을 코딩하는 DNA 합성 주형 서열을 포함하는, PEgRNA의 3' 단부 또는 5' 단부에서의 단일 가닥 연장부, 이는 이어서 상응하는 내인성 가닥을 대체함으로써 내인성 DNA 내로 통합되어, 목적하는 유전자 변화를 설치함.

[0147] 전사 종결인자 - 가이드 RNA 또는 PEgRNA는 분자의 3'에 전사 종결 서열을 포함할 수 있음.

[0148] 상동성 아암

[0149] 용어 "상동성 아암"은 내인성 가닥을 대체함으로써 표적 DNA 부위 내로 통합될, 생성된 리버스 트랜스크립타제-코딩된 단일 가닥 DNA 플랩의 부분을 코딩하는 연장 아암의 부분을 지칭한다. 상동성 아암에 의해 코딩되는 단일 가닥 DNA 플랩의 부분은 표적 DNA 서열의 비-편집된 가닥에 상보적이며, 이는 내인성 가닥의 변위 및 그 자리에서의 단일 가닥 DNA 플랩의 어닐링을 용이하게 함으로써, 편집을 설치한다. 이 성분은 다른 곳에 추가로 정의된다. 상동성 아암은 정의에 의해 본원에 기재된 프라임 편집제의 폴리머라제에 의해 코딩되기 때문에 DNA 합성 주형의 일부이다.

[0150] 숙주 세포

[0151] 본원에 사용된 용어 "숙주 세포"는 본원에 기재된 벡터, 예를 들어 Cas9 또는 Cas9 등가물 및 리버스 트랜스크립타제를 포함하는 융합 단백질을 코딩하는 핵산 분자를 포함하는 벡터를 숙주로서 받고, 복제하고, 발현할 수 있는 세포를 지칭한다.

[0152] 인테인

[0153] 본원에 사용된 용어 "인테인"은 모든 생물 도메인으로부터의 유기체에서 발견되는 자가-프로세싱 폴리펩티드 도

메인을 지칭한다. 인테인 (개재 단백질)은 단백질 스플라이싱으로 공지된 고유한 자가-프로세싱 사건을 수행하며, 여기서 이는 2개의 펩티드 결합의 절단을 통해 더 큰 전구체 폴리펩티드로부터 그 자신을 절제하고, 이 과정에서 새로운 펩티드 결합의 형성을 통해 플랭킹되어 있는 엑스테인 (외부 단백질) 서열을 라이게이션한다. 이러한 재배열은 인테인 유전자가 다른 단백질-코딩 유전자 내에 인 프레임으로 포매되어 발견되기 때문에 번역 후에 (또는 가능하게는 공동-번역으로) 일어난다. 또한, 인테인-매개된 단백질 스플라이싱은 자발적이고; 이는 외부 인자 또는 에너지 공급원을 필요로 하지 않으며, 단지 인테인 도메인의 폴딩만을 필요로 한다. 이 과정은 또한 시스-단백질 스플라이싱으로도 공지되어 있으며, 이는 "분할 인테인"에 의한 트랜스-단백질 스플라이싱의 자연적 과정과 대조된다. 인테인은 엑스테인으로 공지된 플랭킹 단백질 서열이 공동 융합된 전구체 단백질로부터 그 자신의 절제를 촉매하는 자기-스플라이싱 RNA 인트론과 등가인 단백질이다 (문헌 [Perler et al., Nucleic Acids Res. 22:1125-1127 (1994)] 참조) (문헌 [Perler et al., Curr. Opin. Chem. Biol. 1:292-299 (1997); Perler, F. B. Cell 92(1):1-4 (1998); Xu et al., EMBO J. 15(19):5146-5153 (1996)]에서 검토됨).

[0154] 본원에 사용된 용어 "단백질 스플라이싱"은 전구체 단백질의 내부 영역 (인테인)이 절제되고 단백질의 플랭킹 영역 (엑스테인)이 라이게이션되어 성숙 단백질을 형성하는 과정을 지칭한다. 이러한 자연적 과정은 원핵생물 및 진핵생물 둘 다로부터의 수많은 단백질에서 관찰되었다 (Perler, F. B., Xu, M. Q., Paulus, H. Current Opinion in Chemical Biology 1997, 1, 292-299; Perler, F. B. Nucleic Acids Research 1999, 27, 346-347). 인테인 유닛은 단백질 스플라이싱을 촉매하는데 필요한 필수 성분을 함유하고, 종종 인테인 이동성에 참여하는 엔도뉴클레아제 도메인을 함유한다 (Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thomer, J., Belfort, M. Nucleic Acids Research 1994, 22, 1127-1127). 그러나, 생성된 단백질은 연결되지만, 별개의 단백질로서 발현되지 않는다. 단백질 스플라이싱은 또한 트랜스로 수행되어, 별개의 폴리펩티드 상에 발현된 분할 인테인이 자발적으로 조합되어 단일 인테인을 형성한 다음, 단백질 스플라이싱 과정을 겪어 별개의 단백질에 연결될 수 있다.

[0155] 단백질 스플라이싱 메카니즘의 설명은 다수의 인테인-기반 적용으로 이어졌다 (미국 특허 번호 5,496,714 (Comb, et al.); 미국 특허 번호 5,834,247 (Comb, et al.); 문헌 [Camarero and Muir, J. Amer. Chem. Soc., 121:5597-5598 (1999); Chong, et al., Gene, 192:271-281 (1997), Chong, et al., Nucleic Acids Res., 26:5109-5115 (1998); Chong, et al., J. Biol. Chem., 273:10567-10577 (1998); Cotton, et al. J. Am. Chem. Soc., 121:1100-1101 (1999); Evans, et al., J. Biol. Chem., 274:18359-18363 (1999); Evans, et al., J. Biol. Chem., 274:3923-3926 (1999); Evans, et al., Protein Sci., 7:2256-2264 (1998); Evans, et al., J. Biol. Chem., 275:9091-9094 (2000); Iwai and Pluckthun, FEBS Lett. 459:166-172 (1999); Mathys, et al., Gene, 231:1-13 (1999); Mills, et al., Proc. Natl. Acad. Sci. USA 95:3543-3548 (1998); Muir, et al., Proc. Natl. Acad. Sci. USA 95:6705-6710 (1998); Otomo, et al., Biochemistry 38:16040-16044 (1999); Otomo, et al., J. Biolmol. NMR 14:105-114 (1999); Scott, et al., Proc. Natl. Acad. Sci. USA 96:13638-13643 (1999); Severinov and Muir, J. Biol. Chem., 273:16205-16209 (1998); Shingledecker, et al., Gene, 207:187-195 (1998); Southworth, et al., EMBO J. 17:918-926 (1998); Southworth, et al., Biotechniques, 27:110-120 (1999); Wood, et al., Nat. Biotechnol., 17:889-892 (1999); Wu, et al., Proc. Natl. Acad. Sci. USA 95:9226-9231 (1998a); Wu, et al., Biochim Biophys Acta 1387:422-432 (1998b); Xu, et al., Proc. Natl. Acad. Sci. USA 96:388-393 (1999); Yamazaki, et al., J. Am. Chem. Soc., 120:5591-5592 (1998)). 각각의 참고문헌은 본원에 참조로 포함된다.

[0156] 리간드-의존성 인테인

[0157] 본원에 사용된 용어 "리간드-의존성 인테인"은 리간드-결합 도메인을 포함하는 인테인을 지칭한다. 전형적으로, 리간드-결합 도메인은 인테인의 아미노산 서열 내로 삽입되어, 구조 인테인 (N) - 리간드-결합 도메인 - 인테인 (C)을 생성한다. 전형적으로, 리간드-의존성 인테인은 적절한 리간드의 부재 하에 단백질 스플라이싱 활성을 전혀 나타내지 않거나 단지 최소로만 나타내고, 리간드의 존재 하에 단백질 스플라이싱 활성의 현저한 증가를 나타낸다. 일부 실시양태에서, 리간드-의존성 인테인은 리간드의 부재 하에 관찰가능한 스플라이싱 활성을 나타내지 않지만, 리간드의 존재 하에 스플라이싱 활성을 나타낸다. 일부 실시양태에서, 리간드-의존성 인테인은 리간드의 부재 하에 관찰가능한 단백질 스플라이싱 활성을 나타내고, 적절한 리간드의 존재 하에서, 리간드의 부재 하에 관찰된 활성보다 적어도 5배, 적어도 10배, 적어도 50배, 적어도 100배, 적어도 150배, 적어도 200배, 적어도 250배, 적어도 500배, 적어도 1000배, 적어도 1500배, 적어도 2000배, 적어도 2500배, 적어도 5000배, 적어도 10000배, 적어도 20000배, 적어도 25000배, 적어도 50000배, 적어도 100000배, 적어도 500000배 또는 적어도 1000000배 더 큰 단백질 스플라이싱 활성을 나타낸다. 일부 실시양태에서, 활성의

증가는 적어도 1 자릿수, 적어도 2 자릿수, 적어도 3 자릿수, 적어도 4 자릿수 또는 적어도 5 자릿수에 걸쳐 용량 의존성이며, 이는 리간드의 농도를 조정함으로써 인테인 활성의 미세-조정을 가능하게 한다. 적합한 리간드-의존성 인테인은 관련 기술분야에 공지되어 있고, 하기 제공된 것 및 공개된 미국 특허 출원 U.S. 2014/0065711 A1; 문헌 [Mootz et al., "Protein splicing triggered by a small molecule." J. Am. Chem. Soc. 2002; 124, 9044-9045; Mootz et al., "Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo." J. Am. Chem. Soc. 2003; 125, 10561-10569; Buskirk et al., Proc. Natl. Acad. Sci. USA. 2004; 101, 10505-10510); Skretas & Wood, "Regulation of protein activity with small-molecule-controlled inteins." Protein Sci. 2005; 14, 523-532; Schwartz, et al., "Post-translational enzyme activation in an animal via optimized conditional protein splicing." Nat. Chem. Biol. 2007; 3, 50-54; Peck et al., Chem. Biol. 2011; 18 (5), 619-630] (각각의 전체 내용은 본원에 참조로 포함됨)에 기재된 것을 포함한다. 예시적인 서열은 하기와 같다:

명칭	리간드-의존성 인테인의 서열
2-4 인테인:	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO: 8)
3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVSWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYTNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO: 9)
30R3-1 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO: 10)
30R3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHNC (SEQ ID NO: 11)
30R3-3 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVVSWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGDRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH

[0158]

	RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 12)
37R3-1 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYNPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC ((SEQ ID NO: 13)
37R3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 14)
37R3-3 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 15)

[0159]

[0160] 링커

[0161] 본원에 사용된 용어 "링커"는 2개의 다른 분자 또는 모이어티를 연결하는 분자를 지칭한다. 링커는 2개의 융합 단백질을 연결하는 링커의 경우에 아미노산 서열일 수 있다. 예를 들어, Cas9는 아미노산 링커 서열에 의해 리버스 트랜스크립타제에 융합될 수 있다. 링커는 또한 2개의 뉴클레오티드 서열을 함께 연결하는 경우에 뉴클레오티드 서열일 수 있다. 예를 들어, 본 경우에, 전통적인 가이드 RNA는 RT 주형 서열 및 RT 프라이머 결합 부위를 포함할 수 있는 프라임 편집 가이드 RNA의 RNA 연장부에 스페이서 또는 링커 뉴클레오티드 서열을 통해 연결된다. 다른 실시양태에서, 링커는 유기 분자, 기, 중합체 또는 화학적 모이어티이다. 일부 실시양태에서, 링커는 5-100개 아미노산 길이, 예를 들어 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150 또는 150-200개 아미노산 길이이다. 보다 긴 또는 보다 짧은 링커가 또한 고려된다.

[0162] 단리된

[0163] "단리된"은 자연 상태로부터 변경 또는 제거된 것을 의미한다. 예를 들어, 살아있는 동물에 자연적으로 존재하는 핵산 또는 펩티드는 "단리된" 것이 아니지만, 그의 자연 상태의 공존 물질로부터 부분적으로 또는 완전히 분리된 동일한 핵산 또는 펩티드는 "단리된" 것이다. 단리된 핵산 또는 단백질은 실질적으로 정제된 형태로 존재할 수 있거나 또는 비-천연 환경, 예컨대, 예를 들어 숙주 세포에 존재할 수 있다.

[0164] 일부 실시양태에서, 관심 유전자는 단리된 핵산에 의해 코딩된다. 본원에 사용된 용어 "단리된"은 그의 원래 또는 천연 환경 (예를 들어, 자연 발생인 경우에 자연 환경)으로부터 제거되는 본원에 제공된 바와 같은 물질의 특징을 지칭한다. 따라서, 살아있는 동물에 존재하는 자연 발생 폴리뉴클레오티드 또는 단백질 또는 폴리펩티드는 단리된 것이 아니지만, 자연계 내 공존 물질의 일부 또는 모두로부터 인간 개입에 의해 분리된 동일한 폴리뉴클레오티드 또는 폴리펩티드는 단리된 것이다. 따라서, 인공 또는 조작된 물질, 예를 들어 비-자연 발생 핵산 구축물, 예컨대 본원에 기재된 발현 구축물 및 벡터는 또한 단리된 것으로 지칭된다. 물질은 단리되기 위해 정제될 필요는 없다. 따라서, 물질은 벡터의 일부 및/또는 조성물의 일부일 수 있고, 이러한 벡터 또는 조성물은 사실상 물질이 발견된 환경의 일부가 아니라는 점에서 상기 물질은 여전히 단리된 것일 수 있다.

[0165] MS2 태그부착 기술

[0166] 다양한 실시양태에서 (예를 들어, 도 72-73의 실시양태 및 실시예 19에 도시된 바와 같음), 용어 "MS2 태그부착 기술"은 RNA-단백질 상호작용 도메인, 예를 들어 특이적 헤어핀 구조를 특이적으로 인식하고 이에 결합하는 RNA-결합 단백질과 쌍형성된 "RNA-단백질 상호작용 도메인" (일명 "RNA-단백질 동원 도메인 또는 단백질"로도 공지됨)의 조합을 지칭한다. 이들 유형의 시스템은 표적 부위에 결합된 프라임 편집제 복합체에 다양한 기능성을 동원하기 위해 활용될 수 있다. MS2 태그부착 기술은 MS2 박테리오파지 코트 단백질 ("MCP" 또는 "MS2cp")과 파지의 게놈에 존재하는 스템-루프 또는 헤어핀 구조, 즉 "MS2 헤어핀"의 자연적 상호작용에 기초한다. 프라임 편집의 경우에, MS2 태그부착 기술은 MS2 헤어핀을 프라임 편집에 수반되는 목적하는 RNA 분자 (예를 들어, PEgRNA 또는 tPERT) 내로 도입하는 것을 포함하며, 이는 이어서 그 구조를 인식하고 그에 결합하는 RNA-결합 단백질에 대한 특이적 상호작용가능한 결합 표적을 구성한다. MS2 헤어핀의 경우에, 이는 MS2 박테리오파지 코트 단백질 (MCP)에 의해 인식되고 결합된다. 그리고, MCP가 또 다른 단백질 (예를 들어, 리버스 트랜스크립타제 또는 다른 DNA 폴리머라제)에 융합되면, MS2 헤어핀은 프라임 편집 복합체에 의해 점유된 표적 부위에 트랜스로 그 다른 단백질을 "동원"하는 데 사용될 수 있다.

[0167] 본원에 기재된 프라임 편집제는 관심있는 특이적 기능을 프라임 편집제 복합체로 동원하거나 또는 "공동-국재화"하기 위한 임의의 공지된 RNA-단백질 상호작용 도메인을 한 측면으로서 혼입할 수 있다. 다른 모듈 RNA-단백질 상호작용 도메인의 검토는 관련 기술분야에, 예를 들어 문헌 [Johansson et al., "RNA recognition by the MS2 phage coat protein," Sem Virol., 1997, Vol. 8 (3): 176-185; Delebecque et al., "Organization of intracellular reactions with rationally designed RNA assemblies," Science, 2011, Vol. 333:470-474; Mali et al., "Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering," Nat. Biotechnol., 2013, Vol.31: 833-838; 및 Zalatan et al., "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds," Cell, 2015, Vol.160: 339-350]에 기재되어 있고, 이들 각각은 그 전문이 본원에 참조로 포함된다. 다른 시스템은 PCP 단백질을 특이적으로 동원하는 PP7 헤어핀 및 Com 단백질을 특이적으로 동원하는 "com" 헤어핀을 포함한다. 문헌 [Zalatan et al.]을 참조한다.

[0168] MS2 헤어핀 (또는 동등하게 "MS2 압타머"로 지칭됨)의 뉴클레오티드 서열은 GCCAACATGAGGATCACCCATGTCTGCAGGGCC (서열식별번호: 763)이다.

[0169] MCP 또는 MS2cp의 아미노산 서열은 GSASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGEELPVAGWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY (서열식별번호: 764)이다.

[0170] MS2 헤어핀 (또는 "MS2 압타머")은 또한 "RNA 이펙터 동원 도메인"의 한 유형 (또는 동등하게 "RNA-결합 단백질 동원 도메인" 또는 간단히 "동원 도메인")으로서 지칭될 수 있으며, 이는 그것이 PEgRNA 또는 tPERT 내로 설치되는 물리적 구조 (예를 들어, 헤어핀)로, 다른 이펙터 기능 (예를 들어, 다양한 기능을 갖는 RNA-결합 단백질, 예컨대 DNA 폴리머라제 또는 다른 DNA-변형 효소)을 이와 같이 변형된 PEgRNA 또는 rPERT로 효과적으로 동원하고, 이에 따라 이펙터 기능을 프라임 편집 기구에 트랜스로 공동-국재화하기 때문이다. 본 출원은 어떠한 방식으로도 임의의 특정한 RNA 이펙터 동원 도메인에 제한되는 것으로 의도되지 않으며, MS2 헤어핀을 포함한 임의의 이용가능한 이러한 도메인을 포함할 수 있다. 실시예 19 및 도 72(b)는 DNA 합성 도메인 (즉, tPERT 분자) 및 PE2에 융합된 MS2cp 단백질을 포함하는 프라임 편집제에 연결된 MS2 압타머의 사용으로 표적 DNA 부위에 결합된 프라임 편집제 복합체 (MS2cp-PE2:sgRNA 복합체) 및 tPERT 분자의 DNA 합성 도메인의 공동-국재화를 유발하는 것을 도시한다.

[0171] napDNAbp

[0172] 본원에 사용된 용어 "핵산 프로그램가능한 DNA 결합 단백질" 또는 "napDNAbp" (Cas9가 한 예임)는 RNA:DNA 혼성화를 사용하여 DNA 분자 내의 특이적 서열을 표적화하고 그에 결합하는 단백질을 지칭한다. 각각의 napDNAbp는 적어도 1개의 가이드 핵산 (예를 들어, 가이드 RNA)과 회합되며, 이는 napDNAbp를 가이드 핵산 또는 그의 부분 (예를 들어, 가이드 RNA의 프로토스페이서)에 상보적인 DNA 가닥 (즉, 표적 가닥)을 포함하는 DNA 서열에 국재화시킨다. 다시 말해서, 가이드 핵산은 napDNAbp (예를 들어, Cas9 또는 등가물)를 상보적 서열에 국재화시키고 결합하도록 "프로그램화한다".

[0173] 이론에 얽매이지는 않지만, napDNAbp - 가이드 RNA 복합체의 결합 메카니즘은 일반적으로 R-루프를 형성하여

napDNAbp가 이중 가닥 DNA 표적의 풀림을 유도하고, 그에 의해 napDNAbp에 의해 결합된 영역에서 가닥을 분리하는 단계를 포함한다. 이어서, 가이드 RNA 프로토스페이서는 "표적 가닥"에 혼성화된다. 이는 표적 가닥에 상보적인 "비-표적 가닥"을 대체하여, R-루프의 단일 가닥 영역을 형성한다. 일부 실시양태에서, napDNAbp는 1개 이상의 뉴클레아제 활성을 포함하며, 이는 이어서 DNA를 절단하여 다양한 유형의 병변을 남긴다. 예를 들어, napDNAbp는 제1 위치에서 비-표적 가닥을 절단하고/거나 제2 위치에서 표적 가닥을 절단하는 뉴클레아제 활성을 포함할 수 있다. 뉴클레아제 활성에 따라, 표적 DNA는 절단되어 "이중 가닥 파괴"를 형성할 수 있고, 이에 의해 두 가닥이 절단된다. 다른 실시양태에서, 표적 DNA는 단일 부위에서만 절단될 수 있고, 즉 DNA는 한 가닥 상에서 "닉킹"된다. 상이한 뉴클레아제 활성을 갖는 예시적인 napDNAbp는 "Cas9 닉카제" ("nCas9") 및 뉴클레아제 활성을 갖지 않는 탈활성화된 Cas9 ("기능상실 Cas9" 또는 "dCas9")를 포함한다. 이들 및 다른 napDNAbp에 대한 예시적인 서열이 본원에 제공된다.

[0174] 닉카제

[0175] 용어 "닉카제"는 2개의 뉴클레아제 도메인 중 하나가 불활성화된 Cas9를 지칭한다. 이 효소는 표적 DNA의 한 가닥만을 절단할 수 있다.

[0176] 핵 국재화 서열 (NLS)

[0177] 용어 "핵 국재화 서열" 또는 "NLS"는, 예를 들어 핵 수송에 의한 세포 핵 내로의 단백질의 유입을 촉진하는 아미노산 서열을 지칭한다. 핵 국재화 서열은 관련 기술분야에 공지되어 있고, 통상의 기술자에게 명백할 것이다. 예를 들어, NLS 서열은 2001년 5월 31일에 WO/2001/038547로서 공개된 2000년 11월 23일에 출원된 국제 PCT 출원 PCT/EP2000/011690 (Plank et al.)에 기재되어 있으며, 그의 내용은 예시적인 핵 국재화 서열의 개시내용에 대해 본원에 참조로 포함된다. 일부 실시양태에서, NLS는 아미노산 서열 PKKKRKV (서열식별번호: 16) 또는 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (서열식별번호: 17)를 포함한다.

[0178] 핵산 분자

[0179] 본원에 사용된 용어 "핵산"은 뉴클레오티드의 중합체를 지칭한다. 중합체는 천연 뉴클레오시드 (즉, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시구아노신 및 데옥시시티딘), 뉴클레오시드 유사체 (예를 들어, 2-아미노아데노신, 2-티오티미딘, 이노신, 피롤로-피리미딘, 3-메틸 아데노신, 5-메틸시티딘, C5 브로모우리딘, C5 플루오로우리딘, C5 아이오도우리딘, C5 프로피닐 우리딘, C5 프로피닐 시티딘, C5 메틸시티딘, 7 데아자아데노신, 7 데아자구아노신, 8 옥소아데노신, 8 옥소구아노신, O(6) 메틸구아닌, 4-아세틸시티딘, 5-(카르복시히드록시메틸)우리딘, 디히드로우리딘, 메틸슈도우리딘, 1-메틸 아데노신, 1-메틸 구아노신, N6-메틸 아데노신 및 2-티오시티딘), 화학적으로 변형된 염기, 생물학적으로 변형된 염기 (예를 들어, 메틸화 염기), 삽입된 염기, 변형된 당 (예를 들어, 2'-플루오로리보스, 리보스, 2'-데옥시리보스, 2'-O-메틸시티딘, 아라비노스 및 헥소스), 또는 변형된 포스페이트 기 (예를 들어, 포스포로티오에이트 및 5' N 포스포르아미다이트 연결)를 포함할 수 있다.

[0180] PEgRNA

[0181] 본원에 사용된 용어 "프라임 편집 가이드 RNA" 또는 "PEgRNA" 또는 "연장된 가이드 RNA"는 본원에 기재된 프라임 편집 방법 및 조성물을 구현하기 위한 1개 이상의 추가의 서열을 포함하도록 변형된 특수화된 형태의 가이드 RNA를 지칭한다. 본원에 기재된 바와 같이, 프라임 편집 가이드 RNA는 핵산 서열의 1개 이상의 "연장된 영역"을 포함한다. 연장된 영역은 단일 가닥 RNA 또는 DNA를 포함할 수 있지만, 이에 제한되지는 않는다. 추가로, 연장된 영역은 전통적인 가이드 RNA의 3' 단부에서 발생할 수 있다. 다른 배열에서, 연장된 영역은 전통적인 가이드 RNA의 5' 단부에서 발생할 수 있다. 또 다른 배열에서, 연장된 영역은 전통적인 가이드 RNA의 분자내 영역에서, 예를 들어 napDNAbp에 회합 및/또는 결합하는 gRNA 코어 영역에서 발생할 수 있다. 연장된 영역은 (a) 편집될 내인성 표적 DNA와 상동이도록 설계되었고, (b) 내인성 표적 DNA 내로 도입 또는 통합될 적어도 1개의 목적하는 뉴클레오티드 변화 (예를 들어, 전이, 전환, 결실 또는 삽입)를 포함하는 단일 가닥 DNA를 (프라임 편집제의 폴리머라제에 의해) 코딩하는 "DNA 합성 주형"을 포함한다. 연장된 영역은 또한 다른 기능적 서열 요소, 예컨대 비제한적으로 "프라이머 결합 부위" 및 "스페이서 또는 링커" 서열, 또는 다른 구조적 요소, 예컨대 비제한적으로 압타머, 스템 루프, 헤어핀, 토 루프 (예를 들어, 3' 토루프) 또는 RNA-단백질 동원 도메인 (예를 들어, MS2 헤어핀)을 포함할 수 있다. 본원에 사용된 "프라이머 결합 부위"는 R-루프의 닉킹된 DNA로부터 생성된 3' 단부를 갖는 단일 가닥 DNA 서열에 혼성화하는 서열을 포함한다.

[0182] 특정 실시양태에서, PEgRNA는 도 3a에 의해 나타내어지며, 이는 5' 연장 아암, 스페이서 및 gRNA 코어를 갖는

PEgRNA를 보여준다. 5' 연장부는 5'에서 3' 방향으로 리버스 트랜스크립타제 주형, 프라이머 결합 부위 및 링커를 추가로 포함한다. 제시된 바와 같이, 리버스 트랜스크립타제 주형은 또한 "DNA 합성 주형"으로서 보다 광범위하게 지칭될 수 있으며, 여기서 본원에 기재된 프라임 편집제의 폴리머라제는 RT가 아니라 또 다른 유형의 폴리머라제이다.

[0183] 특정의 다른 실시양태에서, PEgRNA는 도 3b에 의해 나타내어지며, 이는 5' 연장 아암, 스페이서 및 gRNA 코어를 갖는 PEgRNA를 보여준다. 5' 연장부는 5'에서 3' 방향으로 리버스 트랜스크립타제 주형, 프라이머 결합 부위 및 링커를 추가로 포함한다. 제시된 바와 같이, 리버스 트랜스크립타제 주형은 또한 "DNA 합성 주형"으로서 보다 광범위하게 지칭될 수 있으며, 여기서 본원에 기재된 프라임 편집제의 폴리머라제는 RT가 아니라 또 다른 유형의 폴리머라제이다.

[0184] 또 다른 실시양태에서, PEgRNA는 도 3d에 의해 나타내어지며, 이는 5'에서 3' 방향으로 스페이서 (1), gRNA 코어 (2) 및 연장 아암 (3)을 갖는 PEgRNA를 보여준다. 연장 아암 (3)은 PEgRNA의 3' 단부에 있다. 연장 아암 (3)은 5'에서 3' 방향으로 "프라이머 결합 부위" (A), "편집 주형" (B) 및 "상동성 아암" (C)을 추가로 포함한다. 연장 아암 (3)은 또한 3' 및 5' 단부에, 동일한 서열 또는 상이한 서열일 수 있는 임의적인 변형제 영역을 포함할 수 있다. 추가로, PEgRNA의 3' 단부는 전사 종결인자 서열을 포함할 수 있다. PEgRNA의 이들 서열 요소는 본원에 추가로 기재되고 정의된다.

[0185] 또 다른 실시양태에서, PEgRNA는 도 3e에 의해 나타내어지며, 이는 5'에서 3' 방향으로 연장 아암 (3), 스페이서 (1) 및 gRNA 코어 (2)를 갖는 PEgRNA를 보여준다. 연장 아암 (3)은 PEgRNA의 5' 단부에 있다. 연장 아암 (3)은 3'에서 5' 방향으로 "프라이머 결합 부위" (A), "편집 주형" (B) 및 "상동성 아암" (C)을 추가로 포함한다. 연장 아암 (3)은 또한 3' 및 5' 단부에, 동일한 서열 또는 상이한 서열일 수 있는 임의적인 변형제 영역을 포함할 수 있다. PEgRNA는 또한 3' 단부에 전사 종결인자 서열을 포함할 수 있다. PEgRNA의 이들 서열 요소는 본원에 추가로 기재되고 정의된다.

[0186] PE1

[0187] 본원에 사용된 바와 같이, "PE1"은 하기 구조: [NLS]-[Cas9(H840A)]-[링커]-[MMLV_RT(wt)] + 목적하는 PEgRNA를 갖는, Cas9(H840A) 및 야생형 MMLV RT를 포함하는 융합 단백질을 포함하는 PE 복합체를 지칭하며, 여기서 PE 융합체는 하기와 같이 제시되는 서열식별번호: 123의 아미노산 서열을 갖는다:

[0188] MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIEDEYRLHETS KEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRL LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRI QHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQR WLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKA YQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPC LRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGP VVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAV TTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEG KEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSS P**SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 123)

[0189] 설명:

[0190] 핵 국재화 서열 (NLS) 상단: (서열식별번호: 124), 하단: (서열식별번호: 133)

[0191] Cas9(H840A) (서열식별번호: 126)

[0192] 33개-아미노산 링커 (서열식별번호: 127)

[0193] M-MLV 리버스 트랜스크립타제 (서열식별번호: 128).

[0194] PE2

[0195] 본원에 사용된 바와 같이, "PE2"는 하기 구조: [NLS]-[Cas9(H840A)]-[링커]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)] + 목적하는 PEgRNA를 갖는, Cas9(H840A) 및 변이체 MMLV RT를 포함하는 융합 단백질을 포함하는 PE 복합체를 지칭하며, 여기서 PE 융합체는 하기와 같이 제시되는 서열식별번호: 134의 아미노산 서열을 갖는다:

MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS**
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI
LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIEDEYRLHETS*
KEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRL
LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT
VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQ
HPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRW
LTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAY
QEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCL
RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPV
VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVT
TETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGK
EIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP
SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 134)

[0196]

[0197] 설명:

[0198] 핵 국재화 서열 (NLS) 상단: (서열식별번호: 124), 하단: (서열식별번호: 133)

[0199] Cas9(H840A) (서열식별번호: 137)

[0200] 33개-아미노산 링커 (서열식별번호: 127)

[0201] M-MLV 리버스 트랜스크립타제 (서열식별번호: 139).

[0202] PE3

[0203] 본원에 사용된 "PE3"은 PE2 플러스 PE2와 복합체화하고 비-편집된 DNA 가닥에 닉을 도입하여 편집된 가닥의 우선적 대체를 유도하는 제2 가닥 닉킹 가이드 RNA를 지칭한다.

[0204] PE3b

[0205] 본원에 사용된 "PE3b"는 PE3을 지칭하지만, 여기서 제2 가닥 닉킹 가이드 RNA는 제2 가닥 닉이 목적하는 편집이 설치될 때까지 도입되지 않도록 시간적 제어를 위해 설계된다. 이는 단지 편집된 가닥에만 매칭되고 원래 대립

유전자에는 매칭되지 않는 스페이서 서열을 갖는 gRNA를 설계함으로써 달성된다. 이하에서 PE3b로 지칭되는 이러한 전략을 사용할 때, 프로토스페이서와 비편집된 대립유전자 사이의 미스매치는 PAM 가닥 상의 편집 사건이 일어날 때까지 sgRNA에 의한 닉킹을 불리하게 할 것이다.

[0206] PE-짧은 형

[0207] 본원에 사용된 "PE-짧은 형"은 C-말단이 말단절단된 리버스 트랜스크립타제에 융합된 PE 구축물을 지칭하고, 하기 아미노산 서열을 갖는다:

MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH**
SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH
MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLL
AQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL
LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
YLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIED*
EYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLG
IKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGL
PPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALH
RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG
YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFN
WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLD
PVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLL
*DTDRVQFGPVVALNPATLLPLPEEGLQHNCLDNSRLIN*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 765)

[0208]

[0209] 설명:

[0210] 핵 국재화 서열 (NLS) 상단: (서열식별번호: 124), 하단: (서열식별번호: 133)

[0211] Cas9(H840A) (서열식별번호: 157)

[0212] 33개-아미노산 링커 1 (서열식별번호: 127)

[0213] M-MLV 말단절단된 리버스 트랜스크립타제

[0214] (서열식별번호: 766)

[0215] 펩티드 태그

[0216] 용어 "펩티드 태그"는 다양한 목적, 예컨대 가시화, 정제, 가용화 및 분리 등을 위해 단백질의 조작을 용이하게 하는 1개 이상의 기능을 단백질 상에 부여하기 위해 단백질 서열에 유전자 융합된 펩티드 아미노산 서열을 지칭한다. 펩티드 태그는 목적 또는 기능에 의해 카테고리화된 다양한 유형의 태그를 포함할 수 있으며, 이는 "친화성 태그" (단백질 정제를 용이하게 하기 위함), "가용화 태그" (단백질의 적절한 폴딩을 보조하기 위함), "크로마토그래피 태그" (단백질의 크로마토그래피 특성을 변경하기 위함), "에피토프 태그" (고친화도 항체에 결합하기 위함), "형광 태그" (세포 또는 시험관내에서 단백질의 가시화를 용이하게 하기 위함)를 포함할 수 있다.

[0217] 폴리머라제

[0218] 본원에 사용된 용어 "폴리머라제"는 뉴클레오티드 가닥을 합성하고 본원에 기재된 프라임 편집제 시스템과 관련하여 사용될 수 있는 효소를 지칭한다. 폴리머라제는 "주형-의존성" 폴리머라제 (즉, 주형 가닥의 뉴클레오티드 염기의 순서에 기초하여 뉴클레오티드 가닥을 합성하는 폴리머라제)일 수 있다. 폴리머라제는 또한 "주형-

비의존성" 폴리머라제 (즉, 주형 가닥의 필요 없이 뉴클레오티드 가닥을 합성하는 폴리머라제)일 수 있다. 폴리머라제는 또한 "DNA 폴리머라제" 또는 "RNA 폴리머라제"로 추가로 카테고리화될 수 있다. 다양한 실시양태에서, 프라임 편집제 시스템은 DNA 폴리머라제를 포함한다. 다양한 실시양태에서, DNA 폴리머라제는 "DNA-의존성 DNA 폴리머라제" (즉, 주형 분자가 DNA의 가닥임)일 수 있다. 이러한 경우에, DNA 주형 분자는 PEgRNA일 수 있으며, 여기서 연장 아암은 DNA의 가닥을 포함한다. 이러한 경우에, PEgRNA는 RNA 부분 (즉, 스페이서 및 gRNA 코어를 포함한 가이드 RNA 성분) 및 DNA 부분 (즉, 연장 아암)을 포함하는 키메라 또는 하이브리드 PEgRNA로 지칭될 수 있다. 다양한 다른 실시양태에서, DNA 폴리머라제는 "RNA-의존성 DNA 폴리머라제" (즉, 주형 분자가 RNA의 가닥임)일 수 있다. 이러한 경우에, PEgRNA는 RNA이며, 즉 RNA 연장부를 포함한다. 용어 "폴리머라제"는 또한 뉴클레오티드의 중합 (즉, 폴리머라제 활성)을 촉매하는 효소를 지칭할 수 있다. 일반적으로, 효소는 폴리뉴클레오티드 주형 서열에 어닐링된 프라이머 (예를 들어, 예컨대 PEgRNA의 프라이머 결합 부위에 어닐링된 프라이머 서열)의 3'-단부에서 합성을 개시할 것이고, 주형 가닥의 5' 단부를 향해 진행할 것이다. "DNA 폴리머라제"는 데옥시뉴클레오티드의 중합을 촉매한다. DNA 폴리머라제와 관련하여 본원에 사용된 용어 DNA 폴리머라제는 "그의 기능적 단편"을 포함한다. "그의 기능적 단편"은 폴리머라제의 전체 미만의 아미노산 서열을 포괄하고, 적어도 하나의 세트의 조건 하에 폴리뉴클레오티드의 중합을 촉매하는 능력을 보유하는 야생형 또는 돌연변이체 DNA 폴리머라제의 임의의 부분을 지칭한다. 이러한 기능적 단편은 별개의 개체로서 존재할 수 있거나 또는 보다 큰 폴리펩티드, 예컨대 융합 단백질의 구성성분일 수 있다.

[0219] 프라임 편집 및 다중-플랩 프라임 편집

[0220] 본원에 사용된 용어 "프라임 편집" 또는 "전형적 프라임 편집"은 napDNAbp, 폴리머라제 (예를 들어, 리버스 트랜스크립타제), 및 표적 DNA 서열 내로 혼입되는 목적하는 새로운 유전자 정보를 코딩하기 위한 (또는 유전자 정보를 결실시키기 위한) DNA 합성 주형을 포함하는 특수화된 가이드 RNA를 사용하는 유전자 편집을 위한 신규 접근법을 지칭한다. 프라임 편집의 특정 실시양태는 다른 도면 중에서 도 1a-1h 및 도 72a-72c의 실시양태에 기재된다. 전형적 프라임 편집은 문헌 [Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149-157 (2019)]의 발명자 공개물에 기재되어 있으며, 이는 그 전문이 본원에 참조로 포함된다.

[0221] 프라임 편집은 핵산 프로그램가능한 DNA 결합 단백질 ("napDNAbp") (이는 폴리머라제와 회합하여 (즉, 융합 단백질의 형태로 또는 달리 napDNAbp와 트랜스로 제공됨) 작동함)을 사용하여 특정된 DNA 부위 내로 새로운 유전자 정보를 직접 기록하는 다재다능한 정확한 게놈 편집 방법인 게놈 편집을 위한 플랫폼을 나타내며, 여기서 프라임 편집 시스템은 표적 부위를 특정하고 또한 가이드 RNA 상에 (예를 들어, 5' 또는 3' 단부에 또는 가이드 RNA의 내부 부분에) 조작된 연장부 (DNA 또는 RNA)에 의해 대체 DNA 가닥의 형태로 목적하는 편집의 합성의 주형이 되는 프라임 편집 (PE) 가이드 RNA ("PEgRNA")로 프로그램화된다. 목적하는 편집 (예를 들어, 단일 핵염기 치환)을 함유하는 대체 가닥은 편집될 표적 부위의 내인성 가닥 (닉 부위의 바로 하류)과 동일한 (또는 상동인) 서열을 공유한다 (목적하는 편집을 포함하는 것은 제외함). DNA 복구 및/또는 복제 기구를 통해, 닉 부위의 하류 내인성 가닥은 목적하는 편집을 함유하는 새로 합성된 대체 가닥에 의해 대체된다. 일부 경우에, 프라임 편집은 "검색-및-대체" 게놈 편집 기술로서 생각될 수 있으며, 이는 프라임 편집제가, 본원에 기재된 바와 같이, 편집될 목적하는 표적 부위를 검색하고 위치를 찾아낼 뿐만 아니라, 동시에, 상응하는 표적 부위 내인성 DNA 가닥 대신에 설치된 목적하는 편집을 함유하는 대체 가닥을 코딩하기 때문이다. 본 개시내용의 프라임 편집제는, 부분적으로, 표적-프라이밍된 역전사 (TPRT) 또는 "프라임 편집"의 메카니즘이 (예를 들어, 도 1a-1f의 다양한 실시양태에 도시된 바와 같이) 높은 효율 및 유전적 유연성으로 정밀 CRISPR/Cas-기반 게놈 편집을 수행하기 위해 활용 또는 적합화될 수 있다는 발견에 관한 것이다. TPRT는 이동성 DNA 요소, 예컨대 포유동물 비-LTR 레트로트랜스포손 및 박테리아 그룹 II 인트론에 의해 자연적으로 사용된다[28,29]. 본 발명자들은 본원에서 Cas 단백질-리버스 트랜스크립타제 융합체 또는 관련 시스템을 사용하여 가이드 RNA에 의해 특이적 DNA 서열을 표적화하고, 표적 부위에서 단일 가닥 닉을 생성하고, 닉킹된 DNA를 가이드 RNA와 통합된 조작된 리버스 트랜스크립타제 주형의 역전사를 위한 프라이머로서 사용하였다. 그러나, 상기 개념이 DNA 폴리머라제 성분으로서 리버스 트랜스크립타제를 사용하는 프라임 편집제로 시작되지만, 본원에 기재된 프라임 편집제는 리버스 트랜스크립타제로 제한되는 것이 아니라, 실질적으로 임의의 DNA 폴리머라제의 사용을 포함할 수 있다. 실제로, 본 출원은 전반에 걸쳐 "리버스 트랜스크립타제"를 사용한 프라임 편집제를 언급할 수 있지만, 리버스 트랜스크립타제는 프라임 편집으로 작동할 수 있는 DNA 폴리머라제의 단지 한 유형인 것으로 본원에 제시된다. 따라서, 본 명세서가 "리버스 트랜스크립타제"를 언급할 때마다, 관련 기술분야의 통상의 기술자는 리버스 트랜스크립타제 대신에 임의의 적합한 DNA 폴리머라제가 사용될 수 있다는 것을 인지하여야 한다. 따라서, 한 측면에서, 프라

임 편집제는 Cas9 (또는 등가의 napDNAbp)를 포함할 수 있으며, 이는 그것을 표적 DNA 내의 프로토스페이서의 상보체에 어닐링하는 스페이서 서열을 함유하는 특수화된 가이드 RNA (즉, PEgRNA)와 회합시킴으로써 DNA 서열을 표적화하도록 프로그램화된다. 특수화된 가이드 RNA는 또한 표적 부위에서 상응하는 내인성 DNA 가닥을 대체하는 데 사용되는 목적하는 유전자 변경을 함유하는 DNA의 대체 가닥을 코딩하는 연장부의 형태로 새로운 유전자 정보를 함유한다. PEgRNA로부터 표적 DNA로 정보를 전달하기 위해, 프라임 편집의 메카니즘은 DNA의 한 가닥 내의 표적 부위를 닉킹하여 3'-히드록실 기를 노출시키는 것을 수반한다. 이어서, 노출된 3'-히드록실 기를 사용하여 표적 부위 내로 직접 PEgRNA 상의 편집-코딩 연장부의 DNA 중합을 프라이밍할 수 있다. 다양한 실시양태에서, 편집을 함유하는 대체 가닥의 중합을 위한 주형을 제공하는 연장부는 RNA 또는 DNA로부터 형성될 수 있다. RNA 연장부의 경우에, 프라임 편집제의 폴리머라제는 RNA-의존성 DNA 폴리머라제 (예컨대, 리버스 트랜스크립타제)일 수 있다. DNA 연장부의 경우에, 프라임 편집제의 폴리머라제는 DNA-의존성 DNA 폴리머라제일 수 있다. 본원에 개시된 프라임 편집제에 의해 형성된 새로 합성된 가닥 (즉, 목적하는 편집을 함유하는 대체 DNA 가닥)은 목적하는 뉴클레오티드 변화 (예를 들어, 단일 뉴클레오티드 변화, 결실 또는 삽입 또는 그의 조합)의 포함을 제외하고는 게놈 표적 서열에 상동일 것이다 (즉, 그와 동일한 서열을 가질 것임). DNA의 새로 합성된 (또는 대체) 가닥은 또한 단일 가닥 DNA 플랩으로서 지칭될 수 있으며, 이는 상보적 상동 내인성 DNA 가닥과의 혼성화에 대해 경쟁함으로써 상응하는 내인성 가닥을 대체할 것이다. 특정 실시양태에서, 시스템은 오류-유발 리버스 트랜스크립타제 효소 (예를 들어, Cas9 도메인과의 융합 단백질로서 제공되거나 또는 Cas9 도메인에 대해 트랜스로 제공됨)의 사용과 조합될 수 있다. 오류-유발 리버스 트랜스크립타제 효소는 단일 가닥 DNA 플랩의 합성 동안 변경을 도입할 수 있다. 따라서, 특정 실시양태에서, 오류-유발 리버스 트랜스크립타제를 이용하여 표적 DNA에 뉴클레오티드 변화를 도입할 수 있다. 시스템과 함께 사용되는 오류-유발 리버스 트랜스크립타제에 따라, 변화는 무작위 또는 비-무작위일 수 있다. 혼성화된 중간체 (내인성 DNA 가닥에 혼성화된 리버스 트랜스크립타제에 의해 합성된 단일 가닥 DNA 플랩을 포함함)의 분해는 생성된 내인성 DNA의 대체 플랩의 제거 (예를 들어, 5' 단부 DNA 플랩 엔도뉴클레아제인 FEN1을 사용함), 합성된 단일 가닥 DNA 플랩의 표적 DNA에 대한 라이게이션, 및 세포 DNA 복구 및/또는 복제 과정의 결과로서 목적하는 뉴클레오티드 변화의 동화를 포함할 수 있다. 주형화된 DNA 합성은 삽입 및 결실을 포함한 임의의 뉴클레오티드의 변형을 위한 단일 뉴클레오티드 정밀도를 제공하기 때문에, 이러한 접근법의 범주는 매우 넓고, 기초 과학 및 요법에서 무수한 적용에 사용될 수 있을 것으로 예측된다.

[0222] 다양한 실시양태에서, 프라임 편집은 표적 DNA 분자 (뉴클레오티드 서열 내의 변화가 도입되는 것이 요구됨)를 프라임 편집 가이드 RNA (PEgRNA)와 복합체화된 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)과 접촉시킴으로써 작동한다. 도 1g를 참조하면, 프라임 편집 가이드 RNA (PEgRNA)는 가이드 RNA의 3' 또는 5' 단부에 또는 가이드 RNA 내의 분자내 위치에 연장부를 포함하고, 목적하는 뉴클레오티드 변화 (예를 들어, 단일 뉴클레오티드 변화, 삽입 또는 결실)를 코딩한다. 단계 (a)에서, napDNAbp/연장된 gRNA 복합체는 DNA 분자와 접촉하고, 연장된 gRNA는 napDNAbp가 표적 유전자좌에 결합하도록 가이드한다. 단계 (b)에서, 표적 유전자좌의 DNA의 가닥 중 하나에 닉을 도입하여 (예를 들어, 뉴클레아제 또는 화학적 작용제에 의해), 표적 유전자좌의 가닥 중 하나에 이용가능한 3' 단부를 생성한다. 특정 실시양태에서, 닉은 R-루프 가닥에 상응하는 DNA의 가닥, 즉 가이드 RNA 서열에 혼성화되지 않는 가닥, 즉 "비-표적 가닥"에서 생성된다. 그러나, 닉은 가닥 중 어느 하나에 도입될 수 있다. 즉, 닉은 R-루프 "표적 가닥" (즉, 연장된 gRNA의 프로토스페이서에 혼성화된 가닥) 또는 "비-표적 가닥" (즉, R-루프의 단일 가닥 부분을 형성하고 표적 가닥에 상보적인 가닥) 내로 도입될 수 있다. 단계 (c)에서, DNA 가닥의 3' 단부 (닉에 의해 형성됨)는 역전사를 프라이밍하기 위해 (즉, "표적-프라이밍된 RT") 가이드 RNA의 연장된 부분과 상호작용한다. 특정 실시양태에서, 3' 단부 DNA 가닥은 가이드 RNA의 연장된 부분 상의 특이적 RT 프라이밍 서열, 즉 PEgRNA 상의 "리버스 트랜스크립타제 프라이밍 서열" 또는 "프라이머 결합 부위"에 혼성화한다. 단계 (d)에서, 프라이밍된 부위의 3' 단부로부터 프라임 편집 가이드 RNA의 5' 단부를 향해 DNA의 단일 가닥을 합성하는 리버스 트랜스크립타제 (또는 다른 적합한 DNA 폴리머라제)가 도입된다. DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)는 napDNAbp에 융합될 수 있거나 또는 대안적으로 napDNAbp에 트랜스로 제공될 수 있다. 이는 목적하는 뉴클레오티드 변화 (예를 들어, 단일 염기 변화, 삽입 또는 결실 또는 그의 조합)를 포함하고, 그 외에는 닉 부위의 또는 그에 인접한 내인성 DNA에 상동인 단일 가닥 DNA 플랩을 형성한다. 단계 (e)에서, napDNAbp 및 가이드 RNA가 방출된다. 단계 (f) 및 (g)는 목적하는 뉴클레오티드 변화가 표적 유전자좌 내로 혼입되도록 하는 단일 가닥 DNA 플랩의 분해에 관한 것이다. 이러한 과정은 3' 단일 가닥 DNA 플랩이 내인성 DNA 서열에 침입하고 혼성화하였을 때 형성되는 상응하는 5' 내인성 DNA 플랩을 제거함으로써, 목적하는 생성물 형성을 향해 구동될 수 있다. 이론에 얽매이지는 않지만, 세포 내인성 DNA 복구 및 복제 과정은 미스매칭된 DNA를 분해하여 뉴클레오티드 변화(들)를 혼입시켜 목적하는 변경된 생성물을

형성한다. 과정은 또한 도 1f에 예시된 바와 같이 "제2 가닥 닉킹"에 의해 생성물 형성을 향해 구동될 수 있다. 이 과정은 하기 유전자 변화: 전환, 전이, 결실 및 삽입 중 적어도 하나 이상을 도입할 수 있다.

[0223] 용어 "프라임 편집제 (PE) 시스템" 또는 "프라임 편집제 (PE)" 또는 "PE 시스템" 또는 "PE 편집 시스템"은 napDNAbp, 리버스 트랜스크립타제, 융합 단백질 (예를 들어, napDNAbp 및 리버스 트랜스크립타제를 포함함), 프라임 편집 가이드 RNA, 및 융합 단백질 및 프라임 편집 가이드 RNA를 포함하는 복합체, 뿐만 아니라 프라임 편집 과정을 편집된 생성물 형성을 향해 구동시키는 것을 돕기 위한 보조 요소, 예컨대 제2 가닥 닉킹 성분 (예를 들어, 제2 가닥 sgRNA) 및 5' 내인성 DNA 플랩 제거 엔도뉴클레아제 (예를 들어, FEN1)를 포함하나 이에 제한되지는 않는, 본원에 기재된 표적-프라이밍된 역전사 (TPRT)를 사용하는 게놈 편집 방법에 수반되는 조성물을 지칭한다.

[0224] 지금까지 기재된 실시양태에서 PEgRNA는 가이드 RNA (그 자체가 스페이서 서열 및 gRNA 코어 또는 스캐폴드를 포함함) 및 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 5' 또는 3' 연장 아암을 포함하는 단일 분자를 구성하지만 (예를 들어, 도 3d 참조), PEgRNA는 또한 가이드 RNA 및 트랜스 프라임 편집제 RNA 주형 (tPERT)으로 구성된 2개의 개별 분자의 형태를 취할 수 있으며, 이는 본질적으로 동일한 분자 내에 연장 아암 (특히, 프라이머 결합 부위 및 DNA 합성 도메인 포함) 및 RNA-단백질 동원 도메인 (예를 들어, MS2 압타머 또는 헤어핀)을 수용하여, tPERT 동원 단백질 (예를 들어, MS2 압타머에 결합하는 MS2cp 단백질)을 포함하는 변형된 프라임 편집제 복합체로 공동-국재화되거나 동원된다. 프라임 편집에 사용될 수 있는 tPERT의 예로서 도 3g 및 도 3h를 참조한다.

[0225] "이중-플랩 프라임 편집 시스템"에서는, 2개의 pegRNA를 사용하여 게놈 부위의 대향 가닥을 표적화하고, 편집된 DNA 서열을 함유하는 2개의 상보적 3' 플랩의 합성을 지시한다 (도 91). 전형적 프라임 편집과는 달리, 상보적인 편집된 가닥이 대신 혼성화에 이용가능하기 때문에, 편집된 DNA 가닥의 쌍 (3' 플랩)이 내인성 게놈 DNA에서 5' 플랩과 직접 경쟁하는 것이 요구되지 않는다. 듀플렉스의 두 가닥이 편집된 DNA로서 합성되기 때문에, 이중-플랩 프라임 편집 시스템은 전형적 프라임 편집에 의해 요구되는 비-편집된 상보적 DNA 가닥의 대체에 대한 필요를 제거한다. 대신, 세포 DNA 복구 기구는 단지 쌍형성된 5' 플랩 (원래 게놈 DNA)만을 절제하고, 쌍형성된 3' 플랩 (편집된 DNA)을 유전자좌 내로 라이게이션시킬 필요가 있다. 따라서, 또한 새로 합성된 DNA 가닥에 게놈 DNA에 상동인 서열을 포함할 필요가 없으므로, 새로운 가닥의 선택적 혼성화를 허용하고 최소 게놈 상동성을 함유하는 편집을 용이하게 한다. DNA의 두 가닥을 절단하는 프라임 편집제의 뉴클레아제-활성 버전은 또한 원래 DNA 서열의 제거를 가속화하는 데 사용될 수 있다.

[0226] 전형적 프라임 편집과 같이, 다중-플랩 프라임 편집 (이중-플랩 및 사중-플랩 프라임 편집 포함)은 핵산 프로그램가능한 DNA 결합 단백질 ("napDNAbp") (이는 폴리머라제와 회합하여 (즉, 융합 단백질의 형태로 또는 달리 napDNAbp와 트랜스로 제공됨) 작동함)을 사용하여 특정된 DNA 부위 내로 새로운 유전자 정보를 직접 기록하는 다재다능한 정확한 게놈 편집 방법이며, 여기서 프라임 편집 시스템은 표적 부위를 특정하고 또한 가이드 RNA 상에 (예를 들어, 5' 또는 3' 단부에 또는 가이드 RNA의 내부 부분에) 조작된 연장부 (DNA 또는 RNA)에 의해 대체 DNA 가닥의 형태로 목적하는 편집의 합성의 주형이 되는 프라임 편집 (PE) 가이드 RNA ("PEgRNA")로 프로그램화된다. 목적하는 편집 (예를 들어, 단일 핵염기 치환)을 함유하는 대체 가닥은 (목적하는 편집을 포함하는 것을 제외하고는) 편집될 표적 부위의 내인성 가닥과 동일한 서열을 공유한다. DNA 복구 및/또는 복제 기구를 통해, 표적 부위의 내인성 가닥은 목적하는 편집을 함유하는 새로 합성된 대체 가닥에 의해 대체된다. 일부 경우에, 프라임 편집은 "검색-및-대체" 게놈 편집 기술로서 생각될 수 있으며, 이는 프라임 편집제가, 본원에 기재된 바와 같이, 편집될 목적하는 표적 부위를 검색하고 위치를 찾아낼 뿐만 아니라, 동시에, 상응하는 표적 부위 내인성 DNA 가닥 대신 설치된 목적하는 편집을 함유하는 대체 가닥을 코딩하기 때문이다.

[0227] 프라임 편집제

[0228] 본원에 기재된 이중 프라임 편집 시스템은 프라임 편집제의 쌍을 포함한다. 용어 "프라임 편집제"는 napDNAbp (예를 들어, Cas9 닉카제) 및 리버스 트랜스크립타제를 포함하는 본원에 기재된 융합 구축물을 지칭하고, PEgRNA (또는 "연장된 가이드 RNA")의 존재 하에 표적 뉴클레오티드 서열 상에서 프라임 편집을 수행할 수 있다. 용어 "프라임 편집제"는 융합 단백질 또는 PEgRNA와 복합체화되고/거나 제2 가닥 닉킹 sgRNA와 추가로 복합체화된 융합 단백질을 지칭할 수 있다. 일부 실시양태에서, 프라임 편집제는 또한 융합 단백질 (napDNAbp에 융합된 리버스 트랜스크립타제), PEgRNA 및 본원에 기재된 바와 같은 비-편집된 가닥의 제2-부위 닉킹 단계를 지시할 수 있는 정규 가이드 RNA를 포함하는 복합체를 지칭할 수 있다. 다른 실시양태에서, "프라이머 편집제"의 리버스 트랜스크립타제 성분은 트랜스로 제공될 수 있다.

[0229] 본원에 기재된 이중-플랩 프라임 편집 시스템은 프라임 편집제의 쌍을 포함한다. 본원에 기재된 사중-플랩 프라임 편집 시스템은 4종의 프라임 편집제를 포함한다.

[0230] 프라이머 결합 부위

[0231] 용어 "프라이머 결합 부위" 또는 "PBS"는 연장 아암의 성분으로서 PEgRNA 상에 (전형적으로 연장 아암의 3' 단부에) 위치하는 뉴클레오티드 서열을 지칭하고, 프라임 편집제에 의한 표적 서열의 Cas9 닉킹 후에 형성된 프라이머 서열에 결합하는 역할을 한다. 다른 곳에 상술된 바와 같이, 프라임 편집제의 Cas9 닉카제 성분이 표적 DNA 서열의 한 가닥을 닉킹하는 경우에, 3'-단부 ssDNA 플랩이 형성되며, 이는 PEgRNA 상의 프라이머 결합 부위에 어닐링하여 역전사를 프라이밍하는 프라이머 서열의 역할을 한다.

[0232] 프로모터

[0233] 용어 "프로모터"는 관련 기술분야에 인식되어 있으며, 세포 전사 기구에 의해 인식되고 하류 유전자의 전사를 개시할 수 있는 서열을 갖는 핵산 분자를 지칭한다. 프로모터는 구성적으로 활성 (프로모터가 주어진 세포 맥락에서 항상 활성임을 의미함)일 수 있거나 또는 조건부로 활성 (프로모터가 특이적 조건의 존재 하에서만 활성임을 의미함)일 수 있다. 예를 들어, 조건부 프로모터는 프로모터 내의 조절 요소와 회합된 단백질을 기본 전사 기구에 연결하는 특이적 단백질의 존재 하에서만 또는 억제 분자의 부재 하에서만 활성일 수 있다. 조건부로 활성인 프로모터의 하위부류는 활성을 위해 소분자 "유도제"의 존재를 필요로 하는 유도성 프로모터이다. 유도성 프로모터의 예는 아라비노스-유도성 프로모터, Tet-on 프로모터 및 타목시펜-유도성 프로모터를 포함하나 이에 제한되지는 않는다. 다양한 구성적, 조건부 및 유도성 프로모터가 통상의 기술자에게 널리 공지되어 있고, 통상의 기술자는 본 발명을 수행하는데 유용한 다양한 이러한 프로모터를 확인할 수 있을 것이며, 이와 관련하여 제한은 없다.

[0234] 프로토스페이서

[0235] 본원에 사용된 용어 "프로토스페이서"는 PAM (프로토스페이서 인접 모티프) 서열에 인접한 DNA 내의 서열 (~20 bp)을 지칭한다. 프로토스페이서는 가이드 RNA의 스페이서 서열과 동일한 서열을 공유한다. 가이드 RNA는 표적 DNA 상의 프로토스페이서 서열의 상보체 (구체적으로, 그의 한 가닥, 즉 표적 DNA 서열의 "비-표적 가닥"에 비해 "표적 가닥")에 어닐링한다. Cas9가 기능하기 위해, 이는 또한 Cas9 유전자의 박테리아 종에 따라 달라지는 특이적 프로토스페이서 인접 모티프 (PAM)를 필요로 한다. 에스. 피오게네스로부터 유래된 가장 통상적으로 사용되는 Cas9 뉴클레아제는 비-표적 가닥 상에서, 게놈 DNA 내의 표적 서열의 바로 하류에서 발견되는 NGG의 PAM 서열을 인식한다. 통상의 기술자는 최신 기술의 문헌이 때때로 "프로토스페이서"를 "스페이서"로서 지칭하기 보다는 가이드 RNA 자체 상의 ~20-nt 표적-특이적 가이드 서열로서 지칭한다는 것을 인지할 것이다. 따라서, 일부 경우에, 본원에 사용된 용어 "프로토스페이서"는 용어 "스페이서"와 상호교환가능하게 사용될 수 있다. "프로토스페이서" 또는 "스페이서"의 출현을 둘러싼 기재의 문맥은 이 용어가 gRNA에 대한 것인지 또는 DNA 표적에 대한 것인지 여부에 관해 독자에게 알리는 것을 도울 것이다.

[0236] 프로토스페이서 인접 모티프 (PAM)

[0237] 본원에 사용된 용어 "프로토스페이서 인접 서열" 또는 "PAM"은 Cas9 뉴클레아제의 중요한 표적화 성분인 대략 2-6개 염기 쌍 DNA 서열을 지칭한다. 전형적으로, PAM 서열은 어느 한 가닥 상에 있고, Cas9 절단 부위의 5'에서 3' 방향으로 하류에 있다. 정규 PAM 서열 (즉, 스트렙토코쿠스 피오게네스의 Cas9 뉴클레아제 또는 SpCas9와 연관된 PAM 서열)은 5'-NGG-3'이며, 여기서 "N"은 임의의 핵염기, 이어서 2개의 구아닌 ("G") 핵염기이다. 상이한 PAM 서열은 상이한 Cas9 뉴클레아제 또는 상이한 유기체로부터의 등가의 단백질과 연관될 수 있다. 추가로, 임의의 주어진 Cas9 뉴클레아제, 예를 들어 SpCas9는 뉴클레아제의 PAM 특이성을 변경시켜 뉴클레아제가 대안적 PAM 서열을 인식하도록 변형될 수 있다.

[0238] 예를 들어, 서열식별번호: 18의 정규 SpCas9 아미노산 서열과 관련하여, PAM 서열은 (a) PAM 특이성을 NGAN 또는 NGNG로 변경시키는 D1135V, R1335Q 및 T1337R "VQR 변이체", (b) PAM 특이성을 NGAG로 변경시키는 D1135E, R1335Q 및 T1337R "EQR 변이체", 및 (c) PAM 특이성을 NGCG로 변경시키는 D1135V, G1218R, R1335E 및 T1337R "VRER 변이체"를 포함한 1개 이상의 돌연변이를 도입함으로써 변형될 수 있다. 추가로, 정규 SpCas9의 D1135E 변이체는 여전히 NGG를 인식하지만, 야생형 SpCas9 단백질과 비교하여 더 선택적이다.

[0239] 또한, 상이한 박테리아 종으로부터의 Cas9 효소 (즉, Cas9 오르토로그)는 다양한 PAM 특이성을 가질 수 있는 것으로 인지될 것이다. 예를 들어, 스타필로코쿠스 아우레우스(*Staphylococcus aureus*)로부터의 Cas9 (SaCas9)

는 NGRRT 또는 NGRRN을 인식한다. 추가로, 네이세리아 메닌기티티스(*Neisseria meningitis*)로부터의 Cas9 (NmCas)는 NNNNGATT를 인식한다. 또 다른 예에서, 스트렙토코쿠스 써모필리스(*Streptococcus thermophilis*)로부터의 Cas9 (StCas9)는 NNAGAAW를 인식한다. 또 다른 예에서, 트레포네마 덴티콜라(*Treponema denticola*)로부터의 Cas9 (TdCas)는 NAAAAC를 인식한다. 이들은 예이고, 제한적인 것으로 의도되지 않는다. 비-SpCas9는 다양한 PAM 서열에 결합하며, 이는 목적하는 표적 절단 부위에 적합한 SpCas9 PAM 서열이 존재하지 않는 경우에 이들은 유용한 것으로 만든다고 추가로 인지될 것이다. 추가로, 비-SpCas9는 이들을 SpCas9보다 더 유용하게 만드는 다른 특징을 가질 수 있다. 예를 들어, 스타필로코쿠스 아우레우스로부터의 Cas9 (SaCas9)는 SpCas9보다 약 1 킬로염기 더 작아서, 아데노-연관 바이러스 (AAV) 내로 패키징될 수 있다. 추가로, 문헌 [Shah et al., "Protospacer recognition motifs: mixed identities and functional diversity," RNA Biology, 10(5): 891-899] (이는 본원에 참조로 포함됨)을 참조할 수 있다.

[0240] 레콤비나제

[0241] 본원에 사용된 용어 "레콤비나제"는 레콤비나제 인식 서열들 사이의 DNA 단편의 절제, 통합, 역위 또는 교환 (예를 들어, 전위)을 유발하는, 레콤비나제 인식 서열들 사이의 DNA의 재조합을 매개하는 부위-특이적 효소를 지칭한다. 레콤비나제는 다음 2종의 별개의 패밀리로 분류될 수 있다: 세린 레콤비나제 (예를 들어, 레솔바제 및 인버타제) 및 티로신 레콤비나제 (예를 들어, 인테그라제). 세린 레콤비나제의 예는 비제한적으로 Hin, Gin, Tn3, β-6, CinH, ParA, γδ, Bxb1, ΦC31, TP901, TG1, ϕBT1, R4, ϕRV1, ϕFC1, MR11, A118, U153 및 gp29를 포함한다. 티로신 레콤비나제의 예는 비제한적으로 Cre, FLP, R, 람다, HK101, HK022 및 pSAM2를 포함한다. 세린 및 티로신 레콤비나제 명칭은 레콤비나제가 DNA를 공격하기 위해 사용하고 가닥 교환 동안 DNA에 공유 연결되는 보존적 친핵성 아미노산 잔기로부터 유래한다. 레콤비나제는 유전자 녹아웃/녹인의 생성 및 유전자 요법 적용을 포함한 수많은 적용을 갖는다. 예를 들어, 문헌 [Brown et al., "Serine recombinases as tools for genome engineering." Methods. 2011;53(4):372-9; Hirano et al., "Site-specific recombinases as tools for heterologous gene integration." Appl. Microbiol. Biotechnol. 2011; 92(2):227-39; Chavez and Calos, "Therapeutic applications of the ΦC31 integrase system." Curr. Gene Ther. 2011;11(5):375-81; Turan and Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." FASEB J. 2011; 25(12):4088-107; Venken and Bellen, "Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and ΦC31 integrase." Methods Mol. Biol. 2012; 859:203-28; Murphy, "Phage recombinases and their applications." Adv. Virus Res. 2012; 83:367-414; Zhang et al., "Conditional gene manipulation: Cre-ating a new biological era." J. Zhejiang Univ. Sci. B. 2012; 13(7):511-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." DNA Repair (Amst). 2012; 1;11(10):781-8]을 참조하며; 이들 각각의 전체 내용은 그 전문이 본원에 참조로 포함된다. 본원에 제공된 레콤비나제는 본 발명의 실시양태에서 사용될 수 있는 레콤비나제의 배타적 예는 아닌 것으로 의도된다. 본 발명의 방법 및 조성물은 새로운 직교 레콤비나제에 대한 데이터베이스를 채굴하거나 또는 정의된 DNA 특이성을 갖는 합성 레콤비나제를 설계함으로써 확장될 수 있다 (예를 들어, 문헌 [Groth et al., "Phage integrases: biology and applications." J. Mol. Biol. 2004; 335, 667-678; Gordley et al., "Synthesis of programmable integrases." Proc. Natl. Acad. Sci. U S A. 2009; 106, 5053-5058] 참조; 각각의 전체 내용은 그 전문이 본원에 참조로 포함됨). 본원에 기재된 방법 및 조성물에 유용한 레콤비나제의 다른 예는 관련 기술분야의 통상의 기술자에게 공지되어 있고, 발견되거나 생성되는 어떠한 새로운 레콤비나제도 본 발명의 상이한 실시양태에 사용될 수 있는 것으로 예상된다. 일부 실시양태에서, 레콤비나제의 촉매 도메인은 뉴클레아제-불활성화된 RNA-프로그램가능한 뉴클레아제 (예를 들어, dCas9 또는 그의 단편)와 융합되어, 레콤비나제 도메인이 핵산 결합 도메인을 포함하지 않거나 또는 표적 핵산에 결합할 수 없도록 한다 (예를 들어, 이러한 레콤비나제 도메인은 특이적 DNA 결합 활성을 갖지 않도록 조작됨). DNA 결합 활성이 결여된 레콤비나제 및 이를 조작하는 방법은 공지되어 있고, 문헌 [Klippel et al., "Isolation and characterisation of unusual gin mutants." EMBO J. 1988; 7: 3983-3989: Burke et al., "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol Microbiol. 2004; 51: 937-948; Olorunniji et al., "Synapsis and catalysis by activated Tn3 resolvase mutants." Nucleic Acids Res. 2008; 36: 7181-7191; Rowland et al., "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." Mol Microbiol. 2009; 74: 282-298; Akopian et al., "Chimeric recombinases with designed DNA sequence recognition." Proc Natl Acad Sci USA. 2003;100: 8688-8691; Gordley et al., "Evolution of programmable zinc finger-recombinases with activity in human cells. J Mol Biol. 2007; 367: 802-813; Gordley et al., "Synthesis of programmable

integrases." Proc Natl Acad Sci USA. 2009;106: 5053-5058; Arnold et al., "Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity." EMBO J. 1999;18: 1407-1414; Gaj et al., "Structure-guided reprogramming of serine recombinase DNA sequence specificity." Proc Natl Acad Sci USA. 2011;108(2):498-503; 및 Proudfoot et al., "Zinc finger recombinases with adaptable DNA sequence specificity." PLoS One. 2011;6(4):e19537] (각각의 전체 내용은 본원에 참조로 포함됨)에 기재된 것을 포함한다. 예를 들어, 레솔바제-인버타제 군의 세린 레콤비나제, 예를 들어 Tn3 및 γδ 레솔바제 및 Hin 및 Gin 인버타제는 자율 촉매 및 DNA-결합 도메인을 갖는 모듈 구조를 갖는다 (예를 들어, 문헌 [Grindley et al., "Mechanism of site-specific recombination." Ann Rev Biochem. 2006; 75: 567-605] 참조, 이의 전체 내용은 참조로 포함됨). 따라서, 이들 레콤비나제의 촉매 도메인은, 예를 들어 어떠한 보조 인자 (예를 들어, DNA 결합 활성)도 필요로 하지 않는 '활성화된' 레콤비나제 돌연변이체의 단리 후 본원에 기재된 바와 같은 뉴클레아제-불활성화된 RNA-프로그램가능한 뉴클레아제 (예를 들어, dCas9 또는 그의 단편)와 재조합될 수 있다 (예를 들어, 문헌 [Klippel et al., "Isolation and characterisation of unusual gin mutants." EMBO J. 1988; 7: 3983-3989: Burke et al., "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. Mol Microbiol. 2004; 51: 937-948; Olorunniji et al., "Synapsis and catalysis by activated Tn3 resolvase mutants." Nucleic Acids Res. 2008; 36: 7181-7191; Rowland et al., "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." Mol Microbiol. 2009; 74: 282-298; Akopian et al., "Chimeric recombinases with designed DNA sequence recognition." Proc Natl Acad Sci USA. 2003;100: 8688-8691] 참조). 추가적으로, N-말단 촉매 도메인 및 C-말단 DNA 결합 도메인을 갖는 많은 다른 천연 세린 레콤비나제가 공지되어 있고 (예를 들어, phiC31 인테그라제, TnpX 트랜스포사제, IS607 트랜스포사제), 그의 촉매 도메인은 본원에 기재된 바와 같은 프로그램가능한 부위-특이적 레콤비나제를 조작하기 위해 공동-선택될 수 있다 (예를 들어, 문헌 [Smith et al., "Diversity in the serine recombinases." Mol Microbiol. 2002;44: 299-307] 참조, 이의 전체 내용은 참조로 포함됨). 유사하게, 티로신 레콤비나제 (예를 들어, Cre, λ 인테그라제)의 코어 촉매 도메인은 공지되어 있고, 본원에 기재된 바와 같은 프로그램가능한 부위-특이적 레콤비나제를 조작하기 위해 유사하게 공동-선택될 수 있다 (예를 들어, 문헌 [Guo et al., "Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse." Nature. 1997; 389:40-46; Hartung et al., "Cre mutants with altered DNA binding properties." J Biol Chem 1998; 273:22884-22891; Shaikh et al., "Chimeras of the Flp and Cre recombinases: Tests of the mode of cleavage by Flp and Cre. J Mol Biol. 2000; 302:27-48; Rongrong et al., "Effect of deletion mutation on the recombination activity of Cre recombinase." Acta Biochim Pol. 2005; 52:541-544; Kilbride et al., "Determinants of product topology in a hybrid Cre-Tn3 resolvase site-specific recombination system." J Mol Biol. 2006; 355:185-195; Warren et al., "A chimeric cre recombinase with regulated directionality." Proc Natl Acad Sci USA. 2008 105:18278-18283; Van Duyne, "Teaching Cre to follow directions." Proc Natl Acad Sci USA. 2009 Jan 6;106(1):4-5; Numrych et al., "A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophage λ." Nucleic Acids Res. 1990; 18:3953-3959; Tirumalai et al., "The recognition of core-type DNA sites by λ integrase." J Mol Biol. 1998; 279:513-527; Aihara et al., "A conformational switch controls the DNA cleavage activity of λ integrase." Mol Cell. 2003; 12:187-198; Biswas et al., "A structural basis for allosteric control of DNA recombination by λ integrase." Nature. 2005; 435:1059-1066; 및 Warren et al., "Mutations in the amino-terminal domain of λ-integrase have differential effects on integrative and excisive recombination." Mol Microbiol. 2005; 55:1104-1112] 참조; 각각의 전체 내용은 참조로 포함됨).

[0242] 레콤비나제 인식 서열

[0243] 본원에 사용된 용어 "레콤비나제 인식 서열" 또는 동등하게 "RRS" 또는 "레콤비나제 표적 서열" 또는 "레콤비나제 부위"는 레콤비나제에 의해 인식되는 뉴클레오티드 서열 표적을 지칭하며, 이는 RRS를 갖는 또 다른 DNA 분자와의 가닥 교환을 겪어 레콤비나제 인식 서열들 사이의 DNA 단편의 절제, 통합, 역위 또는 교환을 유발한다. 다양한 실시양태에서, 다중-가닥 프라임 편집제는 표적 서열에 또는 1개 초과의 표적 서열에 1개 이상의 레콤비나제 부위를 설치할 수 있다. 1개 초과의 레콤비나제 부위가 다중-가닥 프라임 편집제에 의해 설치되는 경우에, 레콤비나제 부위는 인접한 표적 부위 또는 비-인접한 표적 부위 (예를 들어, 개별 염색체)에 설치될 수 있다. 다양한 실시양태에서, 단일 설치된 레콤비나제 부위는 게놈 레콤비나제 부위와 외인적으로 공급된 핵산 분자, 예를 들어 플라스미드 내의 제2 레콤비나제 부위 사이의 레콤비나제-매개된 반응을 위한 "랜딩 부위"로서

사용될 수 있다. 이는 목적하는 핵산 분자의 표적화된 통합을 가능하게 한다. 다른 실시양태에서, 2개의 레콤비나제 부위가 DNA의 인접한 영역 (예를 들어, 25-50 bp, 50-100 bp, 100-200 bp, 200-300 bp, 300-400 bp, 400-500 bp, 500-600 bp, 600-700 bp, 700-800 bp, 800-900 bp, 900-1000 bp, 1000-2000 bp, 2000-3000 bp, 3000-4000 bp, 4000-5000 bp 또는 그 초과만큼 분리됨)에 삽입되는 경우에, 레콤비나제 부위는 개재 서열의 레콤비나제-매개된 절제 또는 역위 또는 동일한 레콤비나제 부위를 갖는 외인성 DNA와의 레콤비나제-매개된 카세트 교환에 사용될 수 있다. 2개 이상의 레콤비나제 부위가 2개의 상이한 염색체 상에 다중-플랩 프라임 편집제에 의해 설치되는 경우에, 개재 서열의 전위는 제1 염색체 위치에서 제2 위치로 발생할 수 있다.

[0244] 재조합하다 또는 재조합

[0245] 핵산 변형 (예를 들어, 게놈 변형)의 문맥에서 용어 "재조합하다" 또는 "재조합"은 2개 이상의 핵산 분자 또는 단일 핵산 분자의 2개 이상의 영역이 레콤비나제 단백질 (예를 들어, 본원에 제공된 본 발명의 레콤비나제 융합 단백질)의 작용에 의해 변형되는 과정을 지칭하는 데 사용된다. 재조합은, 특히, 예를 들어 1개 이상의 핵산 분자 내에 또는 그 사이에 핵산의 삽입, 역위, 절제 또는 전위를 유발할 수 있다.

[0246] 리버스 트랜스크립타제

[0247] 용어 "리버스 트랜스크립타제"는 RNA-의존성 DNA 폴리머라제로서 특징화되는 폴리머라제의 부류를 기재한다. 모든 공지된 리버스 트랜스크립타제는 RNA 주형으로부터 DNA 전사체를 합성하기 위한 프라이머를 필요로 한다. 역사적으로, 리버스 트랜스크립타제는 주로 mRNA를 cDNA로 전사하는 데 사용되어 왔으며, 이는 이어서 추가의 조작을 위해 벡터 내로 클로닝될 수 있다. 조류 골수모구증 바이러스 (AMV) 리버스 트랜스크립타제가 최초로 널리 사용된 RNA-의존성 DNA 폴리머라제였다 (Verma, Biochim. Biophys. Acta 473:1 (1977)). 효소는 5'-3' RNA-지시된 DNA 폴리머라제 활성, 5'-3' DNA-지시된 DNA 폴리머라제 활성 및 RNase H 활성을 갖는다. RNase H는 RNA-DNA 하이브리드에 대한 RNA 가닥에 특이적인 진행성 5' 및 3' 리보뉴클레아제이다 (Perbal, A Practical Guide to Molecular Cloning, New York: Wiley & Sons (1984)). 전사에서의 오류는 리버스 트랜스크립타제에 의해 교정될 수 없으며, 이는 공지된 바이러스 리버스 트랜스크립타제가 교정에 필요한 3'-5' 엑소뉴클레아제 활성이 결여되어 있기 때문이다 (Saunders and Saunders, Microbial Genetics Applied to Biotechnology, London: Croom Helm (1987)). AMV 리버스 트랜스크립타제의 활성 및 그의 연관된 RNase H 활성의 상세한 연구는 문헌 [Berger et al., Biochemistry 22:2365-2372 (1983)]에 제시되었다. 분자 생물학에서 광범위하게 사용되는 또 다른 리버스 트랜스크립타제는 몰로니 뮤린 백혈병 바이러스 (M-MLV)로부터 기원하는 리버스 트랜스크립타제이다. 예를 들어, 문헌 [Gerard, G. R., DNA 5:271-279 (1986) 및 Kotewicz, M. L., et al., Gene 35:249-258 (1985)]을 참조한다. RNase H 활성이 실질적으로 결여된 M-MLV 리버스 트랜스크립타제가 또한 기재되었다. 예를 들어, 미국 특허 번호 5,244,797을 참조한다. 본 발명은 임의의 이러한 리버스 트랜스크립타제 또는 그의 변이체 또는 돌연변이체의 사용을 고려한다.

[0248] 추가로, 본 발명은 오류-유발성인, 즉 오류-유발 리버스 트랜스크립타제 또는 중합 동안 뉴클레오티드의 고충실도 혼입을 지지하지 않는 리버스 트랜스크립타제로 지칭될 수 있는 리버스 트랜스크립타제의 사용을 고려한다. 가이드 RNA와 통합된 RT 주형을 기반으로 하는 단일 가닥 DNA 플랩의 합성 동안, 오류-유발 리버스 트랜스크립타제는 RT 주형 서열과 미스매칭된 1개 이상의 뉴클레오티드를 도입하여, 단일 가닥 DNA 플랩의 오류성 중합을 통해 뉴클레오티드 서열에 변화를 도입할 수 있다. 이어서, 단일 가닥 DNA 플랩의 합성 동안 도입된 이들 오류는 상응하는 내인성 표적 가닥에 대한 혼성화, 내인성 대체된 가닥의 제거, 라이게이션을 통해 및 이어서 1 라운드 이상의 내인성 DNA 복구 및/또는 서열분석 과정을 통해 이중 가닥 분자 내로 통합된다.

[0249] 역전사

[0250] 본원에 사용된 용어 "역전사"는 주형으로서 RNA를 사용하여 DNA 가닥 (즉, 상보적 DNA 또는 cDNA)을 합성하는 효소의 능력을 나타낸다. 일부 실시양태에서, 역전사는 "오류-유발 역전사"일 수 있으며, 이는 그의 DNA 중합 활성에 있어서 오류-유발성인 특정 리버스 트랜스크립타제 효소의 특성을 지칭한다.

[0251] PACE

[0252] 본원에 사용된 용어 "파지-보조 연속 진화 (PACE)"는 바이러스 벡터로서 파지를 사용하는 연속 진화를 지칭한다. PACE 기술의 일반적 개념은, 예를 들어 2010년 3월 11일에 WO 2010/028347로서 공개된, 2009년 9월 8일에 출원된 국제 PCT 출원 PCT/US2009/056194; 2012년 6월 28일에 WO 2012/088381로서 공개된, 2011년 12월 22일에 출원된 국제 PCT 출원 PCT/US2011/066747; 2015년 5월 5일에 허여된 미국 출원, 미국 특허 번호 9,023,594, 2015년 9월 11일에 WO 2015/134121로서 공개된, 2015년 1월 20일에 출원된 국제 PCT 출원

PCT/US2015/012022 및 2016년 10월 20일에 WO 2016/168631로서 공개된, 2016년 4월 15일에 출원된 국제 PCT 출원 PCT/US2016/027795에 기재되어 있으며, 이들 각각의 전체 내용은 본원에 참조로 포함된다.

[0253] 파지

[0254] 용어 "박테리오파지"와 상호교환가능하게 본원에 사용된 용어 "파지"는 박테리아 세포를 감염시키는 바이러스를 지칭한다. 전형적으로, 파지는 유전 물질을 둘러싸는 외부 단백질 캡시드로 이루어진다. 유전 물질은 선형 또는 원형 형태의 ssRNA, dsRNA, ssDNA 또는 dsDNA일 수 있다. 파지 및 파지 벡터는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있고, 본원에 제공된 PACE 방법을 수행하는데 유용한 파지의 비제한적 예는 λ (용원), T2, T4, T7, T12, R17, M13, MS2, G4, P1, P2, P4, Phi X174, N4, Φ6 및 Φ29이다. 특정 실시양태에서, 본 발명에서 이용되는 파지는 M13이다. 추가의 적합한 파지 및 숙주 세포는 관련 기술분야의 통상의 기술자에게 명백할 것이고, 본 발명은 상기 측면에서 제한되지 않는다. 추가의 적합한 파지 및 숙주 세포의 예시적인 기재에 대해서는 문헌 [Elizabeth Kutter and Alexander Sulakvelidze: Bacteriophages: Biology and Applications. CRC Press; 1st edition (December 2004), ISBN: 0849313368; Martha R. J. Clokie and Andrew M. Kropinski: Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions (Methods in Molecular Biology) Humana Press; 1st edition (December, 2008), ISBN: 1588296822; Martha R. J. Clokie and Andrew M. Kropinski: Bacteriophages: Methods and Protocols, Volume 2: Molecular and Applied Aspects (Methods in Molecular Biology) Humana Press; 1st edition (December 2008), ISBN: 1603275649]; 이들 모두는 적합한 파지 및 숙주 세포의 개시내용뿐만 아니라 이러한 파지의 단리, 배양 및 조작을 위한 방법 및 프로토콜에 대해 그 전문이 본원에 참조로 포함됨)을 참조한다.

[0255] 단백질, 펩티드 및 폴리펩티드

[0256] 용어 "단백질", "펩티드" 및 "폴리펩티드"는 본원에서 상호교환가능하게 사용되고, 펩티드 (아미드) 결합에 의해 함께 연결된 아미노산 잔기의 중합체를 지칭한다. 상기 용어는 임의의 크기, 구조 또는 기능의 단백질, 펩티드 또는 폴리펩티드를 지칭한다. 전형적으로, 단백질, 펩티드 또는 폴리펩티드는 적어도 3개 아미노산 길이일 것이다. 단백질, 펩티드 또는 폴리펩티드는 개별 단백질 또는 단백질들의 집합을 지칭할 수 있다. 단백질, 펩티드 또는 폴리펩티드 내의 아미노산 중 1개 이상은, 예를 들어 화학 물질, 예컨대 탄수화물 기, 히드록실 기, 포스페이트 기, 파르네실 기, 이소파르네실 기, 지방산 기, 접합, 관능화 또는 다른 변형을 위한 링커 등의 부가에 의해 변형될 수 있다. 단백질, 펩티드 또는 폴리펩티드는 또한 단일 분자일 수 있거나 또는 다분자 복합체일 수 있다. 단백질, 펩티드 또는 폴리펩티드는 단지 자연 발생 단백질 또는 펩티드의 단편일 수 있다. 단백질, 펩티드 또는 폴리펩티드는 자연 발생, 재조합 또는 합성 또는 그의 임의의 조합일 수 있다. 본원에 제공된 임의의 단백질은 관련 기술분야에 공지된 임의의 방법에 의해 생산될 수 있다. 예를 들어, 본원에 제공된 단백질은 재조합 단백질 발현 및 정제를 통해 생산될 수 있으며, 이는 펩티드 링커를 포함하는 융합 단백질에 특히 적합하다. 재조합 단백질 발현 및 정제를 위한 방법은 널리 공지되어 있고, 문헌 [Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 기재된 것을 포함하며, 이의 전체 내용은 본원에 참조로 포함된다.

[0257] 단백질 스플라이싱

[0258] 본원에 사용된 용어 "단백질 스플라이싱"은 서열인 인테인 (또는 경우에 따라 분할 인테인)이 아미노산 서열 내로부터 절제되고, 아미노산 서열의 나머지 단편인 엑스테인이 아미드 결합을 통해 라이게이션되어 연속 아미노산 서열을 형성하는 과정을 지칭한다. 용어 "트랜스" 단백질 스플라이싱은 인테인이 분할 인테인이고 이들이 상이한 단백질 상에 위치하는 특이적 경우를 지칭한다.

[0259] 제2 가닥 닉킹

[0260] 프라임 편집의 결과로서 형성된 헤테로듀플렉스 DNA (즉, 1개의 편집된 가닥 및 1개의 비-편집된 가닥을 함유함)의 분해는 장기간 편집 결과를 결정한다. 즉, 프라임 편집의 목적은 편집된 가닥을 상보체 내인성 가닥 내로 영구적으로 통합시킴으로써 PE의 중간체로서 형성된 헤테로듀플렉스 DNA (내인성 비-편집된 가닥과 쌍형성된 편집된 가닥)를 분해하는 것이다. "제2 가닥 닉킹"의 접근법은 편집된 가닥의 DNA 분자 내로의 영구적 통합에 유리하게 헤테로듀플렉스 DNA의 분해를 구동하는 것을 돕기 위해 본원에 사용될 수 있다. 본원에 사용된 "제2 가닥 닉킹"의 개념은 제1 닉 (즉, 가이드 RNA의 연장된 부분 상에서의 리버스 트랜스크립타제의 프라이밍에 사용하기 위한 유리 3' 단부를 제공하는 초기 닉 부위)의 하류 위치에, 바람직하게는 비편집된 가닥 상에 제2 닉을 도입하는 것을 지칭한다. 특정 실시양태에서, 제1 닉 및 제2 닉은 대향 가닥 상에 있다. 다른 실시양태

에서, 제1 닉 및 제2 닉은 대향 가닥 상에 있다. 또 다른 실시양태에서, 제1 닉은 비-표적 가닥 (즉, R-루프의 단일 가닥 부분을 형성하는 가닥) 상에 있고, 제2 닉은 표적 가닥 상에 있다. 또 다른 실시양태에서, 제1 닉은 편집된 가닥 상에 있고, 제2 닉은 비편집된 가닥 상에 있다. 제2 닉은 제1 닉의 적어도 5개 뉴클레오티드 하류 또는 제1 닉의 적어도 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 또는 150개 또는 그 초과의 뉴클레오티드 하류에 위치할 수 있다. 특정 실시양태에서, 제2 닉은 비편집된 가닥 상에 PEgRNA-유도된 닉의 부위로부터 약 5-150개 뉴클레오티드 또는 약 5-140 또는 약 5-130 또는 약 5-120 또는 약 5-110 또는 약 5-100 또는 약 5-90 또는 약 5-80 또는 약 5-70 또는 약 5-60 또는 약 5-50 또는 약 5-40 또는 약 5-30 또는 약 5-20 또는 약 5-10 개 떨어져 도입될 수 있다. 한 실시양태에서, 제2 닉은 PEgRNA-유도된 닉으로부터 14-116개 뉴클레오티드 떨어져 도입된다. 이론에 얽매이지는 않지만, 제2 닉은 세포의 내인성 DNA 복구 및 복제 과정을 비편집된 가닥의 대체 또는 편집을 향해 유도하여, 편집된 서열을 두 가닥 상에 영구적으로 설치하고 PE의 결과로서 형성된 헤테로듀플렉스를 분해한다. 일부 실시양태에서, 편집된 가닥은 비-표적 가닥이고, 비편집된 가닥은 표적 가닥이다. 다른 실시양태에서, 편집된 가닥은 표적 가닥이고, 비편집된 가닥은 비-표적 가닥이다.

[0261] 센스 가닥

[0262] 유전학에서, "센스" 가닥은 5'에서 3'로 진행하는 이중 가닥 DNA 내의 절편이며, 이는 3'에서 5'로 진행하는 DNA의 안티센스 가닥 또는 주형 가닥에 상보적이다. 단백질을 코딩하는 DNA 절편의 경우에, 센스 가닥은 mRNA와 동일한 서열을 갖는 DNA의 가닥이며, 이는 전사 동안 안티센스 가닥을 그의 주형으로서 취하고, 결국 (전형적으로, 항상은 아님) 단백질로의 번역을 겪는다. 따라서, 안티센스 가닥은 나중에 단백질로 번역되는 RNA를 담당하는 반면, 센스 가닥은 mRNA의 것과 거의 동일한 구성을 갖는다. dsDNA의 각각의 절편에 대해, 판독하는 방향에 따라 센스 및 안티센스의 2개의 세트가 존재할 가능성이 있다는 것 (센스 및 안티센스는 시각이 상대적이기 때문에)을 주목한다. 궁극적으로, dsDNA의 한 절편 중 어느 가닥이 센스 또는 안티센스로 지칭되는지를 좌우하는 것은 유전자 생성물 또는 mRNA이다.

[0263] PEgRNA의 맥락에서, 제1 단계는 PEgRNA 연장 아암으로부터 주형화된, 5'에서 3' 방향으로 배향된 단일 가닥 상보적 DNA (즉, 혼입되는 3' ssDNA 플랩)의 합성이다. 3' ssDNA 플랩이 센스 가닥으로서 간주되어야 하는지 안티센스 가닥으로서 간주되어야 하는지의 여부는 전사 방향에 좌우되며, 이는 DNA의 두 가닥이 전사를 위한 주형으로서의 역할을 할 수 있다는 것 (그러나 동시에는 아님)이 널리 받아들여지고 있기 때문이다. 따라서, 일부 실시양태에서, 3' ssDNA 플랩 (전체적으로 5'에서 3' 방향으로 진행함)은 코딩 가닥이기 때문에 센스 가닥으로서의 역할을 할 것이다. 다른 실시양태에서, 3' ssDNA 플랩 (전체적으로 5'에서 3' 방향으로 진행함)은 안티센스 가닥 및 이에 따라 전사를 위한 주형으로서의 역할을 할 것이다.

[0264] 서열 상동성

[0265] "상동 서열" 또는 또 다른 서열에 대해 "상동성"을 나타내는 서열은 또 다른 핵산 분자에 대해 적어도 약 65%, 70%, 75%, 80%, 85% 또는 90% 서열 동일성을 나타내는 핵산 분자의 서열을 의미한다. 다른 실시양태에서, 핵산의 "상동 서열"은 참조 핵산에 대해 93%, 95% 또는 98% 서열 동일성을 나타낼 수 있다.

[0266] 서열 상동성 또는 동일성의 백분율이 명시되는 경우에, 2개의 핵산 서열 또는 2개의 폴리펩티드 서열의 맥락에서, 상동성 또는 동일성의 백분율은 일반적으로 최대 상응성을 위해 비교 및 정렬될 때 그의 길이의 한 부분에 걸친 2개 이상의 서열의 정렬을 지칭한다. 달리 언급되지 않는 한, 서열 상동성 또는 동일성은 핵산, 폴리펩티드 또는 그의 부분의 명시된 길이에 걸쳐 평가된다. 일부 실시양태에서, 상동성 또는 동일성은 길이의 기능적 부분 또는 명시된 부분에 걸쳐 평가된다.

[0267] 서열 상동성의 평가를 위한 서열의 정렬은 관련 기술분야에 공지된 알고리즘, 예컨대 문헌 [Altschul et al., J. Mol. Biol. 215:403-410, 1990]에 기재된 베이직 로컬 얼라인먼트 서치 툴 (BLAST) 알고리즘에 의해 수행될 수 있다. BLAST 분석을 수행하기 위한 공개적으로 이용가능한 인터넷 인터페이스는 국립 생물 정보 센터 (National Center for Biotechnology Information)를 통해 접근가능하다. 추가의 공지된 알고리즘은 문헌 [Smith & Waterman, "Comparison of Biosequences", Adv. Appl. Math. 2:482, 1981; Needleman & Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins" J. Mol. Biol. 48:443, 1970; Pearson & Lipman "Improved tools for biological sequence comparison", Proc. Natl. Acad. Sci. USA 85:2444, 1988]에 공개된 것; 또는 이들 또는 유사한 알고리즘의 자동화 구현된 것을 포함한다. 전반적 정렬 프로그램은 또한 대략 동일한 크기의 유사한 서열을 정렬하는 데 사용될 수 있다. 전반적 정렬 프로그램의 예는 NEEDLE (www.ebi.ac.uk/Tools/psa/emboss_needle/에서 이용가능함) (이는 EMBOSS

패키지의 일부임 (Rice P et al., Trends Genet., 2000; 16: 276-277)) 및 GGSEARCH 프로그램 fasta.bioch. Virginia. edu/fasta_www2/fasta_www.cgi?rm= compare&pgm=gnw) (이는 FASTA 패키지의 일부임 (Pearson W and Lipman D, 1988, Proc. Natl. Acad. Sci. USA, 85: 2444-2448))을 포함한다. 이들 프로그램 둘 다는 2개의 서열의 그의 전체 길이를 따른 최적 정렬 (갭 포함)을 발견하는 데 사용되는 니들만-분쉬(Needleman-Wunsch) 알고리즘에 기초한다. 서열 분석의 상세한 논의는 또한 문헌 [Unit 19.3 of Ausubel et al ("Current Protocols in Molecular Biology" John Wiley & Sons Inc, 1994-1998, Chapter 15, 1998)]에서 찾아볼 수 있다.

[0268] 스페이서 서열

[0269] 가이드 RNA 또는 PEgRNA와 관련하여 본원에 사용된 용어 "스페이서 서열"은 표적 DNA 서열 내의 프로토스페이서 서열과 동일한 서열을 공유하는 뉴클레오티드 서열을 함유하는 약 20개 뉴클레오티드의 가이드 RNA 또는 PEgRNA의 부분을 지칭한다. 스페이서 서열은 프로토스페이서 서열의 상보체에 어닐링하여 표적 부위에서 ssRNA/ssDNA 하이브리드 구조 및 내인성 DNA 가닥의 상응하는 R 루프 ssDNA 구조를 형성한다.

[0270] 대상체

[0271] 본원에 사용된 용어 "대상체"는 개별 유기체, 예를 들어 개별 포유동물을 지칭한다. 일부 실시양태에서, 대상체는 인간이다. 일부 실시양태에서, 대상체는 비-인간 포유동물이다. 일부 실시양태에서, 대상체는 비-인간 영장류이다. 일부 실시양태에서, 대상체는 설치류이다. 일부 실시양태에서, 대상체는 양, 염소, 소, 고양이 또는 개이다. 일부 실시양태에서, 대상체는 척추동물, 양서류, 파충류, 어류, 곤충, 파리 또는 선충류이다. 일부 실시양태에서, 대상체는 연구용 동물이다. 일부 실시양태에서, 대상체는 유전자 조작된, 예를 들어 유전자 조작된 비-인간 대상체이다. 대상체는 어느 하나의 성별일 수 있고 임의의 발달 단계일 수 있다.

[0272] 분할 인테인

[0273] 인테인은 가장 빈번하게는 인접 도메인으로서 발견되지만, 일부는 자연 분할 형태로 존재한다. 이러한 경우에, 2개의 단편은 별개의 폴리펩티드로서 발현되고, 이는 스플라이싱, 소위 단백질 트랜스-스플라이싱이 일어나기 전에 회합되어야 한다.

[0274] 예시적인 분할 인테인은 2개의 서브유닛, 즉 DnaE-N 및 DnaE-C를 포함하는 Ssp DnaE 인테인이다. 2개의 상이한 서브유닛은 별개의 유전자, 즉 각각 DnaE-N 및 DnaE-C 서브유닛을 코딩하는 dnaE-n 및 dnaE-c에 의해 코딩된다. DnaE는 시네코시티스 종 PCC6803 내의 자연 발생 분할 인테인이고, 각각 DnaE-N 또는 DnaE-C와의 융합체를 포함하는 2개의 별개의 단백질의 트랜스-스플라이싱을 지시할 수 있다.

[0275] 추가의 자연 발생 또는 조작된 분할-인테인 서열은 관련 기술분야에 공지되어 있거나 또는 본원에 기재된 전체-인테인 서열 또는 관련 기술분야에서 입수가능한 것으로부터 제조될 수 있다. 분할-인테인 서열의 예는 문헌 [Stevens et al., "A promiscuous split intein with expanded protein engineering applications," PNAS, 2017, Vol.114: 8538-8543; Iwai et al., "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme, FEBS Lett, 580: 1853-1858]에서 찾아볼 수 있으며, 이들 각각은 본원에 참조로 포함된다. 추가의 분할 인테인 서열은, 예를 들어 WO 2013/045632, WO 2014/055782, WO 2016/069774 및 EP2877490에서 찾아볼 수 있으며, 이들 각각의 내용은 본원에 참조로 포함된다.

[0276] 또한, 생체내 및 시험관내 트랜스로의 단백질 스플라이싱이 기재되어 있으며 (Shingledecker, et al., Gene 207:187 (1998), Southworth, et al., EMBO J. 17:918 (1998); Mills, et al., Proc. Natl. Acad. Sci. USA, 95:3543-3548 (1998); Lew, et al., J. Biol. Chem., 273:15887-15890 (1998); Wu, et al., Biochim. Biophys. Acta 35732:1 (1998b), Yamazaki, et al., J. Am. Chem. Soc. 120:5591 (1998), Evans, et al., J. Biol. Chem. 275:9091 (2000); Otomo, et al., Biochemistry 38:16040-16044 (1999); Otomo, et al., J. Biolmol. NMR 14:105-114 (1999); Scott, et al., Proc. Natl. Acad. Sci. USA 96:13638-13643 (1999)), 이는 예를 들어 2개의 개별적으로-발현된 절반으로부터 완전한 PE 융합 단백질의 형성에 관하여 도 66 및 67에 제시된 바와 같이, 기능적 생성물을 형성하기 위해 후속적으로 라이게이션을 거치는 2개의 불활성 단편에 대해 단백질로서의 발현 기회를 제공한다.

[0277] 표적 부위

[0278] 용어 "표적 부위"는 본원에 개시된 프라임 편집제 (PE)에 의해 편집되는 핵산 분자 내의 서열을 지칭한다. 표적 부위는 프라임 편집제 (PE) 및 gRNA의 복합체가 결합하는 핵산 분자 내의 서열을 추가로 지칭한다.

[0279] tPERT

[0280] "트랜스 프라임 편집제 RNA 주형 (tPERT)"에 대한 정의를 참조한다.

[0281] 시간적 제2 가닥 닉킹

[0282] 본원에 사용된 용어 "시간적 제2 가닥 닉킹"은 비편집된 가닥에의 제2 닉의 설치가 단지 편집된 가닥에 목적하는 편집이 설치된 후에만 발생하는 것인 제2 가닥 닉킹의 변형을 지칭한다. 이는 이중 가닥 DNA 파괴로 이어질 수 있는 두 가닥 상의 공동 닉을 피한다. 제2 가닥 닉킹 가이드 RNA는 제2 가닥 닉이 목적하는 편집이 설치될 때까지 도입되지 않도록 시간적 제어를 위해 설계된다. 이는 단지 편집된 가닥에만 매칭되고 원래 대립유전자에는 매칭되지 않는 스페이서 서열을 갖는 gRNA를 설계함으로써 달성된다. 이러한 전략을 사용할 때, 프로토스페이서와 비편집된 대립유전자 사이의 미스매치는 PAM 가닥 상의 편집 사건이 일어날 때까지 sgRNA에 의한 닉킹을 불리하게 할 것이다.

[0283] 트랜스 프라임 편집

[0284] 본원에 사용된 용어 "트랜스 프라임 편집" 또는 유사하게 "트랜스 이중 프라임 편집"은 분할 PEgRNA를 이용하는 프라임 편집 또는 이중 프라임 편집의 변형된 형태를 지칭하며, 즉 여기서 PEgRNA는 2개의 별개의 분자: sgRNA 및 트랜스 프라임 편집 RNA 주형 (tPERT)으로 분리된다. sgRNA는 목적하는 게놈 표적 부위에 프라임 편집제를 표적화하는 (또는 보다 일반적으로, 프라임 편집제의 napDNAbp 성분을 표적화하는) 역할을 하는 반면, tPERT는 일단 tPERT가 프라임 편집제 상에 및 tPERT 상에 위치한 결합 도메인의 상호작용에 의해 프라임 편집제에 트랜스로 동원되면 폴리머라제 (예를 들어, 리버스 트랜스크립타제)에 의해 새로운 DNA 서열을 표적 유전자좌에 기록하는 데 사용된다. 한 실시양태에서, 결합 도메인은 RNA-단백질 동원 모이어티, 예컨대 tPERT 상에 위치한 MS2 압타머 및 프라임 편집제에 융합된 MS2cp 단백질을 포함할 수 있다. 트랜스 프라임 편집의 이점은 가이드 RNA로부터 DNA 합성 주형을 분리함으로써, 잠재적으로 보다 긴 길이의 주형을 사용할 수 있다는 것이다.

[0285] 트랜스 프라임 편집의 실시양태는 도 3g 및 3h에 제시된다. 도 3g는 좌측에 트랜스 프라임 편집제 복합체의 조성 ("RP-PE:gRNA 복합체)을 보여주며, 이는 각각의 폴리머라제 (예를 들어, 리버스 트랜스크립타제) 및 rPERT 동원 단백질 (예를 들어, MS2sc)에 융합된 napDNAbp를 포함하고 가이드 RNA와 복합체화된다. 도 3g는 추가로 DNA 합성 주형 및 프라이머 결합 서열을 포함하는 PEgRNA의 연장 아암 특색을 포함하는 별개의 tPERT 분자를 보여준다. tPERT 분자는 또한 RNA-단백질 동원 도메인 (이 경우에, 스템 루프 구조이고, 예를 들어 MS2 압타머일 수 있음)을 포함한다. 도 3h에 기재된 과정에 도시된 바와 같이, RP-PE:gRNA 복합체는 표적 DNA 서열에 결합하여 그를 닉킹한다. 이어서, 동원 단백질 (RP)은 tPERT를 동원하여 DNA 표적 부위에 결합된 프라임 편집제 복합체에 공동-국재화시킴으로써, 프라이머 결합 부위가 닉킹된 가닥 상의 프라이머 서열에 결합하게 하고, 후속적으로 폴리머라제 (예를 들어, RT)가 tPERT의 5'까지 DNA 합성 주형에 대해 단일 가닥의 DNA를 합성하게 한다.

[0286] tPERT가 RNA-단백질 동원 도메인의 5' 단부 상에 PBS 및 DNA 합성 주형을 포함하는 것으로서 도 3g 및 도 3h에 제시되어 있지만, 다른 구성의 tPERT는 RNA-단백질 동원 도메인의 3' 단부 상에 PBS 및 DNA 합성 주형이 위치하도록 설계될 수 있다. 그러나, DNA의 단일 가닥의 합성이 tPERT의 5' 단부에서 자연적으로 종결될 것이고, 따라서 프라임 편집의 DNA 합성 단계 동안 주형으로서 RNA-단백질 동원 도메인의 임의의 부분을 사용할 위험이 없다는 점에서 5' 연장을 갖는 tPERT가 이점을 갖는다.

[0287] 트랜스 프라임 편집제 RNA 주형 (tPERT)

[0288] 본원에 사용된 "트랜스 프라임 편집제 RNA 주형 (tPERT)"은 PEgRNA를 2개의 별개의 분자: 가이드 RNA 및 tPERT 분자로 분리함으로써 작동하는 프라임 편집의 변형된 버전인 트랜스 프라임 편집 또는 이중 프라임 편집에 사용되는 성분을 지칭한다. tPERT 분자는 표적 DNA 부위에서 프라임 편집제 복합체와 공동-국재화되어, 프라이머 결합 부위 및 DNA 합성 주형을 프라임 편집제로 트랜스로 가져오도록 프로그램화된다. 예를 들어, (1) RP-PE:gRNA 복합체 및 (2) RNA-단백질 동원 도메인에 연결된 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 tPERT를 포함하는 2-성분 시스템을 보여주는 트랜스 프라임 편집제 (tPE)의 실시양태에 대해 도 3g를 참조하며, 여기서 RP-PE:gRNA 복합체의 RP (동원 단백질) 성분은 tPERT를 편집될 표적 부위로 동원함으로써 PBS 및 DNA 합성 주형을 프라임 편집제와 트랜스로 회합시킨다. 상기 또 다른 방식에서, tPERT는 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 PEgRNA의 연장 아암을 (모두 또는 일부) 함유하도록 조작된다.

[0289] 전이

[0290] 본원에 사용된 "전이"는 퓨린 핵염기의 상호교환 (A ↔ G) 또는 피리미딘 핵염기의 상호교환 (C ↔ T)을 지칭한

다. 이러한 부류의 상호교환은 유사한 형상의 핵염기를 수반한다. 본원에 개시된 조성물 및 방법은 표적 DNA 분자에서 1개 이상의 전이를 유도할 수 있다. 본원에 개시된 조성물 및 방법은 또한 동일한 표적 DNA 분자에서 전이 및 전환 둘 다를 유도할 수 있다. 이들 변화는 A ↔ G, G ↔ A, C ↔ T 또는 T ↔ C를 수반한다. 왓슨-크릭 쌍형성된 핵염기를 갖는 이중 가닥 DNA의 맥락에서, 전환은 하기 염기 쌍 교환을 지칭한다: A:T ↔ G:C, G:G ↔ A:T, C:G ↔ T:A 또는 T:A ↔ C:G. 본원에 개시된 조성물 및 방법은 표적 DNA 분자에서 1개 이상의 전이를 유도할 수 있다. 본원에 개시된 조성물 및 방법은 또한 동일한 표적 DNA 분자에서 전이 및 전환 둘 다, 뿐만 아니라 결실 및 삽입을 포함한 다른 뉴클레오티드 변화를 유도할 수 있다.

[0291] 전환

[0292] 본원에 사용된 "전환"은 퓨린 핵염기와 피리미딘 핵염기 또는 그 반대의 상호교환을 지칭하며, 따라서 유사하지 않은 형상을 갖는 핵염기의 상호교환을 수반한다. 이들 변화는 T ↔ A, T ↔ G, C ↔ G, C ↔ A, A ↔ T, A ↔ C, G ↔ C 및 G ↔ T를 수반한다. 왓슨-크릭 쌍형성된 핵염기를 갖는 이중 가닥 DNA의 맥락에서, 전환은 하기 염기 쌍 교환을 지칭한다: T:A ↔ A:T, T:A ↔ G:C, C:G ↔ G:C, C:G ↔ A:T, A:T ↔ T:A, A:T ↔ C:G, G:C ↔ C:G 및 G:C ↔ T:A. 본원에 개시된 조성물 및 방법은 표적 DNA 분자에서 1개 이상의 전환을 유도할 수 있다. 본원에 개시된 조성물 및 방법은 또한 동일한 표적 DNA 분자에서 전이 및 전환 둘 다, 뿐만 아니라 결실 및 삽입을 포함한 다른 뉴클레오티드 변화를 유도할 수 있다.

[0293] 치료

[0294] 용어 "치료", "치료하다" 및 "치료하는"은 본원에 기재된 바와 같은 질환 또는 장애 또는 그의 1종 이상의 증상을 역전시키거나, 완화시키거나, 그의 발병을 지연시키거나 또는 그의 진행을 억제하는 것을 목표로 하는 임상 개입을 지칭한다. 본원에 사용된 용어 "치료", "치료하다" 및 "치료하는"은 본원에 기재된 바와 같은 질환 또는 장애 또는 그의 1종 이상의 증상을 역전시키거나, 완화시키거나, 그의 발병을 지연시키거나 또는 그의 진행을 억제하는 것을 목표로 하는 임상 개입을 지칭한다. 일부 실시양태에서, 치료는 1종 이상의 증상이 발생한 후에 및/또는 질환이 진단된 후에 투여될 수 있다. 다른 실시양태에서, 치료는, 예를 들어 증상의 발병을 방지하거나 지연시키거나 또는 질환의 발병 또는 진행을 억제하기 위해 증상의 부재 하에 투여될 수 있다. 예를 들어, 치료는 증상의 발병 전에 (예를 들어, 증상의 병력에 비추어 및/또는 유전적 또는 다른 감수성 인자에 비추어) 감수성 개체에게 투여될 수 있다. 치료는 또한 증상이 해소된 후에, 예를 들어 그의 재발을 방지하거나 지연시키기 위해 계속될 수 있다.

[0295] 트리뉴클레오티드 반복 장애

[0296] 본원에 사용된 "트리뉴클레오티드 반복 장애" (또는 대안적으로, "확장 반복 장애" 또는 "반복부 확장 장애")는 특정 트리뉴클레오티드가 특정 유전자 또는 인트론에서 반복되는 일종의 돌연변이인 "트리뉴클레오티드 반복부 확장"에 의해 유발되는 유전 장애의 세트를 지칭한다. 트리뉴클레오티드 반복은 한 때 게놈에서 흔한 반복인 것으로 생각되었지만, 1990년대에 이들 장애를 명확히 하였다. DNA의 이들 겉보기 '양성' 스트레치는 때때로 확장되어 질환을 유발할 수 있다. 여러 규정된 특색이 트리뉴클레오티드 반복부 확장에 의해 유발된 장애 사이에 공유된다. 첫째로, 돌연변이체 반복은 체세포 및 배선 불안정성 둘 다를 나타내고, 보다 빈번하게는 후속 전달에서 축소되기보다는 확장된다. 둘째로, 보다 이른 연령에서의 발병 및 후속 세대에서의 표현형의 중증도 증가 (예상)는 일반적으로 보다 큰 반복 길이와 상관관계가 있다. 마지막으로, 질환 대립유전자의 모 기원이 종종 예상에 영향을 미칠 수 있으며, 부계 전달이 많은 이들 장애에 대한 보다 큰 확장 위험을 보유한다.

[0297] 트리플렛 확장은 DNA 복제 동안 슬리피지에 의해 유발되는 것으로 생각된다. 이들 영역에서의 DNA 서열의 반복적 성질로 인해, DNA 복제 동안 합성되는 모 가닥과 딸 가닥 사이의 상보적 염기 쌍형성을 유지하면서 '루프 아웃' 구조가 형성될 수 있다. 루프 아웃 구조가 딸 가닥 상의 서열로부터 형성되는 경우에, 이는 반복 수의 증가를 유발할 것이다. 그러나, 루프 아웃 구조가 모 가닥 상에 형성되는 경우에는, 반복 수의 감소가 발생한다. 이들 반복의 확장은 감소보다 더 흔한 것으로 보인다. 일반적으로, 확장이 클수록 질환을 유발하거나 질환의 중증도를 증가시킬 가능성이 크다. 이러한 특성이 트리뉴클레오티드 반복 장애에서 나타나는 예상의 특징을 가져온다. 예상은 발병 연령이 감소하는 경향 및 증상의 중증도가 이들 반복의 확장으로 인해 이환 가족의 후속 세대를 통해 증가하는 경향을 기재한다.

[0298] 뉴클레오티드 반복 장애는 트리플렛 반복이 비-코딩 영역에서 발생하는 것 (즉, 비-코딩 트리뉴클레오티드 반복 장애) 또는 코딩 영역에서 발생하는 것을 포함할 수 있다.

[0299] 본원에 기재된 프라임 편집제 (PE) 시스템은 특히 유약 X 증후군 (FRAXA), 유약 XE MR (FRAXE), 프리드라이히

운동실조 (FRDA), 근긴장성 이영양증 (DM), 척수소뇌성 운동실조 유형 8 (SCA8) 및 척수소뇌성 운동실조 유형 12 (SCA12)를 포함할 수 있는 뉴클레오티드 반복 장애를 치료하는 데 사용될 수 있다.

[0300] 상류

[0301] 본원에 사용된 용어 "상류" 및 "하류"는 5'에서 3' 방향으로 배향되는 핵산 분자 (단일 가닥이든 이중 가닥이든)에 위치한 적어도 2개의 요소의 선형 위치를 정의하는 상대성 용어이다. 특히, 제1 요소는 핵산 분자에서 제2 요소의 상류에 있으며, 여기서 제1 요소는 제2 요소에 대해 5'인 어딘가에 위치한다. 예를 들어, SNP가 닉 부위의 5' 측면 상에 있는 경우에 SNP는 Cas9-유도된 닉 부위의 상류에 있다. 반대로, 제1 요소는 핵산 분자에서 제2 요소의 하류에 있으며, 여기서 제1 요소는 제2 요소에 대해 3'인 어딘가에 위치한다. 예를 들어, SNP가 닉 부위의 3' 측면 상에 있는 경우에 SNP는 Cas9-유도된 닉 부위의 하류에 있다. 핵산 분자는 DNA (이중 또는 단일 가닥), RNA (이중 또는 단일 가닥), 또는 DNA와 RNA의 하이브리드일 수 있다. 분석은 단일 가닥 핵산 분자 및 이중 가닥 분자에 대해 동일하며, 이는 용어 상류 및 하류가 이중 가닥 분자의 어느 가닥이 고려되는지를 선택할 필요가 있다는 것을 제외하고는 핵산 분자의 단일 가닥만을 언급하기 때문이다. 종종, 적어도 2개의 요소의 위치적 상대성을 결정하는 데 사용될 수 있는 이중 가닥 DNA의 가닥은 "센스" 또는 "코딩" 가닥이다. 유전학에서, "센스" 가닥은 5'에서 3'로 진행하는 이중 가닥 DNA 내의 절편이며, 이는 3'에서 5'로 진행하는 DNA의 안티센스 가닥 또는 주형 가닥에 상보적이다. 따라서, 예로서, SNP 핵염기는 SNP 핵염기가 센스 또는 코딩 가닥 상의 프로모터의 3' 측면 상에 존재하는 경우에 게놈 DNA (이중 가닥임) 내의 프로모터 서열의 "하류"에 있다.

[0302] 변이체

[0303] 본원에 사용된 용어 "변이체"는 자연에서 발생하는 것으로부터 벗어난 패턴을 갖는 품질을 나타내는 것을 의미하는 것으로 이해되어야 하며, 예를 들어 변이체 Cas9는 야생형 Cas9 아미노산 서열과 비교하여 아미노산 잔기에서 1개 이상의 변화를 포함하는 Cas9이다. 용어 "변이체"는 참조 서열과 적어도 75% 또는 적어도 80% 또는 적어도 85% 또는 적어도 90% 또는 적어도 95% 또는 적어도 99% 퍼센트 동일성을 갖고 참조 서열과 동일하거나 실질적으로 동일한 기능적 활성 또는 활성들을 갖는 상동 단백질을 포괄한다. 상기 용어는 또한 참조 서열의 돌연변이체, 말단절단물 또는 도메인을 포괄하며, 이는 참조 서열과 동일하거나 실질적으로 동일한 기능적 활성 또는 활성들을 나타낸다.

[0304] 벡터

[0305] 본원에 사용된 용어 "벡터"는 관심 유전자를 코딩하도록 변형될 수 있고, 숙주 세포 내로 진입하여, 숙주 세포 내에서 돌연변이 및 복제된 다음, 복제된 형태의 벡터를 또 다른 숙주 세포 내로 전달할 수 있는 핵산을 지칭한다. 예시적인 적합한 벡터는 바이러스 벡터, 예컨대 레트로바이러스 벡터 또는 박테리오파지 및 사상 파지 및 접합 플라스미드를 포함한다. 추가의 적합한 벡터는 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다.

[0306] 야생형

[0307] 본원에 사용된 용어 "야생형"은 통상의 기술자에 의해 이해되는 관련 기술분야의 용어이고, 돌연변이체 또는 변이체 형태와 구별되는, 자연에서 발생하는 유기체, 균주, 유전자 또는 특징의 전형적 형태를 의미한다.

[0308] 5' 내인성 DNA 플랩

[0309] 본원에 사용된 용어 "5' 내인성 DNA 플랩"은 표적 DNA 내의 PE-유도된 닉 부위의 바로 하류에 위치한 DNA의 가닥을 지칭한다. PE에 의한 표적 DNA 가닥의 닉킹은 닉 부위의 상류 측 상에 3' 히드록실 기 및 닉 부위의 하류 측 상에 5' 히드록실 기를 노출시킨다. 3' 히드록실 기로 끝나는 내인성 가닥은 프라임 편집제의 DNA 폴리머라제를 프라이밍하는 데 사용된다 (예를 들어, 여기서 DNA 폴리머라제는 리버스 트랜스크립타제임). 닉 부위의 하류 측 상에 노출된 5' 히드록실 기로 시작하는 내인성 가닥은 "5' 내인성 DNA 플랩"으로 지칭되고, 궁극적으로 제거되며, PEgRNA의 연장부에 의해 코딩되는 새로 합성된 대체 가닥 (즉, "3' 대체 DNA 플랩")에 의해 대체된다.

[0310] 5' 내인성 DNA 플랩 제거

[0311] 본원에 사용된 용어 "5' 내인성 DNA 플랩 제거" 또는 "5' 플랩 제거"는 RT-합성된 단일 가닥 DNA 플랩이 내인성 DNA에 경쟁적으로 침입하여 혼성화할 때 형성되는 5' 내인성 DNA 플랩의 제거를 지칭하며, 이 과정에서 내인성 가닥을 대체한다. 이러한 내인성 대체된 가닥을 제거하는 것은 목적하는 뉴클레오티드 변화를 포함하는 목적하

는 생성물의 형성을 향해 반응을 구동할 수 있다. 세포 자신의 DNA 복구 효소 (예를 들어, 플랩 엔도뉴클레아제, 예컨대 EXO1 또는 FEN1)는 5' 내인성 플랩의 제거 또는 절제를 촉매할 수 있다. 또한, 숙주 세포는 상기 5' 내인성 플랩의 제거를 촉매하는 1종 이상의 효소 (예를 들어, 플랩 엔도뉴클레아제)를 발현하도록 형질전환되어, 상기 과정을 생성물 형성을 향해 구동할 수 있다. 플랩 엔도뉴클레아제는 관련 기술분야에 공지되어 있고, 문헌 [Patel et al., "Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends," Nucleic Acids Research, 2012, 40(10): 4507-4519 및 Tsutakawa et al., "Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily," Cell, 2011, 145(2): 198-211] (이들 각각은 본원에 참조로 포함됨)에 기재된 바와 같이 찾아볼 수 있다.

[0312] 3' 대체 DNA 플랩

[0313] 본원에 사용된 용어 "3' 대체 DNA 플랩" 또는 간단히 "대체 DNA 플랩"은 프라임 편집제에 의해 합성되고 프라임 편집제 PEgRNA의 연장 아암에 의해 코딩되는 DNA의 가닥을 지칭한다. 보다 특히, 3' 대체 DNA 플랩은 PEgRNA의 폴리머라제 주형에 의해 코딩된다. 3' 대체 DNA 플랩은 편집된 서열 (예를 들어, 단일 뉴클레오티드 변화)도 함유한다는 것을 제외하고는 5' 내인성 DNA 플랩과 동일한 서열을 포함한다. 3' 대체 DNA 플랩은 표적 DNA에 어닐링되어, 5' 내인성 DNA 플랩 (이는 예를 들어 5' 플랩 엔도뉴클레아제, 예컨대 FEN1 또는 EXO1에 의해 절제될 수 있음)을 대체 또는 교체시킨 다음, 라이게이션되어 3' 대체 DNA 플랩의 3' 단부를 내인성 DNA의 노출된 5' 히드록실 단부 (5' 내인성 DNA 플랩의 절제 후에 노출됨)에 연결함으로써, 포스포포디에스테르 결합을 재형성하고 3' 대체 DNA 플랩을 설치하여 1개의 편집된 가닥 및 1개의 비편집된 가닥을 함유하는 헤테로듀플렉스 DNA를 형성한다. DNA 복구 과정은 헤테로듀플렉스를 분해하여 편집된 가닥 내의 정보를 상보적 가닥에 카피함으로써 편집을 DNA 내에 영구적으로 설치한다. 이러한 분해 과정은 비편집된 가닥을 닉킹함으로써, 즉 본원에 기재된 바와 같은 "제2 가닥 닉킹"에 의해 추가로 구동되어 완료될 수 있다.

[0314] 특정 실시양태의 상세한 설명

[0315] 게놈 편집을 위한 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) 시스템의 채택은 생명 과학을 혁신시켰다[1-3]. CRISPR을 사용한 유전자 파괴가 현재 상용화되어 있지만, 단일 뉴클레오티드 편집의 정확한 설치는 다수의 질환-원인 돌연변이를 연구하거나 교정하는데 필요함에도 불구하고 주요 도전과제로 남아있다. 상동성 지정 복구 (HDR)는 이러한 편집을 달성할 수 있지만, 낮은 효율 (종종 <5%), 공여자 DNA 복구 주형에 대한 요건, 및 이중 가닥 DNA 파괴 (DSB) 형성이라는 유해 효과를 겪는다. 최근에, 데이비드 리우 교수 등의 실험실이 DSB 없이 효율적인 단일 뉴클레오티드 편집을 달성하는 염기 편집을 개발하였다. 염기 편집제 (BE)는 CRISPR 시스템을 염기-변형 데아미나제 효소와 조합하여 표적 C·G 또는 A·T 염기 쌍을 각각 A·T 또는 G·C로 전환시킨다[4-6]. 전세계적으로 연구자들에 의해 이미 널리 사용되고 있기는 하지만, 현행 BE는 12개의 가능한 염기 쌍 전환 중 단지 4개만을 가능하게 하고, 작은 삽입 또는 결실은 교정할 수 없다. 더욱이, 염기 편집의 표적화 범주는 표적 염기에 인접한 비-표적 C 또는 A 염기의 편집 ("방관자 편집")에 의해 및 PAM 서열이 표적 염기로부터 15±2 bp에 존재해야 한다는 요건에 의해 제한된다. 따라서, 이들 제한을 극복하는 것은 게놈 편집의 기초 연구 및 치료 용도를 크게 확장시킬 것이다.

[0316] 염기 편집의 많은 이익-즉, 이중 가닥 파괴 및 공여자 DNA 복구 주형의 회피-을 제공하면서 주요 제한을 극복하는 새로운 정밀 편집 접근법 (예를 들어, 전형적 프라임 편집)이 개발되었다. 본원에 기재된 제안된 접근법은 표적-프라이밍된 역전사 (TPRT)를 사용하여 표적 게놈 부위에의 편집된 DNA 가닥의 직접 설치를 달성한다. 본원에 논의된 설계에서, CRISPR 가이드 RNA (gRNA)는 목적하는 뉴클레오티드 변화를 포함하는 단일 가닥 DNA를 코딩하는 리버스 트랜스크립타제 (RT) 주형 서열을 보유하도록 조작될 것이다. CRISPR 뉴클레아제 (Cas9)-닉킹된 표적 부위 DNA는 변형된 gRNA 상의 주형 서열의 역전사를 위한 프라이머로서의 역할을 하여, 임의의 목적하는 뉴클레오티드 편집의 직접 혼입을 가능하게 할 것이다.

[0317] 표적-프라이밍된 역전사 (TPRT)의 메카니즘은 (예를 들어, 도 1a-1f의 다양한 실시양태에 도시된 바와 같이) 높은 효율 및 유전적 유연성으로 정밀 CRISPR/Cas-기반 게놈 편집을 수행하기 위해 활용되거나 적합화될 수 있다. Cas 단백질-리버스 트랜스크립타제 융합체를 사용하여 변형된 가이드 RNA ("연장된 가이드 RNA")에 의해 특이적 DNA 서열을 표적화하고, 표적 부위에 단일 가닥 닉을 생성하고, 닉킹된 DNA를 연장된 가이드 RNA 내로 통합된 조작된 리버스 트랜스크립타제 주형의 역전사를 위한 프라이머로서 사용하는 것을 본원에서 제안하였다. 새로 합성된 가닥은 목적하는 뉴클레오티드 변화 (예를 들어, 단일 뉴클레오티드 변화, 결실 또는 삽입 또는 그의 조합)의 포함을 제외하고는 게놈 표적 서열에 상동일 것이다. DNA의 새로 합성된 가닥은 단일 가닥 DNA 플랩으로

서 지칭될 수 있으며, 이는 상보적 상동 내인성 DNA 가닥과의 혼성화에 대해 경쟁함으로써 상응하는 내인성 가닥을 대체할 것이다. 이러한 혼성화된 중간체의 분해는 생성된 내인성 DNA의 대체 플랩의 제거 (예를 들어, 5' 단부 DNA 플랩 엔도뉴클레아제인 FEN1을 사용함), 합성된 단일 가닥 DNA 플랩의 표적 DNA에 대한 라이게이션, 및 세포 DNA 복구 및/또는 복제 과정의 결과로서 목적하는 뉴클레오티드 변화의 동화를 포함할 수 있다. 주형화된 DNA 합성은 단일 뉴클레오티드 정밀도를 제공하기 때문에, 이러한 접근법의 범주는 매우 넓고, 기초 과학 및 요법에서 무수한 적용에 사용될 수 있을 것으로 예측된다.

[0318] 본 발명은 "다중-플랩 프라임 편집" (예를 들어, "이중-플랩 프라임 편집" 및 "사중-플랩 프라임 편집" 포함)으로 불리는 게놈 편집을 위한 새로운 플랫폼을 기재하고, 문헌 [Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149-157 (2019)] (이는 본원에 참조로 포함됨)에서 본 발명자들에 의해 기재된 바와 같은 "프라임 편집" 또는 "전형적 프라임 편집"의 혁신적인 진보를 나타낸다. 전형적 프라임 편집은 다양한 실시양태에서 동일한 가닥 상의 표적 핵산 내로 통합되는 단일 3' 플랩을 닉 부위에서 중합시키는 반면, 본원에 기재된 다중-플랩 프라임 편집 시스템은 다양한 실시양태에서 상이한 가닥들 상에 3' 플랩의 쌍 또는 다중 쌍을 생성하는 별개의 구축물, 시스템 및 방법론을 수반하며, 상기 쌍이 목적하는 편집을 포함하는 듀플렉스를 형성하고, 예를 들어 게놈 내 특이적 유전자좌 또는 편집 부위에서 표적 핵산 분자 내로 혼입된다. 다양한 측면에서, 3' 플랩의 쌍 또는 다중 쌍은 본원에 기재된 프라임 편집제에 의해 생성되었을 때 서로 어닐링하는 역 상보적 서열을 포함하기 때문에 듀플렉스를 형성한다. 듀플렉스는 인접한 닉 부위 사이에 위치한 내인성 듀플렉스 서열을 자연적으로 대체하는 세포-구동 메카니즘에 의해 표적 부위 내로 혼입된다. 특정 실시양태에서, 새로운 듀플렉스 서열은 1개 이상의 위치 (예를 들어, 인접한 게놈 유전자좌 또는 2개의 상이한 염색체 위치)에 도입될 수 있고, 1개 이상의 관심 서열, 예를 들어 단백질-코딩 서열, 펩티드-코딩 서열 또는 RNA-코딩 서열을 포함할 수 있다. 한 실시양태에서, 다중-플랩 프라임 편집 시스템에 의해 설치된 새로운 듀플렉스 서열은 레콤비나제 부위, 예를 들어 Bxb1 레콤비나제 attB (38 bp) 및/또는 attP (50 bp) 부위, 또는 Hin 레콤비나제, Gin 레콤비나제, Tn3 레콤비나제, β-6 레콤비나제, CinH 레콤비나제, ParA 레콤비나제, γδ 레콤비나제, ΦC31 레콤비나제, TP901 레콤비나제, TG1 레콤비나제, ϕBT1 레콤비나제, R4 레콤비나제, ϕRV1 레콤비나제, ϕFC1 레콤비나제, MR11 레콤비나제, A118 레콤비나제, U153 레콤비나제 및 gp29 레콤비나제, Cre 레콤비나제, FLP 레콤비나제, R 레콤비나제, 람다 레콤비나제, HK101 레콤비나제, HK022 레콤비나제 및 pSAM2 레콤비나제에 의해 인식되는 레콤비나제 부위를 포함할 수 있다.

[0319] 다양한 측면에서, 본 명세서는 두 DNA 가닥을 동시에 편집함으로써 플랩 평형 및 비-편집된 상보적 게놈 DNA 가닥 내로의 편집의 후속 혼입과 연관된 도전과제를 다루는 다중-플랩 프라임 편집 시스템 (예를 들어, 이중 프라임 편집 시스템 및 사중 프라임 편집 시스템 포함)을 기재한다. 예를 들어 이중-플랩 프라임 편집 시스템에서, 2개의 PEgRNA는 게놈 부위의 대향 가닥을 표적화하고 편집된 DNA 서열을 함유하는 2개의 상보적 3' 플랩의 합성을 지시하는 데 사용된다 (도 90). 전형적 프라임 편집과는 달리, 상보적인 편집된 가닥이 대신 혼성화에 이용가능하기 때문에, 편집된 DNA 가닥의 쌍 (3' 플랩)이 내인성 게놈 DNA에서 5' 플랩과 직접 경쟁하는 것이 요구되지 않는다. 듀플렉스의 두 가닥이 편집된 DNA로서 합성되기 때문에, 이중-플랩 프라임 편집 시스템은 전형적 프라임 편집에 의해 요구되는 비-편집된 상보적 DNA 가닥의 대체에 대한 필요를 제거한다. 대신, 세포 DNA 복구 기구는 단지 쌍형성된 5' 플랩 (원래 게놈 DNA)을 절제하고, 쌍형성된 3' 플랩 (편집된 DNA)을 유전자좌 내로 라이게이션시킬 필요가 있다. 따라서, 또한 새로 합성된 DNA 가닥에 게놈 DNA에 상동인 서열을 포함할 필요가 없으므로, 새로운 가닥의 선택적 혼성화를 허용하고 최소 게놈 상동성을 함유하는 편집을 용이하게 한다. DNA의 두 가닥을 절단하는 프라임 편집제의 뉴클레아제-활성 버전은 또한 원래 DNA 서열의 제거를 가속화하는 데 사용될 수 있다. 4개의 PEgRNA를 사용하는 사중-플랩 프라임 편집 시스템은 유사한 이점을 제공한다.

[0320] [1] napDNAbp

[0321] 본원에 기재된 다중-플랩 프라임 편집제는 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)을 포함할 수 있다.

[0322] 한 측면에서, napDNAbp는 적어도 1개의 가이드 핵산 (예를 들어, 가이드 RNA 또는 PEgRNA)과 회합되거나 복합체화될 수 있으며, 이는 napDNAbp를 가이드 핵산 또는 그의 부분 (예를 들어, DNA 표적의 프로토스페이서에 어닐링되는 가이드 RNA의 스페이서)에 상보적인 DNA 가닥 (즉, 표적 가닥)을 포함하는 DNA 서열에 국재화시킨다. 다시 말해서, 가이드 핵산은 napDNAbp (예를 들어, Cas9 또는 등가물)를 DNA 내의 프로토스페이서의 상보적 서열에 국재화시키고 결합하도록 "프로그램화한다".

[0323] 임의의 적합한 napDNAbp가 본원에 기재된 다중-플랩 프라임 편집제에 사용될 수 있다. 다양한 실시양태에서, napDNAbp는 임의의 유형 II, 유형 V, 또는 유형 VI CRISPR-Cas 효소를 포함한 임의의 부류 2 CRISPR-Cas 시스

템일 수 있다. 게놈 편집을 위한 도구로서 CRISPR-Cas의 급속한 개발을 고려할 때, CRISPR-Cas 효소, 예컨대 Cas9 및 Cas9 오르토로그를 기재하고/거나 확인하는 데 사용되는 명명법에서 지속적인 개발이 있어왔다. 본 출원은 기존 및/또는 신규일 수 있는 명명법으로 CRISPR-Cas 효소를 언급한다. 통상의 기술자는 기존 (즉, "레거시")이든 또는 새로운 명명법이든, 사용되는 명명법에 기초하여 본 출원에서 언급되는 특정 CRISPR-Cas 효소를 확인할 수 있을 것이다. CRISPR-Cas 명명법은 문헌 [Makarova et al., "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?," The CRISPR Journal, Vol. 1. No. 5, 2018]에 광범위하게 논의되어 있으며, 그의 전체 내용은 본원에 참조로 포함된다. 본 출원에서 임의의 주어진 경우에 사용된 특정한 CRISPR-Cas 명명법은 어떠한 방식으로든 제한하는 것이 아니며, 통상의 기술자는 어느 CRISPR-Cas 효소가 언급되는지 확인할 수 있을 것이다.

[0324] 예를 들어, 하기 유형 II, 유형 V, 및 유형 VI 부류 2 CRISPR-Cas 효소는 하기 관련 기술분야-인식의 기존 (즉, 레거시) 및 새로운 명칭을 갖는다. 각각의 이들 효소 및/또는 그의 변이체는 본원에 기재된 다중-플랩 프라임 편집제와 함께 사용될 수 있다:

레거시 명명법	현행 명명법*
유형 II CRISPR-Cas 효소	
Cas9	동일
유형 V CRISPR-Cas 효소	
Cpf1	Cas12a
CasX	Cas12e
C2c1	Cas12b1
Cas12b2	동일
C2c3	Cas12c
CasY	Cas12d
C2c4	동일
C2c8	동일
C2c5	동일
C2c10	동일
C2c9	동일
유형 VI CRISPR-Cas 효소	
C2c2	Cas13a
Cas13d	동일
C2c7	Cas13c
C2c6	Cas13b

[0325]

[0326] * 문헌 [Makarova et al., The CRISPR Journal, Vol. 1, No. 5, 2018] 참조

[0327] 이론에 얽매이지는 않지만, 본원에 고려된 특정 napDNAbp의 작용 메카니즘은 R-루프를 형성하여 napDNAbp가 이중 가닥 DNA 표적의 풀림을 유도하고, 그에 의해 napDNAbp에 의해 결합된 영역에서 가닥을 분리하는 단계를 포함한다. 이어서, 가이드 RNA 스페이서는 "표적 가닥"에 프로토스페이서 서열에서 혼성화한다. 이는 표적 가닥에 상보적인 "비-표적 가닥"을 대체하며, 이는 R-루프의 단일 가닥 영역을 형성한다. 일부 실시양태에서, napDNAbp는 1개 이상의 뉴클레아제 활성을 포함하며, 이는 이어서 DNA를 절단하여 다양한 유형의 병변을 남긴다. 예를 들어, napDNAbp는 제1 위치에서 비-표적 가닥을 절단하고/거나 제2 위치에서 표적 가닥을 절단하는 뉴클레아제 활성을 포함할 수 있다. 뉴클레아제 활성에 따라, 표적 DNA는 절단되어 "이중 가닥 파괴"를 형성할 수 있고, 이에 의해 두 가닥이 절단된다. 다른 실시양태에서, 표적 DNA는 단일 부위에서만 절단될 수 있고, 즉, DNA는 한 가닥 상에서 "닉킹"된다. 상이한 뉴클레아제 활성을 갖는 예시적인 napDNAbp는 "Cas9 닉카제" ("nCas9") 및 뉴클레아제 활성을 갖지 않는 탈활성화된 Cas9 ("기능상실 Cas9" 또는 "dCas9")를 포함한다.

[0328] 본원에 개시된 다중-플랩 프라임 편집제와 관련하여 사용될 수 있는 다양한 napDNAbp에 대한 하기 설명은 어떠한 방식으로도 제한하는 것으로 의도되지 않는다. 다중-플랩 프라임 편집제는 정규 SpCas9, 또는 임의의 오르토로그 Cas9 단백질, 또는 임의의 변이체 Cas9 단백질-공지되어 있거나 또는 방향적 진화상 과정 또는 달리 돌연변이유발 과정을 통해 제조 또는 진화될 수 있는 Cas9의 임의의 자연 발생 변이체, 돌연변이체, 또는 달리 조작된 버전 포함-을 포함할 수 있다. 다양한 실시양태에서, Cas9 또는 Cas9 변이체는 닉카제 활성을 갖고, 즉 표적 DNA 서열의 한 가닥만을 절단한다. 다른 실시양태에서, Cas9 또는 Cas9 변이체는 불활성 뉴클레아제를 가지며, 즉 "기능상실" Cas9 단백질이다. 사용될 수 있는 다른 변이체 Cas9 단백질은 정규 SpCas9보다 더 작은 분자량을 갖거나 (예를 들어, 보다 용이한 전달을 위함), 또는 변형되거나 재배열된 1차 아미노산 구조를 갖는

(예를 들어, 원형 순열체 포맷) 것이다.

[0329] 본원에 기재된 다중-플랩 프라임 편집제는 또한 수렴 진화의 결과인 Cas12a (Cpf1) 및 Cas12b1 단백질을 포함한 Cas9 등가물을 포함할 수 있다. 본원에 사용된 napDNAbp (예를 들어, SpCas9, Cas9 변이체, 또는 Cas9 등가물)는 또한 그의 PAM 특이성을 변경/증진시키는 다양한 변형을 함유할 수 있다. 마지막으로, 본 출원은 참조 Cas9 서열, 예컨대 참조 SpCas9 정규 서열 또는 참조 Cas9 등가물 (예를 들어, Cas12a (Cpf1))과 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.9% 서열 동일성을 갖는 임의의 Cas9, Cas9 변이체, 또는 Cas9 등가물을 고려한다.

[0330] napDNAbp는 CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부)-연관 뉴클레아제일 수 있다. 상기 약술된 바와 같이, CRISPR은 이동성 유전 요소 (바이러스, 전위가능 요소 및 접합 플라스미드)에 대한 보호를 제공하는 적응 면역계이다. CRISPR 클러스터는 스페이서, 선행 이동성 요소에 상보적인 서열, 및 표적 침입 핵산을 함유한다. CRISPR 클러스터는 CRISPR RNA (crRNA)로 전사 및 프로세싱된다. 유형 II CRISPR 시스템에서, 프리-crRNA의 교정 프로세싱은 트랜스-코딩된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질을 필요로 한다. tracrRNA는 프리-crRNA의 리보뉴클레아제 3-보조 프로세싱을 위한 가이드로서의 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 원형 dsDNA 표적을 핵산내부분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥이 먼저 핵산내부분해적으로 절단된 다음, 핵산외부분해적으로 3'-5' 트리밍된다. 사실상, DNA-결합 및 절단은 전형적으로 단백질 및 둘 다의 RNA를 필요로 한다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 간단히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 측면을 단일 RNA 종 내로 혼입하도록 조작될 수 있다. 예를 들어, 문헌 [Jinek M. et al., Science 337:816-821 (2012)]을 참조하며, 그의 전체 내용은 본원에 참조로 포함된다.

[0331] 일부 실시양태에서, napDNAbp는 표적 서열의 위치에서, 예컨대 표적 서열 내에서 및/또는 표적 서열의 상보체 내에서 하나 또는 두 가닥의 절단을 지시한다. 일부 실시양태에서, napDNAbp는 표적 서열의 제1 또는 마지막 뉴클레오티드로부터 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500개 또는 그 초과의 염기 쌍 내에서 하나 또는 두 가닥의 절단을 지시한다. 일부 실시양태에서, 벡터는 돌연변이된 napDNAbp가 표적 서열을 함유하는 표적 폴리뉴클레오티드의 하나 또는 두 가닥을 절단하는 능력이 결여되도록 상응하는 야생형 효소에 대해 돌연변이된 napDNAbp를 코딩한다. 예를 들어, 에스. 피오게네스(*S. pyogenes*)로부터의 Cas9의 RuvC I 촉매 도메인에서 아스파르테이트에서 알라닌으로의 치환 (D10A)은 Cas9를 두 가닥을 절단하는 뉴클레아제로부터 닉카제 (단일 가닥을 절단함)로 전환시킨다. Cas9를 닉카제로 만드는 돌연변이의 다른 예는, 비제한적으로, 정규 SpCas9 서열, 또는 다른 Cas9 변이체 또는 Cas9 등가물에서의 등가의 아미노산 위치와 관련하여 H840A, N854A, 및 N863A를 포함한다.

[0332] 본원에 사용된 용어 "Cas 단백질"은 자연으로부터 수득된 전장 Cas 단백질, 자연 발생 Cas 단백질과 상이한 서열을 갖는 재조합 Cas 단백질, 또는 그럼에도 불구하고 개시된 방법에 필요한 모든 또는 유의한 양의 필수적인 기본 기능, 즉 (i) 표적 DNA에 대한 Cas 단백질의 핵산 프로그램가능한 결합의 보유, 및 (ii) 한 가닥 상의 표적 DNA 서열을 닉킹하는 능력을 보유하는 Cas 단백질의 임의의 단편을 지칭한다. 본원에서 고려되는 Cas 단백질은 CRISPR Cas9 단백질, 뿐만 아니라 Cas9 등가물, 변이체 (예를 들어, Cas9 닉카제 (nCas9) 또는 뉴클레아제 불활성 Cas9 (dCas9)) 상동체, 오르토로그, 또는 파라로그 (자연 발생이든 또는 비-자연 발생 (예를 들어, 조작 또는 재조합)이든)를 포괄하고, Cas12a (Cpf1), Cas12e (CasX), Cas12b1 (C2c1), Cas12b2, Cas12c (C2c3), C2c4, C2c8, C2c5, C2c10, C2c9 Cas13a (C2c2), Cas13d, Cas13c (C2c7), Cas13b (C2c6), 및 Cas13b를 포함한 임의의 부류 2 CRISPR 시스템 (예를 들어, 유형 II, V, VI)으로부터의 Cas9 등가물을 포함할 수 있다. 추가의 Cas-등가물은 문헌 [Makarova et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector," Science 2016; 353(6299) 및 Makarova et al., "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?," The CRISPR Journal, Vol. 1. No. 5, 2018]에 기재되어 있으며, 그의 내용은 본원에 참조로 포함된다.

[0333] 용어 "Cas9" 또는 "Cas9 뉴클레아제" 또는 "Cas9 모이어티" 또는 "Cas9 도메인"은 임의의 유기체로부터의 임의의 자연 발생 Cas9, 그의 임의의 자연 발생 Cas9 등가물 또는 기능적 단편, 임의의 유기체로부터의 임의의 Cas9 상동체, 오르토로그, 또는 파라로그, 및 자연 발생 또는 조작된 Cas9의 임의의 돌연변이체 또는 변이체를 포괄한다. 용어 Cas9는 특히 제한적인 것으로 의도되지 않으며, "Cas9 또는 등가물"로 지칭될 수 있다. 예시적인 Cas9 단백질은 본원에 추가로 기재되고/거나 관련 기술분야에 기재되어 있고, 본원에 참조로 포함된다. 본 개

시내용은 본원에 기재된 다중-플랩 프라임 편집제에 사용되는 특정한 Cas9와 관련하여 제한되지 않는다.

[0334] 본원에 언급된 바와 같이, Cas9 뉴클레아제 서열 및 구조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 ["Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); 및 "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)] 참조, 이들 각각의 전체 내용은 본원에 참조로 포함됨).

[0335] Cas9 및 Cas9 등가물의 예는 하기와 같이 제공되지만; 이들 구체적 예는 제한적인 것으로 의도되지 않는다. 본 개시내용의 다중-플랩 프라임 편집제는 임의의 적합한 Cas9 또는 Cas9 등가물을 포함한, 임의의 적합한 napDNAbp를 사용할 수 있다.

[0336] A. 야생형 정규 SpCas9

[0337] 한 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 구축물은 에스. 피오게네스로부터의 "정규 SpCas9" 뉴클레아제를 포함할 수 있으며, 이는 게놈 조작을 위한 도구로서 널리 사용되어 왔고, 부류 2 CRISPR-Cas 시스템의 효소의 유형 II 하위군으로서 카테고리화된다. 이 Cas9 단백질은 2개의 별개의 뉴클레아제 도메인을 함유하는 대형의, 다중-도메인 단백질이다. 점 돌연변이를 Cas9 내로 도입하여 하나 또는 둘 다의 뉴클레아제 활성을 없앨 수 있으며, 이는 각각, sgRNA-프로그램화된 방식으로 DNA에 결합하는 그의 능력을 여전히 보유하는 닉카제 Cas9 (nCas9) 또는 기능상실 Cas9 (dCas9)를 생성한다. 원칙적으로, 또 다른 단백질 또는 도메인에 융합될 경우, Cas9 또는 그의 변이체 (예를 들어, nCas9)는 단순히 적절한 sgRNA와의 공동-발현에 의해 그러한 단백질을 사실상 모든 DNA 서열에 표적화할 수 있다. 본원에 사용된 정규 SpCas9 단백질은 하기 아미노산 서열을 갖는 스트렙토코쿠스 피오게네스(*Streptococcus pyogenes*)로부터의 야생형 단백질을 지칭한다:

설명	서열	SEQ ID NO:
SpCas9 *스트렙토코쿠스 피오게네스* M1 스위스프롯 수탁 번호 Q99ZW2 야생형	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 18
SpCas9 스위스프롯 수탁 번호 Q99ZW2 *스트렙토코쿠스 피오게네스*의 역전사	ATGGATAAAAAATATAGCATTGGCCTGGATATTGGC ACCAACAGCGTGGGCTGGGCGGTGATTACCGATGAA TATAAAGTGCCGAGCAAAAAATTTAAAGTGCTGGGC AACACCGATCGCCATAGCATTAAAAAAAACCTGATT GGCGCGCTGCTGTTTGATAGCGGCGAAACCGCGGAA GCGACCCGCCTGAAACGCACCGCGCGCCGCCGCTAT ACCCGCCGCAAAAACCGCATTTGCTATCTGCAGGAA ATTTTTAGCAACGAAATGGCGAAAGTGGATGATAGC TTTTTTCATCGCCTGGAAGAAAGCTTTCTGGTGGAAG	SEQ ID NO: 19

[0338]

	AAGATAAAAAACATGAACGCCATCCGATTTTTGGCA ACATTGTGGATGAAGTGGCGTATCATGAAAAATATC CGACCATTTATCATCTGCGCAAAAAACTGGTGGATA GCACCGATAAAGCGGATCTGCGCCTGATTTATCTGG CGCTGGCGCATATGATTAAATTTCGCGGCCATTTTCT GATTGAAGGCGATCTGAACCCGGATAACAGCGATGT GGATAAACTGTTTATTCAGCTGGTGCAGACCTATAA CCAGCTGTTTGAAGAAAACCCGATTAACGCGAGCGG CGTGGATGCGAAAGCGATTCTGAGCGCGCGCCTGAG CAAAAGCCGCCGCCTGGAAAACCTGATTGCGCAGCT GCCGGGCGAAAAAAAAAAACGGCCTGTTTGGCAACCT GATTGCGCTGAGCCTGGGCCTGACCCCGAACTTTAA AAGCAACTTTGATCTGGCGGAAGATGCGAAACTGCA GCTGAGCAAAGATACCTATGATGATGATCTGGATAA CCTGCTGGCGCAGATTGGCGATCAGTATGCGGATCT GTTTCTGGCGGCGAAAAACCTGAGCGATGCGATTCT GCTGAGCGATATTCTGCGCGTGAACACCGAAATTAC CAAAGCGCCGCTGAGCGCGAGCATGATTAAACGCTA TGATGAACATCATCAGGATCTGACCCTGCTGAAAGC GCTGGTGCGCCAGCAGCTGCCGGAAAAATATAAAG AAATTTTTTTTGATCAGAGCAAAAACGGCTATGCGG GCTATATTGATGGCGGCGCGAGCCAGGAAGAATTTT ATAAATTTATTAAACCGATTCTGGAAAAAATGGATG GCACCGAAGAACTGCTGGTGAAACTGAACCGCGAA GATCTGCTGCGCAAACAGCGCACCTTTGATAACGGC AGCATTCCGCATCAGATTCATCTGGGCGAACTGCAT GCGATTCTGCGCCGCCAGGAAGATTTTTATCCGTTTC TGAAAGATAACCGCGAAAAAATTGAAAAAATTCTG ACCTTTCGCATTCCGTATTATGTGGGCCCGCTGGCGC GCGGCAACAGCCGCTTTGCGTGGATGACCCGCAAAA GCGAAGAAACCATTACCCCGTGGAACTTTGAAGAAG TGGTGGATAAAGGCGCGAGCGCGCAGAGCTTTATTG AACGCATGACCAACTTTGATAAAAACCTGCCGAACG AAAAAGTGCTGCCGAAACATAGCCTGCTGTATGAAT ATTTTACCGTGTATAACGAACTGACCAAAGTGAAAT ATGTGACCGAAGGCATGCGCAAACCGGCGTTTCTGA GCGGCGAACAGAAAAAAGCGATTGTGGATCTGCTGT TTAAAACCAACCGCAAAGTGACCGTGAAACAGCTGA AAGAAGATTATTTTAAAAAAATTGAATGCTTTGATA GCGTGGAAATTAGCGGCGTGGAAGATCGCTTTAACG CGAGCCTGGGCACCTATCATGATCTGCTGAAAATTA TTAAAGATAAAGATTTTCTGGATAACGAAGAAAACG AAGATATTCTGGAAGATATTGTGCTGACCCTGACCC TGTTTGAAGATCGCGAAATGATTGAAGAACGCCTGA AAACCTATGCGCATCTGTTTGATGATAAAGTGATGA AACAGCTGAAACGCCGCCGCTATACCGGCTGGGGCC GCCTGAGCCGCAAACTGATTAACGGCATTCGCGATA AACAGAGCGGCAAAACCATTCTGGATTTTCTGAAAA GCGATGGCTTTGCGAACCGCAACTTTATGCAGCTGA	

[0339]

	TTCATGATGATAGCCTGACCTTTAAAGAAGATATTC AGAAAGCGCAGGTGAGCGGCCAGGGCGATAGCCTG CATGAACATATTGCGAACCTGGCGGGCAGCCCGGCG ATTAAAAAAGGCATTCTGCAGACCGTGAAAGTGGTG GATGAACTGGTGAAAGTGATGGGCCGCCATAAACCG GAAAACATTGTGATTGAAATGGCGCGCGAAAACCA GACCACCCAGAAAGGCCAGAAAAACAGCCGCGAAC GCATGAAACGCATTGAAGAAGGCATTAAAGAACTG GGCAGCCAGATTCTGAAAGAACATCCGGTGGAAAA CACCCAGCTGCAGAACGAAAAACTGTATCTGTATTA TCTGCAGAACGGCCGCGATATGTATGTGGATCAGGA ACTGGATATTAACCGCCTGAGCGATTATGATGTGGA TCATATTGTGCCGCAGAGCTTTCTGAAAGATGATAG CATTGATAACAAAGTGCTGACCCGCAGCGATAAAAA CCGCGGCAAAAGCGATAACGTGCCGAGCGAAGAAG TGGTGAAAAAAATGAAAAACTATTGGCGCCAGCTGC TGAACGCGAAACTGATTACCCAGCGCAAATTTGATA ACCTGACCAAAGCGGAACGCGGCGGCCTGAGCGAA CTGGATAAAGCGGGCTTTATTAAACGCCAGCTGGTG GAAACCCGCCAGATTACCAAACATGTGGCGCAGATT CTGGATAGCCGCATGAACACCAAATATGATGAAAAC GATAAACTGATTCGCGAAGTGAAAGTGATTACCCTG AAAAGCAAACTGGTGAGCGATTTTCGCAAAGATTTT CAGTTTTATAAAGTGCGCGAAATTAACAACTATCAT CATGCGCATGATGCGTATCTGAACGCGGTGGTGGGC ACCGCGCTGATTAAAAAATATCCGAAACTGGAAAGC GAATTTGTGTATGGCGATTATAAAGTGTATGATGTG CGCAAAATGATTGCGAAAAGCGAACAGGAAATTGG CAAAGCGACCGCGAAATATTTTTTTTATAGCAACAT TATGAACTTTTTTAAAACCGAAATTACCCTGGCGAA CGGCGAAATTCGCAAACGCCCGCTGATTGAAACCAA CGGCGAAACCGGCGAAATTGTGTGGGATAAAGGCC GCGATTTTGCGACCGTGCGCAAAGTGCTGAGCATGC CGCAGGTGAACATTGTGAAAAAAACCGAAGTGCAG ACCGGCGGCTTTAGCAAAGAAAGCATTCTGCCGAAA CGCAACAGCGATAAACTGATTGCGCGCAAAAAAGA TTGGGATCCGAAAAAATATGGCGGCTTTGATAGCCC GACCGTGGCGTATAGCGTGCTGGTGGTGGCGAAAGT GGAAAAAGGCAAAAGCAAAAAACTGAAAAGCGTGA AAGAACTGCTGGGCATTACCATTATGGAACGCAGCA GCTTTGAAAAAAACCCGATTGATTTTCTGGAAGCGA AAGGCTATAAAGAAGTGAAAAAAGATCTGATTATTA AACTGCCGAAATATAGCCTGTTTGAACTGGAAAACG GCCGCAAACGCATGCTGGCGAGCGCGGGCGAACTG CAGAAAGGCAACGAACTGGCGCTGCCGAGCAAATA TGTGAACTTTCTGTATCTGGCGAGCCATTATGAAAA ACTGAAAGGCAGCCCGGAAGATAACGAACAGAAAC AGCTGTTTGTGGAACAGCATAAACATTATCTGGATG AAATTATTGAACAGATTAGCGAATTTAGCAAACGCG	

[0340]

	TGATTCTGGCGGATGCGAACCTGGATAAAGTGCTGA GCGCGTATAACAAACATCGCGATAAACCGATTCGCG AACAGGCGGAAAACATTATTCATCTGTTTACCCTGA CCAACCTGGGCGCGCCGGCGGCGTTTAAATATTTTG ATACCACCATTGATCGCAAACGCTATACCAGCACCA AAGAAGTGCTGGATGCGACCCTGATTCATCAGAGCA TTACCGGCCTGTATGAAACCCGCATTGATCTGAGCC AGCTGGGCGGCGAT	

[0341]

[0342] 본원에 기재된 다중-플랩 프라임 편집제는 정규 SpCas9, 또는 상기 제공된 야생형 Cas9 서열과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 그의 임의의 변이체를 포함할 수 있다. 이들 변이체는 하기를 포함하는, 스위스프롯 수탁 번호 Q99ZW2 (서열식별번호: 18) 엔트리로 보고된 임의의 공지된 돌연변이를 포함하는, 1개 이상의 돌연변이를 함유하는 SpCas9 변이체를 포함할 수 있다:

[0343]

SpCas9 돌연변이 (정규 SpCas9 서열, SEQ ID NO: 18의 아미노산 서열 대비)	기능/특징 (보고된 바와 같음) (유니프롯KB - Q99ZW2 (CAS9_STRPT1) 엔트리 참조 – 본원에 참조로 포함됨)
D10A	프로토스페이서 가닥을 절단하는 (그러나 비-프로토스페이서 가닥은 절단하지 않는) 닉카제 돌연변이체
S15A	감소된 DNA 절단 활성
R66A	감소된 DNA 절단 활성
R70A	DNA 절단 없음
R74A	감소된 DNA 절단
R78A	감소된 DNA 절단
97-150 결실	뉴클레아제 활성 없음
R165A	감소된 DNA 절단
175-307 결실	약 50% 감소된 DNA 절단
312-409 결실	뉴클레아제 활성 없음
E762A	닉카제
H840A	비-프로토스페이서 가닥은 절단하지만 프로토스페이서 가닥은 절단하지 않는 닉카제 돌연변이체
N854A	닉카제
N863A	닉카제
H982A	감소된 DNA 절단
D986A	닉카제
1099-1368 결실	뉴클레아제 활성 없음
R1333A	감소된 DNA 결합

[0344] 본 개시내용에 사용될 수 있는 다른 야생형 SpCas9 서열은 하기를 포함한다:

설명	서열	SEQ ID NO:
SpCas9 *스트렙토코쿠스 피오게네스* MGAS1882 야생형 NC_017053.1	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCA CAAATAGCGTCGGATGGGCGGTGATCACTGATGATTA TAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAAT ACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGG CTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGAC TCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGT CGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTC AAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATC GACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAA GCATGAACGTCATCCTATTTTTGGAAATATAGTAGATG AAGTTGCTTATCATGAGAAATATCCAACTATCTATCAT CTGCGAAAAAAAATTGGCAGATTCTACTGATAAAGCGG ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATT AAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAA TCCTGATAATAGTGATGTGGACAAACTATTTATCCAGT TGGTACAAATCTACAATCAATTATTTGAAGAAAACCCT ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTC TGCACGATTGAGTAAATCAAGACGATTAGAAAATCTC ATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTT TGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTA ATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAA TTACAGCTTTCAAAAGATACTTACGATGATGATTTAGA TAATTTATTGGCGCAAATTGGAGATCAATATGCTGATT TGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTA CTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTA AGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGAT GAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGT TCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTT TTTGATCAATCAAAAAACGGATATGCAGGTTATATTGA TGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATC AAACCAATTTTAGAAAAAATGGATGGTACTGAGGAAT TATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAA GCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAA GAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAA GATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATG TTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGG ATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGA ATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAA TCATTTATTGAACGCATGACAAACTTTGATAAAAATCT TCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTT ATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTC AAATATGTTACTGAGGGAATGCGAAAACCAGCATTTC TTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTC TTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAA AAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAG	SEQ ID NO: 20

[0345]

	TGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTT CATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAA GATAAAGATTTTTTGGATAATGAAGAAAATGAAGATA TCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAA GATAGGGGGATGATTGAGGAAAGACTTAAAACATATG CTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAA CGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAA ATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAA ACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAA TCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGA CATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGG ACAAGGCCATAGTTTACATGAACAGATTGCTAACTTA GCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGAC TGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGG CATAAGCCAGAAAATATCGTTATTGAAATGGCACGTG AAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCG AGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGA ATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAA AATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTA TCTACAAAATGGAAGAGACATGTATGTGGACCAAGAA TTAGATATTAATCGTTTAAGTGATTATGATGTCGATCA CATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAG ACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGG TAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAA AAGATGAAAAACTATTGGAGACAACTTCTAAACGCCA AGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAA GCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTG GTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATC ACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGA ATACTAAATACGATGAAAATGATAAACTTATTCGAGA GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTG ACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAG ATTAACAATTACCATCATGCCCATGATGCGTATCTAAA TGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAA AACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTT TATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGA AATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTA ATATCATGAACTTCTTCAAAACAGAAATTACACTTGCA AATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTA ATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCG AGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCC AAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGG CGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAAT TCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATC CAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGC TTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGG AAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAG GGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAA TCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAA	

[0346]

	GTTAAAAAAGACTTAATCATTAAACTACCTAAATATA GTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTG GCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGG CTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCT AGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATA ACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCA TTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTT CTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAA GTTCTTAGTGCATATAACAAACATAGAGACAAACCAA TACGTGAACAAGCAGAAAATATTATTCATTTATTTACG TTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTT TGATACAACAATTGATCGTAAACGATATACGTCTACA AAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCAT CACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGC TAGGAGGTGACTGA	
SpCas9 *스트렙토코쿠스 피오게네스* MGAS1882 야생형 NC_017053.1	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNT DRHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRRKN RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLA LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFE ENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGL FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG AYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIE ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKV MGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSD NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR	SEQ ID NO: 21

[0347]

	DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST KEVLDATLIHQSITGLYETRIDLSQLGGD	
SpCas9 *스트렙토코쿠스 피오게네스* 야생형 SWBC2D7W014	ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCAC TAATTCCGTTGGATGGGCTGTCATAACCGATGAATACA AAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACAC AGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCC TCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCG CCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCA ATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAAC ATGAACGGCACCCCATCTTTGGAAACATAGTAGATGA GGTGGCATATCATGAAAAGTACCCAACGATTTATCAC CTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGG ACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATA AAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAA TCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGT TAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTA GCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCT GATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTG TTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACC AAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCA AATTGCAGCTTAGTAAGGACACGTACGATGACGATCT CGACAATCTACTGGCACAAATTGGAGATCAGTATGCG GACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAAT CCTCCTATCTGACATACTGAGAGTTAATACTGAGATTA CCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTA CGATGAACATCACCAAGACTTGACACTTCTCAAGGCC CTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAA TATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTAT ATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGT TTATCAAACCCATATTAGAGAAGATGGATGGGACGGA AGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGC GAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACA TCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAA GGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGT GAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTA CTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCG CATGGATGACAAGAAAGTCCGAAGAAACGATTACTCC ATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCA GCTCAATCGTTCATCGAGAGGATGACCAACTTTGACA AGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAG TTTACTTTACGAGTATTTCACAGTGTACAATGAACTCA CGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACC CGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTA GATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTA AGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATG CTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGAT	SEQ ID NO: 22

[0348]

	TTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAG ATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGA ATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACC CTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAA AAACATACGCTCACCTGTTCGACGATAAGGTTATGAA ACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGA TTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGC AAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGAC GGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGA TGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCA CAGGTTTCCGGACAAGGGGACTCATTGCACGAACATA TTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGC ATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTA AGGTCATGGGACGTCACAAACCGGAAAACATTGTAAT CGAGATGGCACGCGAAAATCAAACGACTCAGAAGGG GCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGA AGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAG GAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGA AACTTTACCTCTATTACCTACAAAATGGAAGGGACATG TATGTTGATCAGGAACTGGACATAAACCGTTTATCTGA TTACGACGTCGATCACATTGTACCCCAATCCTTTTTGA AGGACGATTCAATCGACAATAAAGTGCTTACACGCTC GGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGC GAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGC AGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTT CGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCT GAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGT GGAAACCCGCCAAATCACAAAGCATGTTGCACAGATA CTAGATTCCCGAATGAATACGAAATACGACGAGAACG ATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAA GTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAAT TCTATAAAGTTAGGGAGATAAATAACTACCACCATGC GCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCAC TCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGT GTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGA TCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAG CCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTA AGACGGAAATCACTCTGGCAAACGGAGAGATACGCAA ACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAA ATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGA GAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAA GAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAA TCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGC TCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGC TTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGT GGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAA GTCAGTCAAAGAATTATTGGGGATAACGATTATGGAG CGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGA GGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATA	

[0349]

	ATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAA TGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTT CAAAAGGGGAACGAACTCGCACTACCGTCTAAATACG TGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTG AAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTT TTGTTGAGCAGCACAAACATTATCTCGACGAAATCATA GAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAG CTGATGCCAATCTGGACAAAGTATTAAGCGCATACAA CAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAA AATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCT CCAGCCGCATTCAAGTATTTTGACACAACGATAGATCG CAAACGATACACTTCTACCAAGGAGGTGCTAGACGCG ACACTGATTCACCAATCCATCACGGGATTATATGAAAC TCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCC CCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAG ACCATGACGGTGATTATAAAGATCATGACATCGATTA CAAGGATGACGATGACAAGGCTGCAGGA	
SpCas9 *스트렙토코쿠스 피오게네스* 야생형 SWBC2D7W014의 코딩된 생성물	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL	SEQ ID NO: 23

[0350]

	FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVS SDYKDHDGDYKDHDIDYKDDDDKAAG	
SpCas9 *스트렙토코쿠스 피오게네스* M1GAS 야생형 NC_002737.2	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCA CAAATAGCGTCGGATGGGCGGTGATCACTGATGAATA TAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAAT ACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGG CTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGAC TCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGT CGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTC AAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATC GACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAA GCATGAACGTCATCCTATTTTTGGAAATATAGTAGATG AAGTTGCTTATCATGAGAAATATCCAACTATCTATCAT CTGCGAAAAAAAATTGGTAGATTCTACTGATAAAGCGG ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATT AAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAA TCCTGATAATAGTGATGTGGACAAACTATTTATCCAGT TGGTACAAACCTACAATCAATTATTTGAAGAAAACCCT ATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTC TGCACGATTGAGTAAATCAAGACGATTAGAAAATCTC ATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATT TGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTA ATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAA TTACAGCTTTCAAAAGATACTTACGATGATGATTTAGA TAATTTATTGGCGCAAATTGGAGATCAATATGCTGATT TGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTA CTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAA GGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATG AACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTT CGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT TTGATCAATCAAAAAACGGATATGCAGGTTATATTGAT GGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCA AACCAATTTTAGAAAAAATGGATGGTACTGAGGAATT ATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAG CAACGGACCTTTGACAACGGCTCTATTCCCCATCAAAT TCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAG AAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAG ATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTT GGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGAT GACTCGGAAGTCTGAAGAAACAATTACCCCATGGAAT TTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATC ATTTATTGAACGCATGACAAACTTTGATAAAAATCTTC CAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTAT GAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAA ATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTT CAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTC AAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAG	SEQ ID NO: 24

[0351]

	AAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTT GAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATT AGGTACCTACCATGATTTGCTAAAAATTATTAAAGATA AAGATTTTTTGGATAATGAAGAAAATGAAGATATCTT AGAGGATATTGTTTTAACATTGACCTTATTTGAAGATA GGGAGATGATTGAGGAAAGACTTAAAACATATGCTCA CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTC GCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTG ATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAA TATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGC AATTTTATGCAGCTGATCCATGATGATAGTTTGACATT TAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAA GGCGATAGTTTACATGAACATATTGCAAATTTAGCTGG TAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAA AAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCA TAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAA AATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAG AGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATT AGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAAT ACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCT CCAAAATGGAAGAGACATGTATGTGGACCAAGAATTA GATATTAATCGTTTAAGTGATTATGATGTCGATCACAT TGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACA ATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAA ATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAG ATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGT TAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCT GAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTT TTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACT AAGCATGTGGCACAAATTTTGGATAGTCGCATGAATA CTAAATACGATGAAAATGATAAACTTATTCGAGAGGT TAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACT TCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATT AACAATTACCATCATGCCCATGATGCGTATCTAAATGC CGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAAC TTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTAT GATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAA TAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAAT ATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAA TGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAAT GGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAG ATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAA GTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCG GATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTC GGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCA AAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTA TTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAA TCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGA TCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCC	

[0352]

[0353]

	GATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTT AAAAAAGACTTAATCATTAAACTACCTAAATATAGTCT TTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCT AGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTC TGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACG AACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTA TTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTA AGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTT CTTAGTGCATATAACAAACATAGAGACAAACCAATAC GTGAACAAGCAGAAAATATTATTCATTTATTTACGTTG ACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGA TACAACAATTGATCGTAAACGATATACGTCTACAAAA GAAGTTTTAGATGCCACTCTTATCCATCAATCCATCAC TGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAG GAGGTGACTGA	
SpCas9 *스트렙토코쿠스 피오게네스* M1GAS 야생형 NC_002737.2 의 코딩된 생성물 (정규 Q99ZW2 야생형과 100% 동일)	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH	SEQ ID NO: 25

[0354]

	RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	

[0355] 본원에 기재된 다중-플랩 프라임 편집제는 임의의 상기 SpCas9 서열, 또는 그와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 그의 임의의 변이체를 포함할 수 있다.

[0356] B. 야생형 Cas9 오르토로그

[0357] 다른 실시양태에서, Cas9 단백질은 에스. 피오게네스로부터의 정규 Cas9와 상이한 또 다른 박테리아 종으로부터의 야생형 Cas9 오르토로그일 수 있다. 예를 들어, 하기 Cas9 오르토로그가 본 명세서에 기재된 다중-플랩 프라임 편집제 구축물과 관련하여 사용될 수 있다. 또한, 하기 오르토로그 중 임의의 것에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 임의의 변이체 Cas9 오르토로그가 또한 본 발명의 다중-플랩 프라임 편집제와 함께 사용될 수 있다.

설명	서열
LfCas9 *락토바실루스 페르멘툼* 야생형 *진뱅크: SNX31424.1* 1	MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERR TFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPTK NQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVMNIYKLRE AMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFN VLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLE VKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSS ETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRY WTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKERGFDLEKGL KKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQ AKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQK ATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLL NEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKT VKASDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDK VDDNRYQTDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRY KGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFKEIVDQPVFKEQIDQL NQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVG NAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEK APDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDNSL DNRVLTSRKENNKKSDQVPAKLYAAKMKPYWNQLLKQGLITQRKFEN LTKDVDQNIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTEIIETR AGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSF FVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTR DEVAKSFDRLLNMKYMLVSKEVHDRSDQLYGATIVTAKESGKLTSPIEI KKNRLVDLYGAYTNGTSAFMTIIKFTGNKPKYKVIGIPTTSAASLKRAGK PGSESYNQELHRIIKSNPKVKKGFEIVVPHVSYGQLIVDGDCKFTLASPTV QHPATQLVLSKKSLETISSGYKILKDKPAIANERLIRVFDEVVGQMNRYF TIFDQRSNRQKVADARDKFLSLPTESKYEGAKKVQVGKTEVITNLLMGL HANATQGDLKVLGLATFGFFQSTTGLSLSEDTMIVYQSPTGLFERRICLK DI (SEQ ID NO: 26)
SaCas9 *스타필로코쿠스 아우레우스* 야생형 진뱅크: AYD60528.1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ

[0358]

설명	서열
	SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 27)
SaCas9 *스타필로코쿠스 아우레우스*	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRS KRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQ KLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEE KYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKP TLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENA ELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSP VVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP FNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRY ATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYK HHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDIT DDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEV NSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIE VNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSK KHPQIIKK (SEQ ID NO: 28)
StCas9 *스트렙토코쿠스 써모필루스* 유니프롯KB/스위스-프롯: G3ECR1.2 야생형	MLFNKCIIISINLDFSNKEKCMTKPYSIGLDIGTNSVGWAVITDNYKVPSK KMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRI LYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYH DEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKN NDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLF PGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLL GYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEH KEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLK NLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQA KFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNF EDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVR FIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIE LKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIK QRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLI

[0359]

설명	서열
	DDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEVVKSLPGSPAI KKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLK RLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTG DDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDFPSLEVV KKRKTFWYQLLKSKLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETR QITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYK VREINDFHHAHDAYLNAVIASALLKKYPKLEPEFVYGDYPKYNSFRERK SATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLA TVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNEN LVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEKGAKKKITNVLEFQGISI LDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELSDGSRRMLASILSTN NKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEEL FYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSERKG LFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRID LAKLGEG (SEQ ID NO: 29)
LcCas9 *락토바실루스* *크리스파투스* NCBI 참조 서열: WP_133478044.1 야생형	MKIKNYNLALTPSTSAVGHVEVDDDLNILEPVHHQKAIGVAKFGEGETA EARRLARSARRTTKRRANRINHYFNEIMKPEIDKVDPLMFDRIKQAGLSP LDERKEFRTVIFDRPNIASYYHNQFPTIWHLQKYLMITDEKADIRLIYWA LHSLLKHRGHFFNTTPMSQFKPGKLNLKDDMLALDDYNDLEGLSFAVA NSPEIEKVIKDRSMHKKEKIAELKKLIVNDVPDKDLAKRNNKIITQIVNAI MGNSFHLNFIFDMDLDKLTSKAWSFKLDDPELDTKFDAISGSMTDNQIGI FETLQKIYSAISLLDILNGSSNVVDAKNALYDKHKRDLNLYFKFLNTLPD EIAKTLKAGYTLYIGNRKKDLLAARKLLKVNVAKNFSQDDFYKLINKEL KSIDKQGLQTRFSEKVGELVAQNNFLPVQRSSDNVFIPYQLNAITFNKILE NQGKYYDFLVKPNPAKKDRKNAPYELSQLMQFTIPYYVGPLVTPEEQV KSGIPKTSRFAWMVRKDNGAITPWNFYDKVDIEATADKFIKRSIAKDSY LLSELVLPKHSLLYEKYEVFNELSNVSLDGKKLSGGVKQILFNEVFKKTN KVNTSRILKALAKHNIPGSKITGLSNPEEFTSSLQTYNAWKKYFPNQIDNF AYQQDLEKMIEWSTVFEDHKILAKKLDEIEWLDDDQKKFVANTRLRGW GRLSKRLLTGLKDNYGKSIMQRLETTKANFQQIVYKPEFREQIDKISQAA AKNQSLEDILANSYTSPSNRKAIRKTMSVVDEYIKLNHGKEPDKIFLMFQ RSEQEKGKQTEARSKQLNRILSQLKADKSANKLFSKQLADEFSNAIKKS KYKLNDKQYFYFQQLGRDALTGEVIDYDELYKYTVLHIIPRSKLTDDSQ NNKVLTKYKIVDGSVALKFGNSYSDALGMPIKAFWTELNRLKLIPKGKL LNLTTDFSTLNKYQRDGYIARQLVETQQIVKLLATIMQSRFKHTKIIEVR NSQVANIRYQFDYFRIKNLNEYYRGFDAYLAAVVGTYLYKVYPKARRL FVYGQYLKPKKTNQENQDMHLDSEKKSQGFNFLWNLLYGKQDQIFVN GTDVIAFNRKDLITKMNTVYNYKSQKISLAIDYHNGAMFKATLFPRNDR DTAKTRKLIPKKKDYDTDIYGGYTSNVDGYMLLAEIIKRDGNKQYGFYG VPSRLVSELDTLKKTRYTEYEEKLKEIIKPELGVDLKKIKKIKILKNKVPF NQVIIDKGSKFFITSTSYRWNYRQLILSAESQQTLMDLVVDPDFSNHKAR KDARKNADERLIKVYEEILYQVKNYMPMFVELHRCYEKLVDAQKTFKS LKISDKAMVLNQILILLHSNATSPVLEKLGYHTRFTLGKKHNLISENAVL VTQSITGLKENHVSIKQML (SEQ ID NO: 30)
PdCas9	MTNEKYSIGLDIGTSSIGFAVVNDNNRVIRVKGKNAIGVRLFDEGKAAA DRRSFRTTRRSFRTTRRRLSRRRWRLKLLREIFDAYITPVDEAFFIRLKES NLSPKDSKKQYSGDILFNDRSDKDFYEKYPTIYHLRNALMTEHRKFDVR EIYLAIHHIMKFRGHFLNATPANNFKVGRLNLEEKFEELNDIYQRVFPDE

[0360]

설명	서열
페디오코쿠스 담노수스 NCBI 참조 서열: WP_062913273.1 야생형	SIEFRTDNLEQIKEVLLDNKRSRADRQRTLVSDIYQSSEDKDIEKRNKAV ATEILKASLGNKAKLNVITNVEVDKEAAKEWSITFDSESIDDDLAKIEGQ MTDDGHEIIEVLRSLYSGITLSAIVPENHTLSQSMVAKYDLHKDHLKLFK KLINGMTDTKKAKNLRAAYDGYIDGVKGKVLPQEDFYKQVQVNLDDS AEANEIQTYIDQDIFMPKQRTKANGSIPHQLQQQELDQIIENQKAYYPWL AELNPNPDKKRQQLAKYKLDELVTFRVPYYVGPMITAKDQKNQSGAEF AWMIRKEPGNITPWNFDQKVDRMATANQFIKRMTTTDTYLLGEDVLPA QSLLYQKFEVLNELNKIRIDHKPISIEQKQQIFNDLFKQFKNVTIKHLQDY LVSQGQYSKRPLIEGLADEKRFNSSLSTYSDLCGIFGAKLVEENDRQEDL EKIIEWSTIFEDKKIYRAKLNDLTWLTDDQKEKLATKRYQGWGRLSRKL LVGLKNSEHRNIMDILWITNENFMQIQAEPDFAKLVTDANKGMLEKTDS QDVINDLYTSPQNKKAIRQILLVVHDIQNAMHGQAPAKIHVEFARGEER NPRRSVQRQRQVEAAYEKVSNELVSAKVRQEFKEAINNKRDFKDRLFL YFMQGGIDIYTGKQLNIDQLSSYQIDHILPQAFVKDDSLTNRVLTNENQV KADSVPIDIFGKKMLSVWGRMKDQGLISKGKYRNLTMNPENISAHTENG FINRQLVETRQVIKLAVNILADEYGDSTQIISVKADLSHQMREDFELLKN RDVNDYHHAFDAYLAAFIGNYLLKRYPKLESYFVYGDFKKFTQKETKM RRFNFIYDLKHCDQVVNKETGEILWTKDEDIKYIRHLFAYKKILVSHEVR EKRGALYNQTIYKAKDDKGSGQESKKLIRIKDDKETKIYGGYSGKSLAY MTIVQITKKNKVSYRVIGIPTLALARLNKLENDSTENNGELYKIIKPQFTH YKVDKKNGEIIETTDDFKIVVSKVRFQQLIDDAGQFFMLASDTYKNNAQ QLVISNNALKAINNTNITDCPRDDLERLDNLRLDSAFDEIVKKMDKYFSA YDANNFREKIRNSNLIFYQLPVEDQWENNKITELGKRTVLTRILQGLHAN ATTTDMSIFKIKTPFGQLRQRSGISLSENAQLIYQSPTGLFERRVQLNKIK (SEQ ID NO: 31)
FnCas9 *푸소박테리움 누클레아툼* NCBI 참조 서열: WP_060798984.1	MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLRFNKKDMWGSRLFEE AKTAAERRVQRNSRRRLKRRKWRLNLLEEIFSNEILKIDSNFFRRLKESSL WLEDKSSKEKFTLFNDDNYKDYDFYKQYPTIFHLRNELIKNPEKKDIRLV YLAIHSIFKSRGHFLFEGQNLKEIKNFETLYNNLIAFLEDNGINKIIDKNNI EKLEKIVCDSKKGLKDKEKEFKEIFNSDKQLVAIFKLSVGSSVSLNDLFD TDEYKKGEVEKEKISFREQIYEDDKPIYYSILGEKIELLDIAKTFYDFMVL NNILADSQYISEAKVKLYEEHKKDLKNLKYIIRKYNKGNYDKLFKDKNE NNYSAYIGLNKEKSKKEVIEKSRLKIDDLIKNIKGYLPKVEEIEEKDKAIF NKILNKIELKTILPKQRISDNGTLPYQIHEAELEKILENQSKYYDFLNYEE NGIITKDKLLMTFKFRIPYYVGPLNSYHKDKGGNSWIVRKEEGKILPWNF EQKVDIEKSAEEFIKRMTNKCTYLNGEDVIPKDTFLYSEYVILNELNKVQ VNDEFLNEENKRKIIDELFKENKKVSEKKFKEYLLVKQIVDGTIELKGVK DSFNSNYISYIRFKDIFGEKLNLDIYKEISEKSILWKCLYGDDKKIFEKKIK NEYGDILTKDEIKKINTFKFNNWGRLSEKLLTGIEFINLETGECYSSVMDA LRRTNYNLMELLSSKFTLQESINNENKEMNEASYRDLIEESYVSPSLKRAI FQTLKIYEEIRKITGRVPKKVFIEMARGGDESMKNKKIPARQEQLKKLYD SCGNDIANFSIDIKEMKNSLISYDNNSLRQKKLYLYYLQFGKCMYTGREI DLDRLLQNNDTYDIDHIYPRSKVIKDDSFDNLVLVLKNENAEKSNEYPV KKEIQEKMKSFWRFLKEKNFISDEKYKRLTGKDDFELRGFMARQLVNV RQTTKEVGKILQQIEPEIKIVYSKAEIASSFREMFDFIKVRELNDTHHAKD AYLNIVAGNVYNTKFTEKPYRYLQEIKENYDVKKIYNYDIKNAWDKEN SLEIVKKNMEKNTVNITRFIKEKKGQLFDLNPIKKGETSNEIISIKPKVYN GKDDKLNEKYGYYKSLNPAYFLYVEHKEKNKRIKSFERVNLVDVNNIK	

[0361]

설명	서열
	DEKSLVKYLIENKKLVEPRVIKKVYKRQVILINDYPYSIVTLDSNKLMDF ENLKPLFLENKYEKILKNVIKFLEDNQGKSEENYKFIYLKKKDRYEKNET LESVKDRYNLEFNEMYDKFLEKLDSKDYKNYMNNKKYQELLDVKEKFI KLNLFDKAFTLKSFLDLFNRKTMADFSKVGLTKYLGKIQKISSNVLSKNE LYLLEESVTGLFVKKIKL (SEQ ID NO: 32)
EcCas9 *엔테로코쿠스 세코룸* NCBI 참조 서열: WP_047338501.1 야생형	RRKQRIQILQELLGEEVLKTDPGFFHRMKESRYVVEDKRTLDGKQVELP YALFVDKDYTDKEYYKQFPTINHLIVYLMTTSDTPDIRLVYLALHYYMK NRGNFLHSGDINNVKDINDILEQLDNVLETFLDGWNLKLKSYVEDIKNIY NRDLGRGERKKAFVNTLGAKTKAEKAFCSLISGGSTNLAELFDDSSLKEI ETPKIEFASSSLEDKIDGIQEALEDRFAVIEAAKRLYDWKTLTDILGDSSS LAEARVNSYQMHHEQLLELKSLVKEYLDRKVFQEVFVSLNVANNYPAY IGHTKINGKKKELEVKRTKRNDFYSYVKKQVIEPIKKKVSDEAVLTKLSE IESLIEVDKYLPLQVNSDNGVIPYQVKLNELTRIFDNLENRIPVLRENRDK IIKTFKFRIPYYVGSLNGVVKNGKCTNWMVRKEEGKIYPWNFEDKVDLE ASAEQFIRRMTNKCTYLVNEDVLPKYSLLYSKYLVLSELNNLRIDGRPLD VKIKQDIYENVFKKNRKVTLKKIKKYLLKEGIITDDDELSGLADDVKSSL TAYRDFKEKLGHLDLSEAQMENIILNITLFGDDKKLLKKRLAALYPFIDD KSLNRIATLNYRDWGRLSERFLSGITSVDQETGELRTIIQCMYETQANLM QLLAEPYHFVEAIEKENPKVDLESISYRIVNDLYVSPAVKRQIWQTLLVIK DIKQVMKHDPERIFIEMARENQESKKTKSRKQVLSEVYKKAKEYEHLFE KLNSLTEEQLRSKKIYLYFTQLGKCMYSGEPIDFENLVSANSNYDIDHIYP QSKTIDDSFNNIVLVKKSLNAYKSNHYPIDKNIRDNEKVKTLWNTLVSK GLITKEKYERLIRSTPFSDEELAGFIARQLVETRQSTKAVAEILSNWFPESE IVYSKAKNVSNFRQDFEILKVRELNDCHHAHDAYLNIVVGNAYHTKFTN SPYRFIKNKANQEYNLRKLLQKVNKIESNGVVAWVGQSENNPGTIATVK KVIRRNTVLISRMVKEVDGQLFDLTLMKKGKGQVPIKSSDERLTDISKY GGYNKATGAYFTFVKSKKRGKVVRSFEYVPLHLSKQFENNNELLKEYIE KDRGLTDVEILIPKVLINSLFRYNGSLVRITGRGDTRLLLVHEQPLYVSNS FVQQLKSVSSYKLKKSENDNAKLTKTATEKLSNIDELYDGLLRKLDLPIY SYWFSSIKEYLVESRTKYIKLSIEEKALVIFEILHLFQSDAQVPNLKILGLS TKPSRIRIQKNLKDTDKMSIIHQSPSGIFEHEIELTSL (SEQ ID NO: 33)
AhCas9 *아나에로스 티페스 하드루스* NCBI 참조 서열: WP_044924278.1 야생형	MQNGFLGITVSSEQVGWAVTNPKYELERASRKDLWGVRLFDKAETAED RRMFRTNRRLNQRKKNRIHYLRDIFHEEVNQKDPNFFQQLDESNFCEDD RTVEFNFDTNLYKNQFPTVYHLRKYLMETKDKPDIRLVYLAFSKFMKN RGHFLYKGNLGEVMDFENSMKGFCESLEKFNIDFPTLSDEQVKEVRDIL CDHKIAKTVKKKNIITITKVKSKTAKAWIGLFCGCSVPVKVLFQDIDEEIV TDPEKISFEDASYDDYIANIEKGVGIYYEAIVSAKMLFDWSILNEILGDHQ LLSDAMIAEYNKHHDDLKRLQKIIKGTGSRELYQDIFINDVSGNYVCYV GHAKTMSSADQKQFYTFLKNRLKNVNGISSEDAEWIDTEIKNGTLLPKQ TKRDNSVIPHQLQLREFELILDNMQEMYPFLKENREKLLKIFNFVIPYYV GPLKGVVRKGESTNWMVPKKDGVIHPWNFDEMVDKEASAECFISRMT GNCSYLFNEKVLPKNSLLYETFEVLNELNPLKINGEPISVELKQRIYEQLF LTGKKVTKKSLTKYLIKNGYDKDIELSGIDNEFHSNLKSHIDFEDYDNLS DEEVEQIILRITVFEDKQLLKDYLNREFVKLSEDERKQICSLSYKGWGNL SEMLLNGITVTDSNGVEVSVMDMLWNTNLNLMQILSKKYGYKAEIEHY NKEHEKTIYNREDLMDYLNIPPAQRRKVNQLITIVKSLKKTYGVPNKIFF KISREHQDDPKRTSSRKEQLKYLYKSLKSEDEKHLMKELDELNDHELSN DKVYLYFLQKGRCIYSGKKLNLSRLRKSNYQNDIDYIYPLSAVNDRSMN	

[0362]

설명	서열
	NKVLTGIQENRADKYTYFPVDSEIQKKMKGFWMELVLQGFMTKEKYFR LSRENDFSKSELVSFIEREISDNQQSGRMIASVLQYYFPESKIVFVKEKLIS SFKRDFHLISSYGHNHLQAAKDAYITIVVGNVYHTKFTMDPAIYFKNHK RKDYDLNRLFLENISRDGQIAWESGPYGSIQTVRKEYAQNHIAVTKRVV EVKGGLFKQMPLKKGHGEYPLKTNDPRFGNIAQYGGYTNVTGSYFVLV ESMEKGKKRISLEYVPVYLHERLEDDPGHKLLKEYLVDHRKLNHPKILL AKVRKNSLLKIDGFYYRLNGRSGNALILTNAVELIMDDWQTKTANKISG YMKRRAIDKKARVYQNEFHIQELEQLYDFYLDKLKNGVYKNRKNNQA ELIHNEKEQFMELKTEDQCVLLTEIKKLFVCSPMQADLTLIGGSKHTGMI AMSSNVTKADFAVIAEDPLGLRNKVIYSHKGEK (SEQ ID NO: 34)
KvCas9 *칸들레리아 비툴리나* NCBI 참조 서열: WP_031589969.1 야생형	MSQNNNKIYNIGLDIGDASVGWAVVDEHYNLLKRHGKHMWGSRLFTQ ANTAVERRSSRSTRRRYNKRRERIRLLREIMEDMVLDVDPTFFIRLANVS FLDQEDKKDYLKENYHSNYNLFIDKDFNDKTYYDKYPTIYHLRKHLCES KEKEDPRLIYLALHHIVKYRGNFLYEGQKFSMDVSNIEDKMIDVLRQFN EINLFEYVEDRKKIDEVLNVLKEPLSKKHKAEKAFALFDTTKDNKAAYK ELCAALAGNKFNVTKMLKEAELHDEDEKDISFKFSDATFDDAFVEKQPL LGDCVEFIDLLHDIYSWVELQNILGSAHTSEPSISAAMIQRYEDHKNDLK LLKDVIRKYLPKKYFEVFRDEKSKKNNYCNYINHPSKTPVDEFYKYIKK LIEKIDDPDVKTILNKIELESFMLKQNSRTNGAVPYQMQLDELNKILENQ SVYYSDLKDNEDKIRSILTFRIPYYFGPLNITKDRQFDWIIKKEGKENERIL PWNANEIVDVDKTADEFIKRMRNFCTYFPDEPVMAKNSLTVSKYEVLN EINKLRINDHLIKRDMKDKMLHTLFMDHKSISANAMKKWLVKNQYFSN TDDIKIEGFQKENACSTSLTPWIDFTKIFGKINESNYDFIEKIIYDVTVFED KKILRRRLKKEYDLDEEKIKKILKLKYSGWSRLSKKLLSGIKTKYKDSTR TPETVLEVMERTNMNLMQVINDEKLGFKKTIDDANSTSVSGKFSYAEVQ ELAGSPAIKRGIWQALLIVDEIKKIMKHEPAHVYIEFARNEDEKERKDSF VNQMLKLYKDYDFEDETEKEANKHLKGEDAKSKIRSERLKLYYTQMG KCMYTGKSLDIDRLDTYQVDHIVPQSLLKDDSIDNKVLVLSSENQRKLD DLVIPSSIRNKMYGFWEKLFNNKIISPKKFYSLIKTEFNEKDQERFINRQIV ETRQITKHVAQIIDNHYENTKVVTVRADLSHQFRERYHIYKNRDINDFHH AHDAYIATILGTYIGHRFESLDAKYIYGEYKRIFRNQKNKGKEMKKNND GFILNSMRNIYADKDTGEIVWDPNYIDRIKKCFYYKDCFVTKKLEENNG TFFNVTVLPNDTNSDKDNTLATVPVNKYRSNVNKYGGFSGVNSFIVAIK GKKKKGKKVIEVNKLTGIPLMYKNADEEIKINYLKQAEDLEEVQIGKEIL KNQLIEKDGGLYYIVAPTEIINAKQLILNESQTKLVCEIYKAMKYKNYDN LDSEKIIDLYRLLINKMELYYPEYRKQLVKKFEDRYEQLKVISIEEKCNII KQILATLHCNSSIGKIMYSDFKISTTIGRLNGRTISLDDISFIAESPTGMYSK KYKL (SEQ ID NO: 35)
EfCas9 *엔테로코쿠스 파에칼리스* NCBI 참조 서열: WP_016631044.1	MRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFF ARLQESFLVPEDKKWHRHPIFAKLEDEVAYHETYPTIYHLRKKLADSSE QADLRLIYLALAHIVKYRGHFLIEGKLSTENTSVKDQFQQFMVIYNQTFV NGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQEKANGLFGQF LKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVF LAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIR ENCPDEYDNLFKNEQKDGYAGYIAHAGKVSQLKFYQYVKKIIQDIAGAE YFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQE KIEQLVTFRIPYYVGPLSKGDASTFAWLKRQSEEPIRPWNLQETVDLDQS

[0363]

설명	서열
야생형	ATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKA NFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFN ASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFK GQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGV SKHYNRNFMQLINDSQLSFKNAIQKAQSSEHEETLSETVNELAGSPAIKK GIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEK AMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDMYTGDELSLHRLS HYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAY WEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNV AGILDQRYNAKSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQD AYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLL RFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVEVQKGGFS KESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIK QEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRRL LASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEF QEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFN AMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSPTGLYETRRKVVD (SEQ ID NO: 36)
스타필로코쿠스 아우레우스 Cas9	KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKL SEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFI DTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAEL LDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAI NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVK RSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTN ERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNY EVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETF KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATR GLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHA EDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYK EIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDE KNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKC YEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMI DITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQ IIKKG (SEQ ID NO: 37)
게오바실루스 써모데니트리피칸스 Cas9	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRL ARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRV EALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQ SILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQ REYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPK ATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHD VRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVY

[0364]

설명	서열
	GKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLA DKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGY TFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHI ELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIV KFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKV LVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLR LHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVN GRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRR EQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEK LESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQ LDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGE LGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTI DMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIK TAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKY QVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL (SEQ ID NO: 38)
ScCas9 *에스 카니스* 1375 AA 159.2 kDa	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLM GALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSF FQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPE KADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTP NFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDA ILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAE IFKDDTKNGYAGYVDGIGIKHRKRTTKLATQEEFYKFIKPILEKMDGAEEL LAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKI EKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQS FIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGF SNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIK ELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDS RMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFAT VRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKY GGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGF LEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQ HLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKV NSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRL RYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD (SEQ ID NO: 39)

[0365]

[0366] 본원에 기재된 다중-플랩 프라임 편집제는 임의의 상기 Cas9 오르토로그 서열, 또는 그와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 그의 임의의 변이체를 포함할 수 있다.

[0367] napDNAbp는 임의의 적합한 상동체 및/또는 오르토로그 또는 자연 발생 효소, 예컨대 Cas9를 포함할 수 있다. Cas9 상동체 및/또는 오르토로그는 에스. 피오게네스 및 에스. 써모필루스를 포함하나 이에 제한되지는 않는 다양한 종에서 기재되어 있다. 바람직하게는, Cas 모이어티는 닉카제로서 구성되고 (예를 들어, 돌연변이유발되거나, 재조합적으로 조작되거나, 또는 달리 자연으로부터 수득됨), 즉 표적 이중 가닥 DNA의 단일 가닥만을 절단할 수 있다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이고, 이러한 Cas9 뉴클레아제 및 서열은 그의 전체 내용이 본원에 참조로 포함되는 문헌 [Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737]에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다. 일부 실시양태에서, Cas9 뉴클레아제는 불활성 (예를 들어, 불활성화된) DNA 절단 도메인을 가지며, 즉 Cas9는 닉카제이다. 일부 실시양태에서, Cas9 단백질은 표 3의 변이체 중 어느 하나에 의해 제공된 바와 같은 Cas9 단백질의 아미노산 서열과 적어도 80% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 단백질은 상기 표에서 Cas9 오르토로그 중 어느 하나에 의해 제공된 바와 같은 Cas9 단백질의 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다.

[0368] C. 기능상실 Cas9 변이체

[0369] 특정 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제는 Cas9의 둘 다의 뉴클레아제 도메인, 즉 RuvC 도메인 (비-프로토스페이서 DNA 가닥을 절단함) 및 HNH 도메인 (프로토스페이서 DNA 가닥을 절단함)을 불활성화시키는 1개 이상의 돌연변이로 인해 뉴클레아제 활성을 갖지 않는 기능상실 Cas9, 예를 들어 기능상실 SpCas9를 포함할 수 있다. 뉴클레아제 불활성화는 코딩된 단백질 또는 그와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 그의 임의의 변이체의 아미노산 서열에서 1개 이상의 치환 및/또는 결실을 발생시키는 1개 이상의 돌연변이로 인한 것일 수 있다.

[0370] 본원에 사용된 용어 "dCas9"는 뉴클레아제-불활성 Cas9 또는 뉴클레아제-기능상실 Cas9, 또는 그의 기능적 단편을 지칭하고, 임의의 유기체로부터의 임의의 자연 발생 dCas9, 그의 임의의 자연 발생 dCas9 등가물 또는 기능적 단편, 임의의 유기체로부터의 임의의 dCas9 상동체, 오르토로그, 또는 파라로그, 및 자연 발생 또는 조작된 dCas9의 임의의 돌연변이체 또는 변이체를 포괄한다. 용어 dCas9는 특히 제한적인 것으로 의도되지 않으며, "dCas9 또는 등가물"로 지칭될 수 있다. 예시적인 dCas9 단백질 및 dCas9 단백질을 제조하는 방법은 본원에 추가로 기재되고/거나 관련 기술분야에 기재되어 있고, 본원에 참조로 포함된다.

[0371] 다른 실시양태에서, dCas9는 Cas9 뉴클레아제 활성을 불활성화시키는 1개 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 상응하거나, 또는 부분적으로 또는 전체적으로 그를 포함한다. 다른 실시양태에서, 내인성 Cas9 뉴클레아제 활성의 완전 또는 부분 불활성화를 초래할 수 있는, D10A 및 H840A 이외의 돌연변이를 갖는 Cas9 변이체가 제공된다 (예를 들어, 각각 nCas9 또는 dCas9). 이러한 돌연변이는, 예로서, D10 및 H840에서의 다른 아미노산 치환, 또는 야생형 서열, 예컨대 스트렙토코쿠스 피오게네스로부터의 Cas9 (NCBI 참조 서열: NC_017053.1)와 관련하여 Cas9의 뉴클레아제 도메인 내의 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인에서의 치환)을 포함한다. 일부 실시양태에서, Cas9의 변이체 또는 상동체 (예를 들어, 스트렙토코쿠스 피오게네스로부터의 Cas9의 변이체 (NCBI 참조 서열: NC_017053.1 (서열식별번호: 20)))가 제공되며, 이는 NCBI 참조 서열: NC_017053.1과 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나, 또는 적어도 약 99.9% 동일하다. 일부 실시양태에서, NC_017053.1 (서열식별번호: 20)보다 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30개 아미노산, 약 40개 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 아미노산 또는 그 초과만큼 더 짧거나 더 긴 아미노산 서열을 갖는 dCas9의 변이체 (예를 들어, NCBI 참조 서열: NC_017053.1 (서열식별번호: 20)의 변이체)가 제공된다.

[0372] 한 실시양태에서, 기능상실 Cas9는 Q99ZW2의 정규 SpCas9 서열을 기초로 할 수 있고, D10X 및 H810X (여기서 X는 임의의 아미노산일 수 있음) 치환 (밑줄 및 볼드체)을 포함하는 하기 서열을 가질 수 있거나, 또는 그에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 서열식별번호: 40의 변이체일 수 있다.

[0373] 한 실시양태에서, 기능상실 Cas9는 Q99ZW2의 정규 SpCas9 서열을 기초로 할 수 있고, D10A 및 H810A 치환 (밑줄 및 볼드체)을 포함하는 하기 서열을 가질 수 있거나, 또는 그에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 서열식별번호: 41의 변이체일 수 있다.

설명	서열	SEQ ID NO:
기능상실 Cas9 또는 dCas9 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, D10**X** 및 H810**X** 포함 여기서 "X"는 임의의 아미노산임	MDKKYSIGL**X**IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 40
기능상실 Cas9 또는 dCas9 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9,	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE	SEQ ID NO: 41

[0374]

설명	서열	SEQ ID NO:
D10A 및 H810A 포함	DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	

[0375]

[0376] D. Cas9 닉카제 변이체

[0377] 한 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제는 Cas9 닉카제를 포함한다. 용어 "Cas9 닉카제" 또는 "nCas9"는 이중 가닥 DNA 분자 표적에 단일 가닥 파괴를 도입할 수 있는 Cas9의 변이체를 지칭한다. 일부 실시양태에서, Cas9 닉카제는 단일 기능적 뉴클레아제 도메인만을 포함한다. 야생형 Cas9 (예를 들어, 정규 SpCas9)는 2개의 별개의 뉴클레아제 도메인, 즉 RuvC 도메인 (비-프로토스페이서 DNA 가닥을 절단함) 및 HNH 도메인 (프로토스페이서 DNA 가닥을 절단함)을 포함한다. 한 실시양태에서, Cas9 닉카제는 RuvC 뉴클레아제 활성을 불활성화시키는 RuvC 도메인 내의 돌연변이를 포함한다. 예를 들어, 아스파르테이트 (D) 10, 히스티딘 (H) 983, 아스파르테이트 (D) 986, 또는 글루타메이트 (E) 762에서의 돌연변이는 RuvC 뉴클레아제 도메인의 기능 상실 돌연변이 및 기능적 Cas9 닉카제의 생성으로서 보고되어 있다 (예를 들어, 문헌 [Nishimasu et al., "Crystal structure of Cas9 in complex with guide RNA and target DNA," Cell 156(5), 935-949], 이는 본원에 참조로 포함됨). 따라서, RuvC 도메인 내의 닉카제 돌연변이는 D10X, H983X, D986X, 또는 E762X를 포함할 수 있으며, 여기서 X는 야생형 아미노산 이외의 임의의 아미노산이다. 특정 실시양태에서, 닉카제는 D10A, 또는 H983A, 또는 D986A, 또는 E762A, 또는 그의 조합일 수 있다.

[0378] 다양한 실시양태에서, Cas9 닉카제는 RuvC 뉴클레아제 도메인에 돌연변이를 가질 수 있고, 하기 아미노산 서열 중 하나, 또는 그와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 아미노산 서열을 갖는 그의 변이체를 가질 수 있다.

설명	서열	SEQ ID NO:
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, D10**X** 포함, 여기서 X는 임의의 대안적 아미노산임	MDKKYSIGL**X**IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 42
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, E762X 포함, 여기서 X는 임의의	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN	SEQ ID NO: 43

[0379]

설명	서열	SEQ ID NO:
대안적 아미노산임	REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIXMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, H983X 포함, 여기서 X는 임의의 대안적 아미노산임	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD	SEQ ID NO: 44

[0380]

설명	서열	SEQ ID NO:
	ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH**X** AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, D986X 포함, 여기서 X는 임의의 대안적 아미노산임	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AH**X**AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 45
Cas9 닉카제	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI	SEQ ID NO: 46

[0381]

설명	서열	SEQ ID NO:
스트렙토코쿠스 피오게네스 Q99ZW2 Cas9, D10**A** 포함	CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, E762A 포함	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT	SEQ ID NO: 47

[0382]

설명	서열	SEQ ID NO:
	YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVI**A**MARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, H983A 포함	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH**A** AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV	SEQ ID NO: 48

[0383]

설명	서열	SEQ ID NO:
	QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, D986A 포함	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHAAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 49

[0384]

[0385] 또 다른 실시양태에서, Cas9 닉카제는 HNH 뉴클레아제 활성을 불활성화시키는 HNH 도메인 내의 돌연변이를 포함한다. 예를 들어, 히스티딘 (H) 840 또는 아스파라긴 (R) 863에서의 돌연변이는 HNH 뉴클레아제 도메인의 기능 상실 돌연변이 및 기능적 Cas9 닉카제의 생성으로서 보고되어 있다 (예를 들어, 문헌 [Nishimasu et al., "Crystal structure of Cas9 in complex with guide RNA and target DNA," Cell 156(5), 935-949], 이는 본원에 참조로 포함됨). 따라서, HNH 도메인 내의 닉카제 돌연변이는 H840X 및 R863X를 포함할 수 있으며, 여기서 X는 야생형 아미노산 이외의 임의의 아미노산이다. 특정 실시양태에서, 닉카제는 H840A 또는 R863A 또는 그의 조합일 수 있다.

[0386] 다양한 실시양태에서, Cas9 닉카제는 HNH 뉴클레아제 도메인에 돌연변이를 가질 수 있고, 하기 아미노산 서열 중 하나, 또는 그와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 아미노산 서열을 갖는 그의 변이체를 가질 수 있다.

설명	서열	SEQ ID NO:
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, H840**X** 포함, 여기서 X는 임의의 대안적 아미노산임	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 50
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, H840**A** 포함	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN	SEQ ID NO: 51

[0387]

설명	서열	SEQ ID NO:
	REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKN**R**GKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, R863X 포함, 여기서 X는 임의의 대안적 아미노산임	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN**X**GKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD	SEQ ID NO: 52

[0388]

설명	서열	SEQ ID NO:
	ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 닉카제 *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, R863**A** 포함	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN**A**GKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 53

[0389]

[0390] 일부 실시양태에서, N-말단 메티오닌이 Cas9 닉카제로부터, 또는 본원에 개시되거나 고려되는 임의의 Cas9 변이체, 오르토로그, 또는 등가물로부터 제거된다. 예를 들어, 메티오닌-마이너스 Cas9 닉카제는 하기 서열, 또는 그와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 아미노산 서열을 갖는 그의 변이체를 포함한다.

설명	서열
Cas9 닉카제 (Met 마이너스) *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, H840**X** 포함, 여기서 X는 임의의 대안적 아미노산임	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG LYETRIDLSQLGGD (SEQ ID NO: 54)
Cas9 닉카제 (Met 마이너스) *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, H840**A** 포함	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV

[0391]

	DQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG LYETRIDLSQLGGD (SEQ ID NO: 55)
Cas9 닉카제 (Met 마이너스) *스트렙토코쿠스 피오게네스* Q99ZW2 Cas9, R863X 포함, 여기서 X는 임의의 대안적 아미노산임	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNXGKSDNVPS EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG LYETRIDLSQLGGD(SEQ ID NO: 56)
Cas9 닉카제 (Met 마이너스) *스트렙토코쿠스 피오게네스*	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL

[0392]

Q99ZW2 Cas9, R863A 포함	VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNAGKSDNVPS EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG LYETRIDLSQLGGD (SEQ ID NO: 57)

[0393]

[0394] E. 다른 Cas9 변이체

[0395] 기능상실 Cas9 및 Cas9 닉카제 변이체 이외에도, 본원에 사용된 Cas9 단백질은 또한 임의의 야생형 Cas9, 또는 돌연변이체 Cas9 (예를 들어, 기능상실 Cas9 또는 Cas9 닉카제), 또는 단편 Cas9, 또는 원형 순열체 Cas9, 또는 본원에 개시되거나 관련 기술분야에 공지된 Cas9의 다른 변이체를 포함한 임의의 참조 Cas9 단백질과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일한 다른 "Cas9 변이체"를 포함할 수 있다. 일부 실시양태에서, Cas9 변이체는 참조 Cas9와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 또는 그 초과의 아미노산 변화를 가질 수 있다. 일부 실시양태에서, Cas9 변이체는 참조 Cas9의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함하며, 이러한 단편은 야생형 Cas9의 상응하는 단편에 대해 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 96% 동일하거나, 적어도 약 97% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나, 또는 적어도 약 99.9% 동일하다. 일부 실시양태에서, 단편은 상응하는 야생형 Cas9 (예를 들어, 서열식별번호: 18)의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%이다.

[0396] 일부 실시양태에서, 개시내용은 또한, 그의 기능성을 보유하고 임의의 본원에 개시된 Cas9 단백질의 단편인 Cas9 단편을 활용할 수 있다. 일부 실시양태에서, Cas9 단편은 적어도 100개 아미노산 길이이다. 일부 실시양태에서, 단편은 적어도 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250개, 또는 적어도 1300개 아미노산 길이이다.

[0397] 다양한 실시양태에서, 본원에 개시된 다중-플랩 프라임 편집제는 하기와 같이 기재된 Cas9 변이체 중 하나, 또는 임의의 참조 Cas9 변이체와 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일한 그의 Cas9 변이체를 포함할 수 있다.

[0398] F. 작은-크기의 Cas9 변이체

[0399] 일부 실시양태에서, 본원에 고려되는 다중-플랩 프라임 편집제는 정규 SpCas9 서열보다 더 작은 분자량의 Cas9 단백질을 포함할 수 있다. 일부 실시양태에서, 보다 작은-크기의 Cas9 변이체는 예를 들어 발현 벡터, 나노입자, 또는 다른 전달 수단에 의한 세포로의 전달을 용이하게 할 수 있다. 특정 실시양태에서, 보다 작은-크기의 Cas9 변이체는 부류 2 CRISPR-Cas 시스템의 유형 II 효소로서 카테고리화된 효소를 포함할 수 있다. 일부 실시양태에서, 보다 작은-크기의 Cas9 변이체는 부류 2 CRISPR-Cas 시스템의 유형 V 효소로서 카테고리화된 효소를 포함할 수 있다. 다른 실시양태에서, 보다 작은-크기의 Cas9 변이체는 부류 2 CRISPR-Cas 시스템의 유형 VI 효

소로서 카테고리화된 효소를 포함할 수 있다.

[0400] 정규 SpCas9 단백질은 1368개 아미노산 길이이고, 158 킬로달톤의 예측 분자량을 갖는다. 본원에 사용된 용어 "작은-크기의 Cas9 변이체"는 적어도 1300개 미만의 아미노산, 또는 적어도 1290개 미만의 아미노산, 또는 1280개 미만의 아미노산, 또는 1270개 미만의 아미노산, 또는 1260개 미만의 아미노산, 또는 1250개 미만의 아미노산, 또는 1240개 미만의 아미노산, 또는 1230개 미만의 아미노산, 또는 1220개 미만의 아미노산, 또는 1210개 미만의 아미노산, 또는 1200개 미만의 아미노산, 또는 1190개 미만의 아미노산, 또는 1180개 미만의 아미노산, 또는 1170개 미만의 아미노산, 또는 1160개 미만의 아미노산, 또는 1150개 미만의 아미노산, 또는 1140개 미만의 아미노산, 또는 1130개 미만의 아미노산, 또는 1120개 미만의 아미노산, 또는 1110개 미만의 아미노산, 또는 1100개 미만의 아미노산, 또는 1050개 미만의 아미노산, 또는 1000개 미만의 아미노산, 또는 950개 미만의 아미노산, 또는 900개 미만의 아미노산, 또는 850개 미만의 아미노산, 또는 800개 미만의 아미노산, 또는 750개 미만의 아미노산, 또는 700개 미만의 아미노산, 또는 650개 미만의 아미노산, 또는 600개 미만의 아미노산, 또는 550개 미만의 아미노산, 또는 500개 미만의 아미노산이지만, 적어도 약 400개 초과의 아미노산이고, Cas9 단백질의 요구되는 기능을 보유하는 임의의 Cas9 변이체-자연 발생, 조작, 또는 그 밖의 것-를 지칭한다. Cas9 변이체는 부류 2 CRISPR-Cas 시스템의 유형 II, 유형 V, 또는 유형 VI 효소로서 카테고리화된 것을 포함할 수 있다.

[0401] 다양한 실시양태에서, 본원에 개시된 다중-플랩 프라임 편집제는 하기와 같이 기재된 작은-크기의 Cas9 변이체 중 하나, 또는 임의의 참조 작은-크기의 Cas9 단백질과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일한 그의 Cas9 변이체를 포함할 수 있다.

설명	서열	SEQ ID NO:
SaCas9 *스타필로코쿠스 아우레우스* 1053 AA 123 kDa	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEA NVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLL TDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLE RLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLD QSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEML MGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDE NEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIA KILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHN LSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQK EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL ARENKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENA KYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDH IIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDS KISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSV QKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSI NGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIF KEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEI FITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRK DDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYH HDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLS LKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKC YEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNN DLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSI KKYSTDILGNLYEVKSKKHPQIIKK	SEQ ID NO: 58
NmeCas9 *엔. 메닌기티디스* 1083 AA 124.5 kDa	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDL GVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRL LRTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAAL DRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGA LLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQR SDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIE TLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTA ERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSK LTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKA YHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTD EDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPL MEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIR NPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSF KDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKD ILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDAALPF SRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSR EWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLND TRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLR GFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRF	SEQ ID NO: 59

[0402]

설명	서열	SEQ ID NO:
	VRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQE VMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVH EYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVSVL RVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDD PAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWV RNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGI LPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITK KARMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGV KTALSFQKYQIDELGKEIRPCRLKKRPPVR	
CjCas9 *씨. 제주니* 984 AA 114.9 kDa	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKT GESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLN YEDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFA RVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLAN YQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQ SFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALK DFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNN LKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLS DDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIA KDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNIS FKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLP AFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVH KINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECE KLGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEK MLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFE AFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQK NFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLN DTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLH HAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISEL DYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGAL HEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVK NGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAV ARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEP EFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANE KEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK	SEQ ID NO: 60
GeoCas9 *쥐.* *스테아로써* *모필루스* 1087 AA 127 kDa	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENP QTGESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILT KEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLL HLAKRRGFKSNRKSERSNKENSTMLKHIEENRAILSSYRT VGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFS KQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGF CTFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLT DEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYD RGESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFL PIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLA NKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEV YSSACERAGYTFTGPKKKQKTMLLPNIPPIANPVVMRAL TQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKK EQDENRKKNETAIRQLMEYGLTLNPTGHDIVKFKLWSEQ	SEQ ID NO: 61

[0403]

설명	서열	SEQ ID NO:
	NGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTN KVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNK QFSKKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFA NFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKN REESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKEL AKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNY DDQKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDER SGKIQTVVKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQ RLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKN QVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMD IMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRI ELPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISH DHRFSLRGVGSRTLKRFEKYQVDVLGNIYKVRGEKRVG LASSAHSKPGKTIRPLQSTRD	
LbaCas12a *엘. 박테리아* 1228 AA 143.9 kDa	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVED EKRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYIS LFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLF KKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNREN MFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHE VQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIG GFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVL SDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLE KLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKW NAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQ EYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFV LEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKET NRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKD KFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAI MDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPK VFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCH KLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREV EEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDK SHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASL KKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFS EDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGI DRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTD YHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKI CELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKM LIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISS FDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYG NRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQ GDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDV DFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNI ARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTS VKH	SEQ ID NO: 62
BhCas12b	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILK LIRQEAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQK	SEQ ID NO: 63

[0404]

설명	서열	SEQ ID NO:
비. 허사시이 1108 AA 130.4 kDa	CNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEE EKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPI VKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNL KVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQE QLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSE KYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNH PEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVR FEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTE SGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYK DESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRI YFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEW IKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEV VDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKS REVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREK RVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAF LKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNID EIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLN ALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPAC QIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQ GEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQ DNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLS KDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAY QVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVN AGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLM LYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSIS TIEDDSSKQSM	

[0405]

[0406] G. Cas9 등가물

[0407] 일부 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제는 임의의 Cas9 등가물을 포함할 수 있다. 본원에 사용된 용어 "Cas9 등가물"은 그의 아미노산 1차 서열 및/또는 그의 3차원 구조가 진화상 관점에서 상이할 수 있고/거나 관련되지 않을 수 있음에도 불구하고 본 발명의 다중-플랩 프라임 편집제에서 Cas9와 동일한 기능을 제공하는 임의의 napDNAbp 단백질을 포괄하는 광범위한 용어이다. 따라서, Cas9 등가물은 진화적으로 관련된, 본원에 기재되거나 또는 본원에 포괄되는 임의의 Cas9 오르토로그, 상동체, 돌연변이체, 또는 변이체를 포함하는 한편, Cas9 등가물은 또한 Cas9와 동일하거나 유사한 기능을 갖도록 수렴 진화 과정을 통해 진화될 수 있지만, 아미노산 서열 및/또는 3차원 구조와 관련하여 임의의 유사성을 반드시 갖는 것은 아닌 단백질을 포괄한다. 본원에 기재된 다중-플랩 프라임 편집제는 Cas9 등가물이 수렴 진화를 통해 발생한 단백질에 기초할 수 있음에도 불구하고 Cas9와 동일하거나 유사한 기능을 제공할 임의의 Cas9 등가물을 포괄한다. 예를 들어, Cas9가 CRISPR-Cas 시스템의 유형 II 효소를 지칭하는 경우에, Cas9 등가물은 CRISPR-Cas 시스템의 유형 V 또는 유형 VI 효소를 지칭할 수 있다.

[0408] 예를 들어, Cas12e (CasX)는 보고된 바로는 Cas9와 동일한 기능을 갖지만 수렴 진화를 통해 진화된 Cas9 등가물이다. 따라서, 문헌 [Liu et al., "CasX enzymes comprises a distinct family of RNA-guided genome editors," Nature, 2019, Vol.566: 218-223]에 기재된 Cas12e (CasX) 단백질은 본원에 기재된 다중-플랩 프라임 편집제와 함께 사용되는 것으로 고려된다. 또한, Cas12e (CasX)의 임의의 변이체 또는 변형이 고려될 수 있고, 본 개시내용의 범주 내에 있다.

[0409] Cas9는 매우 다양한 종에서 진화된 박테리아 효소이다. 그러나, 본원에 고려되는 Cas9 등가물은 또한 박테리아와 상이한 단일-세포 원핵 미생물의 역 및 계를 구성하는 고세균으로부터 수득될 수 있다.

[0410] 일부 실시양태에서, Cas9 등가물은 Cas12e (CasX) 또는 Cas12d (CasY)를 지칭할 수 있으며, 이는 예를 들어 그의 전체 내용이 본원에 참조로 포함되는 문헌 [Burstein et al., "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21]에 기재되어 있다. 게놈-분해 메타게놈학을 사용하여, 생물 고세균 역에서 최초로 보고된 Cas9를 포함한 다수의 CRISPR-Cas 시스템이 확인되었다. 이러한 분기성 Cas9 단백질은 거의 연구되지 않은 나노고세균에서 활성 CRISPR-Cas 시스템의 일부로서 발견되었다. 박테리아에서, 2종의 이전에 알려지지 않은 시스템, CRISPR-Cas12e 및 CRISPR-Cas12d가 발견되었으며, 이들은 지금까지 발견된 가장 치밀한 시스템 중의 것이다. 일부 실시양태에서, Cas9는 Cas12e, 또는 Cas12e의 변이체를 지칭한다. 일부 실시양태에서, Cas9는 Cas12d, 또는 Cas12d의 변이체를 지칭한다. 다른 RNA-가이드된 DNA 결합

단백질이 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)로서 사용될 수 있고, 이는 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다. 또한, 문헌 [Liu et al., "CasX enzymes comprises a distinct family of RNA-guided genome editors," Nature, 2019, Vol.566: 218-223]을 참조한다. 이들 Cas9 등가물 중 임의의 것이 고려된다.

[0411] 일부 실시양태에서, Cas9 등가물은 자연 발생 Cas12e (CasX) 또는 Cas12d (CasY) 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, napDNAbp는 자연 발생 Cas12e (CasX) 또는 Cas12d (CasY) 단백질이다. 일부 실시양태에서, napDNAbp는 야생형 Cas 모이어티 또는 본원에 제공된 임의의 Cas 모이어티와 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다.

[0412] 다양한 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질은 비제한적으로 Cas9 (예를 들어, dCas9 및 nCas9), Cas12e (CasX), Cas12d (CasY), Cas12a (Cpf1), Cas12b1 (C2c1), Cas13a (C2c2), Cas12c (C2c3), 아르고노트, 및 Cas12b1을 포함한다. Cas9와 상이한 PAM 특이성을 갖는 핵산 프로그램가능한 DNA-결합 단백질의 한 예는 프레보텔라(Prevotella) 및 프란시셀라(Francisella) 1로부터의 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (즉, Cas12a (Cpf1))이다. Cas9와 유사하게, Cas12a (Cpf1)도 또한 부류 2 CRISPR 이펙터이지만, 이는 유형 II 하위군보다는 효소의 유형 V 하위군의 구성원이다. Cas12a (Cpf1)는 강건한 DNA 간섭을 매개하는 것으로 제시되어 있고, 특색이 Cas9와 구별된다. Cas12a (Cpf1)는 tracrRNA가 결여된 단일 RNA-가이드된 엔도뉴클레아제이고, 이는 T-풍부 프로토스페이서-인접 모티프 (TTN, TTTN, 또는 YTN)를 이용한다. 또한, Cpf1은 엇갈린 DNA 이중 가닥 파괴를 통해 DNA를 절단한다. 16종의 Cpf1-패밀리 단백질 중에서, 아시다미노코쿠스(*Acidaminococcus*) 및 라크노스피라세아에(*Lachnospiraceae*)로부터의 2종의 효소는 인간 세포에서 효율적인 게놈-편집 활성을 갖는 것으로 제시된다. Cpf1 단백질은 관련 기술분야에 공지되어 있고, 이전에, 예를 들어 문헌 [Yamano et al., "Crystal structure of Cpf1 in complex with guide RNA and target DNA." Cell (165) 2016, p. 949-962]에 기재되었으며; 그의 전체 내용은 본원에 참조로 포함된다.

[0413] 또 다른 실시양태에서, Cas 단백질은 Cas12a, Cas12b1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (또한 Csn1 및 Csx12로도 공지됨), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, 그의 상동체, 또는 그의 변형된 버전을 포함하나 이에 제한되지는 않고, 바람직하게는 닉카제 돌연변이 (예를 들어, 서열식별번호: 18의 야생형 Cas9 폴리펩티드의 D10A 돌연변이에 상응하는 돌연변이)를 포함하는 임의의 CRISPR 연관 단백질을 포함할 수 있다.

[0414] 다양한 다른 실시양태에서, napDNAbp는 하기 단백질 중 임의의 것일 수 있다: Cas9, Cas12a (Cpf1), Cas12e (CasX), Cas12d (CasY), Cas12b1 (C2c1), Cas13a (C2c2), Cas12c (C2c3), GeoCas9, CjCas9, Cas12g, Cas12h, Cas12i, Cas13b, Cas13c, Cas13d, Cas14, Csn2, xCas9, SpCas9-NG, 원형 순열 Cas9, 또는 아르고노트 (Ago) 도메인, 또는 그의 변이체.

[0415] 예시적인 Cas9 등가물 단백질 서열은 하기를 포함할 수 있다:

설명	서열
AsCas12a (이전에 Cpf1로 공지됨) *아시다미 노코쿠스 종(균주 BV3L6)* 유니프롯KB U2UMQ6	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKEL KPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQA TYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTV TTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPK FKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQ TQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIP LFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNE LNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKE KVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTL KKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEM EPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAI LFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAK MIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKF QTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQY KDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKG HHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHR LGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVI TKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP ETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERV AARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFK SKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQ FTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFL EGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQF DAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSN ILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGV CFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISN QDWLAYIQELRN (SEQ ID NO: 64)
AsCas12a 닉카제 (예를 들어, R1226A)	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKEL KPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQA TYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTV TTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPK FKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQ TQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIP LFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNE LNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKE KVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTL KKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEM EPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAI LFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAK MIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKF QTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQY KDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKG HHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHR LGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVI TKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP

[0416]

	ETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERV AARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFK SKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQ FTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFL EGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQF DAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSN ILPKLLENDDSHAIDTMVALIRSVLQMANSNAATGEDYINSPVRDLNGV CFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISN QDWLAYIQELRN (SEQ ID NO: 65)
LbCas12a (이전에 Cpf1로 공지됨) *라크노스피라세아에 박테리움 GAM79* Ref Seq. WP_119623382.1	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQ QELKEIMDDYYRTFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKI QNEKRKEICCYFTSDKRFKDLFNAKLITDILPNFIKDNKEYTEEEKAEKE QTRVLFQRFATAFTNYFNQRRNNFSEDNISTAISFRIVNENSEIHLQNMR AFQRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYSVDFYDRELTQPGIE YYNGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFR FESDQEVYDALNEFIKTMKKKEIIRRCVHLGQECDDYDLGKIYISSNKYE QISNALYGSWDTIRKCIKEEYMDALPGKGEKKEEKAEAAAKKEEYRSIA DIDKIISLYGSEMDRTISAKKCITEICDMAGQISIDPLVCNSDIKLLQNKEK TTEIKTILDSFLHVYQWGQTFIVSDIIEKDSYFYSELEDVLEDFEGITTLYN HVRSYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAILLMRDQKF YLGIFNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITSR SGQETYKPSKHILDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHPD WKNYDFHFSDTKDYEDISGFYREVEMQGYQIKWTYISADEIQKLDEKG QIFLFQIYNKDFSVHSTGKDNLHTMYLKNLFSEENLKDIVLKLNGEAELF FRKASIKTPIVHKKGSVLVNRSYTQTVGNKEIRVSIPEEYYTEIYNYLNHI GKGKLSSEAQRYLDEGKIKSFTATKDIVKNYRYCCDHYFLHLPITINFKA KSDVAVNERTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSFN IVNGYDYQQKLKDREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQL VVKYNAVVAMEDLNYGFKTGRFKVERQVYQKFETMLIEKLNYLVFKD REVCEEGGVLRGYQLTYIPESLKKVGKQCGFIFYVPAGYTSKIDPTTGFV NLFSFKNLTNRESRQDFVGKFDEIRYDRDKKMFEFSFDYNNYIKKGTILA STKWKVYTNGTRLKRIVVNGKYTSQSMEVELTDAMEKMLQRAGIEYH DGKDLKGQIVEKGIEAEIIDIFRLTVQMRNSRSESEDREYDRLISPVLNDK GEFFDTATADKTLPQDADANGAYCIALKGLYEVKQIKENWKENEQFPR NKLVQDNKTWFDFMQKKRYL (SEQ ID NO: 66)
PcCas12a - 이전에 Cpf1로 공지됨 *프레보텔라 코프리* Ref Seq. WP_119227726.1	MAKNFEDFKRLYSLSKTLRFEAKPIGATLDNIVKSGLLDEDEHRAASYV KVKKLIDEYHKVFIDRVLDDGCLPLENKGNNNSLAEYYESYVSRAQDE DAKKKFKEIQQNLRSVIAKKLTEDKAYANLFGNKLIESYKDKEDKKKII DSDLIQFINTAESTQLDSMSQDEAKELVKEFWGFVTYFYGFFDNRKNMY TAEEKSTGIAYRLVNENLPKFIDNIEAFNRAITRPEIQENMGVLYSDFSEY LNVESIQEMFQLDYYNMLLTQKQIDVYNAIIGGKTDDEHDVKIKGINEYI NLYNQQHKDDKLPKLKALFKQILSDRNAISWLPEEFNSDQEVLNAIKDC YERLAENVLGDKVLKSLLGSLADYSLDGIFIRNDLQLTDISQKMFGNWG VIQNAIMQNIKRVAPARKHKESEEDYEKRIAGIFKKADSFSISYINDCLNE ADPNNAYFVENYFATFGAVNTPTMQRENLFALVQNAYTEVAALLHSDY PTVKHLAQDKANVSKIKALLDAIKSLQHFVKPLLGKGDESDKDERFYGE LASLWAELDTVTPLYNMIRNYMTRKPYSQKKIKLNFENPQLLGGWDAN KEKDYATIILRRNGLYYLAIMDKDSRKLLGKAMPSDGECYEKMVYKFF KDVTTMIPKCSTQLKDVQAYFKVNTDDYVLNSKAFNKPLTITKEVFDLN

[0417]

	NVLYGKYKKFQKGYLTATGDNVGYTHAVNVWIKFCMDFLNSYDSTCI YDFSSLKPESYLSLDAFYQDANLLLYKLSFARASVSYINQLVEEGKMYL FQIYNKDFSEYSKGTPNMHTLYWKALFDERNLADVVYKLNGQAEMFY RKKSIENTHPTHPANHPILNKNKDNKKKESLFDYDLIKDRRYTVDKFMF HVPITMNFKSVGSENINQDVKAYLRHADDMHIIGIDRGERHLLYLVVIDL QGNIKEQYSLNEIVNEYNGNTYHTNYHDLLDVREEERLKARQSWQTIEN IKELKEGYLSQVIHKITQLMVRYHAIVVLEDLSKGFMRSRQKVEKQVYQ KFEKMLIDKLNYLVDKKTDVSTPGGLLNAYQLTCKSDSSQKLGKQSGF LFYIPAWNTSKIDPVTGFVNLLDTHSLNSKEKIKAFFSKFDAIRYNKDKK WFEFNLDYDKFGKKAEDTRTKWTLCTRGMRIDTFRNKEKNSQWDNQE VDLTTEMKSLLEHYYIDIHGNLKDAISAQTDKAFFTGLLHILKLTLQMRN SITGTETDYLVSPVADENGIFYDSRSCGNQLPENADANGAYNIARKGLM LIEQIKNAEDLNNVKFDISNKAWLNFAQQKPYKNG (SEQ ID NO: 67)
ErCas12a - 이전에 Cpf1로 공지됨 *유박테리움 렉탈래* Ref Seq. WP_119223642.1	MFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKDYFKNRA NCFSANDISSSSCHRIVNDNAEIFFSNALVYRRIVKNLSNDDINKISGDMK DSLKEMSLEEIYSYEKYGEFITQEGISFYNDICGKVNLFMNLYCQKNKEN KNLYKLRKLHKQILCIADTSYEVPYKFESDEEVYQSVNGFLDNISSKHIV ERLRKIGENYNGYNLDKIYIVSKFYESVSQKTYRDWETINTALEIHYNNI LPGNGKSKADKVKKAVKNDLQKSITEINELVSNYKLCPDDNIKAETYIH EISHILNNFEAQELKYNPEIHLVESELKASELKNVLDVIMNAFHWCSVFM TEELVDKDNNFYAELEEIYDEIYPVISLYNLVRNYVTQKPYSTKKIKLNF GIPTLADGWSKSKEYSNNAIILMRDNLYYLGIFNAKNKPDKKIIEGNTSE NKGDYKKMIYNLLPGPNKMIPKVFLSSKTGVETYKPSAYILEGYKQNKH LKSSKDFDITFCHDLIDYFKNCIAIHPEWKNFGFDFSDTSTYEDISGFYRE VELQGYKIDWTYISEKDIDLLQEKGQLYLFQIYNKDFSKKSSGNDNLHT MYLKNLFSEENLKDIVLKLNGEAEIFFRKSSIKNPIIHKKGSILVNRTYEA EEKDQFGNIQIVRKTIPENIYQELYKYFNDKSDKELSDEAAKLKNVVGH HEAATNIVKDYRYTYDKYFLHMPITINFKANKTSFINDRILQYIAKEKDL HVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDYQIKLKQQEGARQI ARKEWKEIGKIKEIKEGYLSLVIHEISKMVIKYNAIIAMEDLSYGFKKGRF KVERQVYQKFETMLINKLNYLVFKDISITENGGLLKGYQLTYIPDKLKN VGHQCGCIFYVPAAYTSKIDPTTGFVNIFKFKDLTVDAKREFIKKFDSIRY DSDKNLFCFTFDYNNFITQNTVMSKSSWSVYTYGVRIKRRFVNGRFSNE SDTIDITKDMEKTLEMTDINWRDGHDLRQDIIDYEIVQHIFEIFKLTVQM RNSLSELEDRDYDRLISPVLNENNIFYDSAKAGDALPKDADANGAYCIA LKGLYEIKQITENWKEDGKFSRDKLKISNKDWFDFIQNKRYL (SEQ ID NO: 68)
CsCas12a - 이전에 Cpf1로 공지됨 *클로스트리디움 종 AF34-10BH*	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQ QELKEIMDDYYRAFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKI QNEKRKEICCYFTSDKRFKDLFNAKLITDILPNFIKDNKEYTEEEKAEKE QTRVLFQRFATAFTNYFNQRRNNFSEDNISTAISFRIVNENSEIHLQNMR AFQRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYLVDFYDRVLTQPGIE YYNGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFR FESDQEVYDALNEFIKTMKEKEIICRCVHLGQKCDDYDLGKIYISSNKYE QISNALYGSWDTIRKCIKEEYMDALPGKGEKKEEKAEAAAKKEEYRSIA DIDKIISLYGSEMDRTISAKKCITEICDMAGQISTDPLVCNSDIKLLQNKE KTTEIKTILDSFLHVYQWGQTFIVSDIIEKDSYFYSELEDVLEDFEGITTLY NHVRSYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAILLMRDQK FYLGIFNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITS

[0418]

Ref Seq. WP_118538418.1	RSGQETYKPSKHILDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHP DWKNYDFHFSDTKDYEDISGFYREVEMQGYQIKWTYISADEIQKLDEK GQIFLFQIYNKDFSVHSTGKDNLHTMYLKNLFSEENLKDIVLKLNGEAEL FFRKASIKTPVVHKKGSVLVNRSYTQTVGDKEIRVSIPEEYYTEIYNYLN HIGRGKLSTEAQRYLEERKIKSFTATKDIVKNYRYCCDHYFLHLPITINFK AKSDIAVNERTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSF NIVNGYDYQQKLKDREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQ LVVKYNAVVAMEDLNYGFKTGRFKVERQVYQKFETMLIEKLHYLVFK DREVCEEGGVLRGYQLTYIPESLKKVGKQCGFIFYVPAGYTSKIDPTTGF VNLFSFKNLTNRESRQDFVGKFDEIRYDRDKKMFEFSFDYNNYIKKGTM LASTKWKVYTNGTRLKRIVVNGKYTSQSMEVELTDAMEKMLQRAGIE YHDGKDLKGQIVEKGIEAEIIDIFRLTVQMRNSRSESEDREYDRLISPVLN DKGEFFDTATADKTLPQDADANGAYCIALKGLYEVKQIKENWKENEQF PRNKLVQDNKTWFDFMQKKRYL (SEQ ID NO: 69)
BhCas12b *바실루스 히사시이* Ref Seq. WP_095142515.1	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHH EQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILREL YEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNL KIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPI VKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVE KEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLR GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKK ENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRF EERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKV DIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDR DHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNF KPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVD QKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNL KLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDEL IQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLY GISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALK EDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPY EERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKT GSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGG EKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDG QTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSS SELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFG KLERILISKLTNQYSISTIEDDSSKQSM (SEQ ID NO: 70)
ThCas12b *테르모모나스 하드로테르말리스* Ref Seq. WP_072754838	MSEKTTQRAYTLRLNRASGECAVCQNNSCDCWHDALWATHKAVNRG AKAFGDWLLTLRGGLCHTLVEMEVPAKGNNPPQRPTDQERRDRRVLLA LSWLSVEDEHGAPKEFIVATGRDSADDRAKKVEEKLREILEKRDFQEHEI DAWLQDCGPSLKAHIREDAVWVNRRALFDAAVERIKTLTWEEAWDFL EPFFGTQYFAGIGDGKDKDDAEGPARQGEKAKDLVQKAGQWLSARFGI GTGADFMSMAEAYEKIAKWASQAQNGDNGKATIEKLACALRPSEPPTL DTVLKCISGPGHKSATREYLKTLDKKSTVTQEDLNQLRKLADEDARNC RKKVGKKGKKPWADEVLKDVENSCELTYLQDNSPARHREFSVMLDHA ARRVSMAHSWIKKAEQRRRQFESDAQKLKNLQERAPSAVEWLDRFCES RSMTTGANTGSGYRIRKRAIEGWSYVVQAWAEASCDTEDKRIAAARKV QADPEIEKFGDIQLFEALAADEAICVWRDQEGTQNPSILIDYVTGKTAEH NQKRFKVPAYRHPDELRHPVFCDFGNSRWSIQFAIHKEIRDRDKGAKQD

[0419]

	TRQLQNRHGLKMRLWNGRSMTDVNLHWSSKRLTADLALDQNPNPNPT EVTRADRLGRAASSAFDHVKIKNVFNEKEWNGRLQAPRAELDRIAKLE EQGKTEQAEKLRKRLRWYVSFSPCLSPSGPFIVYAGQHNIQPKRSGQYA PHAQANKGRARLAQLILSRLPDLRILSVDLGHRFAAACAVWETLSSDAF RREIQGLNVLAGGSGEGDLFLHVEMTGDDGKRRTVVYRRIGPDQLLDN TPHPAPWARLDRQFLIKLQGEDEGVREASNEELWTVHKLEVEVGRTVP LIDRMVRSGFGKTEKQKERLKKLRELGWISAMPNEPSAETDEKEGEIRSI SRSVDELMSSALGTLRLALKRHGNRARIAFAMTADYKPMPGGQKYYFH EAKEASKNDDETKRRDNQIEFLQDALSLWHDLFSSPDWEDNEAKKLWQ NHIATLPNYQTPEEISAELKRVERNKKRKENRDKLRTAAKALAENDQLR QHLHDTWKERWESDDQQWKERLRSLKDWIFPRGKAEDNPSIRHVGGLS ITRINTISGLYQILKAFKMRPEPDDLRKNIPQKGDDELENFNRRLLEARDR LREQRVKQLASRIIEAALGVGRIKIPKNGKLPKRPRTTVDTPCHAVVIESL KTYRPDDLRTRRENRQLMQWSSAKVRKYLKEGCELYGLHFLEVPANYT SRQCSRTGLPGIRCDDVPTGDFLKAPWWRRAINTAREKNGGDAKDRFL VDLYDHLNNLQSKGEALPATVRVPRQGGNLFIAGAQLDDTNKERRAIQ ADLNAAANIGLRALLDPDWRGRWWYVPCKDGTSEPALDRIEGSTAFND VRSLPTGDNSSRRAPREIENLWRDPSGDSLESGTWSPTRAYWDTVQSRV IELLRRHAGLPTS (SEQ ID NO: 71)
LsCas12b *라세엘라 사카리* WP_132221894.1	MSIRSFKLKLKTKSGVNAEQLRRGLWRTHQLINDGIAYYMNWLVLLRQ EDLFIRNKETNEIEKRSKEEIQAVLLERVHKQQQRNQWSGEVDEQTLLQ ALRQLYEEIVPSVIGKSGNASLKARFFLGPLVDPNNKTTKDVSKSGPTPK WKKMKDAGDPNWVQEYEKYMAERQTLVRLEEMGLIPLFPMYTDEVG DIHWLPQASGYTRTWDRDMFQQAIERLLSWESWNRRVRERRAQFEKKT HDFASRFSESDVQWMNKLREYEAQQEKSLEENAFAPNEPYALTKKALR GWERVYHSWMRLDSAASEEAYWQEVATCQTAMRGEFGDPAIYQFLAQ KENHDIWRGYPERVIDFAELNHLQRELRRAKEDATFTLPDSVDHPLWVR YEAPGGTNIHGYDLVQDTKRNLTLILDKFILPDENGSWHEVKKVPFSLA KSKQFHRQVWLQEEQKQKKREVVFYDYSTNLPHLGTLAGAKLQWDRN FLNKRTQQQIEETGEIGKVFFNISVDVRPAVEVKNGRLQNGLGKALTVL THPDGTKIVTGWKAEQLEKWVGESGRVSSLGLDSLSEGLRVMSIDLGQ RTSATVSVFEITKEAPDNPYKFFYQLEGTEMFAVHQRSFLLALPGENPPQ KIKQMREIRWKERNRIKQQVDQLSAILRLHKKVNEDERIQAIDKLLQKV ASWQLNEEIATAWNQALSQLYSKAKENDLQWNQAIKNAHHQLEPVVG KQISLWRKDLSTGRQGIAGLSLWSIEELEATKKLLTRWSKRSREPGVVK RIERFETFAKQIQHHINQVKENRLKQLANLIVMTALGYKYDQEQKKWIE VYPACQVVLFENLRSYRFSFERSRRENKKLMEWSHRSIPKLVQMQGELF GLQVADVYAAYSSRYHGRTGAPGIRCHALTEADLRNETNIIHELIEAGFI KEEHRPYLQQGDLVPWSGGELFATLQKPYDNPRILTLHADINAAQNIQK RFWHPSMWFRVNCESVMEGEIVTYVPKNKTVHKKQGKTFRFVKVEGS DVYEWAKWSKNRNKNTFSSITERKPPSSMILFRDPSGTFFKEQEWVEQK TFWGKVQSMIQAYMKKTIVQRMEE (SEQ ID NO: 72)
DtCas12b *데술포나 트로붐 티오디스 뮤탄스*	MVLGRKDDTAELRRALWTTHEHVNLAVAEVERVLLRCRGRSYWTLDR RGDPVHVPESQVAEDALAMAREAQRRNGWPVVGEDEEILLALRYLYEQ IVPSCLLDDLGKPLKGDAQKIGTNYAGPLFDSDTCRRDEGKDVACCGPF HEVAGKYLGALPEWATPISKQEFDGKDASHLRFKATGGDDAFFRVSIEK ANAWYEDPANQDALKNKAYNKDDWKKEKDKGISSWAVKYIQKQLQL GQDPRTEVRRKLWLELGLLPLFIPVFDKTMVGNLWNRLAVRLALAHLL SWESWNHRAVQDQALARAKRDELAALFLGMEDGFAGLREYELRRNESI

[0420]

WP_031386437	KQHAFEPVDRPYVVSGRALRSWTRVREEWLRHGDTQESRKNICNRLQD RLRGKFGDPDVFHWLAEDGQEALWKERDCVTSFSLLNDADGLLEKRK GYALMTFADARLHPRWAMYEAPGGSNLRTYQIRKTENGLWADVVLLS PRNESAAVEEKTFNVRLAPSGQLSNVSFDQIQKGSKMVGRCRYQSANQ QFEGLLGGAEILFDRKRIANEQHGATDLASKPGHVWFKLTLDVRPQAPQ GWLDGKGRPALPPEAKHFKTALSNKSKFADQVRPGLRVLSVDLGVRSF AACSVFELVRGGPDQGTYFPAADGRTVDDPEKLWAKHERSFKITLPGEN PSRKEEIARRAAMEELRSLNGDIRRLKAILRLSVLQEDDPRTEHLRLFME AIVDDPAKSALNAELFKGFGDDRFRSTPDLWKQHCHFFHDKAEKVVAE RFSRWRTETRPKSSSWQDWRERRGYAGGKSYWAVTYLEAVRGLILRW NMRGRTYGEVNRQDKKQFGTVASALLHHINQLKEDRIKTGADMIIQAA RGFVPRKNGAGWVQVHEPCRLILFEDLARYRFRTDRSRRENSRLMRWS HREIVNEVGMQGELYGLHVDTTEAGFSSRYLASSGAPGVRCRHLVEEDF HDGLPGMHLVGELDWLLPKDKDRTANEARRLLGGMVRPGMLVPWDG GELFATLNAASQLHVIHADINAAQNLQRRFWGRCGEAIRIVCNQLSVDG STRYEMAKAPKARLLGALQQLKNGDAPFHLTSIPNSQKPENSYVMTPTN AGKKYRAGPGEKSSGEEDELALDIVEQAEELAQGRKTFFRDPSGVFFAP DRWLPSEIYWSRIRRRIWQVTLERNSSGRQERAEMDEMPY (SEQ ID NO: 73)

[0421]

[0422] 본원에 기재된 다중-플랩 프라임 편집제는 또한 가이드 뉴클레오티드 서열-프로그램가능한 DNA-결합 단백질 도메인으로서 사용될 수 있는 Cas12a (Cpf1) (dCpf1) 변이체를 포함할 수 있다. Cas12a (Cpf1) 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레아제 도메인을 갖지만, HNH 엔도뉴클레아제 도메인은 갖지 않고, Cas12a (Cpf1)의 N-말단은 Cas9의 알파-나선 인식 로브를 갖지 않는다. Cas12a (Cpf1)의 RuvC-유사 도메인이 둘 다의 DNA 가닥을 절단하는 것을 담당하고, RuvC-유사 도메인의 불활성화가 Cas12a (Cpf1) 뉴클레아제 활성을 불활성화시킨다는 것이 문헌 [Zetsche et al., Cell, 163, 759-771, 2015] (이는 본원에 참조로 포함됨)에 제시되었다.

[0423] 일부 실시양태에서, napDNAbp는 미생물 CRISPR-Cas 시스템의 단일 이펙터이다. 미생물 CRISPR-Cas 시스템의 단일 이펙터는, 비제한적으로, Cas9, Cas12a (Cpf1), Cas12b1 (C2c1), Cas13a (C2c2), 및 Cas12c (C2c3)를 포함한다. 전형적으로, 미생물 CRISPR-Cas 시스템은 부류 1 및 부류 2 시스템으로 나뉜다. 부류 1 시스템은 멀티-서브유닛 이펙터 복합체를 갖는 반면, 부류 2 시스템은 단일 단백질 이펙터를 갖는다. 예를 들어, Cas9 및 Cas12a (Cpf1)는 부류 2 이펙터이다. Cas9 및 Cas12a (Cpf1)에 더하여, 3종의 별개의 부류 2 CRISPR-Cas 시스템 (Cas12b1, Cas13a, 및 Cas12c)이 문헌 [Shmakov et al., "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", Mol. Cell, 2015 Nov 5; 60(3): 385-397]에 기재되어 있으며, 그의 전체 내용은 본원에 참조로 포함된다.

[0424] 시스템 중 2개, Cas12b1 및 Cas12c의 이펙터는 Cas12a와 관련된 RuvC-유사 엔도뉴클레아제 도메인을 함유한다. 제3 시스템인 Cas13a는 2개의 예측 HEPN RNase 도메인을 갖는 이펙터를 함유한다. 성숙 CRISPR RNA의 생산은, Cas12b1에 의한 CRISPR RNA의 생산과는 달리, tracrRNA-비의존성이다. Cas12b1은 DNA 절단을 위해 CRISPR RNA 및 tracrRNA 둘 다에 의존한다. 박테리아 Cas13a는 그의 RNA-활성화된 단일 가닥 RNA 분해 활성과 구별되는 CRISPR RNA 성숙을 위한 고유한 RNase 활성을 보유하는 것으로 제시되었다. 이들 RNase 기능은 서로 상이하고, Cas12a의 CRISPR RNA-프로세싱 거동과 상이하다. 예를 들어, 문헌 [East-Seletsky, et al., "Two distinct RNase activities of CRISPR-Cas13a enable guide-RNA processing and RNA detection", Nature, 2016 Oct 13;538(7624):270-273]을 참조하며, 그의 전체 내용은 본원에 참조로 포함된다. 렙토트리키아 샤히이 (Leptotrichia shahii)의 Cas13a의 시험관내 생화학적 분석은 Cas13a가 단일 CRISPR RNA에 의해 가이드되고, 상보적 프로토스페이서를 보유하는 ssRNA 표적을 절단하도록 프로그램화될 수 있다는 것을 보여주었다. 2개의 보존된 HEPN 도메인 내의 촉매 잔기는 절단을 매개한다. 촉매 잔기에서의 돌연변이는 촉매 불활성 RNA-결합 단백질을 생성한다. 예를 들어, 문헌 [Abudayyeh et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector", Science, 2016 Aug 5; 353(6299)]을 참조하고, 그의 전체 내용은 본원에 참조로 포함된다.

[0425] 알리시클로바실루스 아시도테레스트리스(*Alicyclobacillus acidoterrestris*) Cas12b1 (AacC2c1)의 결정 구조는 키메라 단일-분자 가이드 RNA (sgRNA)와의 복합체로 보고되었다. 예를 들어, 문헌 [Liu et al., "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", Mol. Cell, 2017 Jan 19;65(2):310-322]을 참조하고, 그의 전체 내용은 본원에 참조로 포함된다. 결정 구조는 또한 3원 복합체로서 표적 DNA에 결합된 알리시클로바실루스 아시도테레스트리스 C2c1로 보고되었다. 예를 들어, 문헌 [Yang et al., "PAM-dependent

Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", Cell, 2016 Dec 15;167(7):1814-1828]을 참조하고, 그의 전체 내용은 본원에 참조로 포함된다. AacC2c1의 촉매 적격 입체형태는, 표적 및 비-표적 DNA 가닥 둘 다와 함께, 단일 RuvC 촉매 포켓 내에 독립적으로 위치되어 포획되었고, 여기서 C2c1-매개된 절단은 표적 DNA의 엇갈린 7-뉴클레오티드 파괴를 유발한다. C2c1 3원 복합체와 이전에 확인된 Cas9 및 Cpf1 대응물 사이의 구조 비교는 CRISPR-Cas9 시스템에 의해 사용되는 메카니즘의 다양성을 입증한다.

[0426] 일부 실시양태에서, napDNAbp는 C2c1, C2c2, 또는 C2c3 단백질일 수 있다. 일부 실시양태에서, napDNAbp는 C2c1 단백질이다. 일부 실시양태에서, napDNAbp는 Cas13a 단백질이다. 일부 실시양태에서, napDNAbp는 Cas12c 단백질이다. 일부 실시양태에서, napDNAbp는 자연 발생 Cas12b1 (C2c1), Cas13a (C2c2), 또는 Cas12c (C2c3) 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, napDNAbp는 자연 발생 Cas12b1 (C2c1), Cas13a (C2c2), 또는 Cas12c (C2c3) 단백질이다.

[0427] H. Cas9 원형 순열체

[0428] 다양한 실시양태에서, 본원에 개시된 다중-플랩 프라임 편집제는 Cas9의 원형 순열체를 포함할 수 있다.

[0429] 용어 "원형 순열 Cas9" 또는 Cas9의 "원형 순열체" 또는 "CP-Cas9"는 Cas9 단백질 (예를 들어, 야생형 Cas9 단백질)의 N-말단 및 C-말단이 국소로 재배열된 것을 의미하는, 원형 순열체 변이체로서 발생하거나 또는 조작되도록 변형된 임의의 Cas9 단백질 또는 그의 변이체를 지칭한다. 이러한 원형 순열 Cas9 단백질 또는 그의 변이체는 가이드 RNA (gRNA)와 복합체화되는 경우에 DNA에 결합하는 능력을 보유한다. 문헌 [Oakes et al., "Protein Engineering of Cas9 for enhanced function," Methods Enzymol, 2014, 546: 491-511 및 Oakes et al., "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," Cell, January 10, 2019, 176: 254-267]을 참조하고, 이들 각각은 본원에 참조로 포함된다. 본 개시내용은 임의의 이전에 공지된 CP-Cas9를 고려하거나, 또는 생성된 원형 순열 단백질이 가이드 RNA (gRNA)와 복합체화되는 경우 DNA에 결합하는 능력을 보유하는 한, 새로운 CP-Cas9를 사용한다.

[0430] 임의의 변이체, 오르토로그, 또는 자연 발생 Cas9 또는 그의 등가물을 포함한 본원에 기재된 Cas9 단백질 중 임의의 것은 원형 순열 변이체로서 재구성될 수 있다.

[0431] 다양한 실시양태에서, Cas9의 원형 순열체는 하기 구조를 가질 수 있다:

[0432] N-말단-[원래의 C-말단] - [임의적인 링커] - [원래의 N-말단]-C-말단.

[0433] 예로서, 본 개시내용은 정규 에스. 피오게네스 Cas9 (유니프롯KB - Q99ZW2 (CAS9_STRP1)의 1368개 아미노산 (넘버링은 서열식별번호: 18에서의 아미노산 위치에 기초함))의 하기 원형 순열체를 고려한다:

[0434] N-말단-[1268-1368]-[임의적인 링커]-[1-1267]-C-말단;

[0435] N-말단-[1168-1368]-[임의적인 링커]-[1-1167]-C-말단;

[0436] N-말단-[1068-1368]-[임의적인 링커]-[1-1067]-C-말단;

[0437] N-말단-[968-1368]-[임의적인 링커]-[1-967]-C-말단;

[0438] N-말단-[868-1368]-[임의적인 링커]-[1-867]-C-말단;

[0439] N-말단-[768-1368]-[임의적인 링커]-[1-767]-C-말단;

[0440] N-말단-[668-1368]-[임의적인 링커]-[1-667]-C-말단;

[0441] N-말단-[568-1368]-[임의적인 링커]-[1-567]-C-말단;

[0442] N-말단-[468-1368]-[임의적인 링커]-[1-467]-C-말단;

[0443] N-말단-[368-1368]-[임의적인 링커]-[1-367]-C-말단;

[0444] N-말단-[268-1368]-[임의적인 링커]-[1-267]-C-말단;

[0445] N-말단-[168-1368]-[임의적인 링커]-[1-167]-C-말단;

[0446] N-말단-[68-1368]-[임의적인 링커]-[1-67]-C-말단; 또는

[0447] N-말단-[10-1368]-[임의적인 링커]-[1-9]-C-말단, 또는 다른 Cas9 단백질 (다른 Cas9 오르토로그, 변이체 등

포함)의 상응하는 원형 순열체.

[0448] 특정한 실시양태에서, 원형 순열체 Cas9는 하기 구조 (에스. 피오게네스 Cas9 (유니프롯KB - Q99ZW2 (CAS9_STRP1)의 1368개 아미노산 (넘버링은 서열식별번호: 18에서의 아미노산 위치에 기초함)에 기초함)를 갖는다:

[0449] N-말단-[102-1368]-[임의적인 링커]-[1-101]-C-말단;

[0450] N-말단-[1028-1368]-[임의적인 링커]-[1-1027]-C-말단;

[0451] N-말단-[1041-1368]-[임의적인 링커]-[1-1043]-C-말단;

[0452] N-말단-[1249-1368]-[임의적인 링커]-[1-1248]-C-말단; 또는

[0453] N-말단-[1300-1368]-[임의적인 링커]-[1-1299]-C-말단, 또는 다른 Cas9 단백질 (다른 Cas9 오르토로그, 변이체 등 포함)의 상응하는 원형 순열체.

[0454] 또 다른 실시양태에서, 원형 순열체 Cas9는 하기 구조 (에스. 피오게네스 Cas9 (유니프롯KB - Q99ZW2 (CAS9_STRP1)의 1368개 아미노산 (넘버링은 서열식별번호: 18에서의 아미노산 위치에 기초함)에 기초함)를 갖는다:

[0455] N-말단-[103-1368]-[임의적인 링커]-[1-102]-C-말단;

[0456] N-말단-[1029-1368]-[임의적인 링커]-[1-1028]-C-말단;

[0457] N-말단-[1042-1368]-[임의적인 링커]-[1-1041]-C-말단;

[0458] N-말단-[1250-1368]-[임의적인 링커]-[1-1249]-C-말단; 또는

[0459] N-말단-[1301-1368]-[임의적인 링커]-[1-1300]-C-말단, 또는 다른 Cas9 단백질 (다른 Cas9 오르토로그, 변이체 등 포함)의 상응하는 원형 순열체.

[0460] 일부 실시양태에서, 원형 순열체는 직접적으로 또는 링커, 예컨대 아미노산 링커를 사용하여 Cas9의 C-말단 단편을 Cas9의 N-말단 단편에 연결함으로써 형성될 수 있다. 일부 실시양태에서, C-말단 단편은 Cas9의 아미노산 (예를 들어, 아미노산 약 1300-1368)의 C-말단 95% 이상, 또는 Cas9 (예를 들어, 서열식별번호: 77-86 중 어느 하나)의 C-말단 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 또는 5% 또는 그 초과에 상응할 수 있다. N-말단 부분은 Cas9의 아미노산 (예를 들어, 아미노산 약 1-1300)의 N-말단 95% 이상, 또는 Cas9의 (예를 들어, 서열식별번호: 18의) N-말단 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 또는 5% 또는 그 초과에 상응할 수 있다.

[0461] 일부 실시양태에서, 원형 순열체는 직접적으로 또는 링커, 예컨대 아미노산 링커를 사용하여 Cas9의 C-말단 단편을 Cas9의 N-말단 단편에 연결함으로써 형성될 수 있다. 일부 실시양태에서, N-말단으로 재배열된 C-말단 단편은 Cas9의 아미노산 (예를 들어, 서열식별번호: 18의 아미노산 1012-1368)의 C-말단 30% 이하를 포함하거나 또는 그에 상응한다. 일부 실시양태에서, N-말단으로 재배열된 C-말단 단편은 Cas9 (예를 들어, 서열식별번호: 18의 Cas9)의 아미노산의 C-말단 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 또는 1%를 포함하거나 또는 그에 상응한다. 일부 실시양태에서, N-말단으로 재배열된 C-말단 단편은 Cas9 (예를 들어, 서열식별번호: 18의 Cas9)의 C-말단 410개 잔기 이하를 포함하거나 또는 그에 상응한다. 일부 실시양태에서, N-말단으로 재배열된 C-말단 부분은 Cas9 (예를 들어, 서열식별번호: 18의 Cas9)의 C-말단 410, 400, 390, 380, 370, 360, 350, 340, 330, 320, 310, 300, 290, 280, 270, 260, 250, 240, 230, 220, 210, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 20, 또는 10개의 잔기를 포함하거나 또는 그에 상응한다. 일부 실시양태에서, N-말단으로 재배열된 C-말단 부분은 Cas9 (예를 들어, 서열식별번호: 18의 Cas9)의 C-말단 357, 341, 328, 120, 또는 69개의 잔기를 포함하거나 또는 그에 상응한다.

[0462] 다른 실시양태에서, 원형 순열체 Cas9 변이체는 서열식별번호: 18의 에스. 피오게네스 Cas9에 기초하여, 하기 방법: (a) 원래의 단백질을 2개의 절반: N-말단 영역 및 C-말단 영역으로 절제하는, Cas9 1차 구조의 내부 아미노산 잔기에 상응하는 원형 순열체 (CP) 부위를 선택하는 것; (b) 원래의 C-말단 영역 (CP 부위 아미노산을 포함함)을 원래의 N-말단 영역에 선행하도록 이동시켜 Cas9 단백질 서열을 변형시킴으로써 (예를 들어, 유전자 조작 기술에 의함), 이제 CP 부위 아미노산 잔기로 시작하는 Cas9 단백질의 새로운 N-말단을 형성하는 것에 기초

한 Cas9 1차 구조의 위상 재배열로서 정의될 수 있다. CP 부위는, 예를 들어 나선형-II 도메인, RuvCIII 도메인, 또는 CTD 도메인을 포함한 Cas9 단백질의 임의의 도메인에 위치할 수 있다. 예를 들어, CP 부위는 (서열식별번호: 18의 에스. 피오게네스 Cas9 대비) 원래의 아미노산 잔기 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247, 1249, 또는 1282에 위치할 수 있다. 따라서, 일단 N-말단으로 재배치되면, 원래의 아미노산 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247, 1249, 또는 1282는 새로운 N-말단 아미노산이 될 것이다. 이들 CP-Cas9 단백질의 명명법은 각각 Cas9-CP^{181}, Cas9-CP^{199}, Cas9-CP^{230}, Cas9-CP^{270}, Cas9-CP^{310}, Cas9-CP^{1010}, Cas9-CP^{1016}, Cas9-CP^{1023}, Cas9-CP^{1029}, Cas9-CP^{1041}, Cas9-CP^{1247}, Cas9-CP^{1249}, 및 Cas9-CP^{1282}로 지칭될 수 있다. 이러한 설명은 서열식별번호: 18로부터 CP 변이체를 제조하는 것으로 제한하는 것으로 의도되지 않으며, 임의의 Cas9 서열에서 이들 위치에 상응하는 CP 부위에서 또는 다른 CP 부위 전체에서 CP 변이체를 제조하기 위해 구현될 수 있다. 이러한 설명은 어떠한 방식으로든 특이적 CP 부위를 제한하는 것으로 의도되지 않는다. 실질적으로 임의의 CP 부위를 사용하여 CP-Cas9 변이체를 형성할 수 있다.

[0463] 서열식별번호: 18의 Cas9에 기초한 예시적인 CP-Cas9 아미노산 서열이 하기 제공되며, 여기서 링커 서열은 밑줄로 표시되고 임의적인 메티오닌 (M) 잔기는 볼드체로 표시된다. 개시내용은 링커 서열을 포함하지 않거나 또는 상이한 링커 서열을 포함하는 CP-Cas9 서열을 제공한다는 것을 인지해야 한다. CP-Cas9 서열은 서열식별번호: 18의 것 이외의 Cas9 서열에 기초할 수 있으며, 본원에 제공된 임의의 예는 제한하는 것으로 의도되지 않는다는 것을 인지해야 한다. 예시적인 CP-Cas9 서열은 하기와 같다:

CP 명칭	서열	SEQ ID NO:
CP1012	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG	SEQ ID NO: 77
CP1028	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGG	SEQ ID NO: 78

[0464]

[0465]

CP 명칭	서열	SEQ ID NO:
	SGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVIT DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF KSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF VYGDYKVYDVRKMIAKSEQ	
CP1041	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGG GSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVL GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYT RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL EKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE	SEQ ID NO: 79

[0466]

CP 명칭	서열	SEQ ID NO:
	LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV RKMIAKSEQEIGKATAKYFFYS	
CP1249	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL GGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTN SVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNR EDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK	SEQ ID NO: 80

CP 명칭	서열	SEQ ID NO:
	YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGS	
CP1300	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST KEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGS GGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKV PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL FELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSAYNKHRD	SEQ ID NO: 81

[0467]

[0468] Cas9 원형 순열체는 본원에 기재된 다중-플랩 프라임 편집 구축물에 유용할 수 있다. Cas9의 N-말단으로 재배열될 수 있는, 서열식별번호: 18의 Cas9에 기초한 Cas9의 예시적인 C-말단 단편이 하기 제공된다. Cas9의 이러한 C-말단 단편은 예시적인 것이며, 제한하는 것으로 의도되지 않는다는 것을 인지해야 한다. 이들 예시적인 CP-Cas9 단편은 하기 서열을 갖는다:

CP 명칭	서열	SEQ ID NO:
CP1012 C-말단 단편	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKY SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYL ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL YETRIDLSQLGGD	SEQ ID NO: 82
CP1028 C-말단 단편	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 83
CP1041 C-말단 단편	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 84
CP1249 C-말단 단편	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL GGD	SEQ ID NO: 85
CP1300 C-말단 단편	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST KEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 86

[0469]

[0470] I. 변형된 PAM 특이성을 갖는 Cas9 변이체

[0471] 본 개시내용의 다중-플랩 프라임 편집제는 또한 변형된 PAM 특이성을 갖는 Cas9 변이체를 포함할 수 있다. 본 개시내용의 일부 측면은 그의 3'-단부에 정규 PAM (5'-NGG-3', 여기서 N은 A, C, G, 또는 T임)을 포함하지 않는 표적 서열에 대해 활성을 나타내는 Cas9 단백질을 제공한다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NGG-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NNG-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NNA-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NNC-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NNT-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NGT-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NGA-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NGC-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NAA-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NAC-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NAT-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다. 또 다른 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NAG-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타낸다.

[0472] 제1 아미노산 잔기 (예를 들어, A)에서 제2 아미노산 잔기 (예를 들어, T)로의 본원에 기재된 임의의 아미노산 돌연변이 (예를 들어, A262T)는 또한 제1 아미노산 잔기에서 제2 아미노산 잔기와 유사한 (예를 들어, 보존된) 아미노산 잔기로의 돌연변이를 포함할 수 있음을 인지해야 한다. 예를 들어, 아미노산의 소수성 측쇄 (예를 들어, 알라닌, 발린, 이소류신, 류신, 메티오닌, 페닐알라닌, 티로신, 또는 트립토판)로의 돌연변이는 상이한 소수성 측쇄 (예를 들어, 알라닌, 발린, 이소류신, 류신, 메티오닌, 페닐알라닌, 티로신, 또는 트립토판)를 갖는

제2 아미노산으로의 돌연변이일 수 있다. 예를 들어, 알라닌에서 트레오닌으로의 돌연변이 (예를 들어, A262T 돌연변이)는 또한 알라닌에서 트레오닌과 크기 및 화학적 특성이 유사한 아미노산, 예를 들어 세린으로의 돌연변이일 수 있다. 또 다른 예로서, 양으로 하전된 측쇄를 갖는 아미노산 (예를 들어, 아르기닌, 히스티딘, 또는 리신)의 돌연변이는 상이한 양으로 하전된 측쇄를 갖는 제2 아미노산 (예를 들어, 아르기닌, 히스티딘, 또는 리신)으로의 돌연변이일 수 있다. 또 다른 예로서, 극성 측쇄를 갖는 아미노산 (예를 들어, 세린, 트레오닌, 아스파라긴, 또는 글루타민)의 돌연변이는 상이한 극성 측쇄를 갖는 제2 아미노산 (예를 들어, 세린, 트레오닌, 아스파라긴, 또는 글루타민)으로의 돌연변이일 수 있다. 추가의 유사한 아미노산 쌍은 하기: 페닐알라닌 및 티로신; 아스파라긴 및 글루타민; 메티오닌 및 시스테인; 아스파르트산 및 글루탐산; 및 아르기닌 및 리신을 포함하나 이에 제한되지는 않는다. 통상의 기술자는 이러한 보존적 아미노산 치환이 단백질 구조에 대해 미미한 효과를 가질 가능성이 있고, 기능을 손상시키지 않으면서 잘 용인될 가능성이 있음을 인식할 것이다. 일부 실시양태에서, 하나의 아미노산에서 트레오닌으로의 본원에 제공된 아미노산 돌연변이 중 임의의 것은 세린으로의 아미노산 돌연변이일 수 있다. 일부 실시양태에서, 하나의 아미노산에서 아르기닌으로의 본원에 제공된 아미노산 돌연변이 중 임의의 것은 리신으로의 아미노산 돌연변이일 수 있다. 일부 실시양태에서, 하나의 아미노산에서 이소류신으로의 본원에 제공된 아미노산 돌연변이 중 임의의 것은 알라닌, 발린, 메티오닌, 또는 류신으로의 아미노산 돌연변이일 수 있다. 일부 실시양태에서, 하나의 아미노산에서 리신으로의 본원에 제공된 아미노산 돌연변이 중 임의의 것은 아르기닌으로의 아미노산 돌연변이일 수 있다. 일부 실시양태에서, 하나의 아미노산에서 아스파르트산으로의 본원에 제공된 아미노산 돌연변이 중 임의의 것은 글루탐산 또는 아스파라긴으로의 아미노산 돌연변이일 수 있다. 일부 실시양태에서, 하나의 아미노산에서 발린으로의 본원에 제공된 아미노산 돌연변이 중 임의의 것은 알라닌, 이소류신, 메티오닌, 또는 류신으로의 아미노산 돌연변이일 수 있다. 일부 실시양태에서, 하나의 아미노산에서 글리신으로의 본원에 제공된 아미노산 돌연변이 중 임의의 것은 알라닌으로의 아미노산 돌연변이일 수 있다. 그러나, 추가의 보존된 아미노산 잔기가 통상의 기술자에 의해 인식될 것이고, 다른 보존된 아미노산 잔기로의 아미노산 돌연변이 중 임의의 것이 또한 본 개시내용의 범주 내에 있다는 것을 인지해야 한다.

[0473] 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NAA-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타내는 돌연변이의 조합을 포함한다. 일부 실시양태에서, 돌연변이의 조합은 표 1에 열거된 클론 중 어느 하나에 존재한다. 일부 실시양태에서, 돌연변이의 조합은 표 1에 열거된 클론의 보존적 돌연변이이다. 일부 실시양태에서, Cas9 단백질은 표 1에 열거된 Cas9 클론 중 어느 하나의 돌연변이의 조합을 포함한다.

[0474] 표 1: NAA PAM 클론

야생형 SpCas9 (예를 들어, SEQ ID NO: 18)로부터의 돌연변이
D177N, K218R, D614N, D1135N, P1137S, E1219V, A1320V, A1323D, R1333K
D177N, K218R, D614N, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, G715C, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A367T, K710E, R1114G, D1135N, P1137S, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D861N, D1135N, K1188R, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, V743I, R753G, E762G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, S1274R, A1320V, R1333K
A10T, I322V, S409I, E427G, A589S, R753G, D1135N, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, E757K, G865G, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, R753G, E757K, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, K599R, M631A, R654L, K673E, V743I, R753G, N758H, E762G, D1135N, D1180G, E1219V, Q1221H, Q1256R, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N869S, N1054D, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, L727I, V743I, R753G, E762G, R859S, N946D, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, N1317T, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, K1151E, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1080S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, E630K, R654L, K673E, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, G1223S, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1801S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, M673I, N803S, N869S, G1077D, R1114G, D1135N, V1139A, D1180G, E1219V, Q1221H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, R1114G, D1135N, E1219V, Q1221H, A1320V, R1333K

[0475]

[0476] 일부 실시양태에서, Cas9 단백질은 표 1의 변이체 중 어느 하나에 의해 제공된 바와 같은 Cas9 단백질의 아미노산 서열과 적어도 80% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 단백질은 표 1의 변이체 중 어느 하나에 의해 제공된 바와 같은 Cas9 단백질의 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다.

[0477] 일부 실시양태에서, Cas9 단백질은 서열식별번호: 18에 의해 제공된 바와 같은 스트렙토코쿠스 피오게네스 Cas9와 비교하였을 때, 그의 3' 단부에 정규 PAM (5'-NGG-3')을 포함하지 않는 표적 서열에 대해 증가된 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 동일한 표적 서열에 대한 서열식별번호: 18에 의해 제공된 바와 같은 스트렙토코쿠스 피오게네스 Cas9의 활성과 비교하였을 때, 적어도 5배 증가된, 정규 PAM 서열 (5'-NGG-3')에 바로 인접하지 않은 3' 단부를 갖는 표적 서열에 대한 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 동일한 표적 서열에 대한 서열식별번호: 18에 의해 제공된 바와 같은 스트렙토코쿠스 피오게네스의 활성과 비교하였을 때, 적어도 10배, 적어도 50배, 적어도 100배, 적어도 500배, 적어도 1,000배, 적어도 5,000배, 적어도

10,000배, 적어도 50,000배, 적어도 100,000배, 적어도 500,000배, 또는 적어도 1,000,000배 증가된, 정규 PAM 서열 (5'-NGG-3')에 바로 인접하지 않은 표적 서열에 대한 활성을 나타낸다. 일부 실시양태에서, 표적 서열의 3' 단부는 AAA, GAA, CAA, 또는 TAA 서열에 바로 인접한다. 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NAC-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타내는 돌연변이의 조합을 포함한다. 일부 실시양태에서, 돌연변이의 조합은 표 2에 열거된 클론 중 어느 하나에 존재한다. 일부 실시양태에서, 돌연변이의 조합은 표 2에 열거된 클론의 보존적 돌연변이이다. 일부 실시양태에서, Cas9 단백질은 표 2에 열거된 Cas9 클론 중 어느 하나의 돌연변이의 조합을 포함한다.

[0478] 표 2: NAC PAM 클론

야생형 SPCAS9 (예를 들어, SEQ ID NO: 18)로부터의 돌연변이
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N
T472I, R753G, D1332N, R1335Q, T1337N
D1135N, E1219V, D1332N, R1335Q, T1337N
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N
T472I, R753G, D1332N, R1335Q, T1337N
T472I, R753G, Q771H, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, K1156E, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, V647I, R753G, N803S, K959N, G1030R, I1055E, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, E630G, T638P, V647A, G687R, N767D, N803S, K959N, R1114G, D1135N, E1219V, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, I1057T, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, M631I, T638P, R753G, N803S, K959N, Y1036H, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, V875I, K959N, Y1016C, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N, I1348V
K608R, E627K, T638P, V647I, R654L, R753G, N803S, T804A, K848N, V922A, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R753G, N803S, V922A, K959N, K1014N, V1015A, R1114G, D1135N, K1156N, E1219V, N1252D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, A711T, R753G, K775R, K789E, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
I670S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, K797N, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, G752R, R753G, K797N, N803S, K948E, K959N, V1015A, Y1016S, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
I570T, A589V, K608R, E627K, T638P, V647I, R654L, Q716R, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1207G, E1219V, N1234D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, R654L, Q740R, R753G, N803S, K959N, N990S, T995S, V1015A, Y1036D, R1114G, D1135N, E1207G, E1219V, N1234D, N1266H, D1332N, R1335Q, T1337N

[0479]

I562F, V565D, I570T, K608R, L625S, E627K, T638P, V647I, R654I, G752R, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, N1177S, N1234D, D1332N, R1335Q, T1337N
I562F, I570T, K608R, E627K, T638P, V647I, R753G, E790A, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, D1180E, A1184T, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1015A, R1114G, D1127A, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, L625S, E627K, T638P, V647I, R654I, T703P, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1016A, R1114G, D1135N, E1219V, K1246E, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R654L, K673E, R753G, E790A, N803S, K948E, K959N, R1114G, D1127G, D1135N, D1180E, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, L625S, E627K, T638P, V647I, R654I, I670T, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
E627K, M631V, T638P, V647I, K710E, R753G, N803S, N808D, K948E, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N, S1338T, H1349R

[0480]

[0481] 일부 실시양태에서, Cas9 단백질은 표 2의 변이체 중 어느 하나에 의해 제공된 바와 같은 Cas9 단백질의 아미노산 서열과 적어도 80% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 단백질은 표 2의 변이체 중 어느 하나에 의해 제공된 바와 같은 Cas9 단백질의 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다.

[0482] 일부 실시양태에서, Cas9 단백질은 서열식별번호: 18에 의해 제공된 바와 같은 스트렙토코쿠스 피오게네스 Cas9와 비교하였을 때, 그의 3' 단부에 정규 PAM (5'-NGG-3')을 포함하지 않는 표적 서열에 대해 증가된 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 동일한 표적 서열에 대한 서열식별번호: 18에 의해 제공된 바와 같은 스트렙토코쿠스 피오게네스 Cas9의 활성과 비교하였을 때, 적어도 5배 증가된, 정규 PAM 서열 (5'-NGG-3')에 바로 인접하지 않은 3' 단부를 갖는 표적 서열에 대한 활성을 나타낸다. 일부 실시양태에서, Cas9 단백질은 동일한 표적 서열에 대한 서열식별번호: 18에 의해 제공된 바와 같은 스트렙토코쿠스 피오게네스의 활성과 비교하였을 때, 적어도 10배, 적어도 50배, 적어도 100배, 적어도 500배, 적어도 1,000배, 적어도 5,000배, 적어도 10,000배, 적어도 50,000배, 적어도 100,000배, 적어도 500,000배, 또는 적어도 1,000,000배 증가된, 정규 PAM 서열 (5'-NGG-3')에 바로 인접하지 않은 표적 서열에 대한 활성을 나타낸다. 일부 실시양태에서, 표적 서열의 3' 단부는 AAC, GAC, CAC, 또는 TAC 서열에 바로 인접한다.

[0483] 일부 실시양태에서, Cas9 단백질은 그의 3'-단부에 5'-NAT-3' PAM 서열을 포함하는 표적 서열에 대해 활성을 나타내는 돌연변이의 조합을 포함한다. 일부 실시양태에서, 돌연변이의 조합은 표 3에 열거된 클론 중 어느 하나에 존재한다. 일부 실시양태에서, 돌연변이의 조합은 표 3에 열거된 클론의 보존적 돌연변이이다. 일부 실시양태에서, Cas9 단백질은 표 3에 열거된 Cas9 클론 중 어느 하나의 돌연변이의 조합을 포함한다.

[0484] 표 3: NAT PAM 클론

야생형 SPCAS9 (예를 들어, SEQ ID NO: 18)로부터의 돌연변이
K961E, H985Y, D1135N, K1191N, E1219V, Q1221H, A1320A, P1321S, R1335L
D1135N, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
V743I, R753G, E790A, D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, V748I, V743I, R753G, D853E, V922A, R1114G D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1286K, P1321S, D1322G, R1335L
M631L, R654L, R753G, K797E, D853E, V922A, D1012A, R1114G D1135N, G1218S, E1219V, Q1221H, P1249S, N1317K, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, D596Y, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, Q1256R, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, K710E, V750A, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, K649R, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, K1156E, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, I1057G, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1308D, P1321S, D1322G, R1335L
M631L, R654L, R753G, D853E, V922A, R1114G, Y1131C, D1135N, E1150V, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, D853E, I1057V, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, I1057V, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L

[0485]

[0486] 본원에 개시된 다중-플랩 프라임 편집제와 관련하여 사용될 수 있는 다양한 napDNAbp에 대한 상기 설명은 어떠한 방식으로도 제한하는 것으로 의도되지 않는다. 다중-플랩 프라임 편집제는 정규 SpCas9, 또는 임의의 오르토로그 Cas9 단백질, 또는 임의의 변이체 Cas9 단백질-공지되어 있거나 또는 방향적 진화상 과정 또는 달리 돌연변이유발 과정을 통해 제조 또는 진화될 수 있는 Cas9의 임의의 자연 발생 변이체, 돌연변이체, 또는 달리 조작된 버전 포함-을 포함할 수 있다. 다양한 실시양태에서, Cas9 또는 Cas9 변이체는 닉카제 활성을 갖고, 즉 표적 DNA 서열의 가닥만을 절단한다. 다른 실시양태에서, Cas9 또는 Cas9 변이체는 불활성 뉴클레아제를 가지며, 즉 "기능상실" Cas9 단백질이다. 사용될 수 있는 다른 변이체 Cas9 단백질은 정규 SpCas9보다 더 작은 분자량을 갖거나 (예를 들어, 보다 용이한 전달을 위함), 또는 변형되거나 재배열된 1차 아미노산 구조를 갖는 (예를 들어, 원형 순열체 포맷) 것이다. 본원에 기재된 다중-플랩 프라임 편집제는 또한 수렴 진화의 결과인 Cas12a/Cpf1 및 Cas12b 단백질을 포함한 Cas9 등가물을 포함할 수 있다. 본원에 사용된 napDNAbp (예를 들어, SpCas9, Cas9 변이체, 또는 Cas9 등가물)는 또한 그의 PAM 특이성을 변경/증진시키는 다양한 변형을 함유할 수 있다. 마지막으로, 본 출원은 참조 Cas9 서열, 예컨대 참조 SpCas9 정규 서열 또는 참조 Cas9 등가물 (예를 들어, Cas12a/Cpf1)과 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.9% 서열 동일성을 갖는 임의의 Cas9, Cas9 변이체, 또는 Cas9 등가물을 고려한다.

[0487] 특정한 실시양태에서, 확장된 PAM 능력을 갖는 Cas9 변이체는 하기 아미노산 서열을 갖는 SpCas9 (H840A) VRQR (서열식별번호: 87)이며, 서열식별번호: 51의 SpCas9 (H840A) 대비 V, R, Q, R 치환은 밑줄표시된 볼드체로 제시된다. 또한, SpCas9 (H840)에서의 메티오닌 잔기를 SpCas9 (H840A) VRQR에 대해 제거하였다:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL
LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH
SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI
LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG
SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD
YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 87)

[0488]

[0489] 또 다른 특정한 실시양태에서, 확장된 PAM 능력을 갖는 Cas9 변이체는 하기 아미노산 서열을 갖는 SpCas9 (H840A) VRER이며, 서열식별번호: 51의 SpCas9 (H840A) 대비 V, R, E, R 치환은 밑줄표시된 볼드체로 제시된다. 또한, SpCas9 (H840)에서의 메티오닌 잔기를 SpCas9 (H840A) VRER에 대해 제거하였다:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL
LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH
SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI
LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG
SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD
YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 88)

[0490]

[0491] 일부 실시양태에서, 비-정규 PAM 서열과 함께 기능하는 napDNAbp는 아르고노트 단백질이다. 이러한 핵산 프로그램가능한 DNA 결합 단백질의 한 예는 나트로노박테리움 그레고리(Natronobacterium gregoryi)로부터의 아르고노트 단백질 (NgAgo)이다. NgAgo는 ssDNA-가이드된 엔도뉴클레아제이다. NgAgo는 이를 그의 표적 부위로 가이드하기 위해 ~24개 뉴클레오티드의 5' 인산화된 ssDNA (gDNA)에 결합하고, 이는 gDNA 부위에서 DNA 이중 가닥 파괴를 만들 것이다. Cas9와 대조적으로, NgAgo-gDNA 시스템은 프로토스페이서-인접 모티프 (PAM)를 필요로 하지 않는다. 뉴클레아제 불활성 NgAgo (dNgAgo)를 사용하는 것은 표적화될 수 있는 염기를 크게 확장시킬 수 있다. NgAgo의 특징화 및 사용은 문헌 [Gao et al., Nat Biotechnol., 2016 Jul;34(7):768-73. PubMed PMID: 27136078; Swarts et al., Nature. 507(7491) (2014):258-61; 및 Swarts et al., Nucleic Acids Res. 43(10) (2015):5120-9]에 기재되어 있고, 이들 각각은 본원에 참조로 포함된다.

[0492] 일부 실시양태에서, napDNAbp는 아르고노트 단백질의 원핵 상동체이다. 아르고노트 단백질의 원핵 상동체는 공지되어 있고, 예를 들어 문헌 [Makarova K., et al., "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements", Biol Direct. 2009 Aug 25;4:29. doi: 10.1186/1745-6150-4-29]에 기재되어 있고, 그의 전체 내용은 본원에 참조로 포함된다. 일부 실시양태에서, napDNAbp는 마리니토가 피에조필라(Marinitoga piezophila) 아르고노트(MpAgo) 단백질이다. CRISPR-연관 마리니토가 피에조필라 아르고노트 (MpAgo) 단백질은 5'-인산화된 가이드를 사용하여 단일 가닥 표적 서열을 절단한다. 5' 가이드는 모든 공지된 아르고노트에 의해 사용된다. MpAgo-RNA 복합체의 결정 구조는 5' 포스페이트 상호작용을 차단하는 잔기를 포함하는 가이드 가닥 결합 부위를 보여준다. 이 데이터는 5'-히드록실화된 가이드에 대한 비정규 특이성을 갖는 아르고노트 하위부류의 진화를 시사한다. 예를 들어, 문헌 [Kaya et al., "A bacterial Argonaute with noncanonical guide RNA specificity", Proc Natl Acad Sci U S A. 2016 Apr 12;113(15):4057-62]을 참조하고, 그의 전체 내용은 본원에 참조로 포함된다. 다른 아르고노트 단백질이 사용될 수 있고, 본 개시내용의 범주 내에 있다는 것을 인지해야 한다.

[0493] 개시내용의 일부 측면은 상이한 PAM 특이성을 갖는 Cas9 도메인을 제공한다. 전형적으로, Cas9 단백질, 예컨대 에스. 피오게네스로부터의 Cas9 (spCas9)는 특정한 핵산 영역에 결합하기 위해 정규 NGG PAM 서열을 필요로 한다. 이는 게놈 내의 목적하는 염기를 편집하는 능력을 제한할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집 융합 단백질은 정확한 위치에 배치될 필요가 있을 수 있고, 예를 들어 여기서 표적 염기는 PAM의 대략 15개 염기 상류인 4개 염기 영역 (예를 들어, "편집 윈도우") 내에 배치된다. 문헌 [Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016)]을 참조하고, 그의 전체 내용은 본원에 참조로 포함된다. 따라서, 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 정규 (예를 들어, NGG) PAM 서열을 함유하지 않는 뉴클레오티드 서열에 결합할 수 있는 Cas9 도메인을 함유할 수 있다. 비-정규 PAM 서열에 결합하는 Cas9 도메인은 관련 기술분야에 기재되어 있고, 통상의 기술자에게 명백할 것이다. 예를 들어, 비-정규 PAM 서열에 결합하는 Cas9 도메인은 문헌 [Kleinstiver, B. P., et al., "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" Nature 523, 481-485 (2015); 및 Kleinstiver, B. P., et al., "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition" Nature Biotechnology 33, 1293-1298 (2015)]에 기재되어 있고; 각각의 전체 내용은 본원에 참조로 포함된다.

[0494] 예를 들어, 변경된 PAM 특이성을 갖는 napDNAbp 도메인, 예컨대 하기 아미노산 서열을 갖는, 야생형 프란시셀라 노비시다(Francisella novicida) Cpf1 (D917, E1006, 및 D1255) (서열식별번호: 74)과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 도메인:

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKY
HQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKD
SEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFK
GWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAIN
YEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGK
FVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDD
SDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
SQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLA
LEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQ
ASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELA
NIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYL
GVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPS
EDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFS
DTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKG
RPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNK
DNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDV
HILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSAR
KDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQV
YQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPA
GFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIK
AAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKN
MPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 74)

[0495]

[0496] 변경된 PAM 특이성을 갖는 추가의 napDNAbp 도메인, 예컨대 하기 아미노산 서열을 갖는, 야생형 게오바실루스 써모데니트리피칸스(Geobacillus thermodenitrificans) Cas9 (서열식별번호: 75)와 적어도 80%, 적어도 85%,

적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 도메인:

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARR
RLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDEL
ARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFS
LHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFA
SKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERR
LIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHK
IRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENL
ADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGP
KKKQKTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFD
ERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQ
PIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLGSER
WQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIRE
HLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDI
ARVTAFYQRREQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNY
DNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLD
KTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKII
DTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEPN
KPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSN
GGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAG
ETIRPL (SEQ ID NO: 75)

[0497]

[0498] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 정규 (NGG) PAM 서열을 필요로 하지 않는 핵산 프로그램가능한 DNA 결합 단백질이다. 일부 실시양태에서, napDNAbp는 아르고노트 단백질이다. 이러한 핵산 프로그램가능한 DNA 결합 단백질의 한 예는 나트로노박테리움 그레고리(Natronobacterium gregoryi)로부터의 아르고노트 단백질 (NgAgo)이다. NgAgo는 ssDNA-가이드된 엔도뉴클레아제이다. NgAgo는 이를 그의 표적 부위로 가이드하기 위해 ~24개 뉴클레오티드의 5' 인산화된 ssDNA (gDNA)에 결합하고, 이는 gDNA 부위에서 DNA 이중 가닥 파괴를 만들 것이다. Cas9와 대조적으로, NgAgo-gDNA 시스템은 프로토스페이서-인접 모티프 (PAM)를 필요로 하지 않는다. 뉴클레아제 불활성 NgAgo (dNgAgo)를 사용하는 것은 표적화될 수 있는 염기를 크게 확장시킬 수 있다. NgAgo의 특징화 및 사용은 문헌 [Gao et al., Nat Biotechnol., 34(7): 768-73 (2016), PubMed PMID: 27136078; Swarts et al., Nature, 507(7491): 258-61 (2014); 및 Swarts et al., Nucleic Acids Res. 43(10) (2015): 5120-9]에 기재되어 있고, 이들 각각은 본원에 참조로 포함된다. 나트로노박테리움 그레고리 아르고노트의 서열이 서열식별번호: 76에서 제공된다.

[0499] 개시된 융합 단백질은 하기 아미노산 서열을 갖는, 야생형 나트로노박테리움 그레고리 아르고노트 (서열식별번호: 76)와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 napDNAbp 도메인을 포함할 수 있다:

MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYIT
LWKNTTPKDVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTT
DEDETFAGGEPLDHHLDDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYT
LTATDGAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLD
CGVRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLT
CDEFDLHERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTY
RPRRGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVE
TRRQGHGDDAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQA
FAERLDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFS
KGIVNPPESFEVAVVLPEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSS
PESISLNVAGAIDPSEVDAAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQ
MAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPE
DGASGQINIAATATAVYKDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVT
GESPTHIVIHRDGFMNEDLDPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQYDTP
VKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLL
SQSHIQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL (SEQ ID NO:
76)

[0500]

[0501] 또한, 임의의 이용가능한 방법을 이용하여 변이체 또는 돌연변이체 Cas9 단백질을 수득하거나 구축할 수 있다. 본원에 사용된 용어 "돌연변이"는 서열, 예를 들어 핵산 또는 아미노산 서열 내의 잔기의 또 다른 잔기로의 치환, 또는 서열 내의 1개 이상의 잔기의 결실 또는 삽입을 지칭한다. 돌연변이는 전형적으로 원래 잔기를 확인한 다음, 서열 내의 잔기의 위치를 확인하고, 새로 치환된 잔기의 정체를 확인하는 것에 의해 본원에 기재된다. 본원에 제공된 아미노산 치환 (돌연변이)을 제조하는 다양한 방법은 관련 기술분야에 널리 공지되어 있고, 예를

들어 문헌 [Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 의해 제공된다. 돌연변이는 다양한 카테고리, 예컨대 단일 염기 다형성, 미세중복 영역, indel, 및 역위를 포함할 수 있고, 이는 어떠한 방식으로도 제한하는 것으로 의도되지 않는다. 돌연변이는 단백질 활성을 감소시키거나 제거하는 돌연변이의 정상적인 결과인 "기능 상실" 돌연변이를 포함할 수 있다. 대부분의 기능 상실 돌연변이는 열성이며, 이는 이형접합체에서 제2 염색체 카피가 그의 존재가 돌연변이의 효과를 보상하는 완전 기능적 단백질을 코딩하는 유전자의 비돌연변이된 버전을 보유하기 때문이다. 돌연변이는 또한 정상 상태에서는 달리 존재하지 않는 단백질 또는 세포에 비정상적 활성을 부여하는 것인 "기능 획득" 돌연변이를 포괄한다. 많은 기능 획득 돌연변이는 코딩 영역보다는 조절 서열에 존재하고, 따라서 다수의 결과를 가질 수 있다. 예를 들어, 돌연변이는 1개 이상의 유전자가 잘못된 조직에서 발현되게 할 수 있고, 이들 조직은 정상적으로는 결여되어 있는 기능을 얻는다. 그의 속성으로 인해, 기능 획득 돌연변이는 통상적으로 우성이다.

[0502] 돌연변이는 부위-지정 돌연변이유발을 사용하여 참조 Cas9 단백질 내로 도입될 수 있다. 관련 기술분야에 공지된 부위-지정 돌연변이유발의 보다 오래된 방법은 단일 가닥 DNA 주형의 단리를 가능하게 하는 벡터, 예컨대 M13 박테리오파지 벡터 내로의 돌연변이시킬 서열의 서브-클로닝에 의존한다. 이들 방법에서, 돌연변이유발 프라이머 (즉, 돌연변이시킬 부위에 어닐링할 수 있지만 돌연변이시킬 부위에 1개 이상의 미스매칭된 뉴클레오티드를 보유할 수 있는 프라이머)를 단일 가닥 주형에 어닐링시킨 후, 돌연변이유발 프라이머의 3' 단부로부터 출발하여 주형의 상보체를 중합시킨다. 이어서, 생성된 듀플렉스로 숙주 박테리아를 형질전환시키고, 목적하는 돌연변이에 대해 플라크를 스크리닝한다. 보다 최근에, 부위-지정 돌연변이유발은 단일 가닥 주형을 필요로 하지 않는 이점을 갖는 PCR 방법론을 사용하였다. 또한, 서브클로닝을 필요로 하지 않는 방법이 개발되었다. PCR-기반 부위-지정 돌연변이유발이 수행될 때 여러 이슈가 고려되어야 한다. 첫번째로, 이들 방법에서, 폴리머라제에 의해 도입된 바람직하지 않은 돌연변이의 확장을 방지하기 위해 PCR 사이클의 수를 감소시키는 것이 바람직하다. 두번째로, 반응에서 지속되는 비-돌연변이된 모 분자의 수를 감소시키기 위해 선택이 사용되어야 한다. 세번째로, 단일 PCR 프라이머 세트의 사용을 허용하기 위해 연장-길이 PCR 방법이 바람직하다. 그리고 네번째로, 일부 열안정성 폴리머라제의 비-주형-의존성 말단 연장 활성으로 인해, PCR-생성된 돌연변이체 생성물의 평활-단부 라이게이션 전에 단부-폴리싱 단계를 절차에 혼입시키는 것이 종종 필요하다.

[0503] 돌연변이는 또한 방향적 진화 과정, 예컨대 파지-보조 연속 진화 (PACE) 또는 파지-보조 비연속 진화 (PANCE)에 의해 도입될 수 있다. 본원에 사용된 용어 "파지-보조 연속 진화 (PACE)"는 바이러스 벡터로서 파지를 사용하는 연속 진화를 지칭한다. PACE 기술의 일반적 개념은, 예를 들어 2010년 3월 11일에 WO 2010/028347로서 공개된, 2009년 9월 8일에 출원된 국제 PCT 출원 PCT/US2009/056194; 2012년 6월 28일에 WO 2012/088381로서 공개된, 2011년 12월 22일에 출원된 국제 PCT 출원 PCT/US2011/066747; 2015년 5월 5일에 허여된 미국 출원, 미국 특허 번호 9,023,594, 2015년 9월 11일에 WO 2015/134121로서 공개된, 2015년 1월 20일에 출원된 국제 PCT 출원 PCT/US2015/012022, 및 2016년 10월 20일에 WO 2016/168631로서 공개된, 2016년 4월 15일에 출원된 국제 PCT 출원 PCT/US2016/027795에 기재되어 있으며, 이들 각각의 전체 내용은 본원에 참조로 포함된다. 변이체 Cas9는 또한 파지-보조 비-연속 진화 (PANCE)에 의해 수득될 수 있고, 이는 본원에 사용된 바와 같이, 바이러스 벡터로서 파지를 사용하는 비-연속 진화를 지칭한다. PANCE는 새로운 이. 콜라이(E. coli) 숙주 세포를 가로질러 진화시킬 관심 유전자를 함유하는 진화 '선택 파지' (SP)의 연속 플라스크 전달을 사용하여 숙주 이. 콜라이 내부의 유전자는 일정하게 유지되게 하는 한편 SP 내에 함유된 유전자가 연속적으로 진화하게 하는, 신속한 생체내 방향적 진화를 위한 단순화된 기술이다. 연속 플라스크 전달은 오랫동안 미생물의 실험실 진화를 위한 폭넓게 접근가능한 접근법으로서의 역할을 하였고, 보다 최근에는 박테리오파지 진화를 위한 유사한 접근법이 개발되었다. PANCE 시스템은 PACE 시스템보다 더 낮은 엄격도를 특색으로 한다.

[0504] Cas9 또는 Cas9 등가물에 관한 상기 언급된 참고문헌 중 임의의 것은, 그렇게 이미 언급되지 않았다면, 그 전문이 본원에 참조로 포함된다.

[0505] J. 분할 PE 전달을 위한 나뉘어진 napDNAbp 도메인

[0506] 다양한 실시양태에서, 본원에 기재된 프라임 편집제는 세포 내부에서 (수동 조립에 의해, 또는 분할 인테인 서열을 사용하는 것과 같은 능동 조립에 의해) 재구성된 프라임 편집제로 조립되는 2개 이상의 단편으로서 세포에 전달될 수 있다. 일부 경우에, 자기 조립은 2개 이상의 프라임 편집제 단편이 세포 내부에서 공유적으로 또는 비-공유적으로 회합되어 프라임 편집제를 재구성하는 수동적인 것일 수 있다. 다른 경우에, 자기-조립은 각각의 단편 상에 설치된 이량체화 도메인에 의해 촉매될 수 있다. 이량체화 도메인의 예는 본원에 기재되어 있다. 또 다른 경우에, 자기-조립은 각각의 프라임 편집제 단편 상에 설치된 분할 인테인 서열에 의해 촉매될 수

있다.

[0507] 분할 PE 전달은 상이한 전달 접근법의 다양한 크기 제약을 다루는데 유리할 수 있다. 예를 들어, 전달 접근법은 바이러스-기반 전달 방법, 메신저 RNA-기반 전달 방법, 또는 RNP-기반 전달 (리보핵단백질-기반 전달)을 포함할 수 있다. 그리고, 각각의 이들 전달 방법은 프라임 편집제를 보다 작은 조각으로 분할함으로써 보다 효율적이고/거나 효과적일 수 있다. 일단 세포 내부로 들어가면, 보다 작은 조각은 기능적 프라임 편집제로 조립될 수 있다. 분할 수단에 따라, 나뉘어진 프라임 편집제 단편은 비-공유 방식 또는 공유 방식으로 재조립되어 프라임 편집제를 재형성할 수 있다. 한 실시양태에서, 프라임 편집제는 1개 이상의 분할 부위에서 2개 이상의 단편으로 분할될 수 있다. 단편은 (분할되는 것 외에) 비변형될 수 있다. 단편이 세포에 전달되면 (예를 들어, 리보핵단백질 복합체의 직접 전달에 의해 또는 핵산 전달 - 예를 들어, mRNA 전달 또는 바이러스 벡터 기반 전달에 의해), 단편은 공유적으로 또는 비-공유적으로 재회합되어 프라임 편집제를 재구성할 수 있다. 또 다른 실시양태에서, 프라임 편집제는 1개 이상의 분할 부위에서 2개 이상의 단편으로 분할될 수 있다. 각각의 단편은 이량체화 도메인을 포함하도록 변형될 수 있고, 이에 의해 형성되는 각각의 단편은 이량체화 도메인에 커플링된다. 일단 세포 내에 전달되거나 발현되면, 상이한 단편의 이량체화 도메인은 서로 회합하고 결합하여, 상이한 프라임 편집제 단편을 함께 모아 기능적 프라임 편집제를 재형성한다. 또 다른 실시양태에서, 프라임 편집제 단편은 분할 인테인을 포함하도록 변형될 수 있다. 일단 세포 내에 전달되거나 발현되면, 상이한 단편의 분할 인테인 도메인은 서로 회합하고 결합한 다음, 트랜스-스플라이싱을 겪고, 이는 각각의 단편으로부터 분할-인테인 도메인의 절제, 및 단편 사이의 펩티드 결합의 동반 형성을 유발하며, 그에 의해 프라임 편집제가 복원된다.

[0508] 한 실시양태에서, 프라임 편집제는 분할-인테인 접근법을 사용하여 전달될 수 있다.

[0509] 분할 부위의 위치는 napDNAbp 도메인, 폴리머라제 도메인 (예를 들어, RT 도메인), napDNAbp 도메인과 폴리머라제 도메인을 연결하는 링커 도메인 내를 포함한, 프라임 편집제 및 그 안의 임의의 도메인 내의 잔기의 임의의 1개 이상의 쌍 사이에 위치할 수 있다.

[0510] 도 66에 도시된 한 실시양태에서, 프라임 편집제 (PE)는 napDNAbp 내의 분할 부위에서 나뉘어진다.

[0511] 특정 실시양태에서, napDNAbp는 하기와 같은 서열식별번호: 18의 정규 SpCas9 폴리펩티드이다:

SpCas9 *스트렙토코쿠스 피오게네스* M1 스위스프롯 수탁 번호 Q99ZW2 야생형 1368 AA	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 18

[0512]

[0513] 특정 실시양태에서, SpCas9는 서열식별번호: 18의 정규 SpCas9의 잔기 1과 2 사이, 또는 2와 3 사이, 또는 3과 4 사이, 또는 4와 5 사이, 또는 5와 6 사이, 또는 6과 7 사이, 또는 7과 8 사이, 또는 8과 9 사이, 또는 9와 10 사이, 또는 잔기 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 1000-1100, 1100-1200, 1200-1300, 또는 1300-1368 사이의 임의의 위치에 위치하는 잔기의 임의의 2개 쌍 사이에 위치하는 분할 부위에서 2개의 단편으로 분할된다.

[0514] 특정 실시양태에서, napDNAbp는 서열식별번호: 18의 정규 SpCas9의 위치 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 1000-1100, 1100-1200, 1200-1300, 또는 1300-1368 사이의 임의의 위치에 위치하는 잔기의 임의의 2개의 쌍에 상응하는 잔기의 쌍에 위치하는 분할 부위에서 2개의 단편으로 분할된다.

[0515] 특정 실시양태에서, SpCas9는 서열식별번호: 18의 정규 SpCas9의 잔기 1과 2 사이, 또는 2와 3 사이, 또는 3과 4 사이, 또는 4와 5 사이, 또는 5와 6 사이, 또는 6과 7 사이, 또는 7과 8 사이, 또는 8과 9 사이, 또는 9와 10 사이, 또는 잔기 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 1000-1100, 1100-1200, 1200-1300, 또는 1300-1368 사이의 임의의 위치에 위치하는 잔기의 임의의 2개 쌍 사이에 위치하는 분할 부위에서 2개의 단편으로 분할된다. 특정 실시양태에서, 분할 부위는 1개 이상의 폴리펩티드 결합 부위 (즉, "분할 부위 또는 분할-인테인 분할 부위")에 위치하고, 분할 인테인에 융합된 다음, 개별적으로-코딩된 융합 단백질로서 세포에 전달된다. 분할-인테인 융합 단백질 (즉, 단백질 절반)이 세포 내에서 발현되면, 단백질은 트랜스-스플라이싱을 거쳐, 연결된 분할-인테인 서열의 동반 제거와 함께 완전한 또는 전체 PE를 형성한다.

[0516] 예를 들어, 도 66에 제시된 바와 같이, N-말단 엑스테인은 제1 분할-인테인 (예를 들어, N 인테인)에 융합될 수 있고, C-말단 엑스테인은 제2 분할-인테인 (예를 들어, C 인테인)에 융합될 수 있다. N-말단 엑스테인은 C-말

단 엑스테인에 융합되어 세포 내에서 N 인테인 및 C 인테인의 자기-회합 시 napDNAbp 도메인 및 폴리머라제 도메인 (예를 들어, RT 도메인)을 포함하는 전체 프라임 편집제 융합 단백질을 재형성하고, 이어서 그의 자기-절제, 및 전체 프라임 편집제 (PE)의 N-말단 엑스테인 및 C-말단 엑스테인 부분 사이의 펩티드 결합의 동반 형성이 이어진다.

[0517] 분할-인테인을 사용한 분할-PE 전달 전략을 이용하기 위해, 프라임 편집제는 1개 이상의 분할 부위에서 나뉘어져서 프라임 편집제의 적어도 2개의 별개의 절반을 생성할 필요가 있으며, 이들 각각은 각각의 절반이 분할-인테인 서열에 융합되는 경우에 세포 내부에서 재연결될 수 있다.

[0518] 특정 실시양태에서, 프라임 편집제는 단일 분할 부위에서 분할된다. 특정의 다른 실시양태에서, 프라임 편집제는 2개의 분할 부위, 또는 3개의 분할 부위, 또는 4개의 분할 부위, 또는 그 초과에서 분할된다.

[0519] 바람직한 실시양태에서, 프라임 편집제는 단일 분할 부위에서 분할되어 프라임 편집제의 2개의 별개의 절반을 생성하고, 이들 각각은 분할 인테인 서열에 융합될 수 있다.

[0520] 예시적인 분할 인테인은 2개의 서브유닛, 즉 DnaE-N 및 DnaE-C를 포함하는 Ssp DnaE 인테인이다. 2개의 상이한 서브유닛은 별개의 유전자, 즉 각각 DnaE-N 및 DnaE-C 서브유닛을 코딩하는 dnaE-n 및 dnaE-c에 의해 코딩된다. DnaE는 시네코시티스 종 PCC6803의 자연 발생 분할 인테인이고, 각각 DnaE-N 또는 DnaE-C와의 융합체를 포함한 2개의 별개의 단백질의 트랜스-스플라이싱을 지시할 수 있다.

[0521] 추가의 자연 발생 또는 조작된 분할-인테인 서열은 관련 기술분야에 공지되어 있거나, 또는 본원에 기재된 전체-인테인 서열 또는 관련 기술분야에서 입수가능한 것으로부터 제조될 수 있다. 분할-인테인 서열의 예는 문헌 [Stevens et al., "A promiscuous split intein with expanded protein engineering applications," PNAS, 2017, Vol.114: 8538-8543; Iwai et al., "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme, FEBS Lett, 580: 1853-1858]에서 찾아볼 수 있으며, 이들 각각은 본원에 참조로 포함된다. 추가의 분할 인테인 서열은, 예를 들어 WO 2013/045632, WO 2014/055782, WO 2016/069774, 및 EP2877490에서 찾아볼 수 있으며, 이들 각각의 내용은 본원에 참조로 포함된다.

[0522] 또한, 생체내 및 시험관내 트랜스로의 단백질 스플라이싱이 기재되어 있으며 (Shingledecker, et al., Gene 207:187 (1998), Southworth, et al., EMBO J. 17:918 (1998); Mills, et al., Proc. Natl. Acad. Sci. USA, 95:3543-3548 (1998); Lew, et al., J. Biol. Chem., 273:15887-15890 (1998); Wu, et al., Biochim. Biophys. Acta 35732:1 (1998b), Yamazaki, et al., J. Am. Chem. Soc. 120:5591 (1998), Evans, et al., J. Biol. Chem. 275:9091 (2000); Otomo, et al., Biochemistry 38:16040-16044 (1999); Otomo, et al., J. Biolmol. NMR 14:105-114 (1999); Scott, et al., Proc. Natl. Acad. Sci. USA 96:13638-13643 (1999)), 이는 예를 들어 2개의 개별적으로-발현된 절반으로부터 완전한 PE 융합 단백질의 형성에 관하여 도 66 및 67에 제시된 바와 같이, 기능적 생성물을 형성하기 위해 후속적으로 라이게이션을 거치는 2개의 불활성 단편에 대해 단백질로서의 발현 기회를 제공한다.

[0523] 본원에 기재된 다양한 실시양태에서, 연속 진화 방법 (예를 들어, PACE)은 염기 편집제의 제1 부분을 진화시키는데 사용될 수 있다. 제1 부분은 단일 성분 또는 도메인, 예를 들어 Cas9 도메인, 데아미나제 도메인, 또는 UGI 도메인을 포함할 수 있다. 이어서, 개별적으로 진화된 성분 또는 도메인은 분할-인테인 폴리펩티드 도메인과 함께 진화된 부분 및 나머지 비-진화된 부분 둘 다를 개별적으로 발현시킴으로써 세포 내에서 염기 편집제의 나머지 부분에 융합될 수 있다. 제1 부분은 보다 넓게는 본원에 기재된 연속 진화 방법을 사용하여 진화시키는 것을 목적으로 하는 염기 편집제의 임의의 제1 아미노산 부분을 포함할 수 있다. 이러한 실시양태에서 제2 부분은 본원의 방법을 사용하여 진화되지 않는 염기 편집제의 나머지 아미노산 부분을 지칭할 것이다. 염기 편집제의 진화된 제1 부분 및 제2 부분은 각각 세포에서 분할-인테인 폴리펩티드 도메인과 함께 발현될 수 있다. 세포의 천연 단백질 스플라이싱 메카니즘은 진화된 제1 부분 및 비-진화된 제2 부분을 재조립하여 단일 융합 단백질 진화된 염기 편집제를 형성할 것이다. 진화된 제1 부분은 단일 융합 단백질의 N- 또는 C-말단 부분을 포함할 수 있다. 유사한 방식으로, 제2 직교 트랜스-스플라이싱 인테인 쌍의 사용은 진화된 제1 부분이 단일 융합 단백질의 내부 부분을 포함하도록 할 수 있다.

[0524] 따라서, 본원에 기재된 염기 편집제의 진화된 및 비-진화된 성분 중 임의의 것은 세포 내에서 진화된 및 비-진화된 성분을 포함하는 완전한 염기 편집제의 형성을 용이하게 하기 위해 분할-인테인 태그와 함께 발현될 수 있다.

[0525] 단백질 스플라이싱 과정의 메카니즘은 매우 상세하게 연구되었고 (Chong, et al., J. Biol. Chem. 1996, 271,

22159-22168; Xu, M-Q & Perler, F. B. EMBO Journal, 1996, 15, 5146-5153), 보존된 아미노산은 인테인 및 엑스테인 스플라이싱 지점에서 발견되었다 (Xu, et al., EMBO Journal, 1994, 13 5517-522). 본원에 기재된 구축물은 제1 유전자의 5'-말단 (예를 들어, 염기 편집제의 진화된 부분)에 융합된 인테인 서열을 함유한다. 적합한 인테인 서열은 단백질 스플라이싱 요소를 함유하는 것으로 공지된 단백질 중 임의의 것으로부터 선택될 수 있다. 모든 공지된 인테인을 함유하는 데이터베이스는 월드 와이드 웹에서 찾아볼 수 있다 (Perler, F. B. Nucleic Acids Research, 1999, 27, 346-347). 인테인 서열은 제2 유전자의 3' 단부에서 5' 단부에 융합된다. 이러한 유전자를 특정 소기관으로 표적화하기 위해, 펩티드 신호가 유전자의 코딩 서열에 융합될 수 있다. 제2 유전자 후에, 인테인-유전자 서열은 동일한 세포에서 다중 단백질의 발현을 위해 목적하는 만큼 자주 반복될 수 있다. 다중-인테인 함유 구축물의 경우, 상이한 공급원으로부터의 인테인 요소를 사용하는 것이 유용할 수 있다. 발현될 마지막 유전자의 서열 후에, 전사 종결 서열이 삽입되어야 한다. 한 실시양태에서, 변형된 인테인 스플라이싱 유닛은 인테인으로부터 엑스테인의 절제를 촉매할 수 있을뿐만 아니라 또한 엑스테인의 라이게이션을 방지할 수 있도록 설계된다. 피로코쿠스(Pyrococcus) 종 GB-D DNA 폴리머라제에서의 C-말단 엑스테인 접합부의 돌연변이유발은 엑스테인 및 인테인의 절단을 유도하지만 엑스테인의 후속 라이게이션은 방지하는 변경된 스플라이싱 요소를 생성하는 것으로 밝혀졌다 (Xu, M-Q & Perler, F. B. EMBO Journal, 1996, 15, 5146-5153). 세린 538의 알라닌 또는 글리신으로의 돌연변이는 절단을 유도하였지만 라이게이션은 방지하였다. 다른 인테인 스플라이싱 유닛에서의 등가 잔기의 돌연변이는 또한 인테인에 대한 C-말단 엑스테인 접합부에서의 아미노산의 보존으로 인해 엑스테인 라이게이션을 방지할 것이다. 엔도뉴클레아제 도메인을 함유하지 않는 바람직한 인테인은 미코박테리움 크세노피(Mycobacterium xenopi) GyrA 단백질이다 (Telenti, et al., J. Bacteriol. 1997, 179, 6378-6382). 그 밖의 것이 자연에서 발견되었거나 또는 엔도뉴클레아제 도메인을 엔도뉴클레아제 함유 인테인으로부터 제거함으로써 인공적으로 생성되었다 (Chong, et al., J. Biol. Chem. 1997, 272, 15587-15590). 바람직한 실시양태에서, 인테인은, 미코박테리움 크세노피 GyrA 단백질로부터의 인테인과 같이, 스플라이싱 기능을 수행하는데 필요한 최소 개수의 아미노산으로 이루어지도록 선택된다 (Telenti, A., et al., J. Bacteriol. 1997, 179, 6378-6382). 대안적 실시양태에서, 엔도뉴클레아제 활성이 없는 인테인, 예컨대 미코박테리움 크세노피 GyrA 단백질로부터의 인테인 또는 엔도뉴클레아제 도메인을 제거하도록 변형된 사카로미세스 세레비지아에(Saccharomyces cerevisiae) VMA 인테인으로부터 선택된다 (Chong, 1997). 인테인 스플라이싱 유닛의 추가의 변형은 절단 반응의 반응 속도가 변경되도록 하여, 스플라이싱 유닛의 유전자 서열을 단순히 변형시키는 것에 의해 단백질 투여량이 제어되도록 할 수 있다.

[0526] 인테인은 또한 2개의 개별적으로 전사되고 번역되는 유전자에 의해 코딩된 2개의 단편으로서 존재할 수 있다. 이들 소위 분할 인테인은 자기-회합하고, 단백질-스플라이싱 활성을 트랜스로 촉매한다. 분할 인테인은 다양한 시아노박테리아 및 고세균에서 확인되었지만 (Caspi et al, Mol Microbiol. 50: 1569-1577 (2003); Choi J. et al, J Mol Biol. 556: 1093-1106 (2006.); Dassa B. et al, Biochemistry. 46:322-330 (2007.); Liu X. and Yang J., J Biol Chem. 275:26315-26318 (2003); Wu H. et al., Proc Natl Acad Sci USA. £5:9226-9231 (1998.); 및 Zettler J. et al, FEBS Letters. 553:909-914 (2009)), 지금까지 진핵생물에서는 발견되지 않았다. 최근에, 환경 메타게놈 데이터의 생물정보학적 분석은 신규 게놈 배열을 갖는 26개의 상이한 유전자좌를 밝혀내었다. 각각의 유전자좌에서, 보존된 효소 코딩 영역은 분할 인테인에 의해 개재되고, 인테인 서브도메인을 코딩하는 섹션 사이에 자립형 엔도뉴클레아제 유전자가 삽입된다. 이들 중에서, 5개의 유전자좌가 완전히 조립되었다: DNA 헬리카제 (gp41-1, gp41-8); 이노신-5'-모노포스페이트 데히드로게나제 (IMPDH-1); 및 리보뉴클레오티드 리덕타제 촉매 서브유닛 (NrdA-2 및 NrdJ-1). 이러한 파괴된 유전자 조직은 주로 파지에 존재하는 것으로 보인다 (Dassa et al., Nucleic Acids Research. 57:2560-2573 (2009)).

[0527] 분할 인테인 Npu DnaE는 단백질 트랜스-스플라이싱 반응에 대해 보고된 최고 속도를 갖는 것으로 특징화되었다. 또한, Npu DnaE 단백질 스플라이싱 반응은 상이한 엑스테인 서열, 6 내지 37℃의 온도, 및 최대 6M의 우레아의 존재와 관련하여 강건하고 고수율인 것으로 간주된다 (Zettler J. et al., FEBS Letters. 553:909-914 (2009); Iwai I. et al., FEBS Letters 550: 1853-1858 (2006)). 예상된 바와 같이, 이들 인테인의 N-도메인에서 Cys1 Ala 돌연변이가 도입되었을 때, 초기 N에서 S- 아실 이동이 차단되었고, 따라서 단백질 스플라이싱이 차단되었다. 불행하게도, C-말단 절단 반응도 또한 거의 완전히 억제되었다. N-말단 절단가능한 펩티드 결합에서의 아실 이동에 대한 C-말단 스플라이스 접합부에서의 아스파라긴 고리화의 의존성은 자연 분할 DnaE 인테인 대립유전자에 공통적인 고유한 특성인 것으로 보인다 (Zettler J. et al., FEBS Letters. 555:909-914 (2009)).

[0528] 단백질 스플라이싱의 메카니즘은 전형적으로 4개의 단계를 갖는다 [29-30]: 1) 상류 펩티드 결합을 파괴하고 N-엑스테인과 인테인의 제1 아미노산 (Cys 또는 Ser)의 측쇄 사이에 에스테르 결합을 형성하는, 인테인 N-말단에서의 N-S 또는 N-O 아실 이동; 2) N-엑스테인을 C-엑스테인의 제1 아미노산 (Cys, Ser, 또는 Thr)의 측쇄에 연

결하는 새로운 에스테르 결합을 형성하는, N-엑스테인을 인테인 C-말단으로 재배치하는 에스테르교환; 3) 인테인과 C-엑스테인 사이의 펩티드 결합을 파괴하는 Asn 고리화; 및 4) 에스테르 결합을 N-엑스테인과 C-엑스테인 사이의 펩티드 결합으로 대체하는 S-N 또는 O-N 아실 이동.

[0529] 분할 인테인에 의해 촉매되는 단백질 트랜스-스플라이싱은 전적으로 단백질 라이게이션을 위한 효소적 방법을 제공한다 [31]. 분할-인테인은 본질적으로 각각 N-인테인 및 C-인테인으로 명명되는 2개의 조각으로 분할된 인접 인테인 (예를 들어 미니-인테인)이다. 분할 인테인의 N-인테인 및 C-인테인은 비-공유적으로 회합되어 활성 인테인을 형성하고, 인접 인테인과 본질적으로 동일한 방식으로 스플라이싱 반응을 촉매할 수 있다. 분할 인테인은 자연에서 발견되었고, 또한 실험실에서 조작되었다 [31-35]. 본원에 사용된 용어 "분할 인테인"은 N-말단 및 C-말단 서열이 트랜스-스플라이싱 반응에 대해 기능적인 인테인으로 비-공유적으로 재회합 또는 재구성될 수 있는 별개의 분자가 되도록 N-말단 및 C-말단 아미노산 서열 사이에 1개 이상의 펩티드 결합 파괴가 존재하는 임의의 인테인을 지칭한다. 임의의 촉매 활성 인테인 또는 그의 단편은 본 발명의 방법에 사용하기 위한 분할 인테인을 유도하는 데 사용될 수 있다. 예를 들어, 한 측면에서 분할 인테인은 진핵 인테인으로부터 유래될 수 있다. 또 다른 측면에서, 분할 인테인은 박테리아 인테인으로부터 유래될 수 있다. 또 다른 측면에서, 분할 인테인은 고세균 인테인으로부터 유래될 수 있다. 바람직하게는, 이와 같이 유래된 분할 인테인은 트랜스-스플라이싱 반응을 촉매하는데 필수적인 아미노산 서열만을 보유할 것이다.

[0530] 본원에 사용된 "N-말단 분할 인테인 (In)"은 트랜스-스플라이싱 반응에 대해 기능적인 N-말단 아미노산 서열을 포함하는 임의의 인테인 서열을 지칭한다. 따라서, In은 또한 트랜스-스플라이싱이 발생할 때 스플라이싱되는 서열을 포함한다. In은 자연 발생 인테인 서열의 N-말단 부분의 변형인 서열을 포함할 수 있다. 예를 들어, In은, 추가의 및/또는 돌연변이된 잔기의 포함이 In을 트랜스-스플라이싱에서 비-기능적으로 만들지 않는 한, 이러한 추가의 아미노산 잔기 및/또는 돌연변이된 잔기를 포함할 수 있다. 바람직하게는, 추가의 및/또는 돌연변이된 잔기의 포함은 In의 트랜스-스플라이싱 활성을 개선 또는 증진시킨다.

[0531] 본원에 사용된 "C-말단 분할 인테인 (Ic)"은 트랜스-스플라이싱 반응에 대해 기능적인 C-말단 아미노산 서열을 포함하는 임의의 인테인 서열을 지칭한다. 한 측면에서, Ic는 4 내지 7개의 인접 아미노산 잔기를 포함하고, 이 중 적어도 4개의 아미노산은 그가 유래된 인테인의 마지막 β-가닥으로부터의 것이다. 따라서, Ic는 또한 트랜스-스플라이싱이 발생할 때 스플라이싱되는 서열을 포함한다. Ic는 자연 발생 인테인 서열의 C-말단 부분의 변형인 서열을 포함할 수 있다. 예를 들어, Ic는, 추가의 및/또는 돌연변이된 잔기의 포함이 In을 트랜스-스플라이싱에서 비-기능적으로 만들지 않는 한, 이러한 추가의 아미노산 잔기 및/또는 돌연변이된 잔기를 포함할 수 있다. 바람직하게는, 추가의 및/또는 돌연변이된 잔기의 포함은 Ic의 트랜스-스플라이싱 활성을 개선 또는 증진시킨다.

[0532] 본 발명의 일부 실시양태에서, Ic 또는 In에 연결된 펩티드는 특히 형광 기, 비오틴, 폴리에틸렌 글리콜 (PEG), 아미노산 유사체, 비천연 아미노산, 포스페이트 기, 글리코실 기, 방사성동위원소 표지, 및 제약 분자를 포함한 추가의 화학적 모이어티를 포함할 수 있다. 다른 실시양태에서, Ic에 연결된 펩티드는 특히 케톤, 알데히드, Cys 잔기 및 Lys 잔기를 포함하는 1개 이상의 화학적 반응성 기를 포함할 수 있다. 분할 인테인의 N-인테인 및 C-인테인은 비-공유적으로 회합되어 활성 인테인을 형성하고, "인테인-스플라이싱 폴리펩티드 (ISP)"가 존재하는 경우에 스플라이싱 반응을 촉매할 수 있다. 본원에 사용된 "인테인-스플라이싱 폴리펩티드 (ISP)"는 Ic, In, 또는 둘 다가 분할 인테인으로부터 제거될 때 남는 분할 인테인의 아미노산 서열의 부분을 지칭한다. 특정 실시양태에서, In은 ISP를 포함한다. 또 다른 실시양태에서, Ic는 ISP를 포함한다. 또 다른 실시양태에서, ISP는 In에도 Ic에도 공유 연결되지 않은 별개의 펩티드이다.

[0533] 분할 인테인은 미니-인테인의 구조에서 발견되는 ~12개의 보존된 베타-가닥 사이의 비구조화된 루프 또는 개재 아미노산 서열 내의 1개 이상의 분할 부위를 조작함으로써 인접 인테인으로부터 생성될 수 있다 [25-28]. 분할의 생성이 단백질 스플라이싱 활성이 상실되는데 충분한 정도로 인테인, 특히 구조화된 베타-가닥의 구조를 파괴하지 않을 것이라면, 베타-가닥 사이의 영역 내의 분할 부위의 위치에 일부 유연성이 존재할 수 있다.

[0534] 단백질 트랜스-스플라이싱에서, 하나의 전구체 단백질은 N-엑스테인 부분에 이어 N-인테인으로 이루어지고, 또 다른 전구체 단백질은 C-인테인 부분에 이어 C-엑스테인으로 이루어지고, 트랜스-스플라이싱 반응 (N- 및 C-인테인에 의해 함께 촉매됨)은 2개의 인테인 서열을 절제하고, 2개의 엑스테인 서열을 펩티드 결합으로 연결한다. 효소 반응인 단백질 트랜스-스플라이싱은 매우 낮은 (예를 들어, 마이크로몰) 농도의 단백질로 작용할 수 있고, 생리학적 조건 하에 수행될 수 있다.

[0535] [2] 다른 프로그램가능한 뉴클레아제

[0536] 본원에 기재된 다양한 실시양태에서, 다중-플랩 프라임 편집제는 napDNAbp, 예컨대 Cas9 단백질을 포함한다. 이들 단백질은, 가이드 RNA (또는 PEgRNA)의 스페이서 부분에 상보적인 서열을 보유하고 또한 요구되는 PAM 서열을 보유하는 DNA 상의 표적 부위로 Cas9 단백질을 가이드하는 gRNA (또는 경우에 따라 PEgRNA)와 복합체화되는 방식에 의해 "프로그램가능하다". 그러나, 본원에서 고려되는 특정 실시양태에서, napDNAbp는 상이한 유형의 프로그램가능한 단백질, 예컨대 아연 핑거 뉴클레아제 또는 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)로 치환될 수 있다.

[0537] 도 1j는 napDNAbp (예를 들어, SpCas9 닉카제)를 임의의 프로그램가능한 뉴클레아제 도메인, 예컨대 아연 핑거 뉴클레아제 (ZFN) 또는 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)로 대체한, 본원에서 고려되는 프라임 편집의 이러한 변경을 도시한다. 따라서, 적합한 뉴클레아제가 반드시 핵산 표적화 분자 (예컨대 가이드 RNA)에 의해 "프로그램화"될 필요는 없고, 오히려 DNA-결합 도메인, 예컨대 특히 뉴클레아제의 특이성을 규정함으로써 프로그램화될 수 있는 것으로 고려된다. napDNAbp 모이어티를 사용한 프라임 편집에서와 같이, 이러한 대안적 프로그램가능한 뉴클레아제는 표적 DNA의 한 가닥만이 절단되도록 변형되는 것이 바람직하다. 다시 말해서, 프로그램가능한 뉴클레아제는 바람직하게는 닉카제로서 기능해야 한다. 일단 프로그램가능한 뉴클레아제가 선택되면 (예를 들어, ZFN 또는 TALEN), 추가의 기능성을 시스템 내로 조작하여 이것이 프라임 편집-유사 메카니즘에 따라 작동하도록 할 수 있다. 예를 들어, 프로그램가능한 뉴클레아제는 RNA 또는 DNA 연장 아암을 (예를 들어, 화학적 링커를 통해) 그에 커플링시킴으로써 변형될 수 있으며, 여기서 연장 아암은 프라이머 결합 부위 (PBS) 및 DNA 합성 주형을 포함한다. 프로그램가능한 뉴클레아제는 또한 폴리머라제에 (예를 들어, 화학적 또는 아미노산 링커를 통해) 커플링될 수 있으며, 그의 속성은 연장 아암이 DNA인지 또는 RNA인지에 좌우될 것이다. RNA 연장 아암의 경우에, 폴리머라제는 RNA-의존성 DNA 폴리머라제 (예를 들어, 리버스 트랜스크립타제)일 수 있다. DNA 연장 아암의 경우에, 폴리머라제는 DNA-의존성 DNA 폴리머라제 (예를 들어, Pol I, Pol II, 또는 Pol III를 포함한 원핵 폴리머라제, 또는 Pol a, Pol b, Pol g, Pol d, Pol e, 또는 Pol z를 포함한 진핵 폴리머라제)일 수 있다. 시스템은 또한 프로그램가능한 뉴클레아제에 대한 융합체로서 부가되거나 또는 트랜스로 부가되어 반응을 전체로서 용이하게 하기 위한 다른 기능성을 포함할 수 있다 (예를 들어, (a) 절단 부위에서 DNA를 풀어내어 3' 단부를 갖는 절단 가닥을 프라이머로서 이용가능하게 하는 헬리카제, (b) 절단 가닥 상의 내인성 가닥을 제거하도록 도와 합성된 가닥으로 내인성 가닥을 대체하는 쪽으로 반응을 구동하는 FEN1, 또는 (c) 대향 가닥 상에 제2 부위 닉을 생성하여, 비-편집된 가닥의 선호되는 세포 복구를 통해 합성 복구의 통합을 구동하는 것을 도울 수 있는 nCas9:gRNA 복합체). napDNAbp에 의한 프라임 편집과 유사한 방식으로, 달리 프로그램가능한 뉴클레아제와의 이러한 복합체를 사용하여 관심 편집을 보유하는 DNA의 새로 합성된 대체 가닥을 합성한 다음, 이를 DNA의 표적 부위 내로 영구적으로 설치할 수 있다.

[0538] DNA의 표적 부위에 선택적으로 결합하도록 프로그램화될 수 있고, 폴리머라제, 및 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 RNA 또는 DNA 연장 아암을 특이적 닉 부위에 공동-국재화하도록 상기 기재된 방식으로 추가로 변형될 수 있는 대안적 프라임 편집제 시스템을 구축하기 위해 napDNAbp:gRNA 복합체 대신 사용될 수 있는 적합한 대안적 프로그램가능한 뉴클레아제가 관련 기술분야에 널리 공지되어 있다. 예를 들어, 도 1j에 나타내어진 바와 같이, 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)는 본원에 기재된 다중-플랩 프라임 편집 방법 및 물질의 조성물에서 프로그램가능한 뉴클레아제로서 사용될 수 있다. TALEN은 TAL 이펙터 DNA 결합 도메인을 DNA 절단 도메인에 융합시킴으로써 생성된 인공 제한 효소이다. 이들 시약은 효율적이고, 프로그램가능하고, 특이적인 DNA 절단을 가능하게 하고, 이는 계내 게놈 편집을 위한 강력한 도구를 나타낸다. 전사 활성인자-유사 이펙터 (TALE)는 실제로 임의의 DNA 서열에 결합하도록 신속하게 조작될 수 있다. 본원에 사용된 용어 TALEN은 광범위하고, 또 다른 TALEN으로부터의 보조 없이 이중 가닥 DNA를 절단할 수 있는 단량체 TALEN을 포함한다. 용어 TALEN은 또한 함께 작용하여 동일한 부위에서 DNA를 절단하도록 조작된 한 쌍의 TALEN의 하나 또는 둘 다의 구성원을 지칭하는 데 사용된다. 함께 작용하는 TALEN은 좌측-TALEN 및 우측-TALEN으로 지칭될 수 있으며, 이는 DNA의 방향성을 지칭한다. 미국 일련 번호 12/965,590; 미국 일련 번호 13/426,991 (미국 특허 번호 8,450,471); 미국 일련 번호 13/427,040 (미국 특허 번호 8,440,431); 미국 일련 번호 13/427,137 (미국 특허 번호 8,440,432); 및 미국 일련 번호 13/738,381을 참조하며, 이들 모두는 그 전문이 본원에 참조로 포함된다. 또한, TALEN은 WO 2015/027134, US 9,181,535, 문헌 [Boch et al., "Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors", Science, vol. 326, pp. 1509-1512 (2009), Bogdanove et al., TAL Effectors: Customizable Proteins for DNA Targeting, Science, vol. 333, pp. 1843-1846 (2011), Cade et al., "Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs", Nucleic Acids Research, vol. 40, pp. 8001-8010 (2012), 및 Cermak et al., "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting", Nucleic Acids

Research, vol. 39, No. 17, e82 (2011)]에 기재되어 있고, 이들 각각은 본원에 참조로 포함된다.

[0539] 도 1j에 나타낸 바와 같이, 아연 핑거 뉴클레아제도 또한 napDNAbp, 예컨대 Cas9 닉카제 대신 다중-플랩 프라임 편집에 사용하기 위한 대안적 프로그램가능한 뉴클레아제로서 사용될 수 있다. TALEN과 마찬가지로, ZFN 단백질은 이들이 닉카제로서 기능하도록 변형될 수 있으며, 즉, 본원에 기재된 다중-플랩 프라임 편집제와 함께 사용되는 napDNAbp와 유사한 방식으로 표적 DNA의 오직 한 가닥만을 절단하도록 ZFN이 조작될 수 있다. ZFN 단백질은 관련 기술분야에, 예를 들어, 문헌 [Carroll et al., "Genome Engineering with Zinc-Finger Nucleases," Genetics, Aug 2011, Vol. 188: 773-782; Durai et al., "Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells," Nucleic Acids Res, 2005, Vol. 33: 5978-90; 및 Gaj et al., "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering," Trends Biotechnol. 2013, Vol.31: 397-405]에 광범위하게 기재되어 있으며, 이들 각각은 그 전문이 본원에 참조로 포함된다.

[0540] [3] 폴리머라제 (예를 들어, 리버스 트랜스크립타제)

[0541] 다양한 실시양태에서, 본원에 개시된 다중-플랩 프라임 편집제 시스템은 폴리머라제 (예를 들어, DNA-의존성 DNA 폴리머라제 또는 RNA-의존성 DNA 폴리머라제, 예컨대 리버스 트랜스크립타제), 또는 그의 변이체를 포함하며, 이는 napDNAbp 또는 다른 프로그램가능한 뉴클레아제와의 융합 단백질로서 제공될 수 있거나, 또는 트랜스로 제공될 수 있다.

[0542] 임의의 폴리머라제가 본원에 개시된 다중-플랩 프라임 편집제에 사용될 수 있다. 폴리머라제는 야생형 폴리머라제, 기능적 단편, 돌연변이체, 변이체, 또는 말단절단된 변이체 등일 수 있다. 폴리머라제는 진핵, 원핵, 고세균, 또는 바이러스 유기체로부터의 야생형 폴리머라제를 포함할 수 있고/거나, 폴리머라제는 유전자 조작, 돌연변이유발, 방향적 진화-기반 과정에 의해 변형될 수 있다. 폴리머라제는 T7 DNA 폴리머라제, T5 DNA 폴리머라제, T4 DNA 폴리머라제, 클레나우 단편 DNA 폴리머라제, DNA 폴리머라제 III 등을 포함할 수 있다. 폴리머라제는 또한 열안정성일 수 있고, Taq, Tne, Tma, Pfu, Tfl, Tth, 스토펠 단편, VENT® 및 DEEPVENT® DNA 폴리머라제, KOD, Tgo, JDF3, 및 그의 돌연변이체, 변이체 및 유도체를 포함할 수 있다 (미국 특허 번호 5,436,149; 미국 특허 번호 4,889,818; 미국 특허 번호 4,965,185; 미국 특허 번호 5,079,352; 미국 특허 번호 5,614,365; 미국 특허 번호 5,374,553; 미국 특허 번호 5,270,179; 미국 특허 번호 5,047,342; 미국 특허 번호 5,512,462; WO 92/06188; WO 92/06200; WO 96/10640; 문헌 [Barnes, W. M., Gene 112:29-35 (1992); Lawyer, F. C., et al., PCR Meth. Appl. 2:275-287 (1993); Flaman, J.-M, et al., Nuc. Acids Res. 22(15):3259-3260 (1994)] 참조, 이들 각각은 참조로 포함됨). 보다 긴 핵산 분자 (예를 들어, 약 3-5 Kb 길이보다 긴 핵산 분자)의 합성을 위해, 적어도 2종의 DNA 폴리머라제가 사용될 수 있다. 특정 실시양태에서, 폴리머라제 중 1종은 3' 엑소뉴클레아제 활성이 실질적으로 결여될 수 있고, 다른 것은 3' 엑소뉴클레아제 활성을 가질 수 있다. 이러한 쌍형성은 동일하거나 상이한 폴리머라제를 포함할 수 있다. 3' 엑소뉴클레아제 활성이 실질적으로 결여된 DNA 폴리머라제의 예는 Taq, Tne(엑소-), Tma(엑소-), Pfu(엑소-), Pwo(엑소-), 엑소-KOD 및 Tth DNA 폴리머라제, 및 그의 돌연변이체, 변이체 및 유도체를 포함하나 이에 제한되지는 않는다.

[0543] 바람직하게는, 본원에 개시된 다중-플랩 프라임 편집제에 사용가능한 폴리머라제는 "주형-의존성" 폴리머라제이다 (폴리머라제는 프라임 편집 동안 합성 하의 DNA 가닥의 서열을 특정하기 위해 DNA 합성 주형에 의존하도록 의도되기 때문). 본원에 사용된 용어 "주형 DNA 분자"는, 예를 들어 PEgRNA의 DNA 합성 주형의 프라이머 연장 반응에서 DNA 폴리머라제에 의해 상보적 핵산 가닥이 합성되는 핵산의 가닥을 지칭한다.

[0544] 본원에 사용된 용어 "주형 의존성 방식"은 프라이머 분자의 주형 의존성 연장 (예를 들어, DNA 폴리머라제에 의한 DNA 합성)을 수반하는 과정을 지칭하는 것으로 의도된다. 용어 "주형 의존성 방식"은 폴리뉴클레오티드의 새로 합성되는 가닥의 서열이 상보적 염기 쌍형성의 널리 공지된 규칙에 의해 좌우되는 RNA 또는 DNA의 폴리뉴클레오티드 합성을 지칭한다 (예를 들어, 문헌 [Watson, J. D. et al., In: Molecular Biology of the Gene, 4th Ed., W. A. Benjamin, Inc., Menlo Park, Calif. (1987)] 참조). 용어 "상보적"은 염기-쌍형성을 통한 2개의 폴리뉴클레오티드 가닥의 영역 사이 또는 2개의 뉴클레오티드 사이의 서열 상보성의 광범위한 개념을 지칭한다. 아데닌 뉴클레오티드는 티민 또는 우라실인 뉴클레오티드와 특이적 수소 결합 ("염기 쌍형성")을 형성할 수 있는 것으로 공지되어 있다. 유사하게, 시토신 뉴클레오티드는 구아닌 뉴클레오티드와 염기 쌍형성을 할 수 있는 것으로 공지되어 있다. 따라서, 프라임 편집의 경우에, DNA 합성 주형에 대해 프라임 편집제의 폴리머라제에 의해 합성된 DNA의 단일 가닥은 DNA 합성 주형의 서열에 "상보적"인 것으로 언급된다고 언급될 수 있다.

[0545] A. 예시적인 폴리머라제

[0546] 다양한 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제는 폴리머라제를 포함한다. 개시내용은 임의의 자연 발생 유기체 또는 바이러스로부터 수득되거나, 또는 상업적 또는 비-상업적 공급원으로부터 수득된 임의의 야생형 폴리머라제를 고려한다. 추가로, 개시내용의 다중-플랩 프라임 편집제에 사용가능한 폴리머라제는 임의의 자연 발생 돌연변이체 폴리머라제, 조작된 돌연변이체 폴리머라제, 또는 기능을 보유하는 말단절단된 변이체를 포함한 다른 변이체 폴리머라제를 포함할 수 있다. 본원에 사용가능한 폴리머라제는 또한 특정 아미노산 치환, 예컨대 본원에 구체적으로 개시된 것을 함유하도록 조작될 수 있다. 특정의 바람직한 실시양태에서, 개시내용의 다중-플랩 프라임 편집제에 사용가능한 폴리머라제는 주형-기반 폴리머라제이며, 즉 이들은 주형-의존성 방식으로 뉴클레오티드 서열을 합성한다.

[0547] 폴리머라제는 뉴클레오티드 가닥을 합성하고 본원에 기재된 다중-플랩 프라임 편집제 시스템과 관련하여 사용될 수 있는 효소이다. 폴리머라제는 바람직하게는 "주형-의존성" 폴리머라제 (즉, 주형 가닥의 뉴클레오티드 염기의 순서에 기초하여 뉴클레오티드 가닥을 합성하는 폴리머라제)이다. 특정 구성에서, 폴리머라제는 또한 "주형-비의존성" (즉, 주형 가닥의 필요 없이 뉴클레오티드 가닥을 합성하는 폴리머라제)일 수 있다. 폴리머라제는 또한 "DNA 폴리머라제" 또는 "RNA 폴리머라제"로 추가로 카테고리화될 수 있다. 다양한 실시양태에서, 다중-플랩 프라임 편집제 시스템은 DNA 폴리머라제를 포함한다. 다양한 실시양태에서, DNA 폴리머라제는 "DNA-의존성 DNA 폴리머라제" (즉, 주형 분자가 DNA의 가닥임)일 수 있다. 이러한 경우에, DNA 주형 분자는 PEgRNA일 수 있으며, 여기서 연장 아암은 DNA의 가닥을 포함한다. 이러한 경우에, PEgRNA는 RNA 부분 (즉, 스페이서 및 gRNA 코어를 포함한 가이드 RNA 성분) 및 DNA 부분 (즉, 연장 아암)을 포함하는 키메라 또는 하이브리드 PEgRNA로 지칭될 수 있다. 다양한 다른 실시양태에서, DNA 폴리머라제는 "RNA-의존성 DNA 폴리머라제" (즉, 주형 분자가 RNA의 가닥임)일 수 있다. 이러한 경우에, PEgRNA는 RNA, 즉 RNA 연장부를 포함하는 RNA이다. 용어 "폴리머라제"는 또한 뉴클레오티드의 중합 (즉, 폴리머라제 활성)을 촉매하는 효소를 지칭할 수 있다. 일반적으로, 효소는 폴리뉴클레오티드 주형 서열에 어닐링된 프라이머 (예를 들어, 예컨대 PEgRNA의 프라이머 결합 부위에 어닐링된 프라이머 서열)의 3'-단부에서 합성을 개시할 것이고, 주형 가닥의 5' 단부쪽으로 진행할 것이다. "DNA 폴리머라제"는 데옥시뉴클레오티드의 중합을 촉매한다. DNA 폴리머라제와 관련하여 본원에 사용된 용어 DNA 폴리머라제는 "그의 기능적 단편"을 포함한다. "그의 기능적 단편"은 폴리머라제의 전체 미만의 아미노산 서열을 포괄하고, 적어도 1종의 세트의 조건 하에 폴리뉴클레오티드의 중합을 촉매하는 능력을 보유하는 야생형 또는 돌연변이체 DNA 폴리머라제의 임의의 부분을 지칭한다. 이러한 기능적 단편은 별개의 개체로서 존재할 수 있거나, 또는 보다 큰 폴리펩티드, 예컨대 융합 단백질의 구성성분일 수 있다.

[0548] 일부 실시양태에서, 폴리머라제는 박테리오파지로부터의 것일 수 있다. 박테리오파지 DNA 폴리머라제는 일반적으로 5'에서 3' 엑소뉴클레아제 활성이 없는데, 이는 이러한 활성이 별개의 폴리펩티드에 의해 코딩되기 때문이다. 적합한 DNA 폴리머라제의 예는 T4, T7, 및 phi29 DNA 폴리머라제이다. 상업적으로 입수가능한 효소는 T4 (다수의 공급원, 예를 들어 에피센터(Epicentre)로부터 입수가능함) 및 T7 (다수의 공급원, 예를 들어 비변형된 것의 경우 에피센터, 및 3'에서 5' 엑소 T7 "시퀴나제" DNA 폴리머라제의 경우 USB로부터 입수가능함)이다.

[0549] 다른 실시양태에서, 폴리머라제는 고세균 폴리머라제이다. 고세균에서 확인된 2가지 상이한 부류의 DNA 폴리머라제가 존재한다: 1. 패밀리 B/pol I 유형 (피로코쿠스 푸리오수스(Pyrococcus furiosus)로부터의 Pfu의 상동체) 및 2. pol II 유형 (피. 푸리오수스 DP1/DP2 2-서브유닛 폴리머라제의 상동체). 둘 다의 부류로부터의 DNA 폴리머라제는 연관된 5'에서 3' 엑소뉴클레아제 활성이 자연 결여되고 3'에서 5' 엑소뉴클레아제 (교정) 활성을 보유하는 것으로 밝혀졌다. 적합한 DNA 폴리머라제 (pol I 또는 pol II)는 목적하는 검정 온도와 유사한 최적 성장 온도를 갖는 고세균으로부터 유래될 수 있다.

[0550] 열안정성 고세균 DNA 폴리머라제는 피로코쿠스(Pyrococcus) 종 (푸리오수스(furiosus), 종 GB-D, 보에시이(woesii), 아비씨(abyssi), 호르코쉬이(horikoshii)), 써모코쿠스(Thermococcus) 종 (코다카라엔시스(kodakaraensis) KOD1, 리토랄리스(litoralis), 종 9 디그리스 노스-7(9 degrees North-7), 종 JDF-3, 고르고나리우스(gorgonarius)), (피로딕티움 오쿨툼(Pyrodictium occultum)), 및 아르카에오글로부스 풀기두스(Archaeoglobus fulgidus)로부터 단리된다.

[0551] 폴리머라제는 또한 유박테리아 종으로부터의 것일 수 있다. 유박테리아 DNA 폴리머라제의 3종의 부류, pol I, II, 및 III이 존재한다. Pol I DNA 폴리머라제 패밀리의 효소는 5'에서 3' 엑소뉴클레아제 활성을 보유하고, 특정 구성원은 또한 3'에서 5' 엑소뉴클레아제 활성을 나타낸다. Pol II DNA 폴리머라제는 5'에서 3' 엑소뉴클레아제 활성이 자연 결여되어 있지만, 3'에서 5' 엑소뉴클레아제 활성을 나타낸다. Pol III DNA 폴리머라제는 세포의 주요 복제 DNA 폴리머라제를 나타내고, 다중 서브유닛으로 구성된다. pol III 촉매 서브유닛은 5'에서 3' 엑소뉴클레아제 활성이 결여되어 있지만, 일부 경우에 3'에서 5' 엑소뉴클레아제 활성은 동일한 폴리펩티드

에 위치한다.

[0552] 다양한 상업적으로 입수가능한 Pol I DNA 폴리머라제가 존재하며, 이들 중 일부는 5'에서 3' 엑소뉴클레아제 활성을 감소시키거나 없애도록 변형되었다.

[0553] 적합한 열안정성 pol I DNA 폴리머라제는 써무스(Thermus) 종 및 써모토가 마리티마(Thermotoga maritima) 예컨대 써무스 아쿠아티쿠스(Thermus aquaticus) (Taq), 써무스 써모필루스(Thermus thermophilus) (Tth) 및 써모토가 마리티마(Thermotoga maritima) (Tma UlTma)를 포함한 다양한 고온성 유박테리아로부터 단리될 수 있다.

[0554] 상기 열거된 것과 관련된 추가의 유박테리아는 문헌 [Thermophilic Bacteria (Kristjansson, J. K., ed.) CRC Press, Inc., Boca Raton, Fla., 1992]에 기재되어 있다.

[0555] 본 발명은 그의 내용의 전문이 본원에 참조로 포함되는 미국 특허 번호 5,677,152, 6,479,264 및 6,183,998에 개시된 방법에 따라 화학적으로 변형된 키메라 또는 비-키메라 DNA 폴리머라제를 추가로 제공한다.

[0556] 상기 열거된 것과 관련된 추가의 고세균 DNA 폴리머라제는 하기 참고문헌에 기재되어 있다: [Archaea: A Laboratory Manual (Robb, F. T. and Place, A. R., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1995 및 Thermophilic Bacteria (Kristjansson, J. K., ed.) CRC Press, Inc., Boca Raton, Fla., 1992].

[0557] B. 예시적인 리버스 트랜스크립타제

[0558] 다양한 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제는 폴리머라제로서 리버스 트랜스크립타제를 포함한다. 개시내용은 임의의 자연 발생 유기체 또는 바이러스로부터 수득되거나, 또는 상업적 또는 비-상업적 공급원으로부터 수득된 임의의 야생형 리버스 트랜스크립타제를 고려한다. 추가로, 개시내용의 다중-플랩 프라임 편집제에 사용가능한 리버스 트랜스크립타제는 임의의 자연 발생 돌연변이체 RT, 조작된 돌연변이체 RT, 또는 기능을 보유하는 말단절단된 변이체를 포함한 다른 변이체 RT를 포함할 수 있다. RT는 또한 특정 아미노산 치환, 예컨대 본원에 구체적으로 개시된 것을 함유하도록 조작될 수 있다.

[0559] 리버스 트랜스크립타제는 전형적으로 RNA- 및 DNA-의존성 DNA 중합 활성을 비롯한 3가지 효소적 활성, 및 RNA-DNA 하이브리드에서 RNA의 절단을 촉매하는 RNaseH 활성을 갖는 다기능성 효소이다. 리버스 트랜스크립타제의 일부 돌연변이체는 RNaseH 모이어티를 무력화시켜 mRNA에 대한 의도하지 않은 손상을 방지한다. 주형으로서 mRNA를 사용하여 상보적 DNA (cDNA)를 합성하는 이들 효소가 RNA 바이러스에서 최초로 확인되었다. 후속적으로, 바이러스 입자, 세포 또는 조직으로부터 리버스 트랜스크립타제가 직접 단리 및 정제되었다. (예를 들어, 문헌 [Kacian et al., 1971, Biochim. Biophys. Acta 46: 365-83; Yang et al., 1972, Biochem. Biophys. Res. Comm. 47: 505-11; Gerard et al., 1975, J. Virol. 15: 785-97; Liu et al., 1977, Arch. Virol. 55 187-200; Kato et al., 1984, J. Virol. Methods 9: 325-39; Luke et al., 1990, Biochem. 29: 1764-69 and Le Grice et al., 1991, J. Virol. 65: 7004-07] 참조, 이들 각각은 참조로 포함됨). 보다 최근에, 열안정성, 충실도 및 활성과 같은 개선된 특성에 대한 연구에서 돌연변이체 및 융합 단백질이 생성되었다. 관련 기술분야에 공지되어 있거나 또는 관련 기술분야에 공지된 방법을 사용하여 제조될 수 있는 리버스 트랜스크립타제의 야생형, 변이체, 및/또는 돌연변이체 형태 중 임의의 것이 본원에서 고려된다.

[0560] 리버스 트랜스크립타제 (RT) 유전자 (또는 그 안에 함유된 유전자 정보)는 다수의 상이한 공급원으로부터 수득될 수 있다. 예를 들어, 유전자는 레트로바이러스로 감염된 진핵 세포로부터, 또는 레트로바이러스 게놈의 부분 또는 전체를 함유하는 다수의 플라스미드로부터 수득될 수 있다. 또한, RT 유전자를 함유하는 메신저 RNA-유사 RNA는 레트로바이러스로부터 수득될 수 있다. RT에 대한 공급원의 예는 몰로니 뮤린 백혈병 바이러스 (M-MLV 또는 MLVRT); 인간 T-세포 백혈병 바이러스 유형 1 (HTLV-1); 소 백혈병 바이러스 (BLV); 라우스 육종 바이러스 (RSV); 인간 면역결핍 바이러스 (HIV); 효모, 예컨대 사카로미세스(Saccharomyces), 뉴로스포라(Neurospora), 드로소필라(Drosophila); 영장류; 및 설치류를 포함하나 이에 제한되지는 않는다. 예를 들어, 문헌 [Weiss, et al., 미국 특허 번호 4,663,290 (1987); Gerard, G. R., DNA:271-79 (1986); Kotewicz, M. L., et al., Gene 35:249-58 (1985); Tanese, N., et al., Proc. Natl. Acad. Sci. (USA):4944-48 (1985); Roth, M. J., et al., J. Biol. Chem. 260:9326-35 (1985); Michel, F., et al., Nature 316:641-43 (1985); Akins, R. A., et al., Cell 47:505-16 (1986), EMBO J. 4:1267-75 (1985); 및 Fawcett, D. F., Cell 47:1007-15 (1986)]을 참조한다 (각각은 그 전문이 본원에 참조로 포함됨).

[0561] 야생형 RT

[0562] 본원에 개시된 다중-플랩 프라임 편집제와 함께 사용하기 위한 예시적인 효소는 M-MLV 리버스 트랜스크립타제 및 RSV 리버스 트랜스크립타제를 포함할 수 있으나 이에 제한되지는 않는다. 리버스 트랜스크립타제 활성을 갖는 효소는 상업적으로 입수가능하다. 특정 실시양태에서, 리버스 트랜스크립타제는 다중-플랩 프라임 편집제(PE) 시스템의 다른 성분에 트랜스로 제공된다. 즉, 리버스 트랜스크립타제는 개별 성분으로서 발현되거나 또는 달리 제공되며, 즉 napDNAbp와의 융합 단백질로서가 아니다.

[0563] 관련 기술분야의 통상의 기술자는 몰로니 뮤린 백혈병 바이러스 (M-MLV); 인간 면역결핍 바이러스 (HIV) 리버스 트랜스크립타제 및 조류 육종-백혈증 바이러스 (ASLV) 리버스 트랜스크립타제, 예컨대 비제한적으로 라우스 육종 바이러스 (RSV) 리버스 트랜스크립타제, 조류 골수모구증 바이러스 (AMV) 리버스 트랜스크립타제, 조류 적모구증 바이러스 (AEV) 헬퍼 바이러스 MCAV 리버스 트랜스크립타제, 조류 골수구종증 바이러스 MC29 헬퍼 바이러스 MCAV 리버스 트랜스크립타제, 조류 세망내피증 바이러스 (REV-T) 헬퍼 바이러스 REV-A 리버스 트랜스크립타제, 조류 육종 바이러스 UR2 헬퍼 바이러스 UR2AV 리버스 트랜스크립타제, 조류 육종 바이러스 Y73 헬퍼 바이러스 YAV 리버스 트랜스크립타제, 라우스 연관 바이러스 (RAV) 리버스 트랜스크립타제, 및 골수모구증 연관 바이러스 (MAV) 리버스 트랜스크립타제를 포함하나 이에 제한되지는 않는 야생형 리버스 트랜스크립타제가 본원에 기재된 대상 방법 및 조성물에 적합하게 사용될 수 있다는 것을 인식할 것이다.

[0564] 예시적인 야생형 RT 효소는 하기와 같다:

설명	서열
리버스 트랜스크립타제 (M-MLV RT) 야생형 몰로니 뮤린 백혈병 바이러스 PE1 (본원에 개시된 프라임 편집제 1 융합 단백질)에 사용됨	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP(SEQ ID NO: 700)
리버스 트랜스크립타제 몰로니 뮤린 백혈병 바이러스 REF SEQ. AAA66622.1	AFPLERPDWDYTTQAGRNHLVHYRQLLLAGLQNAGRSPTNL AKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDPGQETNVS MSFIWQSAPDIGRKLGRLEDLKSKTLGDLVREAEKIFNKRETP EEREERIRRETEEKEERRRTVDEQKEKERDRRRHREMSKLLAT VVIGQEQDRQEGERKRPQLDKDQCAYCKEKGHWAKDCPKKP RGPRGPRPQTSLLTLGDXGGQGQDPPPEPRITLKVGGQPVTFL VDTGAQHSVLTQNPGPLSDKSAWVQGATGGKRYRWTTDRKV HLATGKVTHSFLHVPDCPYPLLGRDLLTKLKAQIHFEGSGAQV VGPMGQPLQVLTLNIEDEYRLHETSKEPDVSLGFTWLSDFPQA WAESGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKP HIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREV NKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRL HPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEAL HRDLADFR(SEQ ID NO: 90)
리버스 트랜스크립타제 고양이 백혈병 바이러스 REF SEQ. NP955579.1	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAH CQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGIL KPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTV PNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEW RDPEIGLSGQLTWTRLPQGFKNSPTLFDEALHSDLADFRVRYP ALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAK KAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQV REFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLA FEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGP WKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLG QPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVH FGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLP DADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSA QRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRR

[0565]

설명	서열
	GLLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQ AKGNRLADDTAKKAATETHSSLTVL SEQ ID NO: 91)
리버스 트랜스크립타제 HIV-1 RT, 쇄 A REF SEQ. ITL3-A	PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKE GKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTA FTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQ NPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTP DKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQ KLVGKLNWASQIYPGIKVRQLXKLLRGTKALTEVIPLTEEAEL ELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQ EPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGK TPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLW YQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNRGRQKVV TLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPD QSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSA GIRKV(SEQ ID NO: 92) 문헌 [MARTINELLI ET AL., VIROLOGY, 1990, 174(1): 135-144] 참조, 이는 본원에 참조로 포함됨
리버스 트랜스크립타제 HIV-1 RT, 쇄 B REF SEQ. ITL3-B	PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKE GKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTA FTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQ NPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTP DKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQ KLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELE LAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEP FKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTP KFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQ LEKEPIVGAETF(SEQ ID NO: 93) 문헌 [STAMMERS ET AL., J. MOL. BIOL., 1994, 242(4): 586-588] 참조, 이는 본원에 참조로 포함됨
리버스 트랜스크립타제 라우스 육종 바이러스 RT REF SEQ. ACL14945	TVALHLAIPLKWKPNHTPVWIDQWPLPEGKLVALTQLVEKEL QLGHIEPSLSCWNTPVFVIRKASGSYRLLHDLRAVNAKLVPFG AVQQGAPVLSALPRGWPLMVLDLKDCFFSIPLAEQDREAFAF TLPSVNNQAPARRFQWKVLPQGMTCSPTICQLIVGQILEPLRL KHPSLRMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTIS PDKVQKEPGVQYLGYKLGSTYAAPVGLVAEPRIATLWDVQKL VGSLQWLRPALGIPPRLRGPFYEQLRGSDPNEAREWNLDMK MAWREIVQLSTTAALERWDPALPLEGAVARCEQGAIGVLGQG LSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTFGK EVDILLLPACFRDELPLPEGILLALRGFAGKIRSSDTPSIFDIARP LHVSLKVRVTDHPVPGPTVFTDASSSTHKGVVVWREGPRWEI KEIADLGASVQQLEARAVAMALLLWPTTPTNVVTDSAFVAKM LLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSEVPG FFTEGNDVADSQATFQAYPLREAKDLHTALHIGPRALSKACNI

[0566]

설명	서열
	SMQQAREVVQTCPHCNSAPALEAGVNPRGLGPLQIWQTDFTL EPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAAQHHWATVI AVLGRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQG QAMVERANRLLKDKIRVLAEGDGFMKRIPTSKQGELLAKAM YALNHFERGENTKTPIQKHWRPTVLTEGPPVKIRIETGEWEKG WNVLVWGRGYAAVKNRDTDKVIWVPSRKVKPDIAQKDEVT KKDEASPLFA(SEQ ID NO: 94) 문헌 [YASUKAWA ET AL., J. BIOCHEM. 2009, 145(3): 315-324] 참조, 이는 본원에 참조로 포함됨
리버스 트랜스크립타제 콜리플라워 모자이크 바이러스 RT REF SEQ. AGT42196	MMDHLLQKTQIQNQTEQVMNITNPNSIYIKGRLYFKGYKKIEL HCFVDTGASLCIASKFVIPEEHWINAERPIMVKIADGSSITINKV CRDIDLIIAGEIFHIPTVYQQESGIDFIIGNNFCQLYEPFIQFTDRV IFTKDRTYPVHIAKLTRAVRVGTEGFLESMKKRSKTQQPEPVNI STNKIAILSEGRRLSEEKLFITQQRMQKIEELLEKVCSENPLDP NKTKQWMKASIKLSDPSKAIKVKPMKYSPMDREEFDKQIKEL LDLKVIKPSKSPHMAPAFLVNNEAEKRRGKKRMVVNYKAMN KATVGDAYNLPNKDELLTLIRGKKIFSSFDCKSGFWQVLLDQD SRPLTAFTCPQGHYEWNVVPFGLKQAPSIFQRHMDEAFRVFR KFCCVYVDDILVFSNNEEDHLLHVAMILQKCNQHGIILSKKKA QLFKKKINFLGLEIDEGTHKPQGHILEHINKFPDTLEDKKQLQR FLGILTYASDYIPKLAQIRKPLQAKLKENVPWKWTKEDTLYM QKVKKNLQGFPPLHHPLPEEKLIIETDASDDYWGGMLKAIKIN EGTNTELICRYASGSFKAAEKNYHSNDKETLAVINTIKKFSIYL TPVHFLIRTDNTHFKSFVNLNYKGDSKLGRNIRWQAWLSHYS FDVEHIKGTDNHFADFLSREFNRVNS(SEQ ID NO: 95) 문헌 [FARZADFAR ET AL., VIRUS GENES, 2013, 47(2): 347-356] 참조, 이는 본원에 참조로 포함됨
리버스 트랜스크립타제 클레브시엘라 뉴모니아에 REF SEQ. RFF81513.1	MKEKISKIDKNFYTDIFIKTSFQNEFEAGGVIPPIAKNQVSTISN KNKTFYSLAHSSPHYSIQTRIEKFLLKNIPLSASSFAFRKERSYL HYLEPHTQNVKYCHLDIVSFFHSIDVNIVRDTFSVYFSDEFLVK EKQSLLDAFMASVTLTAELDGVEKTFIPMGFKSSPSISNIIFRKI DILIQKFCDKNKITYTRYADDLLFSTKKENNILSSTFFINEISSIL SINKFKLNKSKYLYKEGTISLGGYVIENILKDNSSGNIRLSSSKL NPLYKALYEIKKGSSSKHICIKVFNLKLKRFIYKKNKEKFEAKF YSSQLKNKLLGYRSYLLSFVIFHKKYKCINPIFLEKCVFLISEIE SIMNRKF(SEQ ID NO: 96)
리버스 트랜스크립타제 에스케리키아 콜라이 RT REF SEQ. TGH57013	MKITSNNVTAVINGKGWHSINWKKCHQHVKTIQTRIAKAACN QQWRTVGRLQRLLVRSFSARALAVKRVTENSGRKTPGVDGQI WSTPESKWEAIFKLRRKGYKPLPLKRVFIPKSNGKKRPLGIPV MLDRAMQALHLLGLEPVSETNADHNSYGFRPARCTADAIQQ VCNMYSSRNASKWVLEGDIKGCFEHISHEWLLENIPMDKQIL RNWLKAGIIEKSIFSKTLSGTPQGGIISPVLANMALDGLERLLQ NRFGRNRLI(SEQ ID NO: 97)

[0567]

설명	서열
리버스 트랜스크립타제 바실루스 서브틸리스 RT REF SEQ. QBJ66766	MSKIKINYEKYHIKPFPHFDQRIKVNKKVKENLQNPFYIAAHS FYPFIHYKKISYKFKNGTLSSPKERDIFYSGHMDGYIYKHYGEI LNHKYNNTCIGKGIDHVSLAYRNNKMGKSNIHFAAEVINFISE QQQAFIFVSDFSSYFDSLDHAILKEKLIEVLEEQDKLSKDWWN VFKHITRYNWVEKEEVISDLECTKEKIARDKKSRERYYTPAEF REFRKRVNIKSNDTGVGIPQGTAISAVLANVYAIDLDQKLNQY ALKYGGIYRRYSDDIIMVLPMTSDGQDPSNDHVSFIKSVVKRN KVTMGDSKTSVLYYANNNIYEDYQRKRESKMDYLGFSFDGM TVKIREKSLFKYYHRTYKKINSINWASVKKEKKVGRKKLYLL YSHLGRNYKGHGNFISYCKKAHAVFEGNKKIESLINQQIKRH WKKIQKRLVDV(SEQ ID NO: 98)
유박테리움 렉탈레 군 II 인트론 RT	DTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKE HLAKNGETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPT VTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALN IMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIV RKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEM EKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKV NMTKSKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKF KKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTL CKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARR VAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKCV TC(SEQ ID NO: 99)
게오바실루스 스테아로써모필루스 군 II 인트론 RT	ALLERILARDNLITALKRVEANQGAPGIDGVSTDQLRDYIRAH WSTIHAQLLAGTYRPAPVRRVEIPKPGGGTRQLGIPTVVDRLIQ QAILQELTPIFDPDFSSSSFGFRPGRNAHDAVRQAQGYIQEGYR YVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGL KFCRYADDCNIYVKSLRAGQRVKQSIQRFLEKTLKLKVNEEKS AVDRPWKRAFLGFSFTPERKARIRLAPRSIQRLKQRIRQLTNPN WSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTT KTPQLHQALGKTYWTAQGLKSLTQR(SEQ ID NO: 100)

[0568]

[0569] 변이체 및 오류-유발 RT

[0570] 리버스 트랜스크립타제는 RNA 주형으로부터 상보적 DNA (cDNA) 가닥을 합성하는데 필수적이다. 리버스 트랜스크립타제는 상이한 생화학적 활성을 나타내는 별개의 도메인으로 구성된 효소이다. 효소는 하기와 같이 RNA 주형으로부터 DNA의 합성을 촉매한다: 어닐링된 프라이머의 존재 하에, 리버스 트랜스크립타제는 RNA 주형에 결합하여 중합 반응을 개시한다. RNA-의존성 DNA 폴리머라제 활성은 상보적 DNA (cDNA) 가닥을 합성하여, dNTP를 혼입시킨다. RNase H 활성은 DNA:RNA 복합체의 RNA 주형을 분해한다. 따라서, 리버스 트랜스크립타제는 (a) RNA/DNA 하이브리드를 인식하고 이에 결합하는 결합 활성, (b) RNA-의존성 DNA 폴리머라제 활성, 및 (c) RNase H 활성을 포함한다. 또한, 리버스 트랜스크립타제는 일반적으로 그의 열안정성, 프로세싱성 (dNTP 혼입 속도), 및 충실도 (또는 오류율)를 비롯한 다양한 속성을 갖는 것으로 간주된다. 본원에서 고려되는 리버스 트랜스크립타제 변이체는 이들 효소적 활성 (예를 들어, RNA-의존성 DNA 폴리머라제 활성, RNase H 활성, 또는 DNA/RNA 하이브리드-결합 활성) 또는 효소 특성 (예를 들어, 열안정성, 프로세싱성, 또는 충실도) 중 어느 하나 이상에 영향을 미치거나 이를 변화시키는 리버스 트랜스크립타제에 대한 임의의 돌연변이를 포함할 수 있다. 이러한 변이체는 관련 기술분야에서 공공 도메인에서 입수가능하거나, 상업적으로 입수가능하거나, 또는 방향적 진화상 과정 (예를 들어, PACE 또는 PANCE)을 포함한 공지된 돌연변이유발 방법을 사용하여 제조될 수 있다.

[0571] 다양한 실시양태에서, 리버스 트랜스크립타제는 변이체 리버스 트랜스크립타제일 수 있다. 본원에 사용된 "변이체 리버스 트랜스크립타제"는 참조 서열 (예를 들어, 참조 야생형 서열)에 비해 1개 이상의 돌연변이 (단일 돌연변이, 역위, 결실, 삽입, 및 재배열 포함)를 포함하는 임의의 자연 발생 또는 유전자 조작된 변이체를 포함한다. RT는 자연적으로 RNA-의존성 DNA 폴리머라제 활성, 리보뉴클레아제 H 활성, 및 DNA-의존성 DNA 폴리머라제 활성을 비롯한 여러 활성을 갖는다. 집합적으로, 이들 활성은 효소가 단일 가닥 RNA를 이중 가닥 cDNA로 전환시킬 수 있게 한다. 레트로바이러스 및 레트로트랜스포손에서, 이러한 cDNA는 이어서 숙주 게놈 내로 통합될 수 있고, 이로부터 숙주-세포 전사를 통해 새로운 RNA 카피가 제조될 수 있다. 변이체 RT는 이들 활성 중 1종 이상에 영향을 미치는 (이들 활성을 감소 또는 증가시키거나, 또는 이들 활성을 모두 함께 제거하는) 돌연변이를 포함할 수 있다. 또한, 변이체 RT는 RT를 더 안정하게 또는 덜 안정하게 하고, 응집 경향을 낮추고, 정제 및/또는 검출, 및/또는 특성 또는 특징의 다른 변형을 용이하게 하는 1개 이상의 돌연변이를 포함할 수 있다.

[0572] 관련 기술분야의 통상의 기술자는 몰로니 뮤린 백혈병 바이러스 (M-MLV); 인간 면역결핍 바이러스 (HIV) 리버스 트랜스크립타제 및 조류 육종-백혈증 바이러스 (ASLV) 리버스 트랜스크립타제, 예컨대 비제한적으로 라우스 육종 바이러스 (RSV) 리버스 트랜스크립타제, 조류 골수모구증 바이러스 (AMV) 리버스 트랜스크립타제, 조류 적모구증 바이러스 (AEV) 헬퍼 바이러스 MCAV 리버스 트랜스크립타제, 조류 골수구종증 바이러스 MC29 헬퍼 바이러스 MCAV 리버스 트랜스크립타제, 조류 세망내피증 바이러스 (REV-T) 헬퍼 바이러스 REV-A 리버스 트랜스크립타제, 조류 육종 바이러스 UR2 헬퍼 바이러스 UR2AV 리버스 트랜스크립타제, 조류 육종 바이러스 Y73 헬퍼 바이러스 YAV 리버스 트랜스크립타제, 라우스 연관 바이러스 (RAV) 리버스 트랜스크립타제, 및 골수모구증 연관 바이러스 (MAV) 리버스 트랜스크립타제를 포함하나 이에 제한되지는 않는 다른 리버스 트랜스크립타제로부터 유래된 변이체 리버스 트랜스크립타제가 본원에 기재된 대상 방법 및 조성물에 적합하게 사용될 수 있다는 것을 인식할 것이다.

[0573] 변이체 RT를 제조하는 하나의 방법은 유전자 변형에 의한 (예를 들어, 야생형 리버스 트랜스크립타제의 DNA 서열을 변형시키는 것에 의한) 것이다. DNA 서열의 무작위 돌연변이 뿐만 아니라 표적화된 돌연변이를 허용하는 다수의 방법이 관련 기술분야에 공지되어 있다 (예를 들어, 문헌 [Ausubel et al., Short Protocols in Molecular Biology (1995) 3.sup.rd Ed. John Wiley & Sons, Inc.] 참조). 또한, 통상적인 방법 및 PCR-기반 방법 둘 다를 포함한 부위-지정 돌연변이유발을 위한 다수의 상업적으로 입수가능한 키트가 존재한다. 예는 퀵체인지(QuikChange) 부위-지정 돌연변이유발 키트 (애질런트(AGILENT)®), Q5® 부위-지정 돌연변이유발 키트 (뉴 잉글랜드 바이오랩스(NEW ENGLAND BIOLABS)®), 및 진아트(GeneArt)™ 부위-지정 돌연변이유발 시스템 (써모피셔 사이언티픽(THERMOFISHER SCIENTIFIC)®)을 포함한다.

[0574] 또한, 돌연변이체 리버스 트랜스크립타제는 관련 기술분야의 통상의 기술자에게 공지된 방법론에 따라 삽입 돌연변이 또는 말단절단 (N-말단, 내부, 또는 C-말단 삽입 또는 말단절단)에 의해 생성될 수 있다. 본원에 사용된 용어 "돌연변이"는 서열, 예를 들어 핵산 또는 아미노산 서열 내의 잔기의 또 다른 잔기로의 치환, 또는 서열 내의 1개 이상의 잔기의 결실 또는 삽입을 지칭한다. 돌연변이는 전형적으로 원래 잔기를 확인한 다음, 서열 내의 잔기의 위치를 확인하고, 새로 치환된 잔기의 정체를 확인하는 것에 의해 본원에 기재된다. 본원에 제공된 아미노산 치환 (돌연변이)을 제조하는 다양한 방법은 관련 기술분야에 널리 공지되어 있고, 예를 들어 문헌 [Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 의해 제공된다. 돌연변이는 다양한 카테고리, 예컨대 단일 염기 다형성, 미세중복 영역, indel, 및 역위를 포함할 수 있고, 이는 어떠한 방식으로도 제한하는 것으로 의도되지 않는다. 돌연변이는 단백질 활성을 감소시키거나 제거하는 돌연변이의 정상적인 결과인 "기능 상실" 돌연변이를 포함할 수 있다. 대부분의 기능 상실 돌연변이는 열성이며, 이는 이형접합체에서 제2 염색체 카피가 그의 존재가 돌연변이의 효과를 보상하는 완전 기능적 단백질을 코딩하는 유전자의 비돌연변이된 버전을 보유하기 때문이다. 돌연변이는 또한 정상 상태에서는 달리 존재하지 않는 단백질 또는 세포에 비정상적 활성을 부여하는 것인 "기능 획득" 돌연변이를 포괄한다. 많은 기능 획득 돌연변이는 코딩 영역보다는 조절 서열에 존재하고, 따라서 다수의 결과를 가질 수 있다. 예를 들어, 돌연변이는 1개 이상의 유전자가 잘못된 조직에서 발현되게 할 수 있고, 이들 조직은 정상적으로는 결여되어 있는 기능을 얻는다. 그의 속성으로 인해, 기능 획득 돌연변이는 통상적으로 우성이다.

[0575] 관련 기술분야에 공지된 부위-지정 돌연변이유발의 보다 오래된 방법은 단일 가닥 DNA 주형의 단리를 가능하게 하는 벡터, 예컨대 M13 박테리오파지 벡터 내로의 돌연변이시킬 서열의 서브-클로닝에 의존한다. 이들 방법에서, 돌연변이유발 프라이머 (즉, 돌연변이시킬 부위에 어닐링할 수 있지만 돌연변이시킬 부위에 1개 이상의 미스매칭된 뉴클레오티드를 보유할 수 있는 프라이머)를 단일 가닥 주형에 어닐링시킨 후, 돌연변이유발 프라이머의 3' 단부로부터 출발하여 주형의 상보체를 중합시킨다. 이어서, 생성된 듀플렉스로 숙주 박테리아를 형질전환시키고, 목적하는 돌연변이에 대해 플라크를 스크리닝한다.

[0576] 보다 최근에, 부위-지정 돌연변이유발은 단일 가닥 주형을 필요로 하지 않는 이점을 갖는 PCR 방법론을 사용하였다. 또한, 서브클로닝을 필요로 하지 않는 방법이 개발되었다. PCR-기반 부위-지정 돌연변이유발이 수행될 때 여러 이슈가 고려되어야 한다. 첫번째로, 이들 방법에서, 폴리머라제에 의해 도입된 바람직하지 않은 돌연변이의 확장을 방지하기 위해 PCR 사이클의 수를 감소시키는 것이 바람직하다. 두번째로, 반응에서 지속되는 비-돌연변이된 모 분자의 수를 감소시키기 위해 선택이 사용되어야 한다. 세번째로, 단일 PCR 프라이머 세트의 사용을 허용하기 위해 연장-길이 PCR 방법이 바람직하다. 그리고 네번째로, 일부 열안정성 폴리머라제의 비-주형-의존성 말단 연장 활성으로 인해, PCR-생성된 돌연변이체 생성물의 평활-단부 라이게이션 전에 단부-폴리싱 단계를 절차에 혼입시키는 것이 종종 필요하다.

[0577] 1개 이상의 무작위로 위치한 돌연변이를 보유하는 돌연변이체 패널을 생성할 무작위 돌연변이유발 방법이 관련 기술분야에 존재한다. 이어서, 이러한 돌연변이체 패널을 대상으로 하여, 야생형 리버스 트랜스크립타제에 비해 목적하는 특성, 예를 들어 증가된 안정성을 나타내는 것에 관하여 스크리닝할 수 있다.

[0578] 무작위 돌연변이유발 방법의 예는 소위 "오류-유발 PCR 방법"이다. 명칭이 암시하는 바와 같이, 방법은 DNA 폴리머라제가 고충실도 혼입을 지지하지 않는 조건 하에서 주어진 서열을 증폭시킨다. 상이한 DNA 폴리머라제에 대한 오류-유발 혼입을 촉진하는 조건은 다양하지만, 관련 기술분야의 통상의 기술자는 주어진 효소에 대한 이러한 조건을 결정할 수 있다. 많은 DNA 폴리머라제에 대한 증폭의 충실도에서의 핵심 변수는, 예를 들어 완충제 중 2가 금속 이온의 유형 및 농도이다. 따라서, 망가니즈 이온의 사용 및/또는 마그네슘 또는 망가니즈 이온 농도의 변경은 폴리머라제의 오류율에 영향을 미치도록 적용될 수 있다.

[0579] 다양한 측면에서, 다중-플랩 프라임 편집제의 RT는 "오류-유발" 리버스 트랜스크립타제 변이체일 수 있다. 관련 기술분야에 공지되고/거나 이용가능한 오류-유발 리버스 트랜스크립타제가 사용될 수 있다. 리버스 트랜스크립타제는 자연적으로 어떠한 교정 기능도 갖지 않고; 따라서 리버스 트랜스크립타제의 오류율은 교정 활성을 포함하는 DNA 폴리머라제보다 일반적으로 더 높다는 것이 인지될 것이다. 임의의 특정한 리버스 트랜스크립타제의 오류율은 그의 RNA 주형에 대한 DNA의 주형-지정 중합의 정확도를 나타내는 효소의 "충실도"의 특성이다. 고충실도를 갖는 RT는 낮은 오류율을 갖는다. 반대로, 저충실도를 갖는 RT는 높은 오류율을 갖는다. M-MLV-기반 리버스 트랜스크립타제의 충실도는 합성된 15,000 내지 27,000개 뉴클레오티드에서 1개 오류 범위의 오류율을 갖는 것으로 보고되어 있다. 참조로 포함되는 문헌 [Boutabout et al., "DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1," Nucleic Acids Res, 2001, 29: 2217-2222]을 참조한다. 따라서, 본 출원의 목적상, "오류-유발"인 것으로 간주되거나 또는 "오류-유발 충실도"를 갖는 것으로 간주되는 리버스 트랜스크립타제는 합성된 15,000개 뉴클레오티드에서 1개 오류 미만의 오류율을 갖는 것이다.

[0580] 오류-유발 리버스 트랜스크립타제는 또한 출발 RT 효소 (예를 들어, 야생형 M-MLV RT)의 돌연변이유발을 통해 생성될 수 있다. 돌연변이유발 방법은 제한되지 않고, 방향적 진화 과정, 예컨대 파지-보조 연속 진화 (PACE) 또는 파지-보조 비연속 진화 (PANCE)를 포함할 수 있다. 본원에 사용된 용어 "파지-보조 연속 진화 (PACE)"는 바이러스 벡터로서 파지를 사용하는 연속 진화를 지칭한다. PACE 기술의 일반적 개념은, 예를 들어 2010년 3월 11일에 WO 2010/028347로서 공개된, 2009년 9월 8일에 출원된 국제 PCT 출원 PCT/US2009/056194; 2012년 6월 28일에 WO 2012/088381로서 공개된, 2011년 12월 22일에 출원된 국제 PCT 출원 PCT/US2011/066747; 2015년 5월 5일에 허여된 미국 출원, 미국 특허 번호 9,023,594, 2015년 9월 11일에 WO 2015/134121로서 공개된, 2015년 1월 20일에 출원된 국제 PCT 출원 PCT/US2015/012022, 및 2016년 10월 20일에 WO 2016/168631로서 공개된, 2016년 4월 15일에 출원된 국제 PCT 출원 PCT/US2016/027795에 기재되어 있으며, 이들 각각의 전체 내용은 본원에 참조로 포함된다.

[0581] 오류-유발 리버스 트랜스크립타제는 또한 파지-보조 비-연속적 진화 (PANCE)에 의해 수득될 수 있고, 이는 본원에 사용된 바와 같이, 바이러스 벡터로서 파지를 사용하는 비-연속 진화를 지칭한다. PANCE는 새로운 이. 콜라이 숙주 세포를 가로질러 진화시킬 관심 유전자를 함유하는 진화 '선택 파지' (SP)의 연속 플라스크 전달을 사용하여 숙주 이. 콜라이 내부의 유전자는 일정하게 유지되게 하는 한편 SP 내에 함유된 유전자가 연속적으로 진화하게 하는, 신속한 생체내 방향적 진화를 위한 단순화된 기술이다. 연속 플라스크 전달은 오랫동안 미생물의 실험실 진화를 위한 폭넓게 접근가능한 접근법으로서의 역할을 하였고, 보다 최근에는 박테리오파지 진화를 위한 유사한 접근법이 개발되었다. PANCE 시스템은 PACE 시스템보다 더 낮은 엄격도를 특색으로 한다.

[0582] 다른 오류-유발 리버스 트랜스크립타제가 문헌에 기재되어 있고, 이들 각각은 본원의 방법 및 조성물에 사용하기 위해 고려된다. 예를 들어, 오류-유발 리버스 트랜스크립타제는 문헌 [Bebenek et al., "Error-prone Polymerization by HIV-1 Reverse Transcriptase," J Biol Chem, 1993, Vol. 268: 10324-10334 및 Sebastian-Martin et al., "Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases," Scientific Reports, 2018, Vol. 8: 627]에 기재되어 있고, 이들 각각은 참조로 포함된다. 또한, 오류-유발 리버스 트랜스크립타제를 비롯한 리버스 트랜스크립타제는 모두 뉴 잉글랜드 바이오랩스®로부터의 것인 프로토스크립트(ProtoScript)® (II) 리버스 트랜스크립타제, AMV 리버스 트랜스크립타제, 웜스타트(WarmStart)® 리버스 트랜스크립타제, 및 M-MuLV 리버스 트랜스크립타제, 또는 모두 다카라 바이오 유에스에이, 인크.(TAKARA BIO USA, INC.) (이전의 클론테크(CLONTECH))로부터의 것인 AMV 리버스 트랜스크립타제 XL, 스마트스크라이브(SMARTScribe) 리버스 트랜스크립타제, GPR 초순수 MMLV 리버스 트랜스크립타제를 포함하여 상업적 공급업체로부터 입수할 수 있다.

[0583] 본원의 개시내용은 또한 RNaseH 도메인에서 돌연변이를 갖는 리버스 트랜스크립타제를 고려한다. 상기 언급된 바와 같이, 리버스 트랜스크립타제의 내재적 특성 중 하나는 RNA:cDNA 하이브리드의 RNA 주형을 중합과 동시에 절단하는 RNase H 활성이다. RNase H 활성은 RNA 주형이 전장 역전사의 완료 전에 분해될 수 있기 때문에 긴 cDNA의 합성에 바람직하지 않을 수 있다. RNase H 활성은 또한, 아마도 효소의 폴리머라제 활성과 그의 경쟁으로 인해, 역전사 효율을 저하시킬 수 있다. 따라서, 본 개시내용은 변형된 RNaseH 활성을 포함하는 임의의 리버스 트랜스크립타제 변이체를 고려한다.

[0584] 본원의 개시내용은 또한 RNA-의존성 DNA 폴리머라제 도메인에서 돌연변이를 갖는 리버스 트랜스크립타제를 고려한다. 상기 언급된 바와 같이, 리버스 트랜스크립타제의 내재적 특성 중 하나는 RNA:cDNA 하이브리드의 주형 RNA 가닥에 의해 코딩된 바와 같이 신생 cDNA 가닥 내로 핵염기를 혼입시키는 RNA-의존성 DNA 폴리머라제 활성이다. RNA-의존성 DNA 폴리머라제 활성은 증가 또는 감소되어 (즉, 그의 혼입 속도의 관점에서) 효소의 프로세싱성을 증가 또는 감소시킬 수 있다. 따라서, 본 개시내용은 비변형된 버전에 비해 효소의 프로세싱성이 증가 또는 감소되도록 변형된 RNA-의존성 DNA 폴리머라제 활성을 포함하는 임의의 리버스 트랜스크립타제 변이체를 고려한다.

[0585] 또한, 변경된 열안정성 특징을 갖는 리버스 트랜스크립타제 변이체가 본원에서 고려된다. 리버스 트랜스크립타제가 고온을 견디는 능력은 cDNA 합성의 중요한 측면이다. 상승된 반응 온도는 강한 2차 구조 및/또는 높은 GC 함량을 갖는 RNA를 변성시키는 것을 도와, 리버스 트랜스크립타제가 서열을 따라 판독하도록 한다. 그 결과, 보다 높은 온도에서의 역전사는 전장 cDNA 합성 및 보다 높은 수율을 가능하게 하며, 이는 다중-플랩 프라임 편집 과정의 결과로서 3' 플랩 ssDNA의 개선된 생성으로 이어질 수 있다. 야생형 M-MLV 리버스 트랜스크립타제는 전형적으로 37-48℃ 범위의 최적 온도를 갖지만; 49℃, 50℃, 51℃, 52℃, 53℃, 54℃, 55℃, 56℃, 57℃, 58℃, 59℃, 60℃, 61℃, 62℃, 63℃, 64℃, 65℃, 66℃, 및 그 초과를 포함한 48℃ 초과의 보다 높은 온도에서 역전사 활성을 허용하는 돌연변이가 도입될 수 있다.

[0586] 오류-유발 RT, 열안정성 RT, 프로세싱성-증가 RT를 비롯한 본원에서 고려되는 변이체 리버스 트랜스크립타제는 돌연변이유발 또는 진화성 과정을 비롯한 다양한 상용 전략에 의해 조작될 수 있다. 일부 경우에, 변이체는 단일 돌연변이를 도입함으로써 생산될 수 있다. 다른 경우에, 변이체는 1개 초과의 돌연변이를 필요로 할 수 있다. 1개 초과의 돌연변이를 포함하는 그러한 돌연변이체에 대해, 주어진 돌연변이의 효과는 특정한 돌연변이체가 보유하는 다른 돌연변이와 별개로 부위-지정 돌연변이유발에 의한 야생형 유전자에의 확인된 돌연변이의 도입에 의해 평가될 수 있다. 이어서, 이와 같이 생산된 단일 돌연변이체의 스크리닝 검정은 그러한 돌연변이 단독의 효과의 결정을 가능하게 할 것이다.

[0587] 본원에 사용된 변이체 RT 효소는 또한 임의의 야생형 RT, 또는 돌연변이체 RT, 또는 단편 RT, 또는 본원에서 개시되거나 고려되거나 또는 관련 기술분야에 공지된 RT의 다른 변이체를 포함한 임의의 참조 RT 단백질과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일한 다른 "RT 변이체"를 포함할 수 있다.

[0588] 일부 실시양태에서, RT 변이체는 참조 RT와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개, 또는 최대 100개, 또는 최대 200개, 또는 최대 300개, 또는 최대 400개, 또는 최대 500개 또는 그 초과의 아미노산 변화를 가질 수 있다. 일부 실시양태에서, RT 변이체는 참조 RT의 단편을 포함하며, 이러한 단편은 참조 RT의 상응하는 단편에 대해 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 96% 동일하거나, 적어도 약 97% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나, 또는 적어도 약 99.9% 동일하다. 일부 실시양태에서, 단편은 상응하는 야생형 RT (M-MLV 리버스 트랜스크립타제) (예를 들어, 서열식별번호: 89) 또는 서열식별번호: 90-100의 리버스 트랜스크립타제 중 임의의 것의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%이다.

[0589] 일부 실시양태에서, 개시내용은 또한 그의 기능성을 보유하고 임의의 본원에 개시된 RT 단백질의 단편인 RT 단편을 이용할 수 있다. 일부 실시양태에서, RT 단편은 적어도 100개 아미노산 길이이다. 일부 실시양태에서, 단편은 적어도 100, 150, 200, 250, 300, 350, 400, 450, 500, 550개, 또는 최대 600개 또는 그 초과의 아미노

산 길이이다.

[0590] 또 다른 실시양태에서, 개시내용은 또한 충분한 폴리머라제 기능을 여전히 보유하는 말단절단된 변이체를 생성하는, 특정 수의 아미노산만큼 N-말단 또는 C-말단, 또는 둘 다에서 말단절단된 RT 변이체를 이용할 수 있다. 일부 실시양태에서, RT 말단절단된 변이체는 단백질의 N-말단 단부에서 적어도 1, 적어도 2, 적어도 3, 적어도 4, 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 적어도 10, 적어도 11, 적어도 12, 적어도 13, 적어도 14, 적어도 15, 적어도 16, 적어도 17, 적어도 18, 적어도 19, 적어도 20, 적어도 21, 적어도 22, 적어도 23, 적어도 24, 적어도 25, 적어도 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 또는 250개의 아미노산의 말단절단을 갖는다. 다른 실시양태에서, RT 말단절단된 변이체는 단백질의 C-말단 단부에서 적어도 1, 적어도 2, 적어도 3, 적어도 4, 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 적어도 10, 적어도 11, 적어도 12, 적어도 13, 적어도 14, 적어도 15, 적어도 16, 적어도 17, 적어도 18, 적어도 19, 적어도 20, 적어도 21, 적어도 22, 적어도 23, 적어도 24, 적어도 25, 적어도 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 또는 250개의 아미노산의 말단절단을 갖는다. 또 다른 실시양태에서, RT 말단절단된 변이체는 N-말단 및 C-말단 단부에서 동일하거나 상이한 길이의 말단절단을 갖는다.

[0591] 예를 들어, 본원에 개시된 다중-플랩 프라임 편집제는 M-MLV 리버스 트랜스크립타제의 말단절단된 버전을 포함할 수 있다. 이러한 실시양태에서, 리버스 트랜스크립타제는 4개의 돌연변이를 함유한다 (D200N, T306K, W313F, T330P; PE2에 존재하는 L603W 돌연변이는 말단절단으로 인해 더 이상 존재하지 않음에 주목). 이러한 말단절단된 편집제를 코딩하는 DNA 서열은 PE2보다 522 bp 더 작고, 따라서 DNA 서열의 전달이 그의 크기로 인해 도전과제인 적용 (즉, 아데노-연관 바이러스 및 렌티바이러스 전달)에서 잠재적으로 유용하게 한다. 이러한 실시양태는 MMLV-RT(trunc)로 지칭되고, 하기 아미노산 서열을 갖는다:

MMLV-RT(TRUNC)	*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQ APLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGL PPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAA TSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEG QRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMA APLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDNSRLI N* (SEQ ID NO: 766)

[0592]

[0593] 다양한 실시양태에서, 본원에 개시된 다중-플랩 프라임 편집제는 본원에 기재된 RT 변이체, 또는 임의의 참조 Cas9 변이체와 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일한 그의 RT 변이체 중 하나를 포함할 수 있다.

[0594] 또 다른 실시양태에서, 본 발명의 방법 및 조성물은 그 내용이 본원에 참조로 포함되는 문헌 [Effefson et al., "Synthetic evolutionary origin of a proofreading reverse transcriptase," Science, June 24, 2016, Vol. 352: 1590-1593]에 기재된 바와 같은, 리버스 트랜스크립타제로 진화된 DNA 폴리머라제를 이용할 수 있다.

[0595] 특정의 다른 실시양태에서, 리버스 트랜스크립타제는 napDNAbp를 또한 포함하는 융합 단백질의 성분으로서 제공된다. 다시 말해서, 일부 실시양태에서, 리버스 트랜스크립타제는 융합 단백질로서 napDNAbp에 융합된다.

[0596] 다양한 실시양태에서, 변이체 리버스 트랜스크립타제는 서열식별번호: 89로 나타내어진 바와 같은 야생형 M-MLV 리버스 트랜스크립타제로부터 조작될 수 있다.

[0597] 다양한 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 하기 돌연변이: P51L, S67K, E69K, L139P, T197A, D200N, H204R, F209N, E302K, E302R, T306K, F309N, W313F, T330P, L345G, L435G, N454K, D524G, E562Q, D583N, H594Q, L603W, E607K, 또는 D653N 중 1개 이상을 포함하는 변이체 RT를 포함할 수 있다.

[0598] 본 개시내용의 다양한 실시양태에 따른 napDNAbp 단백질에 융합될 수 있거나 또는 개별 단백질로서 제공될 수

있는 일부 예시적인 리버스 트랜스크립타제가 하기 제공된다. 예시적인 리버스 트랜스크립타제는 하기 야생형 효소 또는 부분 효소와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 변이체를 포함한다:

설명	서열 (야생형 대비 변이체 치환)
리버스 트랜스크립타제 (M-MLV RT) 야생형 몰로니 뮤린 백혈병 바이러스 PE1 (본원에 개시된 프라임 편집제 1 융합 단백질)에 사용됨	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 89)
M-MLV RT D200N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 701)
M-MLV RT D200N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL

[0599]

설명	서열 (야생형 대비 변이체 치환)
T330P	LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 702)
M-MLV RT D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 740)
M-MLV RT D200N T330P L603W E69K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQ**K**ARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRA LLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMG QPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDT STLLIENSSP (SEQ ID NO: 703)
M-MLV RT D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN

[0600]

설명	서열 (야생형 대비 변이체 치환)
E302R	SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**R**FLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP(SEQ ID NO: 704)
M-MLV RT D200N T330P L603W E607K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 705)
M-MLV RT D200N T330P L603W L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 706)
M-MLV RT D200N T330P L603W L435G	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM

[0601]

설명	서열 (야생형 대비 변이체 치환)
	GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVI**G**APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 707)
M-MLV RT D200N T330P L603W N454K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 708)
M-MLV RT D200N T330P L603W T306K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLG**K**AGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 709)
M-MLV RT D200N T330P L603W W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL

[0602]

설명	서열 (야생형 대비 변이체 치환)
	GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 710)
M-MLV RT D200N T330P L603W D524G E562Q D583N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRA**Q**LIALTQALKMA EGKKLNVYT**N**SRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 711)
M-MLV RT D200N T330P L603W E302R W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**R**FLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 712)
M-MLV RT D200N T330P L603W E607K L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL

[0603]

설명	서열 (야생형 대비 변이체 치환)
	NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 713)
M-MLV RT P51L S67K T197A H204R E302K F309N W313F T330P L435G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LII**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSP**A**LFDEAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET VMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFN WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTM GQPLVI**G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT **G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDT STLLIENSSP (SEQ ID NO: 714)
M-MLV RT D200N P51L S67K T197A H204R E302K F309N W313F T330P L345G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LII**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSP**A**LF**N**EAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET VMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFN WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTM GQPLVI**G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT **G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDT STLLIENSSP (SEQ ID NO: 715)
M-MLV RT D200N T330P L603W T306K W313F PE2에서의 것	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA

[0604]

설명	서열 (야생형 대비 변이체 치환)
	EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 716)

[0605]

[0606] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 하기 돌연변이: P51X, S67X, E69X, L139X, T197X, D200X, H204X, F209X, E302X, T306X, F309X, W313X, T330X, L345X, L435X, N454X, D524X, E562X, D583X, H594X, L603X, E607X, 또는 D653X 중 1개 이상을 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다.

[0607] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 P51X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 L이다.

[0608] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 S67X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 K이다.

[0609] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 E69X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 K이다.

[0610] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 L139X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 P이다.

[0611] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 T197X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 A이다.

[0612] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 D200X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 N이다.

[0613] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 H204X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 R이다.

[0614] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 F209X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 N이다.

[0615] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 E302X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 K이다.

[0616] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 E302X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 R이다.

[0617] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 T306X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 K이다.

[0618] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 F309X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 N이다.

[0619] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미

노산 위치에서 W313X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 F이다.

[0620] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 T330X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 P이다.

[0621] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 L345X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 G이다.

[0622] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 L435X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 G이다.

[0623] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 N454X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 K이다.

[0624] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 D524X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 G이다.

[0625] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 E562X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 Q이다.

[0626] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 D583X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 N이다.

[0627] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 H594X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 Q이다.

[0628] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 L603X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 W이다.

[0629] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 E607X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 K이다.

[0630] 다양한 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 (RT는 융합 파트너로서 또는 트랜스로 제공됨)는 서열식별번호: 89의 야생형 M-MLV RT에서 또는 또 다른 야생형 RT 폴리펩티드 서열에서의 상응하는 아미노산 위치에서 D653X 돌연변이를 포함하는 변이체 RT를 포함할 수 있으며, 여기서 "X"는 임의의 아미노산일 수 있다. 특정 실시양태에서, X는 N이다.

[0631] 본 개시내용의 다양한 실시양태에 따른 napDNAbp 단백질에 융합될 수 있거나 또는 개별 단백질로서 제공될 수 있는 일부 예시적인 리버스 트랜스크립타제가 하기 제공된다. 예시적인 리버스 트랜스크립타제는 하기 야생형 효소 또는 부분 효소와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 적어도 99% 서열 동일성을 갖는 변이체를 포함한다:

설명	서열 (야생형 대비 변이체 치환)
리버스 트랜스크립타제 (M-MLV RT) 야생형 몰로니 뮤린 백혈병 바이러스 PE1 (본원에 개시된 프라임 편집제 1 융합 단백질)에 사용됨	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 89)
M-MLV RT D200N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 106)
M-MLV RT D200N T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN

[0632]

설명	서열 (야생형 대비 변이체 치환)
	VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 107)
M-MLV RT D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 108)
M-MLV RT E69K D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQ**K**ARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 109)
M-MLV RT D200N T330P L603W E302R	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLR**R**FLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP(S EQ ID NO: 110)

[0633]

설명	서열 (야생형 대비 변이체 치환)
M-MLV RT D200N T330P L603W E607K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 111)
M-MLV RT D200N T330P L603W L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 112)
M-MLV RT D200N T330P L603W L435G	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI**G**A PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 113)
M-MLV RT D200N T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT

[0634]

설명	서열 (야생형 대비 변이체 치환)
L603W N454K	VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 114)
M-MLV RT D200N T330P L603W T306K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLG**K**AGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPD QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQ EGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLP KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSS P (SEQ ID NO: 115)
M-MLV RT D200N T330P L603W W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 116)
M-MLV RT D200N T330P L603W D524G E562Q	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ

[0635]

설명	서열 (야생형 대비 변이체 치환)
D583N	PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRA**Q**LIALTQALKMAEGKKLN VYT**N**SRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 117)
M-MLV RT D200N T330P L603W E302R W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLR**R**FLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 118)
M-MLV RT D200N T330P L603W E607K L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 119)
M-MLV RT P51L S67K T197A H204R E302K F309N W313F T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL II**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SP**A**LFDEAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI

[0636]

설명	서열 (야생형 대비 변이체 치환)
L435G N454K D524G D583N H594Q D653N	**G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPA TLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GSSLL QEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK KLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILALLKALFL PKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDTSTLLIENS SP (SEQ ID NO: 120)
M-MLV RT P51L S67K T197A D200N H204R E302K F309N W313F T330P L345G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL II**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SP**A**LF**N**EAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI **G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVALNPA TLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GSSLL QEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK KLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILALLKALFL PKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDTSTLLIENS SP (SEQ ID NO: 121)
M-MLV RT D200N T330P L603W T306K W313F PE2에서의 것	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 122)

[0637]

[0638] 본원에 기재된 다중-플랩 프라임 편집제 시스템은 하기 미국 특허 (이들 각각은 그 전문이 참조로 포함됨): 미국 특허 번호 10,202,658; 10,189,831; 10,150,955; 9,932,567; 9,783,791; 9,580,698; 9,534,201; 및 9,458,484 중 임의의 것에 기재되어 있거나 개시된 임의의 공중 이용가능한 리버스 트랜스크립타제, 및 돌연변이를 설치하는 것에 대한 공지된 방법 또는 단백질을 진화시키는 것에 대한 공지된 방법을 사용하여 제조될 수 있는 그의 임의의 변이체를 고려한다. 하기 참고문헌은 관련 기술분야의 리버스 트랜스크립타제를 기재한다. 각각의 그의 개시내용은 그 전문이 본원에 참조로 포함된다.

Herzig, E., Voronin, N., Kucherenko, N. & Hizi, A. A Novel Leu92 Mutant of HIV-1 Reverse Transcriptase with a Selective Deficiency in Strand Transfer Causes a Loss of Viral Replication. *J. Virol.* **89,** 8119-8129 (2015).

Mohr, G. *et al.* A Reverse Transcriptase-CasI Fusion Protein Contains a Cas6 Domain Required for Both CRISPR RNA Biogenesis and RNA Spacer Acquisition. *Mol. Cell* **72,** 700-714.e8 (2018).

Zhao, C., Liu, F. & Pyle, A. **M.** An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* **24,** 183-195 (2018).

Zimmerly, S. & Wu, L. An Unexplored Diversity of Reverse Transcriptases in Bacteria. *Microbiol Spectr* **3,** MDNA3-0058-2014 (2015).

Ostertag, E. M. & Kazazian Jr, H. H. Biology of Mammalian LI Retrotransposons. *Annual Review of Genetics* **35,** 501-538 (2001).

Perach, M. & Hizi, A. Catalytic Features of the Recombinant Reverse Transcriptase of Bovine Leukemia Virus Expressed in Bacteria. *Virology* **259,** 176-189 (1999).

Lim, D. *et al.* Crystal structure of the moloney murine leukemia vims RNase H domain. *J. Virol.* **80,** 8379-8389 (2006).

Zhao, C. & Pyle, A. M. Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nature Structural & Molecular Biology* **23,** 558-565 (2016).

Griffiths, D. J. Endogenous retroviruses in the human genome sequence. *Genome Biol.* **2,** REVIEWS 1017 (2001).

[0639]

Baranauskas, A. *el al.* Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng Des Sel* **25,** 657-668 (2012).

Zimmerly, S., Guo, H., Perlman, P. S. & Lambowltz, A. M. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82,** 545-554 (1995).

Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human LI retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87,** 905-916 (1996).

Berkhout, B., Jebbink, M. & Zsíros, J. Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus. *Journal of Virology* **73,** 2365-2375 (1999).

Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M. & Gerard, G. F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res* **16,** 265-277 (1988).

Arezi, B. & Hogrefe, H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res* **37,** 473-481 (2009).

Blain, S. W. & Goff, S. P. Nuclease activities of Moloney murine leukemia vims reverse transcriptase. Mutants with altered substrate specificities. *J. Biol. Chem.* **268,** 23585-23592 (1993).

Xiong, Y. & Eickbush, T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* **9,** 3353-3362 (1990).

Herschhorn, A. & Hizi, A. Retroviral reverse transcriptases. *Cell. Mol. Life Sci.* **67,** 2717-2747 (2010).

Taube, R., Loya, S., Avidan, O., Perach, M. & Hizi, A. Reverse transcriptase of mouse mammary tumour vims: expression in bacteria, purification and biochemical characterization. *Biochem. J.* **329 (Pt 3),** 579-587 (1998).

Liu, M. *et al.* Reverse Transcriptase-Mediated Tropism Switching in Bordetella Bacteriophage. *Science* **295,** 2091-2094 (2002).

Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72,** 595-605 (1993).

Nottingham, R. M. *et al.* RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA* **22,** 597-613 (2016).

[0640]

Telesnitsky, A. & Goff, S. P. RNase H domain mutations affect the interaction between Moloney murine leukemia virus reverse transcriptase and its primer- template. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 1276-1280 (1993).

Halvas, E. K., Svarovskaia, E. S. & Pathak, V. K. Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity. *Journal of Virology* **74,** 10349-10358 (2000).

Nowak, E. *et al.* Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic Acids Res* **41,** 3874-3887 (2013).

Stamos, J. L., Lentzsch, A. M. & Lambowitz, A. M. Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications. *Molecular Cell* **68,** 926-939. e4 (2017).

Das, D. & Georgiadis, M. M. The Crystal Structure of the Monomeric Reverse Transcriptase from Moloney Murine Leukemia Virus. *Structure* **12,** 819-829 (2004).

Avidan, O., Meer, M. E., Oz, I. & Hizi, A. The processivity and fidelity of DNA synthesis exhibited by the reverse transcriptase of bovine leukemia virus. *European Journal of Biochemistry* **269,** 859-867 (2002).

Gerard, G. F. *et al.* The role of template-primer in protection of reverse transcriptase from thermal inactivation. *Nucleic Acids Res* **30,** 3118-3 129 (2002).

Monot, C. *et al.* The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. *PLOS Genetics* **9,** e1003499 (2013).

Mohr, S. *et al.* Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19,** 958-970 (2013).

[0641]

[0642] 리버스 트랜스크립타제에 관한 상기 언급된 참고문헌 중 임의의 것은, 이미 언급되지 않은 한, 그 전문이 본원에 참조로 포함된다.

[0643] [4] 프라임 편집제

[0644] 본 개시내용은 프라임 편집제를 포함하는 시스템을 제공한다. 한 측면에서, 본 개시내용은 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체를 포함하는, 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템으로서, 여기서 각각의 제1 및 제2 프라임 편집제 복합체는 (1) (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 (ii) RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 프라임 편집제; 및 (2) 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함하는 pegRNA를 포함하고, 여기서 제1 프라임 편집제 복합체의 pegRNA의 DNA 합성 주형은 제1 단일 가닥 DNA 서열을 코딩하고, 제2 프라임 편집제 복합체의 pegRNA의 DNA 합성 주형은 제2 단일 가닥 DNA 서열을 코딩하고, 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고, 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 편집될 표적 부위에서의 DNA 서열과 비교하여 편집된 부분을 포함하는 듀플렉스를 형성하고, 이는 편집될 표적 부위 내로 통합되는 것인 시스템을 제공한다.

[0645] 또 다른 측면에서, 본 개시내용은 1개 이상의 이중 가닥 DNA 서열을 편집하기 위한 시스템으로서, 시스템은 하기를 포함하고:

[0646] a) 하기를 포함하는 제1 프라임 편집제 복합체:

[0647] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[0648] ii. 편집될 제1 표적 부위에서 제1 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[0649] b) 하기를 포함하는 제2 프라임 편집제 복합체:

[0650] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는

제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[0651] ii. 편집될 제1 표적 부위에서 제1 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[0652] c) 하기를 포함하는 제3 프라임 편집제 복합체:

[0653] i. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

[0654] ii. 편집될 제2 표적 부위에서 제2 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA);

[0655] d) 하기를 포함하는 제4 프라임 편집제 복합체:

[0656] i. 제4 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

[0657] ii. 편집될 제2 표적 부위에서 제2 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

[0658] 여기서 제1 PEgRNA는 제1 단일 가닥 DNA 서열을 코딩하는 제1 DNA 합성 주형을 포함하고, 제2 PEgRNA는 제2 단일 가닥 DNA 서열을 코딩하는 제2 DNA 합성 주형을 포함하고, 제3 PEgRNA는 제3 단일 가닥 DNA 서열을 코딩하는 제3 DNA 합성 주형을 포함하고, 제4 PEgRNA는 제4 단일 가닥 DNA 서열을 코딩하는 제4 DNA 합성 주형을 포함하고;

[0659] 여기서 제1 및 제3 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고; 여기서 제2 및 제4 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하는 것인 시스템을 제공한다.

[0660] 일부 실시양태에서, 본원에 기재된 시스템에 사용된 프라임 편집제 (예를 들어, 상기 기재된 시스템에서의 제1 프라임 편집제, 제2 프라임 편집제, 제3 프라임 편집제 및/또는 제4 프라임 편집제)는 융합 단백질로서 제공된다. 일부 실시양태에서, 프라임 편집제 융합 단백질은 napDNAbp 및 폴리머라제 (예를 들어, DNA-의존성 DNA 폴리머라제 또는 RNA-의존성 DNA 폴리머라제, 예컨대 리버스 트랜스크립타제)를 포함한다. 일부 실시양태에서, napDNAbp와 폴리머라제는 임의로 링커에 의해 연결되어 융합 단백질을 형성한다. 프라임 편집제 융합 단백질 및 융합 단백질의 추가의 도메인의 다양한 구성이 본원에 추가로 기재된다. 다양한 실시양태에서, 본 개시내용에 의해 제공된 시스템 및 방법은 본원에 기재된 임의의 프라임 편집제 융합 단백질의 사용을 고려한다.

[0661] 일부 실시양태에서, 본원에 기재된 시스템에 사용된 프라임 편집제 복합체는 프라임 편집제 (예를 들어, 상기 기재된 시스템에서의 제1 프라임 편집제, 제2 프라임 편집제, 제3 프라임 편집제 및/또는 제4 프라임 편집제)를 포함하며, 여기서 프라임 편집제 중 1종 이상의 성분은 본 명세서 전반에 걸쳐 추가로 상세하게 기재된 바와 같이 트랜스로 제공된다. 일부 실시양태에서, 프라임 편집제는 트랜스로 발현된 napDNAbp 및 폴리머라제를 포함한다. 일부 실시양태에서, napDNAbp 및 폴리머라제는 1개 이상의 벡터로부터 발현된다 (예를 들어, 두 성분이 동일한 벡터로부터 발현되거나 또는 각각의 성분이 상이한 벡터로부터 발현됨). 특정 실시양태에서, 프라임 편집제는 트랜스로 발현된 본원에 기재된 바와 같은 추가의 성분을 포함한다. 일부 실시양태에서, 본원에 기재된 시스템에 사용된 프라임 편집제는 융합 단백질로서 제공된 1종 이상의 프라임 편집제 및 트랜스로 성분이 제공된 1종 이상의 프라임 편집제 둘 다를 포함할 수 있다.

[0662] 일부 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제 시스템은 napDNAbp 및 폴리머라제 (예를 들어, DNA-의존성 DNA 폴리머라제 또는 RNA-의존성 DNA 폴리머라제, 예컨대 리버스 트랜스크립타제)를 포함하고, 임의로 링커에 의해 연결된 융합 단백질을 고려한다. 본 출원은 단일 융합 단백질로 조합될 임의의 적합한 napDNAbp 및 폴리머라제 (예를 들어, DNA-의존성 DNA 폴리머라제 또는 RNA-의존성 DNA 폴리머라제, 예컨대 리버스 트랜스크립타제)를 고려한다. napDNAbp 및 폴리머라제 (예를 들어, DNA-의존성 DNA 폴리머라제 또는 RNA-의존성 DNA 폴리머라제, 예컨대 리버스 트랜스크립타제)의 예는 각각 본원에 규정된다. 폴리머라제는 관련 기술 분야에 널리 공지되어 있고, 아미노산 서열은 용이하게 이용가능하기 때문에, 본 개시내용은 어떠한 방식으로도 본원에서 확인된 그러한 특정 폴리머라제로 제한되는 것으로 의도되지 않는다.

[0663] 다양한 실시양태에서, 융합 단백질은 임의의 적합한 구조적 구성을 포함할 수 있다. 예를 들어, 융합 단백질은 N-말단에서 C-말단 방향으로 napDNAbp 융합된 폴리머라제 (예를 들어, DNA-의존성 DNA 폴리머라제 또는 RNA-의

존성 DNA 폴리머라제, 예컨대 리버스 트랜스크립타제)를 포함할 수 있다. 다른 실시양태에서, 융합 단백질은 N-말단에서 C-말단 방향으로 폴리머라제 (예를 들어, 리버스 트랜스크립타제) 융합된 napDNAbp를 포함할 수 있다. 융합된 도메인은 임의로 링커, 예를 들어 아미노산 서열에 의해 연결될 수 있다. 다른 실시양태에서, 융합 단백질은 구조 NH_2-[napDNAbp]-[폴리머라제]-COOH; 또는 NH_2-[폴리머라제]-[napDNAbp]-COOH를 포함할 수 있으며, 여기서 각각의 경우의 "]-["는 임의적인 링커 서열의 존재를 나타낸다. 폴리머라제가 리버스 트랜스크립타제인 실시양태에서, 융합 단백질은 구조 NH_2-[napDNAbp]-[RT]-COOH; 또는 NH_2-[RT]-[napDNAbp]-COOH를 포함할 수 있으며, 여기서 각각의 경우의 "]-["는 임의적인 링커 서열의 존재를 나타낸다.

[0664] 예시적인 융합 단백질이 도 14에 도시되어 있으며, 이는 링커 서열을 통해 닉카제 Cas9 ("Cas9(H840A)")에 융합된 MLV 리버스 트랜스크립타제 ("MLV-RT")를 포함하는 융합 단백질을 보여준다. 이러한 예는 본원에 기재된 프라임 편집제 (PE) 시스템에 이용될 수 있는 융합 단백질의 범주를 제한하는 것으로 의도되지 않는다.

[0665] 다양한 실시양태에서, 다중-플랩 프라임 편집제 융합 단백질은 H840A 돌연변이를 포함하는 Cas9 변이체 (즉, Cas9 닉카제) 및 M-MLV RT 야생형, 뿐만 아니라 Cas9 닉카제 도메인의 C-말단을 RT 도메인의 N-말단에 연결하는 N-말단 NLS 서열 (19개 아미노산) 및 아미노산 링커 (32개 아미노산)를 포함하는 하기 아미노산 서열 (본원에서 "PE1"로 지칭됨)을 가질 수 있다. PE1 융합 단백질은 하기 구조를 갖는다: [NLS]-[Cas9(H840A)]-[링커]-[MMLV_RT(wt)]. PE1 및 그의 개별 성분의 아미노산 서열은 하기와 같다:

설명	서열
PE1 융합 단백질 CAS9(H840A)-MMLV_RT(WT)	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGS*TLNIEDEYRLH ETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQ PLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQV KYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPG FAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKAL FLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSS PSGGSKRTADGSEFEPKKKRKV** (SEQ ID NO: 123)

[0666]

설명	서열
	설명: 핵 국재화 서열 (NLS) 상단:(SEQ ID NO: 124), 하단: (SEQ ID NO: 133) **CAS9(H840A)** (SEQ ID NO: 126) **33-**아미노산 링커 (SEQ ID NO: 127) *M-MLV 리버스 트랜스크립타제* (SEQ ID NO: 128)
PE1 - N-말단 NLS	MKRTADGSEFESPKKKRKV (SEQ ID NO: 124)
PE1 - CAS9 (H840A) (MET 마이너스)	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD (SEQ ID NO: 130)
PE1 - 링커 CAS9 도메인과 RT 도메인 사이 (33개 아미노산)	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 127)
PE1 - M-MLV RT	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 132)
PE1 - C-말단 NLS	SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 133)

[0667]

[0668]

[0669] 또 다른 실시양태에서, 다중-플랩 프라임 편집제 융합 단백질은 H840A 돌연변이를 포함하는 Cas9 변이체 (즉, Cas9 닉카제) 및 돌연변이 D200N, T330P, L603W, T306K, 및 W313F를 포함하는 M-MLV RT, 뿐만 아니라 Cas9 닉

카제 도메인의 C-말단을 RT 도메인의 N-말단에 연결하는 N-말단 NLS 서열 (19개 아미노산) 및 아미노산 링커 (33개 아미노산)를 포함하는 하기 아미노산 서열 (본원에서 "PE2"로 지칭됨)을 가질 수 있다. PE2 융합 단백질은 하기 구조를 갖는다: [NLS]-[Cas9(H840A)]-[링커]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)]. PE2의 아미노산 서열은 하기와 같다:

PE2 융합 단백질 CAS9(H840A)-MMLV_RT D200N T330P L603W T306K W313F	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TLNIEDEYRLH ETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQ PLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQV KYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPG FAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKA LFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENS SP**SGGSKRTADGSEFEPKKKRKV** (SEQ ID NO: 134) 설명:

[0670]

	핵 국재화 서열 (NLS) 상단:(SEQ ID NO: 124), 하단: (SEQ ID NO: 133) **CAS9(H840A)** (SEQ ID NO: 137) **33-**아미노산 링커 (SEQ ID NO: 127) *M-MLV 리버스 트랜스크립타제* (SEQ ID NO: 139)
PE2 - N-말단 NLS	MKRTADGSEFESPKKKRKV (SEQ ID NO: 124)
PE2 - CAS9 (H840A) (MET 마이너스)	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD (SEQ ID NO: 141)
PE2 - 링커 CAS9 도메인과 RT 도메인 사이 (33개 아미노산)	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 127)
PE2 - MMLV_RT D200N T330P L603W T306K W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ

[0671]

	QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 143)
PE2 - C-말단 NLS	SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 133)

[0672]

[0673] 또 다른 실시양태에서, 프라임 편집제 융합 단백질은 하기 아미노산 서열을 가질 수 있다:

PE 융합 단백질 MMLV_RT(WT)-32AA-CAS9(H840A)	MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQ AWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYN LLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELD CQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARK ETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAH IHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSP**SGGSSGGSSGSETPGTSES ATPESSGGSSGGSS**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKV LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL QNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 145) 설명:

[0674]

	핵 국재화 서열 (NLS) 상단:(SEQ ID NO: 124), 하단: (SEQ ID NO: 133) **CAS9(H840A)** (SEQ ID NO: 147) **33-아미노산 링커** (SEQ ID NO: 127) *M-MLV 리버스 트랜스크립타제* (SEQ ID NO: 149)
PE 융합 단백질 MMLV_RT(WT)-60AA-CAS9(H840A)	MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQ AWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYN LLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELD CQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARK ETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAH IHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAE ARGNRMADQAARKAAITETPDTSTLLIENSSPS***GGSSGGSSGSETPGTSES ATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSSGGS*DKKYS IGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD LTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE

[0675]

	VLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 150) 설명: 핵 국재화 서열 (NLS) 상단:(SEQ ID NO: 124), 하단: (SEQ ID NO: 133) **CAS9(H840A)**(SEQ ID NO: 153) 아미노산 링커 (SEQ ID NO: 175) *M-MLV 리버스 트랜스크립타제* (SEQ ID NO: 149)
PE 융합 단백질 CAS9(H840A)-FEN1-MMLV_RT D200N T330P L603W T306K W313F	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS QLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**GIQGLAKLIA DVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQNEEGETT SHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSERRAEAE KQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDA PSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLP IQEFHLSRILQELGLNQEQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHK SIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPNE EELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLDDFFKVTGSLSS AKRKEPEPKGSTKKKAKTGAAGKFKRGK**SGGSSGGSSGSETPGTSES ATPESSGGSSGGSS***TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETG*

[0676]

	GMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPC QSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPP SHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRA LLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQP TPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALV KQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHN CLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTET EVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGN RMADQAARKAAITETPDTSTLLIENSSP<u>SGGSKRTADGSEFEPKKKRKV</u>
(SEQ ID NO: 154)

설명:
<u>핵 국재화 서열 (NLS)</u> 상단:(SEQ ID NO: 124), 하단: (SEQ ID NO: 133)
CAS9(H840A) (SEQ ID NO: 157)
<u>33-아미노산 링커 1</u> (SEQ ID NO: 127)
M-MLV 리버스 트랜스크립타제 (SEQ ID NO: 159)
<u>33-아미노산 링커 2</u> (SEQ ID NO: 127)
FEN1 (SEQ ID NO: 161) |

[0677]

[0678] 다른 실시양태에서, 다중-플랩 프라임 편집제 융합 단백질은 변경된 PAM 특이성을 갖는 SaCas9 또는 SpCas9 닉카제, 예컨대 하기 예시적인 서열에 기초할 수 있다:

SACAS9-M-MLV RT 프라임 편집제	MKRTADGSEFESPKKKRKVGKRNYILGLDIGITSVGYGIIDYETR DVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKK LLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLA KRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLE RLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFID TYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPE ELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIEN VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYH DIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQE EIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKL VPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG LPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGK ENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII PRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETF KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFK KERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMF EEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKK PNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLIN KSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGN YLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYE EAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE VNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNL YEVKSKKHPQIIKKGSGGSSGGSSGSETPGTSESATPESSGGSSG GSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGL AVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWR DPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDL ILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQ ICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREF LGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRR PVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAEL IALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSE GKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRM ADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEPKKKR KV (SEQ ID NO: 162)
SPCAS9(H840A)-VRQR-말로니 뮤린 백혈병 바이러스 리버스	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN

[0679]

트랜스크립타제 프라임 편집제	ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK KDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESAT PESSGGSSGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQA WAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKP HIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPT SQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDL ADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLG YRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPT PKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVL TQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAG KLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTD QPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPA GTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIY RRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHS AEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADG SEFEPKKKRKV (SEQ ID NO: 163)
SPCAS9(H840A)-VRER-말로니 뮤린 백혈병 바이러스 리버스	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN

[0680]

트랜스크립타제 프라임 편집제	ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK KDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVL DATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESAT PESSGGSSGGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQA WAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKP HIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVN KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPT SQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDL ADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLG YRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPT PKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVL TQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAG KLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTD QPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPA GTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIY RRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHS AEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADG SEFEPKKKRKV (SEQ ID NO: 164)

[0681]

[0682] 또 다른 실시양태에서, 본원에서 고려되는 다중-플랩 프라임 편집제 융합 단백질은 M-MLV 리버스 트랜스크립타제의 말단절단된 버전에 융합된 Cas9 닉카제 (예를 들어, Cas9 (H840A))를 포함할 수 있다. 이러한 실시양태에서, 리버스 트랜스크립타제는 또한 4개의 돌연변이를 함유한다 (D200N, T306K, W313F, T330P; PE2에 존재하는 L603W 돌연변이는 말단절단으로 인해 더 이상 존재하지 않음에 주목). 이러한 말단절단된 편집제를 코딩하는 DNA 서열은 PE2보다 522 bp 더 작고, 따라서 DNA 서열의 전달이 그의 크기로 인해 도전과제인 적용 (즉, 아데노-연관 바이러스 및 렌티바이러스 전달)에서 잠재적으로 유용하게 한다. 이러한 실시양태는 Cas9(H840A)-MMLV-RT(trunc) 또는 "PE2-짧은 형" 또는 "PE2-trunc"로서 지칭되고, 하기 아미노산 서열을 갖는다:

CAS9(H840A)-MMLV-RT(TRUNC) 또는 PE2-짧은 형	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQLGGD**<u>SGGSSGGSSGSETPGTSE SATPESSGGSSGGSS</u>*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQ AWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHI QRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVE DIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFA FEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQH PDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQI CQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGK AGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEA LVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLP EEGLQHNCLDNSRLIN**<u>SGGSKRTADGSEFEPKKKRKV</u> (SEQ ID NO: 765)

[0683]

	설명: <u>핵 국재화 서열 (NLS)</u> 상단:(SEQ ID NO: 124), 하단: (SEQ ID NO: 133) **CAS9(H840A)** (SEQ ID NO: 157) **<u>33</u>-아미노산 링커 1** (SEQ ID NO: 127) *M-MLV 말단절단된 리버스 트랜스크립타제* (SEQ ID NO: 766)

[0684]

[0685] 다양한 세포주에서의 상이한 표적 부위에 대한 PE2, PE2-trunc, PE3, 및 PE3-trunc의 효율 (즉, "명시된 편집 또는 indel을 갖는 총 서열분석 판독물의 %")을 비교한 막대 그래프를 제공하는 도 75를 참조한다. 데이터는 말단절단된 RT 변이체를 포함하는 프라임 편집제가 약 비-말단절단된 RT 단백질을 포함하는 프라임 편집제만큼 효율적이었음을 보여준다.

[0686] 다양한 실시양태에서, 본원에서 고려되는 다중-플랩 프라임 편집제 융합 단백질은 또한 PE1, PE2, 또는 상기 표시된 프라임 편집제 융합 서열 중 임의의 것과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일한 아미노산 서열을 갖는 상기 개시된 서열의

임의의 변이체를 포함할 수 있다.

[0687] 특정 실시양태에서, 링커는 본 발명의 펩티드 또는 펩티드 도메인 또는 모이어티 중 임의의 것을 연결하는 데 사용될 수 있다 (예를 들어, 리버스 트랜스크립타제에 연결되거나 융합된 napDNAbp).

[0688] [5] 링커 및 다른 도메인

[0689] PE 융합 단백질은 napDNAbp (예를 들어, Cas9 도메인) 및 폴리머라제 도메인 (예를 들어, RT 도메인) 이외의 다양한 다른 도메인을 포함할 수 있다. 예를 들어, napDNAbp가 Cas9이고 폴리머라제가 RT인 경우에, PE 융합 단백질은 Cas9 도메인을 RT 도메인과 연결하는 1개 이상의 링커를 포함할 수 있다. 링커는 또한 다른 기능적 도메인, 예컨대 핵 국재화 서열 (NLS) 또는 FEN1 (또는 다른 플랩 엔도뉴클레아제)을 PE 융합 단백질 또는 그의 도메인에 연결할 수 있다.

[0690] 또한, 트랜스 프라임 편집을 수반하는 실시양태에서, 링커는 tPERT 동원 단백질을 프라임 편집제에, 예를 들어 tPERt 동원 단백질과 napDNAbp 사이를 연결하는 데 사용될 수 있다. 예를 들어, 폴리머라제 도메인 및 동원 단백질 도메인을 napDNAbp에 개별적으로 융합시키는 링커를 포함하는 트랜스 프라임 편집제 (tPE)의 예시적인 개략도에 대해 도 3g를 참조한다.

[0691] A. 링커

[0692] 상기 정의된 바와 같이, 본원에 사용된 용어 "링커"는 2개의 분자 또는 모이어티, 예를 들어 뉴클레아제의 절단 도메인 및 결합 도메인을 연결하는 화학적 기 또는 분자를 지칭한다. 일부 실시양태에서, 링커는 RNA-프로그램 가능한 뉴클레아제의 gRNA 결합 도메인과 폴리머라제 (예를 들어, 리버스 트랜스크립타제)의 촉매 도메인을 연결한다. 일부 실시양태에서, 링커는 dCas9 및 리버스 트랜스크립타제를 연결한다. 전형적으로, 링커는 2개의 기, 분자, 또는 다른 모이어티 사이에 위치하거나 또는 그에 플랭킹하고, 공유 결합을 통해 서로 연결하여, 2개를 연결한다. 일부 실시양태에서, 링커는 아미노산 또는 복수의 아미노산 (예를 들어, 펩티드 또는 단백질)이다. 일부 실시양태에서, 링커는 유기 분자, 기, 중합체, 또는 화학적 모이어티이다. 일부 실시양태에서, 링커는 5-100개 아미노산 길이, 예를 들어, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 또는 150-200개 아미노산 길이이다. 보다 길거나 보다 짧은 링커가 또한 고려된다.

[0693] 링커는 공유 결합만큼 단순할 수 있거나, 또는 많은 원자 길이의 중합체 링커일 수 있다. 특정 실시양태에서, 링커는 폴리펩티드이거나 또는 아미노산을 기초로 한다. 다른 실시양태에서, 링커는 펩티드-유사가 아니다. 특정 실시양태에서, 링커는 공유 결합 (예를 들어, 탄소-탄소 결합, 디술피드 결합, 탄소-헤테로원자 결합 등)이다. 특정 실시양태에서, 링커는 아미드 연결의 탄소-질소 결합이다. 특정 실시양태에서, 링커는 시클릭 또는 비-시클릭, 치환 또는 비치환, 분지형 또는 비분지형 지방족 또는 헤테로지방족 링커이다. 특정 실시양태에서, 링커는 중합체 (예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아미드, 폴리에스테르 등)이다. 특정 실시양태에서, 링커는 아미노알칸산의 단량체, 이량체, 또는 중합체를 포함한다. 특정 실시양태에서, 링커는 아미노알칸산 (예를 들어, 글리신, 에탄산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노부탄산, 5-펜탄산 등)을 포함한다. 특정 실시양태에서, 링커는 아미노헥산산 (Ahx)의 단량체, 이량체, 또는 중합체를 포함한다. 특정 실시양태에서, 링커는 카르보시클릭 모이어티 (예를 들어, 시클로펜탄, 시클로헥산)를 기초로 한다. 다른 실시양태에서, 링커는 폴리에틸렌 글리콜 모이어티 (PEG)를 포함한다. 다른 실시양태에서, 링커는 아미노산을 포함한다. 특정 실시양태에서, 링커는 펩티드를 포함한다. 특정 실시양태에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 특정 실시양태에서, 링커는 페닐 고리를 기초로 한다. 링커는 펩티드로부터 링커로의 친핵체 (예를 들어, 티올, 아미노)의 부착을 용이하게 하기 위해 관능화된 모이어티를 포함할 수 있다. 임의의 친전자체가 링커의 일부로서 사용될 수 있다. 예시적인 친전자체는 활성화된 에스테르, 활성화된 아미드, 마이클 수용자, 알킬 할라이드, 아릴 할라이드, 아실 할라이드, 및 이소티오시아네이트를 포함하나 이에 제한되지는 않는다.

[0694] 일부 다른 실시양태에서, 링커는 아미노산 서열 $(GGGGS)_n$ (서열식별번호: 165), $(G)_n$ (서열식별번호: 166), $(EAAAK)_n$ (서열식별번호: 167), $(GGS)_n$ (서열식별번호: 168), $(SGGS)_n$ (서열식별번호: 169), $(XP)_n$ (서열식별번호: 170), 또는 그의 임의의 조합을 포함하고, 여기서 n은 독립적으로 1 내지 30의 정수이고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 링커는 아미노산 서열 (GGS)n (서열식별번호: 176)을 포함하고, 여기서 n은 1, 3, 또는 7이다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 171)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 172)를

포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSGGSGGS (서열식별번호: 173)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGS (서열식별번호: 174)를 포함한다. 다른 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSSGGS (서열식별번호: 175, 60AA)를 포함한다.

[0695] 특정 실시양태에서, 링커는 본 발명의 펩티드 또는 펩티드 도메인 또는 모이어티 중 임의의 것을 연결하는 데 사용될 수 있다 (예를 들어, 리버스 트랜스크립타제에 연결되거나 융합된 napDNAbp).

[0696] 상기 정의된 바와 같이, 본원에 사용된 용어 "링커"는 2개의 분자 또는 모이어티, 예를 들어 뉴클레아제의 절단 도메인 및 결합 도메인을 연결하는 화학적 기 또는 분자를 지칭한다. 일부 실시양태에서, 링커는 RNA-프로그램 가능한 뉴클레아제의 gRNA 결합 도메인과 레콤비나제의 촉매 도메인을 연결한다. 일부 실시양태에서, 링커는 dCas9 및 리버스 트랜스크립타제를 연결한다. 전형적으로, 링커는 2개의 기, 분자, 또는 다른 모이어티 사이에 위치하거나 또는 그에 플랭킹하고, 공유 결합을 통해 서로 연결하여, 2개를 연결한다. 일부 실시양태에서, 링커는 아미노산 또는 복수의 아미노산 (예를 들어, 펩티드 또는 단백질)이다. 일부 실시양태에서, 링커는 유기 분자, 기, 중합체, 또는 화학적 모이어티이다. 일부 실시양태에서, 링커는 5-100개 아미노산 길이, 예를 들어, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 또는 150-200개 아미노산 길이이다. 보다 길거나 보다 짧은 링커가 또한 고려된다.

[0697] 링커는 공유 결합만큼 단순할 수 있거나, 또는 많은 원자 길이의 중합체 링커일 수 있다. 특정 실시양태에서, 링커는 폴리펩티드이거나 또는 아미노산을 기초로 한다. 다른 실시양태에서, 링커는 펩티드-유사가 아니다. 특정 실시양태에서, 링커는 공유 결합 (예를 들어, 탄소-탄소 결합, 디술피드 결합, 탄소-헤테로원자 결합 등)이다. 특정 실시양태에서, 링커는 아미드 연결의 탄소-질소 결합이다. 특정 실시양태에서, 링커는 시클릭 또는 비-시클릭, 치환 또는 비치환, 분지형 또는 비분지형 지방족 또는 헤테로지방족 링커이다. 특정 실시양태에서, 링커는 중합체 (예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아미드, 폴리에스테르 등)이다. 특정 실시양태에서, 링커는 아미노알칸산의 단량체, 이량체, 또는 중합체를 포함한다. 특정 실시양태에서, 링커는 아미노알칸산 (예를 들어, 글리신, 에탄산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노부탄산, 5-펜탄산 등)을 포함한다. 특정 실시양태에서, 링커는 아미노헥산산 (Ahx)의 단량체, 이량체, 또는 중합체를 포함한다. 특정 실시양태에서, 링커는 카르보시클릭 모이어티 (예를 들어, 시클로펜탄, 시클로헥산)를 기초로 한다. 다른 실시양태에서, 링커는 폴리에틸렌 글리콜 모이어티 (PEG)를 포함한다. 다른 실시양태에서, 링커는 아미노산을 포함한다. 특정 실시양태에서, 링커는 펩티드를 포함한다. 특정 실시양태에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 특정 실시양태에서, 링커는 페닐 고리를 기초로 한다. 링커는 펩티드로부터 링커로의 친핵체 (예를 들어, 티올, 아미노)의 부착을 용이하게 하기 위해 관능화된 모이어티를 포함할 수 있다. 임의의 친전자체가 링커의 일부로서 사용될 수 있다. 예시적인 친전자체는 활성화된 에스테르, 활성화된 아미드, 마이클 수용자, 알킬 할라이드, 아릴 할라이드, 아실 할라이드, 및 이소티오시아네이트를 포함하나 이에 제한되지는 않는다.

[0698] 일부 다른 실시양태에서, 링커는 아미노산 서열 (GGGGS)n (서열식별번호: 165), (G)n (서열식별번호: 166), $(EAAAK)_n$ (서열식별번호: 167), $(GGS)_n$ (서열식별번호: 168), (SGGS)n (서열식별번호: 169), (XP)n (서열식별번호: 170), 또는 그의 임의의 조합을 포함하고, 여기서 n은 독립적으로 1 내지 30의 정수이고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 링커는 아미노산 서열 (GGS)n (서열식별번호: 176)을 포함하고, 여기서 n은 1, 3, 또는 7이다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 171)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 172)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSGGSGGS (서열식별번호: 173)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGS (서열식별번호: 174)를 포함한다.

[0699] 특히, 프라임 편집제 도메인을 서로 연결하기 위해 다양한 실시양태에서 하기 링커가 사용될 수 있다:

[0700] GGS (서열식별번호: 767);

[0701] GGSGGS (서열식별번호: 768);

[0702] GGSGGSGGS (서열식별번호: 769);

[0703] SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 127);

[0704] SGSETPGTSESATPES (서열식별번호: 171);

[0705] SGGSSGGSSGSETPGTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSSGG S (서열식별번호: 175).

[0706] B. 핵 국재화 서열 (NLS)

[0707] 다양한 실시양태에서, PE 융합 단백질은 단백질의 세포 핵으로의 전위를 촉진하는 것을 돕는 1개 이상의 핵 국재화 서열 (NLS)을 포함할 수 있다. 이러한 서열은 관련 기술분야에 널리 공지되어 있고, 하기 예를 포함할 수 있다:

[0708]

설명	서열	SEQ ID NO:
SV40 대형 T-AG의 NLS	PKKKRKV	SEQ ID NO: 16
NLS	MKRTADGSEFESPKKKRKV	SEQ ID NO: 124
NLS	MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	SEQ ID NO: 17
뉴클레오플라스민의 NLS	AVKRPAATKKAGQAKKKKLD	SEQ ID NO: 190
EGL-13의 NLS	MSRRRKANPTKLSENAKKLAKEVEN	SEQ ID NO: 191
C-MYC의 NLS	PAAKRVKLD	SEQ ID NO: 192
TUS-단백질의 NLS	KLKIKRPVK	SEQ ID NO: 193
폴리오마 대형 T-AG의 NLS	VSRKRPRP	SEQ ID NO: 194
D형 간염 바이러스 항원의 NLS	EGAPPAKRAR	SEQ ID NO: 195
뮤린 P53의 NLS	PPQPKKKPLDGE	SEQ ID NO: 196
PE1 및 PE2의 NLS	SGGSKRTADGSEFEPKKKRKV	SEQ ID NO: 133

[0709] 상기 NLS 예는 비제한적이다. PE 융합 단백질은 문헌 [Cokol et al., "Finding nuclear localization signals," EMBO Rep., 2000, 1(5): 411-415 and Freitas et al., "Mechanisms and Signals for the Nuclear Import of Proteins," Current Genomics, 2009, 10(8): 550-7]에 기재된 것 중 임의의 것을 비롯한 임의의 공지된 NLS 서열을 포함할 수 있으며, 이들 각각은 본원에 참조로 포함된다.

[0710] 다양한 실시양태에서, 다중-플랩 프라임 편집제 및 본원에 개시된 프라임 편집제를 코딩하는 구축물은 1개 이상, 바람직하게는 적어도 2개의 핵 국재화 신호를 추가로 포함한다. 특정 실시양태에서, 다중-플랩 프라임 편집제는 적어도 2개의 NLS를 포함한다. 적어도 2개의 NLS를 갖는 실시양태에서, NLS는 동일한 NLS일 수 있거나 또는 상이한 NLS일 수 있다. 추가로, NLS는 다중-플랩 프라임 편집제의 나머지 부분과의 융합 단백질의 일부로서 발현될 수 있다. 일부 실시양태에서, NLS 중 1개 이상은 2부분 NLS ("bpNLS")이다. 특정 실시양태에서, 개시된 융합 단백질은 2개의 2부분 NLS를 포함한다. 일부 실시양태에서, 개시된 융합 단백질은 2개 초과의 2부분 NLS를 포함한다.

[0711] NLS 융합체의 위치는 프라임 편집제의 N-말단, C-말단, 또는 서열 내일 수 있다 (예를 들어, 코딩된 napDNAbp 성분 (예를 들어, Cas9)과 폴리머라제 도메인 (예를 들어, 리버스 트랜스크립타제 도메인) 사이에 삽입됨).

[0712] NLS는 관련 기술분야의 임의의 공지된 NLS 서열일 수 있다. NLS는 또한 핵 국재화를 위한 임의의 추후-발견될 NLS일 수 있다. NLS는 또한 임의의 자연 발생 NLS, 또는 임의의 비-자연 발생 NLS (예를 들어, 1개 이상의 목적하는 돌연변이를 갖는 NLS)일 수 있다.

[0713] 용어 "핵 국재화 서열" 또는 "NLS"는, 예를 들어 핵 수송에 의해 세포 핵 내로의 단백질의 유입을 촉진하는 아미노산 서열을 지칭한다. 핵 국재화 서열은 관련 기술분야에 공지되어 있고, 통상의 기술자에게 명백할 것이다. 예를 들어, NLS 서열은 2001년 5월 31일에 WO/2001/038547로서 공개된 2000년 11월 23일에 출원된 국제 PCT 출원 PCT/EP2000/011690 (Plank et al.)에 기재되어 있으며, 그의 내용은 본원에 참조로 포함된다. 일

부 실시양태에서, NLS는 아미노산 서열 PKKKRKV (서열식별번호: 16), MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (서열식별번호: 17), KRTADGSEFESPKKKRKV (서열식별번호: 3864), 또는 KRTADGSEFEPKKKRKV (서열식별번호: 125)를 포함한다. 다른 실시양태에서, NLS는 아미노산 서열 NLSKRPAAIKKAGQAKKKK (서열식별번호: 136), PAAKRVKLD (서열식별번호: 192), RQRRNELKRSF (서열식별번호: 3934), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (서열식별번호: 3935)를 포함한다.

[0714] 개시내용의 한 측면에서, 다중-플랩 프라임 편집제는 1개 이상의 핵 국재화 신호 (NLS), 바람직하게는 적어도 2개의 NLS로 변형될 수 있다. 특정 실시양태에서, 다중-플랩 프라임 편집제는 2개 이상의 NLS로 변형된다. 개시내용은 개시내용의 시점에 관련 기술분야에 공지된 임의의 핵 국재화 신호, 또는 본 출원 시점 후에 최신 기술에서 확인되거나 또는 달리 이용가능하게 된 임의의 핵 국재화 신호의 사용을 고려한다. 대표적인 핵 국재화 신호는 서열이 발현될 세포의 핵으로 단백질을 지시하는 펩티드 서열이다. 핵 국재화 신호는 우세하게 염기성이고, 단백질의 아미노산 서열의 거의 어디에나 위치할 수 있고, 일반적으로 4개의 아미노산 (문헌 [Autieri & Agrawal, (1998) J. Biol. Chem. 273: 14731-37], 본원에 참조로 포함됨) 내지 8개의 아미노산의 짧은 서열을 포함하고, 전형적으로 리신 및 아르기닌 잔기가 풍부하다 (문헌 [Magin et al., (2000) Virology 274: 11-16], 본원에 참조로 포함됨). 핵 국재화 신호는 종종 프롤린 잔기를 포함한다. 다양한 핵 국재화 신호가 확인되었고, 생물학적 분자를 세포질로부터 세포의 핵으로 수송하는 데 사용되었다. 예를 들어, 문헌 [Tinland et al., (1992) Proc. Natl. Acad. Sci. U.S.A. 89:7442-46; Moede et al., (1999) FEBS Lett. 461:229-34]을 참조하고, 이는 참조로 포함된다. 전위는 현재 핵공 단백질을 수반하는 것으로 생각된다.

[0715] 대부분의 NLS는 3개의 일반적 군으로 분류될 수 있다: (i) SV40 대형 T 항원 NLS에 의해 예시되는 1부분 NLS (PKKKRKV (서열식별번호: 16)); (ii) 가변적 수의 스페이서 아미노산에 의해 분리되고, 크세노푸스(Xenopus) 뉴클레오플라스민 NLS에 의해 예시되는 2개의 염기성 도메인으로 이루어진 2부분 모티프 (KRXXXXXXXXXXKKKL (서열식별번호: 3936)); 및 (iii) 비정규 서열, 예컨대 hnRNP A1 단백질의 M9, 인플루엔자 바이러스 핵단백질 NLS, 및 효모 Gal4 단백질 NLS (Dingwall and Laskey 1991).

[0716] 핵 국재화 신호는 단백질의 아미노산 서열 내의 다양한 지점에서 나타난다. NLS는 단백질의 N-말단, C-말단 및 중심 영역에서 확인되었다. 따라서, 개시내용은 다중-플랩 프라임 편집제의 C-말단, N-말단, 뿐만 아니라 내부 영역에서 1개 이상의 NLS로 변형될 수 있는 다중-플랩 프라임 편집제를 제공한다. 성분 NLS 잔기로서 기능하지 않는 보다 긴 서열의 잔기는 핵 국재화 신호 자체를 예를 들어 장성에 의해 또는 입체적으로 방해하지 않도록 선택되어야 한다. 따라서, NLS-포함 서열의 조성에 대한 엄격한 제한은 없지만, 실제로, 이러한 서열은 길이 및 조성이 기능적으로 제한될 수 있다.

[0717] 본 개시내용은 1개 이상의 NLS를 포함하도록 다중-플랩 프라임 편집제를 변형시키는 임의의 적합한 수단을 고려한다. 한 측면에서, 다중-플랩 프라임 편집제는 그의 N-말단 또는 그의 C-말단 (또는 둘 다)에서 1개 이상의 NLS에 번역상 융합된 프라임 편집제 단백질을 발현하도록, 즉 프라임 편집제-NLS 융합 구축물을 형성하도록 조작될 수 있다. 다른 실시양태에서, 프라임 편집제-코딩 뉴클레오티드 서열은 코딩된 프라임 편집제의 내부 영역에 1개 이상의 NLS를 코딩하는 리딩 프레임이 혼입되도록 유전자 변형될 수 있다. 또한, NLS는 프라임 편집제와 N-말단, C-말단, 또는 내부-부착된 NLS 아미노산 서열 사이에, 예를 들어 단백질의 중심 영역에, 코딩된 다양한 아미노산 링커 또는 스페이서 영역을 포함할 수 있다. 따라서, 본 개시내용은 또한 프라임 편집제 및 1개 이상의 NLS를 포함하는 융합 단백질을 발현시키기 위한 뉴클레오티드 구축물, 벡터, 및 숙주 세포를 제공한다.

[0718] 본원에 기재된 다중-플랩 프라임 편집제는 또한 1개 이상의 링커, 예를 들어, 및 중합체, 아미노산, 핵산, 폴리사카라이드, 화학적, 또는 핵산 링커 요소를 통해 프라임 편집제에 연결된 핵 국재화 신호를 포함할 수 있다. 개시내용의 고려되는 범주 내의 링커는 어떠한 제한을 갖는 것으로 의도되지 않으며, 임의의 적합한 유형의 분자 (예를 들어, 중합체, 아미노산, 폴리사카라이드, 핵산, 지질, 또는 임의의 합성 화학적 링커 도메인)일 수 있고, 프라임 편집제와 1개 이상의 NLS 사이에 결합 (예를 들어, 공유 연결, 수소 결합) 형성을 유발하는 임의의 적합한 전략에 의해 프라임 편집제에 연결될 수 있다.

[0719] C. 플랩 엔도뉴클레아제 (예를 들어, FEN1)

[0720] 다양한 실시양태에서, PE 융합 단백질은 5' 단일 가닥 DNA 플랩의 제거를 촉매하는 효소를 지칭하는 플랩 엔도뉴클레아제 (예를 들어, FEN1)를 1개 이상 포함할 수 있다. 이들은 DNA 복제를 비롯한 세포 과정 동안 형성된 5' 플랩의 제거를 프로세싱하는 자연 발생 효소이다. 본원에 기재된 다중-플랩 프라임 편집 방법은 다중-플랩 프라임 편집 동안 표적 부위에 형성된 내인성 DNA의 5' 플랩을 제거하기 위해 내인성으로 공급된 플랩 엔도뉴클

레아제 또는 트랜스로 제공된 것을 이용할 수 있다. 플랩 엔도뉴클레아제는 관련 기술분야에 공지되어 있고, 문헌 [Patel et al., "Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends," Nucleic Acids Research, 2012, 40(10): 4507-4519 및 Tsutakawa et al., "Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily," Cell, 2011, 145(2): 198-211] (이들 각각은 본원에 참조로 포함됨)에 기재된 바와 같이 찾아볼 수 있다. 예시적인 플랩 엔도뉴클레아제는 하기 아미노산 서열로 표시될 수 있는 FEN1이다:

설명	서열	SEQ ID NO:
FEN1 야생형 (wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 198

[0721]

[0722] 플랩 엔도뉴클레아제는 또한 임의의 FEN1 변이체, 돌연변이체, 또는 다른 플랩 엔도뉴클레아제 오르토로그, 상동체, 또는 변이체를 포함할 수 있다. 비제한적 FEN1 변이체 예는 하기와 같다:

설명	서열	SEQ ID NO:
FEN1 K168R (FEN1 wt 대비)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALV**R**AGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 199
FEN1 S187A (FEN1 wt 대비)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFG **A**PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFV DLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLD PNKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEP NEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRL DDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFK RGK	SEQ ID NO: 200
FEN1 K354R (FEN1 wt 대비)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSA**R**RKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 201
GEN1	MGVNDLWQILEPVKQHIPLRNLGGKTIAVDLSLWVCE AQTVKKMMGSVMKPHLRNLFFRISYLTQMDVKLVFV MEGEPPKLKADVISKRNQSRYGSSGKSWSQKTGRSHF KSVLRECLHMLECLGIPWVQAAGEAEAMCAYLNAGG HVDGCLTNDGDTFLYGAQTVYRNFTMNTKDPHVDCY TMSSIKSKLGLDRDALVGLAILLGCDYLPKGVPGVGKE QALKLIQILKGQSLLQRFNRWNETSCNSSPQLLVTKKL AHCSVCSHPGSPKDHERNGCRLCKSDKYCEPHDYEYC CPCEWHRTEHDRQLSEVENNIKKKACCCEGFPFHEVIQ EFLLNKDKLVKVIRYQRPDLLLFQRFTLEKMEWPNHY ACEKLLVLLTHYDMIERKLGSRNSNQLQPIRIVKTRIRN GVHCFEIEWEKPEHYAMEDKQHGEFALLTIEEESLFEA AYPEIVAVYQKQKLEIKGKKQKRIKPKENNLPEPDEVM	SEQ ID NO: 202

[0723]

	SFQSHMTLKPTCEIFHKQNSKLNSGISPDPTLPQESISAS LNSLLLPKNTPCLNAQEQFMSSLRPLAIQQIKAVSKSLI SESSQPNTSSHNISVIADLHLSTIDWEGTSFSNSPAIQRN TFSHDLKSEVESELSAIPDGFENIPEQLSCESERYTANIK KVLDEDSDGISPEEHLLSGITDLCLQDLPLKERIFTKLSY PQDNLQPDVNLKTLSILSVKESCIANSGSDCTSHLSKDL PGIPLQNESRDSKILKGDQLLQEDYKVNTSVPYSVSNT VVKTCNVRPPNTALDHSRKVDMQTTRKILMKKSVCLD RHSSDEQSAPVFGKAKYTTQRMKHSSQKHNSSHFKES GHNKLSSPKIHIKETEQCVRSYETAENEESCFPDSTKSS LSSLQCHKKENNSGTCLDSPLPLRQRLKLRFQST	
ERCC5	MGVQGLWKLLECSGRQVSPEALEGKILAVDISIWLNQ ALKGVRDRHGNSIENPHLLTLFHRLCKLLFFRIRPIFVF DGDAPLLKKQTLVKRRQRKDLASSDSRKTTEKLLKTF LKRQAIKTAFRSKRDEALPSLTQVRRENDLYVLPPLQE EEKHSSEEEDEKEWQERMNQKQALQEEFFHNPQAIDIE SEDFSSLPPEVKHEILTDMKEFTKRRRTLFEAMPEESDD FSQYQLKGLLKKNYLNQHIEHVQKEMNQQHSGHIRRQ YEDEGGFLKEVESRRVVSEDTSHYILIKGIQAKTVAEV DSESLPSSSKMHGMSFDVKSSPCEKLKTEKEPDATPPSP RTLLAMQAALLGSSSEEELESENRRQARGRNAPAAVD EGSISPRTLSAIKRALDDDEDVKVCAGDDVQTGGPGAE EMRINSSTENSDEGLKVRDGKGIPFTATLASSSVNSAEE HVASTNEGREPTDSVPKEQMSLVHVGTEAFPISDESMI KDRKDRLPLESAVVRHSDAPGLPNGRELTPASPTCTNS VSKNETHAEVLEQQNELCPYESKFDSSLLSSDDETKCK PNSASEVIGPVSLQETSSIVSVPSEAVDNVENVVSFNAK EHENFLETIQEQQTTESAGQDLISIPKAVEPMEIDSEESE SDGSFIEVQSVISDEELQAEFPETSKPPSEQGEEELVGTR EGEAPAESESLLRDNSERDDVDGEPQEAEKDAEDSLHE WQDINLEELETLESNLLAQQNSLKAQKQQQERIAATVT GQMFLESQELLRLFGIPYIQAPMEAEAQCAILDLTDQTS GTITDDSDIWLFGARHVYRNFFNKNKFVEYYQYVDFH NQLGLDRNKLINLAYLLGSDYTEGIPTVGCVTAMEILN EFPGHGLEPLLKFSEWWHEAQKNPKIRPNPHDTKVKK KLRTLQLTPGFPNPAVAEAYLKPVVDDSKGSFLWGKP DLDKIREFCQRYFGWNRTKTDESLFPVLKQLDAQQTQ LRIDSFFRLAQQEKEDAKRIKSQRLNRAVTCMLRKEKE AAASEIEAVSVAMEKEFELLDKAKRKTQKRGITNTLEE SSSLKRKRLSDSKRKNTCGGFLGETCLSESSDGSSSEDA ESSSLMNVQRRTAAKEPKTSASDSQNSVKEAPVKNGG ATTSSSSDSDDDGGKEKMVLVTARSVFGKKRRKLRRA RGRKRKT	SEQ ID NO: 203

[0724]

[0725] 다양한 실시양태에서, 본원에서 고려되는 다중-플랩 프라임 편집제 융합 단백질은 상기 서열 중 임의의 것과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일한 아미노산 서열을 갖는 상기 개시된 서열의 임의의 플랩 엔도뉴클레아제 변이체를 포함할 수 있다.

[0726] 5' 단부 단일 가닥 DNA 플랩의 제거를 용이하게 하기 위해 본 방법에 의해 이용될 수 있는 다른 엔도뉴클레아제는 (1) trex 2, (2) exo1 엔도뉴클레아제 (예를 들어, Keijzers et al., Biosci Rep. 2015, 35(3): e00206)를 포함하나 이에 제한되지는 않는다.

[0727] Trex 2

[0728] 3' 3개의 프라임 복구 엑소뉴클레아제 2 (TREX2) - 인간

[0729] 수탁 번호 NM_080701

MSEAPRAETFVFLDLEATGLPSVEPEIAELSLFAVHRSSLENPEHDESGALVLPRVLD
KLTLCMCPERPFTAKASEITGLSSEGLARCRKAGFDGAVVRTLQAFLSRQAGPICLVA
HNGFDYDFPLLCAELRRLGARLPRDTVCLDTLPALRGLDRAHSHGTRARGRQGYSL
GSLFHRYFRAEPSAAHSAEGDVHTLLLIFLHRAAELLAWADEQARGWAHIEPMYLPP
DDPSLEA (SEQ ID NO: 3865).

[0730]

[0731] 3' 3개의 프라임 복구 엑소뉴클레아제 2 (TREX2) - 마우스

[0732] 수탁 번호 NM_011907

MSEPPRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLP
RVLDKLTLCMCPERPFTAKASEITGLSSESLMHCGKAGFNGAVVRTLQGFLSRQEGPI
CLVAHNGFDYDFPLLCTELQRLGAHLPQDTVCLDTLPALRGLDRAHSHGTRAQGRK
SYSLASLFHRYFQAEPSAAHSAEGDVHTLLLIFLHRAPELLAWADEQARSWAHIEPM
YVPPDGPSLEA (SEQ ID NO: 3866).

[0733]

[0734] 3' 3개의 프라임 복구 엑소뉴클레아제 2 (TREX2) - 래트

[0735] 수탁 번호 NM_001107580

MSEPLRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLP
RVLDKLTLCMCPERPFTAKASEITGLSSEGLMNCRKAAFNDAVVRTLQGFLSRQEGP
ICLVAHNGFDYDFPLLCTELQRLGAHLPRDTVCLDTLPALRGLDRVHSHGTRAQGRK
SYSLASLFHRYFQAEPSAAHSAEGDVNTLLLIFLHRAPELLAWADEQARSWAHIEPM
YVPPDGPSLEA (SEQ ID NO: 3867).

[0736]

[0737] ExoI

[0738] 인간 엑소뉴클레아제 1 (EXO1)은 DNA 미스매치 복구 (MMR), 마이크로-매개된 단부-연결, 상동 재조합 (HR), 및 복제를 포함한 많은 상이한 DNA 대사 과정에 연루되어 왔다. 인간 EXO1은 진핵 뉴클레아제 패밀리인 Rad2/XPG에 속하며, 이는 또한 FEN1 및 GEN1을 포함한다. Rad2/XPG 패밀리는 파지에서 인간까지의 종에 걸쳐 뉴클레아제 도메인에서 보존된다. EXO1 유전자 생성물은 5' 엑소뉴클레아제 및 5' 플랩 활성 둘 다를 나타낸다. 추가적으로, EXO1은 내재적 5' RNase H 활성을 함유한다. 인간 EXO1은 이중 가닥 DNA (dsDNA), 닉, 갭, 슈도 Y 구조를 프로세싱하는데 높은 친화도를 갖고, 그의 유전성 플랩 활성을 사용하여 홀리데이 접합부를 분해할 수 있다. 인간 EXO1은 MMR에 연루되며, MLH1 및 MSH2와 직접 상호작용하는 보존된 결합 도메인을 함유한다. EXO1 핵산분해 활성은 PCNA, MutSα (MSH2/MSH6 복합체), 14-3-3, MRN 및 9-1-1 복합체에 의해 양으로 자극된다.

[0739] 엑소뉴클레아제 1 (EXO1) 수탁 번호 NM_003686 (호모 사피엔스(Homo sapiens) 엑소뉴클레아제 1 (EXO1), 전사체 변이체 3) - 이소형 A

MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYV
GFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVS
EARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITE
DSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDY
LSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLY
QLVFDPIKRKLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYN
PDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVG
VERVISTKGLNLPRKSSIVKRPRSAELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSE
VFVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDC
VSNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDH
IPDKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQ
FRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSNASK
LSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADS
LSTTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKF (SEQ
ID NO: 3868).

[0740]

[0741] 엑소뉴클레아제 1 (EXO1) 수탁 번호 NM_006027 (호모 사피엔스 엑소뉴클레아제 1 (EXO1), 전사체 변이체 3) - 이소형 B

MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPT
DRYVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLRE
GKVSEARECFTRSINITHAMAHKVIKAARSQGVDCLV APYEADAQLAYLNKAGIVQ
AIITEDSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILS
GCDYLSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANN
TFLYQLVFDPIKRKLIPLNA YEDDVDPETLS YAGQYVDDS IALQIALGNKDINTFEQID
DYNPDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPS
TVGVERVIS TKGLNLPRKS SIVKRPRS AELSEDDLLS QYSLSFTKKTKKN SSEGNKSLS
FSEVFVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDS
TDCVSNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIP
GDHIPDKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTA
LQQFRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSN
ASKLSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSS
ADSLSTTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKD
SEKLPPCKKPLSPVRDNIQLTPE AEEDIFNKPEC GRVQRAIFQ (SEQ ID NO: 3869).

[0742]

[0743] 엑소뉴클레아제 1 (EXO1) 수탁 번호 NM_001319224 (호모 사피엔스 엑소뉴클레아제 1 (EXO1), 전사체 변이체 4) - 이소형 C

MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPT
DRYVGFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLRE
GKVSEARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQ
AIITEDSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILS
GCDYLSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANN
TFLYQLVFDPIKRKLIPLNA YEDDVDPETLS YAGQYVDDS IALQIALGNKDINTFEQID
DYNPDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPS
TVGVERVIS TKGLNLPRKS SIVKRPRS ELSEDDLLS QYSLSFTKKTKKN SSEGNKSLSF
SEVFVPDLVN GPTNKKS VSTPPRTRNKFATFLQRKNEES GAVVVPGTRS RFFCS SDST
DCVSNKVSIQPLDET AVTDKENNLHES EYGDQEGKRLVDTD VARN SSDDIPNNHIPG
DHIPDKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTAL
QQFRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSN
ASKLSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSS
ADSLS TTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKD
SEKLPPC KKPLS PVRDNIQLTPE AEEDIFNKPEC GRVQRAIFQ (SEQ ID NO: 3870).

[0744]

[0745] D. 인테인 및 분할-인테인

[0746] 일부 실시양태 (예를 들어, AAV 입자를 사용한 생체내 다중-플랩 프라임 편집제의 전달)에서, 폴리펩티드 (예를 들어, 데아미나제 또는 napDNAbp) 또는 융합 단백질 (예를 들어, 다중-플랩 프라임 편집제)을 N-말단 절반 및 C-말단 절반으로 분할하고, 이를 개별적으로 전달한 다음, 그것을 공동국재화시켜 세포 내에서 완전한 단백질 (또는 경우에 따라 융합 단백질)을 재형성하도록 하는 것이 유리할 수 있음이 이해될 것이다. 단백질 또는 융합 단백질의 별개의 절반은 단백질 트랜스 스플라이싱의 메카니즘에 의한 완전한 단백질 또는 융합 단백질의 재형성을 용이하게 하기 위해 각각 분할-인테인 태그를 포함할 수 있다.

[0747] 분할 인테인에 의해 촉매되는 단백질 트랜스-스플라이싱은 단백질 라이게이션을 위한 전적으로 효소적인 방법을 제공한다. 분할-인테인은 본질적으로 각각 N-인테인 및 C-인테인으로 명명되는 2개의 조각으로 분할된 인접 인테인 (예를 들어 미니-인테인)이다. 분할 인테인의 N-인테인 및 C-인테인은 비-공유적으로 회합되어 활성 인테인을 형성하고, 인접 인테인과 본질적으로 동일한 방식으로 스플라이싱 반응을 촉매할 수 있다. 분할 인테인은 자연에서 발견되었고, 또한 실험실에서 조작되었다. 본원에 사용된 용어 "분할 인테인"은 N-말단 및 C-말단 서열이 트랜스-스플라이싱 반응에 대해 기능적인 인테인으로 비-공유적으로 재회합 또는 재구성될 수 있는 별개의 분자가 되도록 N-말단 및 C-말단 아미노산 서열 사이에 1개 이상의 펩티드 결합 파괴가 존재하는 임의의 인테인을 지칭한다. 임의의 촉매 활성 인테인 또는 그의 단편은 본 발명의 방법에 사용하기 위한 분할 인테인을 유도하는 데 사용될 수 있다. 예를 들어, 한 측면에서 분할 인테인은 진핵 인테인으로부터 유래될 수 있다. 또 다른 측면에서, 분할 인테인은 박테리아 인테인으로부터 유래될 수 있다. 또 다른 측면에서, 분할 인테인은 고세균 인테인으로부터 유래될 수 있다. 바람직하게는, 이와 같이 유래된 분할 인테인은 트랜스-스플라이싱 반응을 촉매하는데 필수적인 아미노산 서열만을 보유할 것이다.

[0748] 본원에 사용된 "N-말단 분할 인테인 (In)"은 트랜스-스플라이싱 반응에 대해 기능적인 N-말단 아미노산 서열을 포함하는 임의의 인테인 서열을 지칭한다. 따라서, In은 또한 트랜스-스플라이싱이 발생할 때 스플라이싱되는 서열을 포함한다. In은 자연 발생 인테인 서열의 N-말단 부분의 변형인 서열을 포함할 수 있다. 예를 들어, In은, 추가의 및/또는 돌연변이된 잔기의 포함이 In을 트랜스-스플라이싱에서 비-기능적으로 만들지 않는 한, 이러한 추가의 아미노산 잔기 및/또는 돌연변이된 잔기를 포함할 수 있다. 바람직하게는, 추가의 및/또는 돌연변이된 잔기의 포함은 In의 트랜스-스플라이싱 활성을 개선 또는 증진시킨다.

[0749] 본원에 사용된 "C-말단 분할 인테인 (Ic)"은 트랜스-스플라이싱 반응에 대해 기능적인 C-말단 아미노산 서열을 포함하는 임의의 인테인 서열을 지칭한다. 한 측면에서, Ic는 4 내지 7개의 인접 아미노산 잔기를 포함하고, 이 중 적어도 4개의 아미노산은 그가 유래된 인테인의 마지막 β-가닥으로부터의 것이다. 따라서, Ic는 또한 트랜스-스플라이싱이 발생할 때 스플라이싱되는 서열을 포함한다. Ic는 자연 발생 인테인 서열의 C-말단 부분의 변형인 서열을 포함할 수 있다. 예를 들어, Ic는, 추가의 및/또는 돌연변이된 잔기의 포함이 In을 트랜스-스플라이싱에서 비-기능적으로 만들지 않는 한, 이러한 추가의 아미노산 잔기 및/또는 돌연변이된 잔기를 포함

할 수 있다. 바람직하게는, 추가의 및/또는 돌연변이된 잔기의 포함은 Ic의 트랜스-스플라이싱 활성을 개선 또는 증진시킨다.

[0750] 본 발명의 일부 실시양태에서, Ic 또는 In에 연결된 펩티드는 특히 형광 기, 비오틴, 폴리에틸렌 글리콜 (PEG), 아미노산 유사체, 비천연 아미노산, 포스페이트 기, 글리코실 기, 방사성동위원소 표지, 및 제약 분자를 포함한 추가의 화학적 모이어티를 포함할 수 있다. 다른 실시양태에서, Ic에 연결된 펩티드는 특히 케톤, 알데히드, Cys 잔기 및 Lys 잔기를 포함하는 1개 이상의 화학적 반응성 기를 포함할 수 있다. 분할 인테인의 N-인테인 및 C-인테인은 비-공유적으로 회합되어 활성 인테인을 형성하고, "인테인-스플라이싱 폴리펩티드 (ISP)"가 존재하는 경우에 스플라이싱 반응을 촉매할 수 있다. 본원에 사용된 "인테인-스플라이싱 폴리펩티드 (ISP)"는 Ic, In, 또는 둘 다가 분할 인테인으로부터 제거될 때 남는 분할 인테인의 아미노산 서열의 부분을 지칭한다. 특정 실시양태에서, In은 ISP를 포함한다. 또 다른 실시양태에서, Ic는 ISP를 포함한다. 또 다른 실시양태에서, ISP는 In에도 Ic에도 공유 연결되지 않은 별개의 펩티드이다.

[0751] 분할 인테인은 미니-인테인의 구조에서 발견되는 ~12개의 보존된 베타-가닥 사이의 비구조화된 루프 또는 개재 아미노산 서열 내의 1개 이상의 분할 부위를 조작함으로써 인접 인테인으로부터 생성될 수 있다. 분할의 생성이 단백질 스플라이싱 활성이 상실되는데 충분한 정도로 인테인, 특히 구조화된 베타-가닥의 구조를 파괴하지 않을 것이라면, 베타-가닥 사이의 영역 내의 분할 부위의 위치에 일부 유연성이 존재할 수 있다.

[0752] 단백질 트랜스-스플라이싱에서, 하나의 전구체 단백질은 N-엑스테인 부분에 이어 N-인테인으로 이루어지고, 또 다른 전구체 단백질은 C-인테인 부분에 이어 C-엑스테인으로 이루어지고, 트랜스-스플라이싱 반응 (N- 및 C-인테인에 의해 함께 촉매됨)은 2개의 인테인 서열을 절제하고, 2개의 엑스테인 서열을 펩티드 결합으로 연결한다. 효소 반응인 단백질 트랜스-스플라이싱은 매우 낮은 (예를 들어, 마이크로몰) 농도의 단백질로 작용할 수 있고, 생리학적 조건 하에 수행될 수 있다.

[0753] 예시적인 서열은 하기와 같다:

명칭	리간드-의존성 인테인의 서열
2-4 인테인:	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 8)
3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDR VAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASM MGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEIL MIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATS SRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRAL DKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHL YSMKYTNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDK FLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHN C (SEQ ID NO: 9)
30R3-1 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 10)
30R3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 11)
30R3-3 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH

[0754]

	RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 12)
37R3-1 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYNPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC ((SEQ ID NO: 13)
37R3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 14)
37R3-3 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 15)

[0755]

[0756] 인테인은 가장 빈번하게는 인접 도메인으로서 발견되지만, 일부는 자연 분할 형태로 존재한다. 이러한 경우에, 2개의 단편은 별개의 폴리펩티드로서 발현되고, 이는 스플라이싱, 소위 단백질 트랜스-스플라이싱이 일어나기 전에 회합되어야 한다.

[0757] 예시적인 분할 인테인은 2개의 서브유닛, 즉 DnaE-N 및 DnaE-C를 포함하는 Ssp DnaE 인테인이다. 2개의 상이한 서브유닛은 별개의 유전자, 즉 각각 DnaE-N 및 DnaE-C 서브유닛을 코딩하는 dnaE-n 및 dnaE-c에 의해 코딩된다. DnaE는 시네코시티스 종 PCC6803의 자연 발생 분할 인테인이고, 각각 DnaE-N 또는 DnaE-C와의 융합체를 포함한 2개의 별개의 단백질의 트랜스-스플라이싱을 지시할 수 있다.

[0758] 추가의 자연 발생 또는 조작된 분할-인테인 서열은 관련 기술분야에 공지되어 있거나, 또는 본원에 기재된 전체-인테인 서열 또는 관련 기술분야에서 입수가능한 것으로부터 제조될 수 있다. 분할-인테인 서열의 예는 문헌 [Stevens et al., "A promiscuous split intein with expanded protein engineering applications," PNAS, 2017, Vol.114: 8538-8543; Iwai et al., "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme, FEBS Lett, 580: 1853-1858]에서 찾아볼 수 있으며, 이들 각각은 본원에 참조로 포함된다. 추가의 분할 인테인 서열은, 예를 들어 WO 2013/045632, WO 2014/055782, WO 2016/069774, 및 EP2877490에서 찾아볼 수 있으며, 이들 각각의 내용은 본원에 참조로 포함된다.

[0759] 또한, 생체내 및 시험관내 트랜스로의 단백질 스플라이싱이 기재되어 있으며 (Shingledecker, et al., Gene 207:187 (1998), Southworth, et al., EMBO J. 17:918 (1998); Mills, et al., Proc. Natl. Acad. Sci. USA, 95:3543-3548 (1998); Lew, et al., J. Biol. Chem., 273:15887-15890 (1998); Wu, et al., Biochim. Biophys. Acta 35732:1 (1998b), Yamazaki, et al., J. Am. Chem. Soc. 120:5591 (1998), Evans, et al., J. Biol. Chem. 275:9091 (2000); Otomo, et al., Biochemistry 38:16040-16044 (1999); Otomo, et al., J. Biolmol. NMR 14:105-114 (1999); Scott, et al., Proc. Natl. Acad. Sci. USA 96:13638-13643 (1999)), 이는 예를 들어 2개의 개별적으로-발현된 절반으로부터 완전한 PE 융합 단백질의 형성에 관하여 도 66 및 67에 제시된 바와 같이, 기능적 생성물을 형성하기 위해 후속적으로 라이게이션을 거치는 2개의 불활성 단편에 대해 단백질로서의 발현 기회를 제공한다.

[0760] E. RNA-단백질 상호작용 도메인

[0761] 다양한 실시양태에서, 2개의 별개의 단백질 도메인 (예를 들어, Cas9 도메인 및 폴리머라제 도메인)은 "RNA-단백질 동원 시스템", 예컨대 "MS2 태그부착 기술"을 사용함으로써 서로 공동국재화되어 기능적 복합체 (2개의 별개의 단백질 도메인을 포함하는 융합 단백질의 기능과 유사함)를 형성할 수 있다. 이러한 시스템은 일반적으로 하나의 단백질 도메인은 "RNA-단백질 상호작용 도메인" (일명 "RNA-단백질 동원 도메인")으로, 다른 것은 RNA-단백질 상호작용 도메인, 예를 들어 특정 헤어핀 구조를 특이적으로 인식하고 그에 결합하는 "RNA-결합 단백질"로 태그부착한다. 이들 유형의 시스템은 다중-플랩 프라임 편집제의 도메인을 공동국재화하기 위해서뿐만 아니라 다중-플랩 프라임 편집제, 예컨대 UGI 도메인에 추가의 기능성을 동원하기 위해 활용될 수 있다. 한 예에서, MS2 태그부착 기술은 MS2 박테리오파지 코트 단백질 ("MCP" 또는 "MS2cp")과 파지의 게놈에 존재하는 스템-루프 또는 헤어핀 구조, 즉 "MS2 헤어핀"의 자연 상호작용에 기초한다. MS2 헤어핀의 경우에, 이는 MS2 박테리오파지 코트 단백질 (MCP)에 의해 인식되고 결합된다. 따라서, 하나의 예시적인 시나리오에서 데아미나제-MS2 융합체는 Cas9-MCP 융합체를 동원할 수 있다.

[0762] 다른 모듈 RNA-단백질 상호작용 도메인의 검토는 관련 기술분야에, 예를 들어 문헌 [Johansson et al., "RNA recognition by the MS2 phage coat protein," Sem Virol., 1997, Vol. 8(3): 176-185; Delebecque et al., "Organization of intracellular reactions with rationally designed RNA assemblies," Science, 2011, Vol. 333: 470-474; Mali et al., "Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering," Nat. Biotechnol., 2013, Vol.31: 833-838; 및 Zalatan et al., "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds," Cell, 2015, Vol.160: 339-350]에 기재되어 있고, 이들 각각은 그 전문이 본원에 참조로 포함된다. 다른 시스템은 PCP 단백질을 특이적으로 동원하는 PP7 헤어핀, 및 Com 단백질을 특이적으로 동원하는 "com" 헤어핀을 포함한다. 문헌 [Zalatan et al.]을 참조한다.

[0763] MS2 헤어핀 (또는 동등하게 "MS2 압타머"로 지칭됨)의 뉴클레오티드 서열은 GCCAACATGAGGATCACCCATGTCTGCAGGGCC (서열식별번호: 763)이다.

[0764] MCP 또는 MS2cp의 아미노산 서열은: GSASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGEELPVAGWRSYLNMELTIPIFATNSD CELIVKAMQGLLKDGNPIPSAIAANSGIY (서열식별번호: 764)이다.

[0765] F. UGI 도메인

[0766] 다른 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제는 1개 이상의 우라실 글리코실라제 억제제 도메인을 포함할 수 있다. 본원에 사용된 용어 "우라실 글리코실라제 억제제 (UGI)" 또는 "UGI 도메인"은 우라실-DNA 글리코실라제 염기-절제 복구 효소를 억제할 수 있는 단백질을 지칭한다. 일부 실시양태에서, UGI 도메인은 야생형 UGI 또는 서열식별번호: 3873에 제시된 바와 같은 UGI를 포함한다. 일부 실시양태에서, 본원에 제공된 UGI 단백질은 UGI의 단편 및 UGI 또는 UGI 단편에 상동인 단백질을 포함한다. 예를 들어, 일부 실시양태에서, UGI 도메인은 서열식별번호: 3873에 제시된 아미노산 서열의 단편을 포함한다. 일부 실시양태에서, UGI 단편은 서열식별번호: 3873에 제시된 바와 같은 아미노산 서열의 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%를 포함하는 아미노산 서열을 포함한다. 일부 실시양태에서, UGI는 서열식별번호: 3873에 제시된 아미노산 서열과 상동인 아미노산 서열, 또는 서열식별번호: 3873에 제시된 아미노산 서열의 단편과 상동인 아미노산 서열을 포함한다. 일부 실시양태에서, UGI 또는 UGI의 단편 또는 UGI 또는 UGI 단편의 상동체를 포함하는 단백질은 "UGI 변이체"로 지칭된다. UGI 변이체는 UGI 또는 그의 단편에 대해 상동성을 공유한다. 예를 들어, UGI 변이체는 야생형 UGI 또는 서열식별번호: 3873에 제시된 바와 같은 UGI와 적어도 70% 동일하거나, 적어도 75% 동일하거나, 적어도 80% 동일하거나, 적어도 85% 동일하거나, 적어도 90% 동일하거나, 적어도 95% 동일하거나, 적어도 96% 동일하거나, 적어도 97% 동일하거나, 적어도 98% 동일하거나, 적어도 99% 동일하거나, 적어도 99.5% 동일하거나, 또는 적어도 99.9% 동일하다. 일부 실시양태에서, UGI 변이체는 UGI의 단편을 포함하며, 이러한 단편은 야생형 UGI 또는 서열식별번호: 3873에 제시된 바와 같은 UGI의 상응하는 단편에 대해 적어도 70% 동일하거나, 적어도 80% 동일하거나, 적어도 90% 동일하거나, 적어도 95% 동일하거나, 적어도 96% 동일하거나, 적어도 97% 동일하거나, 적어도 98% 동일하거나, 적어도 99% 동일하거나, 적어도 99.5% 동일하거나, 또는 적어도 99.9% 동일하다. 일부 실시양태에서, UGI는 하기 아미노산 서열을 포함한다:

[0767] 우라실-DNA 글리코실라제 억제제:

[0768] >sp|P14739|UNGI_BPPB2

[0769] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML (서열식별번호: 3873).

[0770] 본원에 기재된 다중-플랩 프라임 편집제는 본원에 기재된 바와 같은 1개 이상의 링커에 의해 분리될 수 있는 1개 초과의 UGI 도메인을 포함할 수 있다.

[0771] G. 추가의 PE 요소

[0772] 특정 실시양태에서, 본원에 기재된 다중-플랩 프라임 편집제는 염기 복구 억제제를 포함할 수 있다. 용어 "염기 복구 억제제" 또는 "IBR"은 핵산 복구 효소, 예를 들어 염기 절제 복구 효소의 활성을 억제할 수 있는 단백질을 지칭한다. 일부 실시양태에서, IBR은 OGG 염기 절제 복구 억제제이다. 일부 실시양태에서, IBR은 염기 절제 복구 억제제 ("iBER")이다. 예시적인 염기 절제 복구 억제제는 APE1, 엔도 III, 엔도 IV, 엔도 V, 엔도 VIII, Fpg, hOGG1, hNEIL1, T7 엔도I, T4PDG, UDG, hSMUG1, 및 hAAG의 억제제를 포함한다. 일부 실시양태에서, IBR은 엔도 V 또는 hAAG의 억제제이다. 일부 실시양태에서, IBR은 촉매 불활성 글리코실라제 또는 촉매 불활성 디옥시게나제, 또는 옥시다제의 소분자 또는 펩티드 억제제, 또는 그의 변이체일 수 있는 iBER이다. 일부 실시양태에서, IBR은 TDG 억제제, MBD4 억제제 또는 AlkBH 효소의 억제제일 수 있는 iBER이다. 일부 실시양태에서, IBR은 촉매 불활성 TDG 또는 촉매 불활성 MBD4를 포함하는 iBER이다. 예시적인 촉매 불활성 TDG는 서열식별번호: 3872 (인간 TDG)의 N140A 돌연변이체이다.

[0773] 일부 예시적인 글리코실라제가 하기 제공된다. 이들 글리코실라제 도메인 중 임의의 것의 촉매적으로 불활성화된 변이체는 본 개시내용에 제공된 다중-플랩 프라임 편집제의 napDNAbp 또는 폴리머라제 도메인에 융합될 수 있는 iBER이다.

[0774] OGG (인간)

MPARALLPRRMGHRTLASTPALWASIPCPRSELRLDLVLPSGQSFRWREQSPAHWSG
VLADQVWTLTQTEEQLHCTVYRGDKSQASRPTPDELEAVRKYFQLDVTLAQLYHHWGSVD
SHFQEVAQKFQGVRLLRQDPIECLFSFICSSNNNIARITGMVERLCQAFGPRLIQLDDVTYHGF
PSLQALAGPEVEAHLRKLGLGYRARYVSASARAILEEQGGLAWLQQLRESSYEEAHKALCIL
PGVGTKVADCICLMALDKPQAVPVDVHMWHIAQRDYSWHPTTSQAKGPSPQTNKELGNFF
RSLWGPYAGWAQAVLFSADLRQSRHAQEPPAKRRKGSKGPEG (SEQ ID NO: 3937)

[0775]

[0776] MPG (인간)

MVTPALQMKKPKQFCRRMGQKKQRPARAGQPHSSSDAAQAPAEQPHSSSDAAQAP
CPRERCLGPPTTPGPYRSIYFSSPKGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPNGTELRG
RIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCMNISSQGDGACVLL
RALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLAQDEAVW
LERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQDTQA (SEQ

[0777] ID NO: 3938)

[0778] MBD4 (인간)

MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQ
MMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDV
YFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSH
LQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDV
NFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQ
KSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA
KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGE
KIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHD
PWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLR
AKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHD
WLWENHEKLSLS (SEQ ID NO: 3871)

[0779]

[0780] TDG (인간)

MEAENAGSYSLQQAQAFYTFPFQQLMAEAPNMAVVNEQQMPEEVPAPAPA
QEPVQEAPKGRKRKPRTTEPKQPVEPKKPVESKKSGKSAKSKEKQEKITDTFKVKRK
VDRFNGVSEAELLTKTLPDILTFNLDIVIIGINPGLMAAYKGHHYPGPGNHFWKCLFM
SGLSEVQLNHMDDHTLPGKYGIGFTNMVERTTPGSKDLSSKEFREGGRILVQKLQKY
QPRIAVFNGKCIYEIFSKEVFGVKVKNLEFGLQPHKIPDTETLCYVMPSSSARCAQFPR
AQDKVHYYIKLKDLRDQLKGIERNMDVQEVQYTFDLQLAQEDAKKMAVKEEKYDP
GYEAAYGGAYGENPCSSEPCGFSSNGLIESVELRGESAFSGIPNGQWMTQSFTDQIPS
FSNHCGTQEQEEESHA (SEQ ID NO: 3872)

[0781]

[0782] 일부 실시양태에서, 본원에 기재된 융합 단백질은 1개 이상의 이종 단백질 도메인 (예를 들어, 프라임 편집제 성분에 더하여 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 초과의 도메인이거나, 또는 약 그보다 더 많음)을 포함할 수 있다. 융합 단백질은 임의의 추가의 단백질 서열, 및 임의로 임의의 2개의 도메인 사이의 링커 서열을 포함할 수 있다. 존재할 수 있는 다른 예시적인 특색은 국재화 서열, 예컨대 세포질 국재화 서열, 유출 서열, 예컨대 핵 유출 서열, 또는 다른 국재화 서열, 뿐만 아니라 융합 단백질의 가용화, 정제, 또는 검출에 유용한 서열 태그이다.

[0783] 다중-플랩 프라임 편집제 또는 그의 성분 (예를 들어, napDNAbp 도메인, 폴리머라제 도메인, 또는 NLS 도메인)에 융합될 수 있는 단백질 도메인의 예는 비제한적으로 에피토프 태그 및 리포터 유전자 서열을 포함한다. 에피토프 태그의 비제한적 예는 히스티딘 (His) 태그, V5 태그, FLAG 태그, 인플루엔자 헤마글루티닌 (HA) 태그, Myc 태그, VSV-G 태그, 및 티오레독신 (Trx) 태그를 포함한다. 리포터 유전자의 예는 글루타티온-5-트랜스퍼라제 (GST), 양고추냉이 퍼옥시다제 (HRP), 클로람페니콜 아세틸트랜스퍼라제 (CAT), 베타-갈락토시다제, 베타-글루쿠로니다제, 루시페라제, 녹색 형광 단백질 (GFP), HcRed, DsRed, 시안 형광 단백질 (CFP), 황색 형광 단백질 (YFP), 및 청색 형광 단백질 (BFP)을 포함한 자가형광 단백질을 포함하나 이에 제한되지는 않는다. 다중-플랩 프라임 편집제는 말토스 결합 단백질 (MBP), S-태그, Lex A DNA 결합 도메인 (DBD) 융합체, GAL4 DNA 결합 도메인 융합체, 및 단순 포진 바이러스 (HSV) BP16 단백질 융합체를 포함하나 이에 제한되지는 않는, DNA 분자에 결합하거나 또는 다른 세포 분자에 결합하는 단백질 또는 단백질의 단편을 코딩하는 유전자 서열에 융합될 수 있다. 프라임 편집제의 일부를 형성할 수 있는 추가의 도메인은 2011년 3월 10일에 공개된 미국 특허 공개 번호 2011/0059502에 기재되어 있으며, 그 전문은 본원에 참조로 포함된다.

[0784] 개시내용의 한 측면에서, 글루타티온-5-트랜스퍼라제 (GST), 양고추냉이 퍼옥시다제 (HRP), 클로람페니콜 아세틸트랜스퍼라제 (CAT) 베타-갈락토시다제, 베타-글루쿠로니다제, 루시페라제, 녹색 형광 단백질 (GFP), HcRed, DsRed, 시안 형광 단백질 (CFP), 황색 형광 단백질 (YFP), 및 청색 형광 단백질 (BFP)을 포함한 자가형광 단백질을 포함하나 이에 제한되지는 않는 리포터 유전자는 유전자 생성물의 발현의 변경 또는 변형을 측정하기 위한 마커로서 작용하는 유전자 생성물을 코딩하도록 세포 내로 도입될 수 있다. 개시내용의 특정 실시양태에서, 유

전자 생성물은 루시페라제이다. 본 개시내용의 추가 실시양태에서, 유전자 생성물의 발현이 감소된다.

[0785] 본원에 제공된 적합한 단백질 태그는 비오틴 카르복실라제 담체 단백질 (BCCP) 태그, myc-태그, 칼모듈린-태그, FLAG-태그, 헤마글루티닌 (HA)-태그, 히스티딘 태그 또는 His-태그로도 지칭되는 폴리히스티딘 태그, 말토스 결합 단백질 (MBP)-태그, nus-태그, 글루타티온-S-트랜스퍼라제 (GST)-태그, 녹색 형광 단백질 (GFP)-태그, 티오레독신-태그, S-태그, 소프태그 (예를 들어, 소프태그 1, 소프태그 3), 스트렙-태그, 비오틴 리가제 태그, FlAsH 태그, V5 태그, 및 SBP-태그를 포함하나 이에 제한되지는 않는다. 추가의 적합한 서열은 관련 기술분야의 통상의 기술자에게 명백할 것이다. 일부 실시양태에서, 융합 단백질은 1개 이상의 His 태그를 포함한다.

[0786] 본 개시내용의 일부 실시양태에서, 다중-플랩 프라임 편집 시스템의 활성은 PE 시스템의 발현된 성분의 체류 시간, 양, 및/또는 활성을 조정함으로써 일시적으로 조절될 수 있다. 예를 들어, 본원에 기재된 바와 같이, PE는 PE의 세포내 반감기를 변형시킬 수 있는 단백질 도메인과 융합될 수 있다. 2개 이상의 벡터 (예를 들어, 본원에 기재된 성분이 2개 이상의 별개의 벡터 상에 코딩된 벡터 시스템)를 수반하는 특정 실시양태에서, PE 시스템의 활성은 벡터가 전달되는 시기를 제어함으로써 일시적으로 조절될 수 있다. 예를 들어, 일부 실시양태에서, 뉴클레아제 시스템을 코딩하는 벡터는 주형을 코딩하는 벡터 전에 PE를 전달할 수 있다. 다른 실시양태에서, PEgRNA를 코딩하는 벡터는 PE 시스템을 코딩하는 벡터 전에 가이드를 전달할 수 있다. 일부 실시양태에서, PE 시스템 및 PEgRNA를 코딩하는 벡터는 동시에 전달된다. 특정 실시양태에서, 동시에 전달된 벡터는 예를 들어 PE, PEgRNA, 및/또는 제2 가닥 가이드 RNA 성분을 일시적으로 전달한다. 추가 실시양태에서, 벡터 상의 코딩 서열로부터 전사된 RNA (예컨대, 예를 들어 뉴클레아제 전사체)는 RNA의 세포내 반감기를 변형시키고/거나 번역 제어를 조정할 수 있는 적어도 1개의 요소를 추가로 포함할 수 있다. 일부 실시양태에서, RNA의 반감기는 증가될 수 있다. 일부 실시양태에서, RNA의 반감기는 감소될 수 있다. 일부 실시양태에서, 요소는 RNA의 안정성을 증가시킬 수 있다. 일부 실시양태에서, 요소는 RNA의 안정성을 감소시킬 수 있다. 일부 실시양태에서, 요소는 RNA의 3' UTR 내에 있을 수 있다. 일부 실시양태에서, 요소는 폴리아데닐화 신호 (PA)를 포함할 수 있다. 일부 실시양태에서, 요소는 캡, 예를 들어 상류 mRNA 또는 PEgRNA 단부를 포함할 수 있다. 일부 실시양태에서, RNA는 전사 후 세포에서 보다 빠르게 분해되도록 PA를 포함하지 않을 수 있다. 일부 실시양태에서, 요소는 적어도 1개의 AU-풍부 요소 (ARE)를 포함할 수 있다. ARE는 조직 유형, 세포 유형, 시기, 세포 국재화, 및 환경에 의존하는 방식으로 ARE 결합 단백질 (ARE-BP)에 의해 결합될 수 있다. 일부 실시양태에서, 탈안정화 요소는 RNA 붕괴를 촉진하거나, RNA 안정성에 영향을 미치거나, 또는 번역을 활성화시킬 수 있다. 일부 실시양태에서, ARE는 50 내지 150개 뉴클레오티드 길이를 포함할 수 있다. 일부 실시양태에서, ARE는 서열 AUUUA의 적어도 1개의 카피를 포함할 수 있다. 일부 실시양태에서, 적어도 1개의 ARE가 RNA의 3' UTR에 부가될 수 있다. 일부 실시양태에서, 요소는 우드척 간염 바이러스 (WHP)일 수 있다.

[0787] 전사체로부터의 발현을 증진시키기 위한 3차 구조를 생성하는 전사후 조절 요소 (WPRE). 추가 실시양태에서, 요소는, 예를 들어 문헌 [Zufferey et al., J Virol, 73(4): 2886-92 (1999) and Flajolet et al., J Virol, 72(7): 6175-80 (1998)]에 기재된 바와 같이, 전사체로부터의 발현을 증진시킬 수 있는 변형된 및/또는 말단절단된 WPRE 서열이다. 일부 실시양태에서, WPRE 또는 등가물은 RNA의 3' UTR에 부가될 수 있다. 일부 실시양태에서, 요소는 신속- 또는 저속-붕괴 전사체가 풍부한 다른 RNA 서열 모티프로부터 선택될 수 있다.

[0788] 일부 실시양태에서, PE 또는 PEgRNA를 코딩하는 벡터는 PE 시스템에 의한 벡터 상에 존재하는 표적 서열의 절단을 통해 자기-파괴될 수 있다. 절단은 벡터로부터 PE 또는 PEgRNA의 지속적인 전사를 막을 수 있다. 전사가 어느 정도의 시간 동안 선형화된 벡터 상에서 일어날 수 있지만, 세포내 분해에 적용되는 발현된 전사체 또는 단백질은 코딩 벡터의 발현으로부터의 지속적인 공급의 부재 하에 오프-타겟 효과를 생성하는 시간을 덜 가질 것이다.

[0789] [6] PEgRNA

[0790] 본원에 기재된 다중-플랩 프라임 편집 시스템은 임의의 적합한 PEgRNA의 사용을 고려한다. 표적-프라이밍된 역전사 (TPRT)의 메카니즘은 목적하는 뉴클레오티드 변화를 코딩하는 역전사 (RT) 주형 서열을 포함하는 특수하게 구성된 가이드 RNA의 사용을 통해 정밀하고 다재다능한 CRISPR/Cas-기반 게놈 편집을 수행하기 위해 활용되거나 적합화될 수 있다. 본 출원은, RT 주형 서열이 표준 또는 전통적인 가이드 RNA 분자의 연장부로서 제공될 수 있기 때문에, 이러한 특수하게 구성된 가이드 RNA를 "연장된 가이드 RNA" 또는 "PEgRNA"로서 언급한다. 본 출원은 이중-플랩 및 사중-플랩 프라임 편집에 사용하기 위한 연장된 가이드 RNA에 대한 임의의 적합한 구성 또는 배열을 고려한다.

[0791] 이중-플랩 및 다중-플랩 프라임 편집에 사용된 pegRNA의 일반적 설계가 도 92에 제시된다. 이중-플랩 및 다중-

플랩 프라임 편집에 사용된 pegRNA는 전형적 프라임 편집에 사용된 것과 유사한 설계를 갖지만, RT 주형 영역이 표적 유전자좌에 대한 임의의 상동성을 코딩할 필요는 없다. 대신, 2개의 pegRNA는 다양한 실시양태에서 3' 단부가 서로의 역 상보체 서열인 3' 플랩의 합성을 코딩하는 RT 주형을 함유할 수 있다. 3' 플랩 사이의 이러한 상보성은 그의 어닐링 및 내인성 DNA 서열의 의도된 새로운 DNA 서열로의 대체를 촉진한다. 이는 2개의 pegRNA 내의 RT 주형의 5' 영역이 서로에 대해 역 상보체 서열일 것을 필요로 하고, 이러한 상보성의 양은 다를 수 있다 (도 92).

[0792] PEgRNA 아키텍처

[0793] 도 3a는 본원에 개시된 다중-플랩 프라임 편집 시스템에 사용가능한 연장된 가이드 RNA의 한 실시양태를 보여주며, 여기서 전통적인 가이드 RNA (녹색 부분)는 ~20개의 nt 프로토스페이서 서열 및 napDNAbp와 결합하는 gRNA 코어 영역을 포함한다. 이러한 실시양태에서, 가이드 RNA는 5' 단부에 연장된 RNA 절편, 즉 5' 연장부를 포함한다. 이러한 실시양태에서, 5' 연장부는 역전사 주형 서열, 역전사 프라이머 결합 부위, 및 임의적인 5-20개의 뉴클레오티드 링커 서열을 포함한다. 도 1a-1b에 제시된 바와 같이, RT 프라이머 결합 부위는 R-루프의 비-표적 가닥에 닉이 형성된 후에 형성된 유리 3' 단부에 혼성화하여, 리버스 트랜스크립타제의 5'-3' 방향으로의 DNA 중합을 프라이밍한다.

[0794] 도 3b는 본원에 개시된 프라임 편집 시스템에 사용가능한 연장된 가이드 RNA의 또 다른 실시양태를 보여주며, 여기서 전통적인 가이드 RNA (녹색 부분)는 ~20개의 nt 프로토스페이서 서열 및 napDNAbp와 결합하는 gRNA 코어를 포함한다. 이러한 실시양태에서, 가이드 RNA는 3' 단부에 연장된 RNA 절편, 즉 3' 연장부를 포함한다. 이러한 실시양태에서, 3' 연장부는 역전사 주형 서열, 및 역전사 프라이머 결합 부위를 포함한다. 도 1c-1d에 제시된 바와 같이, RT 프라이머 결합 부위는 R-루프의 비-표적 가닥에 닉이 형성된 후에 형성된 유리 3' 단부에 혼성화하여, 리버스 트랜스크립타제의 5'-3' 방향으로의 DNA 중합을 프라이밍한다.

[0795] 도 3c는 본원에 개시된 프라임 편집 시스템에 사용가능한 연장된 가이드 RNA의 또 다른 실시양태를 보여주며, 여기서 전통적인 가이드 RNA (녹색 부분)는 ~20개의 nt 프로토스페이서 서열 및 napDNAbp와 결합하는 gRNA 코어를 포함한다. 이러한 실시양태에서, 가이드 RNA는 gRNA 코어 내의 분자간 위치에서 연장된 RNA 절편, 즉 분자내 연장부를 포함한다. 이러한 실시양태에서, 분자내 연장부는 역전사 주형 서열, 및 역전사 프라이머 결합 부위를 포함한다. RT 프라이머 결합 부위는 R-루프의 비-표적 가닥에 닉이 형성된 후에 형성된 유리 3' 단부에 혼성화하여, 리버스 트랜스크립타제의 5'-3' 방향으로의 DNA 중합을 프라이밍한다.

[0796] 한 실시양태에서, 분자간 RNA 연장부의 위치는 가이드 RNA의 프로토스페이서 서열 내에 있지 않다. 또 다른 실시양태에서, 분자간 RNA 연장부의 위치는 gRNA 코어 내에 있다. 또 다른 실시양태에서, 분자간 RNA 연장부의 위치는 프로토스페이서 서열 내, 또는 프로토스페이서 서열을 파괴하는 위치를 제외하고는 가이드 RNA 분자의 임의의 위치이다.

[0797] 한 실시양태에서, 분자간 RNA 연장부는 프로토스페이서 서열의 3' 단부로부터 하류에 삽입된다. 또 다른 실시양태에서, 분자간 RNA 연장부는 프로토스페이서 서열의 3' 단부의 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드 하류에 삽입된다.

[0798] 다른 실시양태에서, 분자간 RNA 연장부는 gRNA 내로 삽입되며, 이는 Cas9 단백질 또는 그의 등가물 (즉, 상이한 napDNAbp)과 결합 및/또는 상호작용하는, tracrRNA에 상응하거나 또는 그를 포함하는 가이드 RNA의 부분을 지칭한다. 바람직하게는, 분자간 RNA 연장부의 삽입은 tracrRNA 부분과 napDNAbp 사이의 상호작용을 방해하지 않거나 최소로 방해한다.

[0799] RNA 연장부 (이는 적어도 RT 주형 및 프라이머 결합 부위를 포함함, 예를 들어 도 92 참조)의 길이는 임의의 유용한 길이일 수 있다. 다양한 실시양태에서, RNA 연장부는 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴

클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드, 또는 적어도 500개 뉴클레오티드 길이이다.

[0800] RT 주형 서열은 또한 임의의 적합한 길이일 수 있다. 예를 들어, RT 주형 서열은 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드, 또는 적어도 500개 뉴클레오티드 길이일 수 있다.

[0801] 또 다른 실시양태에서, 역전사 프라이머 결합 부위 서열은 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드, 또는 적어도 500개 뉴클레오티드 길이이다.

[0802] 다른 실시양태에서, 임의적인 링커 또는 스페이서 서열은 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 200개 뉴클레오티드, 적어도 300개 뉴클레오티드, 적어도 400개 뉴클레오티드, 또는 적어도 500개 뉴클레오티드 길이이다.

[0803] 특정 실시양태에서, RT 주형 서열은 비-표적 가닥에 상동성이지만 (따라서, 표적 가닥의 상응하는 부위에 상보성이지만), 1개 이상의 뉴클레오티드 변화를 포함하는 단일 가닥 DNA 분자를 코딩한다. 적어도 1개의 뉴클레오티드 변화는 1개 이상의 단일-염기 뉴클레오티드 변화, 1개 이상의 결실, 및 1개 이상의 삽입을 포함할 수 있다.

[0804] 도 1g에 도시된 바와 같이, RT 주형 서열의 합성된 단일 가닥 DNA 생성물은 비-표적 가닥과 상동성이고, 1개 이상의 뉴클레오티드 변화를 함유한다. RT 주형 서열의 단일 가닥 DNA 생성물은 상보적 표적 가닥 서열과 평형 상태로 혼성화하여, 상동 내인성 표적 가닥 서열을 대체한다. 대체된 내인성 가닥은 일부 실시양태에서 5' 내인성 DNA 플랩 종으로 지칭될 수 있다 (예를 들어, 도 1e 참조). 이러한 5' 내인성 DNA 플랩 종은 5' 플랩 엔도뉴클레아제 (예를 들어, FEN1)에 의해 제거될 수 있고, 이제 내인성 표적 가닥에 혼성화된 단일 가닥 DNA 생성물은 라이게이션되어 내인성 서열과 새로 합성된 가닥 사이에 미스매치를 생성할 수 있다. 미스매치는 세포의 선천성 DNA 복구 및/또는 복제 과정에 의해 분해될 수 있다.

[0805] 다양한 실시양태에서, RT 주형 서열의 뉴클레오티드 서열은 5' 플랩 종으로서 대체되고 편집될 부위와 중첩되는 비-표적 가닥의 뉴클레오티드 서열에 상응한다.

[0806] 연장된 가이드 RNA의 다양한 실시양태에서, 역전사 주형 서열은 닉 부위에 인접한 내인성 DNA 서열에 상보적인 단일 가닥 DNA 플랩을 코딩할 수 있으며, 여기서 단일 가닥 DNA 플랩은 목적하는 뉴클레오티드 변화를

포함한다. 단일 가닥 DNA 플랩은 닉 부위에서 내인성 단일 가닥 DNA를 대체할 수 있다. 닉 부위에서 대체된 내인성 단일 가닥 DNA는 5' 단부를 가질 수 있고, 내인성 플랩을 형성할 수 있으며, 이는 세포에 의해 절제될 수 있다. 다양한 실시양태에서, 5' 단부 내인성 플랩의 제거는 상응하는 상보적 DNA 가닥에 대한 단일 가닥 3' DNA 플랩의 혼성화, 및 단일 가닥 3' DNA 플랩에 의해 운반되는 목적하는 뉴클레오티드 변화의 표적 DNA 내로의 혼입 또는 동화를 촉진하기 때문에 5' 단부 내인성 플랩의 절제는 생성물 형성을 구동하는 것을 도울 수 있다.

[0807] 연장된 가이드 RNA의 다양한 실시양태에서, 단일 가닥 DNA 플랩의 세포 복구는 목적하는 뉴클레오티드 변화의 설치를 발생시켜, 목적하는 생성물을 형성한다.

[0808] 또 다른 실시양태에서, 목적하는 뉴클레오티드 변화는 닉 부위의 약 -5 내지 약 +5 사이, 또는 닉 부위의 약 -10 내지 약 +10 사이, 또는 닉 부위의 약 -20 내지 약 +20 사이, 또는 닉 부위의 약 -30 내지 약 +30 사이, 또는 닉 부위의 약 -40 내지 약 +40 사이, 또는 닉 부위의 약 -50 내지 약 +50 사이, 또는 닉 부위의 약 -60 내지 약 +60 사이, 또는 닉 부위의 약 -70 내지 약 +70 사이, 또는 닉 부위의 약 -80 내지 약 +80 사이, 또는 닉 부위의 약 -90 내지 약 +90 사이, 또는 닉 부위의 약 -100 내지 약 +100 사이, 또는 닉 부위의 약 -200 내지 약 +200 사이인 편집 윈도우에서 설치된다.

[0809] 다른 실시양태에서, 목적하는 뉴클레오티드 변화는 닉 부위로부터 약 +1 내지 +2, 또는 닉 부위로부터 약 +1 내지 +3, +1 내지 +4, +1 내지 +5, +1 내지 +6, +1 내지 +7, +1 내지 +8, +1 내지 +9, +1 내지 +10, +1 내지 +11, +1 내지 +12, +1 내지 +13, +1 내지 +14, +1 내지 +15, +1 내지 +16, +1 내지 +17, +1 내지 +18, +1 내지 +19, +1 내지 +20, +1 내지 +21, +1 내지 +22, +1 내지 +23, +1 내지 +24, +1 내지 +25, +1 내지 +26, +1 내지 +27, +1 내지 +28, +1 내지 +29, +1 내지 +30, +1 내지 +31, +1 내지 +32, +1 내지 +33, +1 내지 +34, +1 내지 +35, +1 내지 +36, +1 내지 +37, +1 내지 +38, +1 내지 +39, +1 내지 +40, +1 내지 +41, +1 내지 +42, +1 내지 +43, +1 내지 +44, +1 내지 +45, +1 내지 +46, +1 내지 +47, +1 내지 +48, +1 내지 +49, +1 내지 +50, +1 내지 +51, +1 내지 +52, +1 내지 +53, +1 내지 +54, +1 내지 +55, +1 내지 +56, +1 내지 +57, +1 내지 +58, +1 내지 +59, +1 내지 +60, +1 내지 +61, +1 내지 +62, +1 내지 +63, +1 내지 +64, +1 내지 +65, +1 내지 +66, +1 내지 +67, +1 내지 +68, +1 내지 +69, +1 내지 +70, +1 내지 +71, +1 내지 +72, +1 내지 +73, +1 내지 +74, +1 내지 +75, +1 내지 +76, +1 내지 +77, +1 내지 +78, +1 내지 +79, +1 내지 +80, +1 내지 +81, +1 내지 +82, +1 내지 +83, +1 내지 +84, +1 내지 +85, +1 내지 +86, +1 내지 +87, +1 내지 +88, +1 내지 +89, +1 내지 +90, +1 내지 +90, +1 내지 +91, +1 내지 +92, +1 내지 +93, +1 내지 +94, +1 내지 +95, +1 내지 +96, +1 내지 +97, +1 내지 +98, +1 내지 +99, +1 내지 +100, +1 내지 +101, +1 내지 +102, +1 내지 +103, +1 내지 +104, +1 내지 +105, +1 내지 +106, +1 내지 +107, +1 내지 +108, +1 내지 +109, +1 내지 +110, +1 내지 +111, +1 내지 +112, +1 내지 +113, +1 내지 +114, +1 내지 +115, +1 내지 +116, +1 내지 +117, +1 내지 +118, +1 내지 +119, +1 내지 +120, +1 내지 +121, +1 내지 +122, +1 내지 +123, +1 내지 +124, 또는 +1 내지 +125 사이인 편집 윈도우에서 설치된다.

[0810] 또 다른 실시양태에서, 목적하는 뉴클레오티드 변화는 닉 부위로부터 약 +1 내지 +2, 또는 닉 부위로부터 약 +1 내지 +5, +1 내지 +10, +1 내지 +15, +1 내지 +20, +1 내지 +25, +1 내지 +30, +1 내지 +35, +1 내지 +40, +1 내지 +45, +1 내지 +50, +1 내지 +55, +1 내지 +100, +1 내지 +105, +1 내지 +110, +1 내지 +115, +1 내지 +120, +1 내지 +125, +1 내지 +130, +1 내지 +135, +1 내지 +140, +1 내지 +145, +1 내지 +150, +1 내지 +155, +1 내지 +160, +1 내지 +165, +1 내지 +170, +1 내지 +175, +1 내지 +180, +1 내지 +185, +1 내지 +190, +1 내지 +195, 또는 +1 내지 +200 사이인 편집 윈도우에서 설치된다.

[0811] 다양한 측면에서, 연장된 가이드 RNA는 가이드 RNA의 변형된 버전이다. 가이드 RNA는 자연 발생이거나, 코딩 핵산으로부터 발현되거나, 또는 화학적으로 합성될 수 있다. 가이드 RNA를 수득하거나 또는 달리 합성하고, 관심 게놈 표적 부위의 표적 가닥과 상호작용하고 혼성화하는 프로토스페이서 서열을 포함한 가이드 RNA의 적절한 서열을 결정하는 방법은 관련 기술분야에 널리 공지되어 있다.

[0812] 다양한 실시양태에서, 가이드 RNA 서열의 특정한 설계 측면은 다른 인자, 예컨대 PAM 서열 위치, 표적 서열에서의 퍼센트 G/C 함량, 미세상동성 영역의 정도, 2차 구조 등 중에서, 관심 게놈 표적 부위 (즉, 편집될 목적하는 부위)의 뉴클레오티드 서열 및 본원에 기재된 다중-플랩 프라임 편집 시스템에 존재하는 napDNAbp (예를 들어, Cas9 단백질)의 유형에 좌우될 것이다.

[0813] 일반적으로, 가이드 서열은 표적 서열과 혼성화하고 표적 서열에 대한 napDNAbp (예를 들어, Cas9, Cas9 상동체, 또는 Cas9 변이체)의 서열-특이적 결합을 지시하기 위해 표적 폴리뉴클레오티드 서열과 충분한 상보성을 갖는 임의의 폴리뉴클레오티드 서열이다. 일부 실시양태에서, 가이드 서열과 그의 상응하는 표적 서열 사이

의 상보성의 정도는, 적합한 정렬 알고리즘을 사용하여 최적으로 정렬한 경우에, 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% 또는 그 초과이거나, 또는 약 그보다 더 높다. 최적 정렬은 서열을 정렬하기 위한 임의의 적합한 알고리즘의 사용에 의해 결정될 수 있고, 그의 비제한적 예는 스미스-워터만(Smith-Waterman) 알고리즘, 니들만-분쉬(Needleman-Wunsch) 알고리즘, 버로우스-휠러 변환(Burrows-Wheeler Transform) (예를 들어, 버로우스 휠러 얼라이너(Burrows Wheeler Aligner)), 클러스탈W, 클러스탈 X, BLAT, 노보얼라인(Novoalign) (노보크라프트 테크놀로지스(Novocraft Technologies)), ELAND (일루미나(Illumina), 캘리포니아주 샌디에고), SOAP (soap.genomics.org.cn에서 이용가능함), 및 Maq (maq.sourceforge.net에서 이용가능함)에 기초한 알고리즘을 포함한다. 일부 실시양태에서, 가이드 서열은 약 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75개 또는 그 초과의 뉴클레오티드 길이이거나, 또는 약 그보다 더 길다.

[0814] 일부 실시양태에서, 가이드 서열은 약 75, 50, 45, 40, 35, 30, 25, 20, 15, 12개 미만, 또는 그보다 더 짧은 뉴클레오티드 길이이다. 표적 서열에 대한 다중-플랩 프라임 편집제의 서열-특이적 결합을 지시하는 가이드 서열의 능력은 임의의 적합한 검정에 의해 평가될 수 있다. 예를 들어, 시험될 가이드 서열을 포함한 다중-플랩 프라임 편집제의 성분은, 예컨대 본원에 개시된 다중-플랩 프라임 편집제의 성분을 코딩하는 벡터로의 형질감염에 이어서, 예컨대 본원에 기재된 바와 같은 서베이어 검정에 의한 표적 서열 내의 우선적 절단의 평가에 의해, 상응하는 표적 서열을 갖는 숙주 세포에 제공될 수 있다. 유사하게, 표적 폴리뉴클레오티드 서열의 절단은 시험 튜브에서 표적 서열, 시험될 가이드 서열을 포함한 다중-플랩 프라임 편집제의 성분 및 시험 가이드 서열과 상이한 대조군 가이드 서열을 제공하고, 시험 및 대조군 가이드 서열 반응 사이에서 표적 서열에서의 결합 또는 절단율을 비교함으로써 평가될 수 있다. 다른 검정이 가능하고, 이는 관련 기술분야의 통상의 기술자가 발견할 수 있을 것이다.

[0815] 가이드 서열은 임의의 표적 서열을 표적화하도록 선택될 수 있다. 일부 실시양태에서, 표적 서열은 세포의 게놈 내의 서열이다. 예시적인 표적 서열은 표적 게놈에서 고유한 것을 포함한다. 예를 들어, 에스. 피오게네스 Cas9의 경우, 게놈 내의 고유한 표적 서열은 형태 MMMMMMMMNNNNNNNNNNNNXGG (서열식별번호: 204)의 Cas9 표적 부위를 포함할 수 있다 (여기서 NNNNNNNNNNNNXGG (서열식별번호: 205) (N은 A, G, T, 또는 C이고; X는 임의의 것일 수 있음)). 게놈 내의 고유한 표적 서열은 형태 MMMMMMMMMNNNNNNNNNNNNXGG (서열식별번호: 206)의 에스. 피오게네스 Cas9 표적 부위를 포함할 수 있다 (여기서 NNNNNNNNNNNXGG (서열식별번호: 207) (N은 A, G, T, 또는 C이고; X는 임의의 것일 수 있음)). 에스. 써모필루스 CRISPR1Cas9의 경우, 게놈 내의 고유한 표적 서열은 형태 MMMMMMMMNNNNNNNNNNNNXXAGAAW (서열식별번호: 208)의 Cas9 표적 부위를 포함할 수 있다 (여기서 NNNNNNNNNNNNXXAGAAW (서열식별번호: 209) (N은 A, G, T, 또는 C이고; X는 임의의 것일 수 있고; W는 A 또는 T임)). 게놈 내의 고유한 표적 서열은 형태 MMMMMMMMMNNNNNNNNNNNXXAGAAW (서열식별번호: 210)의 에스. 써모필루스 CRISPR 1 Cas9 표적 부위를 포함할 수 있다 (여기서 NNNNNNNNNNNXXAGAAW (서열식별번호: 211) (N은 A, G, T, 또는 C이고; X는 임의의 것일 수 있고; W는 A 또는 T임)). 에스. 피오게네스 Cas9의 경우, 게놈 내의 고유한 표적 서열은 형태 MMMMMMMMNNNNNNNNNNNNXGGXG (서열식별번호: 212)의 Cas9 표적 부위를 포함할 수 있다 (여기서 NNNNNNNNNNNNXGGXG (서열식별번호: 213) (N은 A, G, T, 또는 C이고; X는 임의의 것일 수 있음)). 게놈 내의 고유한 표적 서열은 형태 MMMMMMMMMNNNNNNNNNNNXGGXG (서열식별번호: 214)의 에스. 피오게네스 Cas9 표적 부위를 포함할 수 있다 (여기서 NNNNNNNNNNNXGGXG (서열식별번호: 215) (N은 A, G, T, 또는 C이고; X는 임의의 것일 수 있음)). 이들 서열 각각에서, "M"은 A, G, T, 또는 C일 수 있고, 서열을 고유한 것으로 확인하는데 고려될 필요는 없다.

[0816] PEgRNA 또는 가이드 RNA 서열에서 본원에 사용된 바와 같이, 달리 나타내지 않는 한, 문자 "T" 또는 "티민"은 PEgRNA 또는 가이드 RNA 서열을 코딩하는 DNA 서열 내의 핵염기를 나타내고, PEgRNA 또는 가이드 RNA의 우라실 (U) 핵염기 또는 관련 기술분야에 공지된 임의의 화학적으로 변형된 우라실 핵염기, 예컨대 5-메톡시우라실을 지칭하는 것으로 의도된다는 것이 인지되어야 한다.

[0817] 일부 실시양태에서, 가이드 서열은 가이드 서열 내의 2차 구조의 정도를 감소시키도록 선택된다. 2차 구조는 임의의 적합한 폴리뉴클레오티드 폴딩 알고리즘에 의해 결정될 수 있다. 일부 프로그램은 최소 깁스 자유 에너지 계산에 기초한다. 이러한 한 알고리즘의 예는 문헌 [Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148)]에 기재된 바와 같은 mFold이다. 폴딩 알고리즘의 또 다른 예는 중심 구조 예측 알고리즘을 사용하는, 비엔나 대학교의 이론 화학 연구소에서 개발된 온라인 웹서버 RNAfold이다 (예를 들어, 문헌 [A. R. Gruber et al., 2008, Cell 106(1): 23-24; 및 PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62] 참조). 추가의 알고리즘은 본원에 참조로 포함되는 미국 출원 일련 번호 61/836,080; 광역 참조 BI-

2013/004A)에서 찾아볼 수 있다.

[0818] 일반적으로, tracr 메이트 서열은 (1) 상응하는 tracr 서열을 함유하는 세포에서 tracr 메이트 서열과 플랭킹된 가이드 서열의 절제; 및 (2) 표적 서열에서의 복합체의 형성 (여기서 복합체는 tracr 서열에 혼성화된 tracr 메이트 서열을 포함함) 중 1가지 이상을 촉진하기 위해 tracr 서열과 충분한 상보성을 갖는 임의의 서열을 포함한다. 일반적으로, 상보성의 정도는 2개의 서열 중 더 짧은 것의 길이에 따른 tracr 메이트 서열 및 tracr 서열의 최적 정렬에 대한 것이다. 최적 정렬은 임의의 적합한 정렬 알고리즘에 의해 결정될 수 있고, 2차 구조, 예컨대 tracr 서열 또는 tracr 메이트 서열 내의 자기-상보성을 추가로 고려할 수 있다. 일부 실시양태에서, 최적으로 정렬될 때 2개 중 더 짧은 것의 길이에 따른 tracr 서열과 tracr 메이트 서열 사이의 상보성의 정도는 약 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99% 또는 그 초과이거나, 또는 약 그보다 더 높다. 일부 실시양태에서, tracr 서열은 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50개 또는 그 초과의 뉴클레오티드 길이이거나, 또는 약 그보다 더 길다. 일부 실시양태에서, tracr 서열 및 tracr 메이트 서열은 단일 전사체 내에 함유되어, 둘 사이의 혼성화는 2차 구조, 예컨대 헤어핀을 갖는 전사체를 생성한다. 헤어핀 구조에 사용하기에 바람직한 루프 형성 서열은 4개 뉴클레오티드 길이이고, 가장 바람직하게는 서열 GAAA를 갖는다. 그러나, 대안적 서열일 수 있는 바와 같이, 더 길거나 더 짧은 루프 서열이 사용될 수 있다. 서열은 바람직하게는 뉴클레오티드 트리플렛 (예를 들어, AAA), 및 추가의 뉴클레오티드 (예를 들어 C 또는 G)를 포함한다. 루프 형성 서열의 예는 CAAA 및 AAAG를 포함한다. 본 발명의 한 실시양태에서, 전사체 또는 전사된 폴리뉴클레오티드 서열은 적어도 2개 이상의 헤어핀을 갖는다. 바람직한 실시양태에서, 전사체는 2, 3, 4 또는 5개의 헤어핀을 갖는다. 본 발명의 추가 실시양태에서, 전사체는 최대 5개의 헤어핀을 갖는다. 일부 실시양태에서, 단일 전사체는 전사 종결 서열을 추가로 포함하고; 바람직하게는 이는 폴리T 서열, 예를 들어 6개의 T 뉴클레오티드이다. 가이드 서열, tracr 메이트 서열, 및 tracr 서열을 포함하는 단일 폴리뉴클레오티드의 추가의 비제한적 예는 하기와 같고 (5'에서 3'으로 열거됨), 여기서 "N"은 가이드 서열의 염기를 나타내고, 제1 블록의 소문자는 tracr 메이트 서열을 나타내고, 제2 블록의 소문자는 tracr 서열을 나타내고, 최종 폴리-T 서열은 전사 종결인자를 나타낸다:

[0819] (1)NNNNNNNNGTTTTTGTACTCTCAAGATTTAGAAATAAATCTTGCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTCGTTATTTAATTTTTT (서열식별번호: 216);

[0820] (2)NNNNNNNNNNNNNNNNNNGTTTTTGTACTCTCAGAAATGCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTCGTTATTTAATTTTTT (서열식별번호: 217);

[0821] (3)NNNNNNNNNNNNNNNNNNNNGTTTTTGTACTCTCAGAAATGCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTTT (서열식별번호: 218);

[0822] (4)NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT (서열식별번호: 219);

[0823] (5)NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGTTTTTTT (서열식별번호: 220); 및

[0824] (6) NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCATTTTTTTT(서열식별번호: 221).

[0825] 일부 실시양태에서, 서열 (1) 내지 (3)은 에스. 써모필루스 CRISPR1로부터의 Cas9와 조합되어 사용된다. 일부 실시양태에서, 서열 (4) 내지 (6)은 에스. 피오게네스로부터의 Cas9와 조합되어 사용된다. 일부 실시양태에서, tracr 서열은 tracr 메이트 서열을 포함하는 전사체와 별개의 전사체이다.

[0826] 본원에 개시된 바와 같은 Cas9 도메인 및 단일 가닥 DNA 결합 단백질을 포함하는 융합 단백질 중 임의의 것을 표적 부위, 예를 들어 편집될 점 돌연변이를 포함하는 부위로 표적화하기 위해, 전형적으로 융합 단백질을 가이드 RNA, 예를 들어 sgRNA와 함께 공동-발현시키는 것이 필요하다는 것은 관련 기술분야의 통상의 기술자에게 명백할 것이다. 본원의 다른 곳에 보다 상세히 설명된 바와 같이, 가이드 RNA는 전형적으로 Cas9 결합을 가능하게 하는 tracrRNA 프레임워크, 및 Cas9:핵산 편집 효소/도메인 융합 단백질에 서열 특이성을 부여하는 가이드 서열을 포함한다.

[0827] 일부 실시양태에서, 가이드 RNA는 구조

5'-[가이드 서열]-
GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU
[0828] UGAAAAAGUGGCACCGAGUCGGUGCUUUUU-3' (SEQ ID NO: 222)

[0829] 를 포함하며, 여기서 가이드 서열은 표적 서열에 상보적인 서열을 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오티드 길이이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특이적 게놈 표적 부위로 표적화하기에 적합한 가이드 RNA의 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다. 이러한 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오티드의 50개 뉴클레오티드 상류 또는 하류 내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 제공된 융합 단백질 중 임의의 것을 특이적 표적 서열에 표적화하는데 적합한 일부 예시적인 가이드 RNA 서열이 본원에 제공된다. 추가의 가이드 서열은 관련 기술분야에 널리 공지되어 있고, 본원에 기재된 다중-플랩 프라임 편집제와 함께 사용될 수 있다.

[0830] 다른 실시양태에서, PEgRNA는 도 3d에 도시된 것을 포함한다.

[0831] 또 다른 실시양태에서, PEgRNA는 도 3e에 도시된 것을 포함할 수 있다.

[0832] 도 3d는 본원에서 고려되고 실시예 2에 규정된 방법론에 따라 설계될 수 있는 PEgRNA의 실시양태의 구조를 제공한다. PEgRNA는 5'에서 3' 방향으로 정렬된 3개의 주요 성분 요소, 즉 스페이서, gRNA 코어, 및 3' 단부의 연장 아암을 포함한다. 연장 아암은 5'에서 3' 방향으로 하기 구조적 요소, 즉: 프라이머 결합 부위 (A), 편집 주형 (B), 및 상동성 아암 (C)으로 추가로 나뉠 수 있다. 또한, PEgRNA는 임의적인 3' 단부 변형제 영역 (e1) 및 임의적인 5' 단부 변형제 영역 (e2)을 포함할 수 있다. 또한, PEgRNA는 PEgRNA의 3' 단부에 전사 종결 신호를 포함할 수 있다 (도시되지 않음). 이들 구조적 요소는 본원에서 추가로 정의된다. PEgRNA의 구조의 도시는 제한적인 것으로 의도되지 않고, 요소의 배열에서의 변경을 포괄한다. 예를 들어, 임의적인 서열 변형제 (e1) 및 (e2)는 제시된 다른 영역 중 임의의 것 내에 또는 그 사이에 위치할 수 있고, 3' 및 5' 단부에 위치하는 것으로 제한되지는 않는다.

[0833] 도 3e는 본원에서 고려되고 실시예 2에 규정된 방법론에 따라 설계될 수 있는 PEgRNA의 또 다른 실시양태의 구조를 제공한다. PEgRNA는 5'에서 3' 방향으로 정렬된 3개의 주요 성분 요소, 즉 스페이서, gRNA 코어, 및 3' 단부의 연장 아암을 포함한다. 연장 아암은 5'에서 3' 방향으로 하기 구조적 요소, 즉: 프라이머 결합 부위 (A), 편집 주형 (B), 및 상동성 아암 (C)으로 추가로 나뉠 수 있다. 또한, PEgRNA는 임의적인 3' 단부 변형제 영역 (e1) 및 임의적인 5' 단부 변형제 영역 (e2)을 포함할 수 있다. 또한, PEgRNA는 PEgRNA의 3' 단부에 전사 종결 신호를 포함할 수 있다 (도시되지 않음). 이들 구조적 요소는 본원에서 추가로 정의된다. PEgRNA의 구조의 도시는 제한적인 것으로 의도되지 않고, 요소의 배열에서의 변경을 포괄한다. 예를 들어, 임의적인 서열 변형제 (e1) 및 (e2)는 제시된 다른 영역 중 임의의 것 내에 또는 그 사이에 위치할 수 있고, 3' 및 5' 단부에 위치하는 것으로 제한되지는 않는다.

[0834] PEgRNA 개선

[0835] PEgRNA는 또한 PEgRNA의 특성 및/또는 특징을 변형시켜 다중-플랩 프라임 편집의 효능을 개선시킬 수 있는 추가의 설계 개선을 포함할 수 있다. 다양한 실시양태에서, 이들 개선은 (1) 부담스러운 서열 요건 없이 보다 긴 PEgRNA의 발현을 가능하게 할, 비-폴리머라제 III (pol III) 프로모터로부터의 기능적 PEgRNA의 효율적인 발현을 가능하게 하는 설계; (2) 효능을 개선시킬 수 있는 코어, Cas9-결합 PEgRNA 스캐폴드에 대한 개선; (3) RT 프로세싱성을 개선시켜 표적화된 게놈 유전자좌에서 보다 긴 서열의 삽입을 가능하게 하는, PEgRNA에 대한 변형; 및 (4) PEgRNA 안정성을 개선시키거나, RT 프로세싱성을 증진시키거나, PEgRNA의 미스폴딩을 방지하거나, 또는 게놈 편집에 중요한 추가의 인자를 동원하는, PEgRNA의 5' 또는 3' 말단에의 RNA 모티프의 부가를 포함하나 이에 제한되지는 않는 다수의 상이한 카테고리 중 1종 이상에 속할 수 있다.

[0836] 한 실시양태에서, PEgRNA는 보다 큰 연장 아암을 갖는 보다 긴-길이의 PEgRNA의 발현을 개선시키기 위해 polIII 프로모터와 함께 설계될 수 있다. sgRNA는 전형적으로 U6 snRNA 프로모터로부터 발현된다. 이러한 프로모터는 pol III을 동원하여 회합된 RNA를 발현시키고, 이는 핵 내에 보유된 짧은 RNA의 발현에 유용하다. 그러나, pol III은 고도로 프로세싱성이 아니고, 수백개 뉴클레오티드 길이보다 더 긴 RNA를 효율적인 게놈 편집에 요구되는 수준으로 발현시킬 수 없다. 추가적으로, pol III은 U 스트레치에서 멈추거나 종결될 수 있으며, 이는 잠재적으로 PEgRNA를 사용하여 삽입될 수 있는 서열 다양성을 제한한다. 폴리머라제 II (예컨대 pCMV) 또는 폴리머라

제 I (예컨대 U1 snRNA 프로모터)을 동원하는 다른 프로모터를 보다 긴 sgRNA를 발현시키는 그의 능력에 대해 검사하였다. 그러나, 이들 프로모터는 전형적으로 부분적으로 전사되고, 이는 발현된 PEgRNA에서 스페이서의 5'의 여분의 서열을 생성할 것이며, 이는 부위-의존성 방식으로 현저하게 감소된 Cas9:sgRNA 활성을 생성하는 것으로 제시되었다. 추가적으로, pol III-전사된 PEgRNA는 단순히 연속 6-7개 U에서 종결될 수 있지만, pol II 또는 pol I로부터 전사된 PEgRNA는 상이한 종결 신호를 필요로 할 것이다. 종종 이러한 신호는 또한 폴리아데닐화를 발생시키며, 이는 핵으로부터 PEgRNA의 바람직하지 않은 수송을 발생시킬 것이다. 유사하게, pol II 프로모터, 예컨대 pCMV로부터 발현된 RNA는 전형적으로 5'-캡핑되며, 이는 또한 그의 핵 유출을 발생시킨다.

[0837] 이전에, 린(Rinn) 및 동료들은 긴-비코딩 RNA- (lncRNA) 태그부착된 sgRNA의 생산을 위해 다양한 발현 플랫폼을 스크리닝하였다[183]. 이들 플랫폼은 pCMV로부터 발현되고 인간으로부터의 MALAT1 ncRNA로부터의 ENE 요소[184], KSHV로부터의 PAN ENE 요소[185], 또는 U1 snRNA로부터의 3' 박스[186]에서 종결하는 RNA를 포함한다. 특히, MALAT1 ncRNA 및 PAN ENE는 폴리A-테일을 보호하는 삼중 나선을 형성한다[184, 187]. 이들 구축물은 또한 RNA 안정성을 증진시킬 수 있다. 이들 발현 시스템은 또한 보다 긴 PEgRNA의 발현을 가능하게 할 것으로 고려된다.

[0838] 또한, 자기-절단 리보자임, 예컨대 해머헤드[188], 피스톨[189], 해체트[189], 헤어핀[190], VS[191], 트위스터[192], 또는 트위스터 시스터[192] 리보자임, 또는 전사된 가이드를 프로세싱하기 위한 다른 자기-절단 요소, 또는 Csy4[193]에 의해 인식되고 또한 가이드의 프로세싱으로 이어지는 헤어핀을 부가하는, PEgRNA의 일부로서 전사될 pol II 프로모터의 부분의 절단을 위한 일련의 방법이 설계되었다. 또한, 다중 ENE 모티프의 혼입은 KSHV PAN RNA 및 요소에 대해 이전에 입증된 바와 같이 개선된 PEgRNA 발현 및 안정성으로 이어질 수 있는 것으로 가설화된다[185]. 또한, PEgRNA를 원형 인트론 RNA (ciRNA)의 형태로 원형화하는 것은 또한 증진된 RNA 발현 및 안정성, 뿐만 아니라 핵 국재화로 이어질 수 있는 것으로 예상된다[194].

[0839] 다양한 실시양태에서, PEgRNA는 하기 서열에 의해 예시된 바와 같은 다양한 상기 요소를 포함할 수 있다.

[0840] 비제한적 예 1 - pCMV, Csy4 헤어핀, PEgRNA, 및 MALAT1 ENE로 이루어진 PEgRNA 발현 플랫폼

[0841]
```
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGA
CTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAG
TCTGTTTTAGGGTCATGAAGGTTTTTCTTTTCCTGAGAAAACAACACGTATTGTTTT
CTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAAAAAGCAAAAGAT
GCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCT
TTGCTTTGACT (SEQ ID NO: 223)
```

[0842] 비제한적 예 2 - pCMV, Csy4 헤어핀, PEgRNA, 및 PAN ENE로 이루어진 PEgRNA 발현 플랫폼

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGA
CTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAG
TCTGTTTTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCA
GACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAAT
TTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAAC
ATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 224)

[0843]

[0844] 비제한적 예 3 - pCMV, Csy4 헤어핀, PEgRNA, 및 3xPAN ENE로 이루어진 PEgRNA 발현 플랫폼

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGT
TACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACG
TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGT
GGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTT
ATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGC
GGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT
CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT
CTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCC
CAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT
ATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTT
TTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGG
TTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGCCTAGAG
CTCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTTA
ATCCATAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 225)

[0845]

[0846] 비제한적 예 4 - pCMV, Csy4 헤어핀, PEgRNA, 및 3' 박스로 이루어진 PEgRNA 발현 플랫폼

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGA
CTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAG
TCTGTTTGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTGGTT
TGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 226)

[0847]

[0848] 비제한적 예 5 - pU1, Csy4 헤어핀, PEgRNA, 및 3' 박스로 이루어진 PEgRNA 발현 플랫폼

CTAAGGACCAGCTTCTTTGGGAGAGAACAGACGCAGGGGCGGGAGGGAAAAAG
GGAGAGGCAGACGTCACTTCCCCTTGGCGGCTCTGGCAGCAGATTGGTCGGTTGA
GTGGCAGAAAGGCAGACGGGGACTGGGCAAGGCACTGTCGGTGACATCACGGAC
AGGGCGACTTCTATGTAGATGAGGCAGCGCAGAGGCTGCTGCTTCGCCACTTGCT
GCTTCACCACGAAGGAGTTCCCGTGCCCTGGGAGCGGGTTCAGGACCGCTGATCG
GAAGTGAGAATCCCAGCTGTGTGTCAGGGCTGGAAAGGGCTCGGGAGTGCGCGG
GGCAAGTGACCGTGTGTGTAAAGAGTGAGGCGTATGAGGCTGTGTCGGGGCAGA
GGCCCAAGATCTCAGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAG
TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTCAGCAAG
TTCAGAGAAATCTGAACTTGCTGGATTTTTGGAGCAGGGAGATGGAATAGGAGCT
TGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTACCTCCAGGAACGGTGC
ACCCACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTT
GTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 227).

[0849]

[0850] 다양한 다른 실시양태에서, PEgRNA는 스캐폴드 또는 코어 서열에 개선을 도입하는 것에 의해 개선될 수 있다. 이는 공지된 것을 도입함으로써 수행될 수 있다.

[0851] 코어, Cas9-결합 PEgRNA 스캐폴드는 PE 활성을 증진시키기 위해 개선될 가능성이 있다. 여러 이러한 접근법은 이미 입증되었다. 예를 들어, 스캐폴드의 제1 쌍형성 요소 (P1)는 GTTTT-AAAAC (서열식별번호: 3939) 쌍형성 요소를 함유한다. 이러한 연속 T는 pol III 일시정지 및 RNA 전사체의 조기 종결을 발생시키는 것으로 밝혀졌다. P1의 이러한 부분에서 T-A 쌍 중 1개의 G-C 쌍으로의 합리적 돌연변이는 sgRNA 활성을 증진시키는 것으로 밝혀졌고, 이는 이러한 접근법이 또한 PEgRNA에 대해 실행가능할 것임을 시사한다[195]. 추가적으로, P1의 길이를 증가시키는 것은 또한 sgRNA 폴딩을 증진시키고 개선된 활성으로 이어지는 것으로 밝혀졌고[195], 이는 이것이 PEgRNA 활성의 개선을 위한 또 다른 방안임을 시사한다. 코어에 대한 예시적인 개선은 하기를 포함할 수 있다:

[0852] P1에 대해 6 nt 연장부를 함유하는 PEgRNA

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGCTCATGAAAATGAGCTAGCAAG
TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTC
TGCCATCAAAGCGTGCTCAGTCTGTTTTTTT (SEQ ID NO: 228)

[0853]

[0854] P1 내에 T-A에서 G-C로의 돌연변이를 함유하는 PEgRNA

GGCCCAGACTGAGCACGTGAGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGC
TAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGC
GTGCTCAGTCTGTTTTTTT (SEQ ID NO: 229)

[0855]

[0856] 다양한 다른 실시양태에서, PEgRNA는 편집 주형 영역에 변형을 도입함으로써 개선될 수 있다. PEgRNA에 의해 주형화되는 삽입의 크기가 증가함에 따라, 엔도뉴클레아제에 의해 분해되거나, 자발적 가수분해를 겪거나, 또는 RT에 의해 역전사될 수 없거나 또는 PEgRNA 스캐폴드의 폴딩 및 후속 Cas9-RT 결합을 파괴하는 2차 구조로 폴딩될 가능성이 더 크다. 따라서, 전체 유전자의 삽입과 같은 대규모 삽입에 영향을 미치기 위해서는 PEgRNA의 주형에 대한 변형이 필요할 가능성이 있다. 이를 위한 일부 전략은 RNA를 분해 또는 가수분해에 대해 보다 저항성이게 하거나 또는 억제성 2차 구조를 채택할 가능성을 낮추는, 합성 또는 반합성 PEgRNA 내의 변형된 뉴클레오티드의 혼입을 포함한다[196]. 이러한 변형은 G-풍부 서열에서 RNA 2차 구조를 감소시킬 8-아자-7-데아자구아노신; 분해를 감소시키고 특정 종류의 RNA 2차 구조를 증진시키는 잠금 핵산 (LNA); RNA 안정성을 증진시키는 2'-O-메틸, 2'-플루오로, 또는 2'-O-메톡시에톡시 변형을 포함할 수 있다. 이러한 변형은 또한 안정성 및 활성을 증진시키기 위해 PEgRNA의 다른 곳에 포함될 수 있다. 대안적으로 또는 추가적으로, PEgRNA의 주형은 목적하는 단백질 생성물을 코딩하고 또한 RT에 의해 언폴딩될 수 있는 간단한 2차 구조를 채택할 가능성이 더 크도록 설계될 수 있다. 이러한 단순한 구조는 열역학적 싱크로서 작용할 것이며, 이는 역전사를 막을 보다 복잡한 구조가 발생할 가능성을 줄인다. 마지막으로, 또한 주형을 2개의 별개의 PEgRNA로 분할할 수 있다. 이러한 설계에서, PE는 전사를 개시하고 또한 Cas9에 융합된 RNA-결합 단백질 또는 PEgRNA 자체 상의 RNA 인식 요소, 예컨대 MS2 압타머를 통해 표적화된 부위에 별개의 주형 RNA를 동원하는 데 사용될 것이다. RT는 이러한 별개의 주형 RNA에 직접 결합할 수 있거나, 또는 제2 주형으로 스와핑하기 전에 원래의 PEgRNA 상에서 역전사를 개시할 수 있다. 이러한 접근법은 긴 주형의 부가 시 PEgRNA의 미스폴딩을 방지함으로써, 또한 아마도 PE-기반 긴 삽입을 억제할 수 있는, 긴 삽입이 이루어지는 게놈으로부터 Cas9의 해리를 필요로 하지 않음으로써 긴 삽입을 가능하

게 할 수 있다.

[0857] 또 다른 실시양태에서, PEgRNA는 PEgRNA의 5' 및 3' 말단에, 또는 심지어 그 사이의 위치에 (예를 들어, gRNA 코어 영역, 또는 스페이서에) 추가의 RNA 모티프를 도입함으로써 개선될 수 있다. 여러 이러한 모티프 - 예컨대 KSHV로부터의 PAN ENE 및 MALAT1로부터의 ENE는 비-pol III 프로모터로부터 보다 긴 PEgRNA의 발현을 종결시키기 위한 가능한 수단으로서 상기 논의되었다. 이들 요소는 폴리A 테일을 에워싸는 RNA 삼중 나선을 형성하여, 이들이 핵 내에 보유되게 한다[184, 187]. 그러나, 말단 뉴클레오티드를 폐쇄하는 PEgRNA의 3' 말단에 복합체 구조를 형성함으로써, 이들 구조는 또한 아마도 PEgRNA의 엑소뉴클레아제-매개된 분해를 방지하는 것을 도울 것이다.

[0858] 3' 말단에 삽입된 다른 구조적 요소도 또한 RNA 안정성을 증진시킬 수 있지만, 비-pol III 프로모터로부터의 종결은 가능하게 하지 않는다. 이러한 모티프는 3' 말단을 폐쇄할 헤어핀 또는 RNA 쿼드로플렉스[197], 또는 3' 말단에서 2'-3'-시클릭 포스페이트의 형성을 발생시키고 또한 잠재적으로 PEgRNA가 엑소뉴클레아제에 의해 분해될 가능성을 낮출 자기-절단 리보자임, 예컨대 HDV[198]를 포함할 수 있다. PEgRNA를 불완전한 스플라이싱을 통해 고리화하도록 유도하여 - ciRNA를 형성하는 것은 - 또한 PEgRNA 안정성을 증가시키고 PEgRNA가 핵 내에 보유되도록 할 수 있다[194].

[0859] 추가의 RNA 모티프는 또한 RT 프로세싱성을 개선시키거나 또는 DNA-RNA 듀플렉스에 대한 RT 결합을 증진시킴으로써 PEgRNA 활성을 증진시킬 수 있다. 그의 동족 레트로바이러스 게놈에서 RT에 의해 결합된 천연 서열의 부가는 RT 활성을 증진시킬 수 있다[199]. 이는 레트로바이러스 게놈 이량체화 및 전사의 개시에 수반되는 천연 프라이머 결합 부위 (PBS), 폴리퓨린 트랙 (PPT), 또는 키싱 루프를 포함할 수 있다[199].

[0860] PEgRNA의 5' 및 3' 말단에서의 이량체화 모티프 - 예컨대 키싱 루프 또는 GNRA 테트라루프/테트라루프 수용체 쌍[200] - 의 부가는 또한 PEgRNA의 효과적인 원형화를 발생시켜 안정성을 개선시킬 수 있다. 추가적으로, 이들 모티프의 부가가 PEgRNA 스페이서 및 프라이머의 물리적 분리를 가능하게 하여, PE 활성을 방해할 스페이서의 폐쇄를 방지할 수 있는 것으로 고려된다. 스페이서 영역에서 또는 프라이머 결합 부위를 따라 작은 토홀드 헤어핀을 형성하는 PEgRNA에 대한 짧은 5' 연장부 또는 3' 연장부는 또한 PEgRNA의 길이를 따라 상보성내 영역의 어닐링, 예를 들어 발생할 수 있는 스페이서와 프라이머 결합 부위 사이의 상호작용과 유리하게 경쟁할 수 있다. 마지막으로, 키싱 루프는 또한 다른 주형 RNA를 게놈 부위로 동원하고 RT 활성을 하나의 RNA로부터 다른 것으로 스와핑하는 것을 가능하게 하는 데 사용될 수 있다. 다양한 2차 구조의 예시적인 실시양태로서, 도 3d 및 도 3e에 도시된 PEgRNA는 제시된 바와 같이 연장 아암의 말단 부분 (즉, e1 및 e2)을 비롯하여 PEgRNA의 임의의 영역 내로 조작될 수 있는 많은 2차 RNA 구조를 열거한다.

[0861] 예시적인 개선은 하기를 포함하나 이에 제한되지는 않는다:

[0862] PEgRNA-HDV 융합체

[0863] GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
TAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGC
GTGCTCAGTCTGGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAA
CATGCTTCGGCATGGCGAATGGGACTTTTTTT (SEQ ID NO: 230)

[0864] PEgRNA-MMLV 키싱 루프

[0865] GGTGGGAGACGTCCCACCGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCA
AGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGC
CATCAAAGCTTCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCTTTTTTT (SEQ ID NO: 231)

[0866] PEgRNA-VS 리보자임 키싱 루프

[0867] GAGCAGCATGGCGTCGCTGCTCACGGCCCAGACTGAGCACGTGAGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACC
GAGTCGGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTCCATCAGTTGACA
CCCTGAGGTTTTTTT (SEQ ID NO: 232)

[0868] PEgRNA-GNRA 테트라루프/테트라루프 수용체

[0869] GCAGACCTAAGTGGUGACATATGGTCTGGGCCCAGACTGAGCACGTGAGTTTTAG
AGCTAUACGTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTUACGAAGTGG
GACCGAGTCGGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTGCATGCGATT
AGAAATAATCGCATGTTTTTTT(SEQ ID NO: 233)

[0870] PEgRNA 주형 스위칭 2차 RNA-HDV 융합체

[0871] TCTGCCATCAAAGCTGCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCGGCCG
GCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCG
AATGGGACTTTTTTT (SEQ ID NO: 234)

[0872] PEgRNA 스캐폴드는 SpCas9 및 프라임 편집제 (PE)가 개선되는 방법과 유사한 방식으로 방향적 진화를 통해 추가로 개선될 수 있다. 방향적 진화는 Cas9 또는 진화된 Cas9 변이체에 의한 PEgRNA 인식을 증진시킬 수 있다. 추가적으로, 상이한 PEgRNA 스캐폴드 서열은 상이한 게놈 유전자좌에서 최적이어서, 해당 부위에서 PE 활성을 증진시키거나, 오프-타겟 활성을 감소시키거나, 또는 둘 다일 가능성이 있다. 마지막으로, 다른 RNA 모티프가 부가된 PEgRNA 스캐폴드의 진화는 비진화된 융합 RNA에 비해 융합된 PEgRNA의 활성을 거의 확실히 개선시킬 것이다. 예를 들어, c-디-GMP-I 압타머 및 해머헤드 리보자임으로 구성된 알로스테릭 리보자임의 진화는 현저하게 개선된 활성으로 이어졌고[202], 이는 진화가 해머헤드-PEgRNA 융합체의 활성을 또한 개선시킬 것임을 시사한다. 또한, Cas9는 현재 일반적으로 sgRNA의 5' 연장을 용인하지 않지만, 방향적 진화는 아마도 이러한 불용인을 완화시켜 추가의 RNA 모티프가 이용되도록 하는 돌연변이를 생성할 수 있을 것이다.

[0873] 본 개시내용은 본원에 개시된 다중-플랩 프라임 편집 시스템의 효능을 추가로 개선시키는 임의의 이러한 방식을 고려한다.

[0874] 다양한 실시양태에서, 연속 T 시리즈는 전사될 PEgRNA의 능력을 제한할 수 있기 때문에 연장 아암으로부터 연속 T 서열의 출현을 제한하는 것이 유리할 수 있다. 예를 들어, 적어도 연속 3개의 T, 적어도 연속 4개의 T, 적어도 연속 5개의 T, 적어도 연속 6개의 T, 적어도 연속 7개의 T, 적어도 연속 8개의 T, 적어도 연속 9개의 T, 적어도 연속 10개의 T, 적어도 연속 11개의 T, 적어도 연속 12개의 T, 적어도 연속 13개의 T, 적어도 연속 14개의 T, 또는 적어도 연속 15개의 T의 스트링은 PEgRNA를 설계할 때 피해야 하거나, 또는 적어도 최종 설계된 서열로부터 제거되어야 한다. 한 실시양태에서, PEgRNA 연장 아암 내에 원치 않는 연속 T 스트링이 포함되는 것을 피할 수 있지만, 연속 A:T 핵염기 쌍이 풍부한 표적 부위를 피할 수 있다.

[0875] 트랜스 프라임 편집을 위한 분할 PEgRNA 설계

[0876] 본 개시내용은 또한 PEgRNA를 2개의 별개의 분자: 가이드 RNA 및 tPERT 분자로 분리함으로써 작업하는 프라임 편집의 변형된 버전을 지칭하는 트랜스 프라임 편집을 고려한다. tPERT 분자는 표적 DNA 부위에서 프라임 편집제 복합체와 공동-국재화되어, 프라이머 결합 부위 및 DNA 합성 주형을 프라임 편집제로 트랜스로 가져오도록 프로그램화된다. 예를 들어, (1) 동원 단백질 (RP)-PE:gRNA 복합체 및 (2) RNA-단백질 동원 도메인 (예를 들어, 스템 루프 또는 헤어핀)에 연결된 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 tPERT를 포함하는 2-성분 시스템을 보여주는 트랜스 프라임 편집제 (tPE)의 실시양태에 대해 도 3g를 참조하며, 여기서 RP-PE:gRNA 복합체의 동원 단백질 성분은 tPERT를 편집될 표적 부위로 동원함으로써 PBS 및 DNA 합성 주형을 프라임 편집제와 트랜스로 회합시킨다. 상기 또 다른 방식에서, tPERT는 프라이머 결합 부위 및 DNA 합성 주형을 포함하는 PEgRNA의 연장 아암을 (모두 또는 일부) 함유하도록 조작된다. 이러한 접근법의 하나의 이점은 PEgRNA의 연장 아암을 가이드 RNA로부터 분리함으로써, 연장 아암의 PBS와 가이드 RNA의 스페이서 서열 사이에 발생하는 경향이 있는 어닐링 상호작용을 최소화한다는 것이다.

[0877] 트랜스 프라임 편집의 주요 특색은 트랜스 프라임 편집제가 tPERT를 DNA 편집의 부위로 동원함으로써, 프라임 편집의 부위에 PEgRNA의 모든 기능을 효과적으로 공동-국재화하는 능력이다. 동원은 RNA-단백질 동원 도메인, 예컨대 MS2 압타머를 tPERT에 설치하고, RNA-단백질 동원 도메인에 특이적으로 결합할 수 있는 상응하는 동원 단백질을 프라임 편집제에 융합시켜 (예를 들어, napDNAbp에 대한 링커를 통해 또는 폴리머라제에 대한 링커를 통해), tPERT 분자를 프라임 편집제 복합체에 동원함으로써 달성될 수 있다. 도 3h에 기재된 과정에 도시된 바와 같이, RP-PE:gRNA 복합체는 표적 DNA 서열에 결합하여 닉킹한다. 이어서, 동원 단백질 (RP)은 tPERT를 동원하여 DNA 표적 부위에 결합된 프라임 편집제 복합체에 공동-국재화시킴으로써, tPERT 상에 위치한 프라이머 결합 부위가 닉킹된 가닥 상의 프라이머 서열에 결합하게 하고, 후속적으로 폴리머라제 (예를 들어, RT)가 tPERT

상에 위치한 DNA 합성 주형에 대해 tPERT의 5' 단부까지 단일 가닥의 DNA를 합성하게 한다.

[0878] 도 3g 및 도 3h에는 tPERT가 RNA-단백질 동원 도메인의 5' 단부 상에 PBS 및 DNA 합성 주형을 포함하는 것으로서 제시되지만, RNA-단백질 동원 도메인의 3' 단부 상에 PBS 및 DNA 합성 주형이 위치하는 다른 구성의 tPERT가 설계될 수 있다. 그러나, DNA의 단일 가닥의 합성이 tPERT의 5' 단부에서 자연적으로 종결될 것이고, 따라서 프라임 편집의 DNA 합성 단계 동안 주형으로서 RNA-단백질 동원 도메인의 임의의 부분을 사용할 위험이 없다는 점에서 5' 연장부를 갖는 tPERT가 이점을 갖는다.

[0879] PEgRNA 설계 방법

[0880] 본 개시내용은 또한 PEgRNA를 설계하는 방법에 관한 것이다.

[0881] 설계의 한 측면에서, 설계 접근법은 프라임 편집이 사용되는 특정한 적용을 고려할 수 있다. 예를 들어, 본원에 예시되고 논의된 바와 같이, 프라임 편집은 비제한적으로, (a) 뉴클레오티드 서열에 돌연변이-교정 변화를 설치하고, (b) 단백질 및 RNA 태그를 설치하고, (c) 관심 단백질 상에 이뮤노에피토프를 설치하고, (d) 단백질에 유도성 이량체화 도메인을 설치하고, (e) 생체분자의 활성을 변경하기 위해 서열을 설치하거나 제거하고, (f) 특이적 유전자 변화를 지시하기 위해 레콤비나제 표적 부위를 설치하고, (g) 오류-유발 RT를 사용함으로써 표적 서열을 돌연변이유발하는 데 사용될 수 있다. 일반적으로 관심 표적 부위에서 뉴클레오티드 서열을 삽입, 변화, 또는 결실시키는 이들 방법에 더하여, 프라임 편집제는 또한 고도로 프로그램가능한 라이브러리를 구축하는데, 뿐만 아니라 세포 데이터 기록 및 계통 추적 연구를 수행하는 데 사용될 수 있다. 이들 다양한 용도에서, 본원에 기재된 바와 같이 임의의 주어진 이들 적용에 특히 유용한 PEgRNA의 제조에 관한 특정한 설계 측면이 존재할 수 있다.

[0882] 프라임 편집의 임의의 특정한 적용 또는 사용을 위해 PEgRNA를 설계하는 경우에, 하기를 포함하나 이에 제한되지는 않는 다수의 고려사항이 고려될 수 있다:

[0883] (a) 표적 서열, 즉 1개 이상의 핵염기 변형을 프라임 편집제로 설치하고자 목적하는 뉴클레오티드 서열;

[0884] (b) 표적 서열 내의 절단 부위의 위치, 즉 프라임 편집제가 단일 가닥 닉을 유도하여 닉의 한 측면 상에 3' 단부 RT 프라이머 서열을 생성하고 닉의 다른 측면 상에 5' 단부 내인성 플랩을 생성할 (이는 궁극적으로 FEN1 또는 그의 등가물에 의해 제거되고 3' ssDNA 플랩에 의해 대체됨) 특이적 핵염기 위치. 절단 부위는 "편집 위치"와 유사한데, 이는 이것이 RNA-의존성 DNA 중합 동안 RT에 의해 연장되게 되는 3' 단부 RT 프라이머 서열을 생성하여 목적하는 편집을 함유하는 3' ssDNA 플랩을 생성하고, 이는 이어서 표적 서열에서 5' 내인성 DNA 플랩을 대체하기 때문이다.

[0885] (c) 이용가능한 PAM 서열 (정규 SpCas9 PAM 부위, 뿐만 아니라 확장된 또는 상이한 PAM 특이성을 갖는 Cas9 변이체 및 등가물에 의해 인식되는 비-정규 PAM 부위 포함);

[0886] (d) 이용가능한 PAM 서열과 표적 서열 내의 절단 부위의 위치 사이의 간격;

[0887] (e) 사용되는 프라임 편집제의 특정한 Cas9, Cas9 변이체, 또는 Cas9 등가물;

[0888] (f) 프라이머 결합 부위의 서열 및 길이;

[0889] (g) 편집 주형의 서열 및 길이;

[0890] (h) 상동성 아암의 서열 및 길이;

[0891] (i) 스페이서 서열 및 길이; 및

[0892] (j) 코어 서열.

[0893] 본 개시내용은 상기 이들 측면을 논의한다.

[0894] 한 실시양태에서, 적합한 PEgRNA, 및 임의로 제2-부위 닉킹을 위한 닉킹-sgRNA 설계 가이드를 설계하기 위한 접근법이 본원에 제공된다. 이러한 실시양태는 상기 고려사항 중 1가지 이상을 고려한, 프라임 편집을 위한 PEgRNA 및 닉킹-sgRNA를 설계하기 위한 단계별 지침 세트를 제공한다. 단계는 도 70a-70i에 제시된 예를 참조한다.

[0895] 1. 표적 서열 및 편집을 규정한다. 목적하는 편집 (점 돌연변이, 삽입, 결실, 또는 그의 조합)의 위치 주위에 집중된 표적 DNA 영역 (~200bp)의 서열을 검색한다. 도 70a를 참조한다.

[0896] 2. 표적 PAM의 위치를 찾아낸다. 목적하는 편집 위치에 근접한 PAM을 확인한다. PAM은 목적하는 편집 위치에 근접한 DNA의 어느 한 가닥 상에서 확인될 수 있다. 편집 위치에 가까운 PAM이 바람직하지만 (즉, 여기서 닉 부위는 편집 위치로부터 30 nt 미만이거나, 또는 편집 위치로부터 닉 부위까지 29 nt, 28 nt, 27 nt, 26 nt, 25 nt, 24 nt, 23 nt, 22 nt, 21 nt, 20 nt, 19 nt, 18 nt, 17 nt, 16 nt, 15 nt, 14 nt, 13 nt, 12 nt, 11 nt, 10 nt, 9 nt, 8 nt, 7 nt, 6 nt, 5 nt, 4 nt, 3 nt, 또는 2 nt 미만임), 닉을 편집 위치로부터 ≥ 30 nt에 배치하는 프로토스페이서 및 PAM을 사용하여 편집을 설치하는 것이 가능하다. 도 70b를 참조한다.

[0897] 3. 닉 부위를 위치시킨다. 고려되는 각각의 PAM에 대해, 상응하는 닉 부위 및 어떤 가닥 상의 것인지 확인한다. Sp Cas9 H840A 닉카제의 경우, 절단은 PAM-함유 가닥에서 NGG PAM에 대해 5'의 세번째 및 네번째 염기 사이에서 일어난다. 모든 편집된 뉴클레오티드는 닉 부위의 3'에 존재해야 하고, 따라서 적절한 PAM은 PAM-함유 가닥 상의 표적 편집에 대해 5'에 닉을 배치해야 한다. 하기 제시된 예에서, 2종의 가능한 PAM이 존재한다. 단순성을 위해, 나머지 단계는 PAM 1만을 사용한 PEgRNA의 설계를 입증할 것이다. 도 70c를 참조한다.

[0898] 4. 스페이서 서열을 설계한다. Sp Cas9의 프로토스페이서는 PAM-함유 가닥 상의 NGG PAM에 대해 5'의 20개 뉴클레오티드에 상응한다. 효율적인 Pol III 전사 개시는 G가 제1 전사된 뉴클레오티드일 것을 요구한다. 프로토스페이서의 제1 뉴클레오티드가 G인 경우에, PEgRNA에 대한 스페이서 서열은 단순히 프로토스페이서 서열이다. 프로토스페이서의 제1 뉴클레오티드가 G가 아닌 경우에, PEgRNA의 스페이서 서열은 G에 이어서 프로토스페이서 서열이 후속한다. 도 70d를 참조한다.

[0899] 5. 프라이머 결합 부위 (PBS)를 설계한다. 출발 대립유전자 서열을 사용하여, PAM-함유 가닥 상의 DNA 프라이머를 확인한다. DNA 프라이머의 3' 단부는 닉 부위의 바로 상류에 있는 뉴클레오티드이다 (즉, Sp Cas9의 경우 NGG PAM에 대해 5'의 네번째 염기). PE2 및 PE3과 함께 사용하기 위한 일반적 설계 원리로서, DNA 프라이머에 대해 12 내지 13개 뉴클레오티드의 상보성을 함유하는 PEgRNA 프라이머 결합 부위 (PBS)가 ~40-60% GC 함량을 함유하는 서열에 대해 사용될 수 있다. 낮은 GC 함량을 갖는 서열의 경우에, 보다 긴 (14- 내지 15-nt) PBS가 시험되어야 한다. 보다 높은 GC 함량을 갖는 서열의 경우에, 보다 짧은 (8- 내지 11-nt) PBS가 시험되어야 한다. 최적의 PBS 서열은 GC 함량에 상관없이 실험적으로 결정되어야 한다. 길이-p PBS 서열을 설계하기 위해, 출발 대립유전자 서열을 사용하여 PAM-함유 가닥에서 닉 부위의 5'의 첫번째 p 뉴클레오티드의 역 상보체를 취한다. 도 70e를 참조한다.

[0900] 6. RT 주형 (또는 DNA 합성 주형)을 설계한다. RT 주형 (또는 폴리머라제가 리버스 트랜스크립타제가 아닌 경우 DNA 합성 주형)은 설계된 편집 및 편집에 인접한 서열에 대한 상동성을 코딩한다. 한 실시양태에서, 이들 영역은 도 3d 및 도 3e의 DNA 합성 주형에 상응하며, 여기서 DNA 합성 주형은 "편집 주형" 및 "상동성 아암"을 포함한다. 최적의 RT 주형 길이는 표적 부위에 기초하여 달라진다. 짧은-범위 편집 (위치 +1 내지 +6)의 경우, 짧은 (9 내지 12 nt), 중간 (13 내지 16 nt), 및 긴 (17 내지 20 nt) RT 주형을 시험하는 것이 권장된다. 긴-범위 편집 (위치 +7 및 그 초과)의 경우, 충분한 3' DNA 플랩 상동성을 허용하기 위해 편집 위치를 지나 적어도 5개의 nt (바람직하게는 10개 이상의 nt) 연장된 RT 주형을 사용하는 것이 권장된다. 긴-범위 편집의 경우, 여러 RT 주형이 기능적 설계를 확인하기 위해 스크리닝되어야 한다. 보다 큰 삽입 및 결실 (≥5 nt)을 위해, RT 주형 내로의 보다 큰 3' 상동성 (~20 nt 이상)의 혼입이 권장된다. 편집 효율은 전형적으로 RT 주형이 역전사된 DNA 생성물에서 마지막 뉴클레오티드로서 G의 합성 (PEgRNA의 RT 주형에서의 C에 상응함)을 코딩하는 경우에 손상된다. 많은 RT 주형이 효율적인 프라임 편집을 지지하기 때문에, RT 주형을 설계할 때 최종 합성된 뉴클레오티드로서 G를 회피하는 것이 권장된다. 길이-r RT 주형 서열을 설계하기 위해, 목적하는 대립유전자 서열을 사용하고, 원래 PAM을 함유하는 가닥에서 닉 부위의 3'의 첫번째 r 뉴클레오티드의 역 상보체를 취한다. SNP 편집과 비교하여, 동일한 길이의 RT 주형을 사용한 삽입 또는 결실 편집은 동일한 상동성을 함유하지 않을 것임을 주목한다. 도 70f를 참조한다.

[0901] 7. 전장 PEgRNA 서열을 조립한다. PEgRNA 성분을 하기 순서 (5'에서 3')로: 스페이서, 스캐폴드, RT 주형 및 PBS로 연결한다. 도 70g를 참조한다.

[0902] 8. PE3에 대한 닉킹-sgRNA를 설계한다. 편집의 상류 및 하류의 비-편집된 가닥 상에서 PAM을 확인한다. 최적의 닉킹 위치는 고도로 유전자좌-의존적이고, 실험적으로 결정되어야 한다. 일반적으로, PEgRNA-유도된 닉에 걸친 위치에 대해 5'에 40 내지 90개의 뉴클레오티드가 배치된 닉은 보다 높은 편집 수율 및 보다 적은 indel로 이어진다. 닉킹 sgRNA는 출발 대립유전자에서 20-nt 프로토스페이서와 매칭되는 스페이서 서열을 가지며, 여기서 프로토스페이서가 G로 시작되지 않는 경우 5'-G가 부가된다. 도 70h를 참조한다.

[0903] 9. PE3b 닉킹-sgRNA를 설계한다. PAM이 상보적 가닥에 존재하고 그의 상응하는 프로토스페이서가 편집을 위해

표적화된 서열과 중첩되는 경우에, 이러한 편집은 PE3b 시스템에 대한 후보일 수 있다. PE3b 시스템에서, 닉킹-sgRNA의 스페이서 서열은 출발 대립유전자가 아닌 목적하는 편집된 대립유전자의 서열과 매칭된다. PE3b 시스템은 편집된 뉴클레오티드(들)가 닉킹-sgRNA 프로토스페이서의 시드 영역 (PAM에 인접한 ~10개의 nt) 내에 속할 때 효율적으로 작동한다. 이는 편집된 가닥의 설치 후까지 상보적 가닥의 닉킹을 방지하여, 표적 DNA에의 결합에 대한 PEgRNA와 sgRNA 사이의 경쟁을 방지한다. PE3b는 또한 두 가닥 상의 동시 닉의 생성을 피하고, 따라서 높은 편집 효율을 유지하면서 indel 형성을 유의하게 감소시킨다. PE3b sgRNA는 필요한 경우에 5' G를 부가하여 목적하는 대립유전자에서 20-nt 프로토스페이서와 매칭되는 스페이서 서열을 가져야 한다. 도 70i를 참조한다.

[0904] 적합한 PEgRNA 및 제2-부위 닉킹 sgRNA를 설계하기 위한 상기 단계별 과정은 어떠한 방식으로도 제한하는 것으로 의도되지 않는다. 본 개시내용은 관련 기술분야의 통상의 기술자에 의해 그로부터 유도될 수 있는 상기 기재된 단계별 과정의 변경을 고려한다.

[0905] [7] 다중-플랩 프라임 편집을 이용하는 적용

[0906] 이중-플랩 및 사중-플랩 프라임 편집 (즉, 다중-플랩 프라임 편집)은 많은 잠재적 적용, 예컨대 펩티드 태그, RNA 태그, 이뮤노에피토프, 이량체화 도메인 및 레콤비나제 표적 부위의 설치를 갖는다. 이중-플랩 프라임 편집의 하나의 이러한 적용은 게놈 내 사용자-명시된 위치에서의 레콤비나제 또는 인테그라제 서열의 설치이다. 도 93은 Bxb1 레콤비나제 attB (38 bp) 및 attP (50 bp) 부위를 인간 게놈의 표적화된 영역 (HEK293T 부위 3 또는 HEK3) 내로 설치하고, 동시에 2개의 닉 부위 사이에 90 bp의 개재 서열을 결실시키는 것을 예시한다. 3' 플랩 사이의 다양한 정도의 상보성이 성공적인 편집을 허용하며, 다만 보다 긴 상보성 서열이 보다 유리한 목적하는 편집 대 indel 비를 생성한다. 다른 이중-플랩 프라임 편집 적용은 펩티드 또는 단백질 서열로의 유전자의 내인성 태그부착, 또는 돌연변이가 엑손 서열 내에 속하는 다수의 변이체를 치환시키는 잠재력을 갖는 새로운 DNA 서열로의 엑손의 대체를 포함한다.

[0907] 이중-플랩 프라임 편집을 사용하여 인간 게놈 내 표적화된 위치에 1 또는 2개의 레콤비나제 부위를 도입할 수 있다. 단일 레콤비나제 부위가 삽입되는 경우에, 이들은 게놈 레콤비나제 부위와 외인적으로 공급된 DNA, 예컨대 플라스미드 내의 제2 레콤비나제 부위 사이의 레콤비나제-매개된 반응을 위한 랜딩 부위로서 사용될 수 있다. 이는 DNA 카고의 표적화된 통합을 가능하게 한다. 2개의 레콤비나제 부위가 DNA의 인접한 영역에 삽입되는 경우에, 레콤비나제 부위의 배향에 따라, 이들은 레콤비나제-매개된 절제 또는 개재 서열의 역위를 위해, 또는 카고 통합을 위한 외인성 DNA와의 레콤비나제-매개된 카세트 교환을 위해 사용될 수 있다. 상이한 염색체 상의 상용성 레콤비나제 부위의 통합은 표적화된 및 방향적인 염색체 전위를 가능하게 한다. 이중-플랩 프라임 편집을 사용하여 인간 게놈 내 다수의 유전자좌에 레콤비나제 부위를 효율적으로 도입할 수 있다 (도 94). 따라서, 이중-플랩 프라임 편집에 의한 레콤비나제 부위 통합과 DNA 레콤비나제 효소의 쌍형성은 많은 유형의 SV 편집 및 DNA 카고의 표적 통합을 달성하기 위한 강력한 접근법을 나타낸다.

[0908] 본원에 기재된 다중-플랩 프라임 편집제 (예를 들어, 도 90에 도시된 실시양태)는 새로운 DNA 서열의 정확한 삽입, 내인성 게놈 DNA 서열의 정확한 결실, 또는 내인성 게놈 DNA 서열의 새로운 DNA 서열로의 대체에 사용될 수 있다.

[0909] 예를 들어 및 본원에 예시되고 논의된 바와 같이, 이중 프라임 편집은 (a) 뉴클레오티드 서열에 돌연변이-교정 변화를 설치하고, (b) 단백질 및 RNA 태그를 설치하고, (c) 관심 단백질 상에 이뮤노에피토프를 설치하고, (d) 단백질에 유도성 이량체화 도메인을 설치하고, (e) 특이적 유전자 변화를 지시하기 위해 레콤비나제 표적 부위를 설치하는 데 사용될 수 있다. 일반적으로 관심 표적 부위에서 뉴클레오티드 서열을 삽입, 변화 또는 결실시키는 이들 방법에 추가로, 이중 프라임 편집제는 또한 고도로 프로그램가능한 라이브러리를 구축하는 데, 뿐만 아니라 세포 데이터 기록 및 계통 추적 연구를 수행하는 데 사용될 수 있다.

[0910] 이중 프라임 편집의 이들 구체적 예시적인 용도는 결코 제한적인 것으로 의도되지 않는다. 본 출원은, 일반적으로, 뉴클레오티드 서열, 예를 들어 게놈 DNA 내 표적 부위에서의 1개 이상의 핵염기의 설치, 결실 및/또는 대체의 일부 형태를 수반하는 이중 프라임 편집을 위한 임의의 용도를 고려한다.

[0911] 이중 프라임 편집을 위한 임의의 예시된 용도를 위해, PE1, PE2, PE3 및 PE3b 또는 PE-짧은 형을 포함한 본원에 개시된 임의의 프라임 편집제를 사용할 수 있다.

[0912] A. 프라임 편집 대 다중-플랩 프라임 편집

[0913] 전형적 프라임 편집

[0914] 다양한 실시양태에서, 프라임 편집 (또는 "프라임 편집")은 표적 DNA 분자 (뉴클레오티드 서열에서 변화가 도입되는 것이 목적됨)를 연장된 가이드 RNA와 복합체화된 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)과 접촉시킴으로써 작동한다. 도 1g를 참조하면, 연장된 가이드 RNA는 가이드 RNA의 3' 또는 5' 단부에, 또는 가이드 RNA 내의 분자내 위치에 연장부를 포함하고, 목적하는 뉴클레오티드 변화 (예를 들어, 단일 뉴클레오티드 변화, 삽입, 또는 결실)를 코딩한다. 단계 (a)에서, napDNAbp/연장된 gRNA 복합체는 DNA 분자와 접촉하고, 연장된 gRNA는 napDNAbp가 표적 유전자좌에 결합하도록 가이드한다. 단계 (b)에서, 표적 유전자좌의 DNA의 가닥 중 하나에 닉을 도입하여 (예를 들어, 뉴클레아제 또는 화학적 작용제에 의함), 표적 유전자좌의 가닥 중 하나에 이용가능한 3' 단부를 생성한다. 특정 실시양태에서, 닉은 R-루프 가닥, 즉 가이드 RNA 서열에 혼성화되지 않은 가닥, 즉 "비-표적 가닥"에 상응하는 DNA의 가닥에서 생성된다. 그러나, 닉은 가닥 중 어느 하나에 도입될 수 있다. 즉, 닉은 R-루프 "표적 가닥" (즉, 연장된 gRNA의 프로토스페이서 서열에 혼성화된 가닥) 또는 "비-표적 가닥" (즉, R-루프의 단일 가닥 부분을 형성하고 표적 가닥에 상보적인 가닥) 내로 도입될 수 있다. 단계 (c)에서, DNA 가닥의 3' 단부 (닉에 의해 형성됨)는 역전사를 프라이밍하기 위해 (즉, "표적-프라이밍된 RT") 가이드 RNA의 연장된 부분과 상호작용한다. 특정 실시양태에서, 3' 단부 DNA 가닥은 가이드 RNA의 연장된 부분 상의 특이적 RT 프라이밍 서열, 즉 "리버스 트랜스크립타제 프라이밍 서열"에 혼성화한다. 단계 (d)에서, 프라이밍된 부위의 3' 단부로부터 연장된 가이드 RNA의 5' 단부쪽으로 DNA의 단일 가닥을 합성하는 리버스 트랜스크립타제가 도입된다 (napDNAbp와의 융합 단백질로서 또는 트랜스로). 이는 목적하는 뉴클레오티드 변화 (예를 들어, 단일 염기 변화, 삽입, 또는 결실, 또는 그의 조합)를 포함하고, 그 외에는 닉 부위의 또는 그와 인접한 내인성 DNA와 상동성인 단일 가닥 DNA 플랩을 형성한다. 단계 (e)에서, napDNAbp 및 가이드 RNA가 방출된다. 단계 (f) 및 (g)는 목적하는 뉴클레오티드 변화가 표적 유전자좌 내로 혼입되도록 하는 단일 가닥 DNA 플랩의 분해에 관한 것이다. 이러한 과정은 일단 3' 단일 가닥 DNA 플랩이 내인성 DNA 서열에 침입하여 혼성화되면 형성되는 상응하는 5' 내인성 DNA 플랩을 제거함으로써 (예를 들어, 트랜스로, 프라임 편집제와의 융합체로서 제공되거나 또는 내인성으로 제공되는 FEN1 또는 유사한 효소에 의함) 목적하는 생성물 형성쪽으로 구동될 수 있다. 이론에 얽매이지는 않지만, 세포 내인성 DNA 복구 및 복제 과정은 미스매칭된 DNA를 분해하여 뉴클레오티드 변화(들)를 혼입시켜 목적하는 변경된 생성물을 형성한다. 과정은 또한 도 1g에 예시된 바와 같은 "제2 가닥 닉킹", 또는 도 1i에 예시되고 본원에 논의된 바와 같은 "일시적 제2 가닥 닉킹"에 의해 생성물 형성쪽으로 구동될 수 있다.

[0915] 프라임 편집의 과정은 하기 유전자 변화: 전환, 전이, 결실, 및 삽입 중 적어도 하나 이상을 도입할 수 있다. 또한, 프라임 편집은 특정한 적용을 위해 구현될 수 있다. 예를 들어, 본원에 예시되고 논의된 바와 같이, 프라임 편집은 (a) 뉴클레오티드 서열에 돌연변이-교정 변화를 설치하고, (b) 단백질 및 RNA 태그를 설치하고, (c) 관심 단백질 상에 이뮤노에피토프를 설치하고, (d) 단백질에 유도성 이량체화 도메인을 설치하고, (e) 생체분자의 활성을 변경하기 위해 서열을 설치하거나 제거하고, (f) 특이적 유전자 변화를 지시하기 위해 레콤비나제 표적 부위를 설치하고, (g) 오류-유발 RT를 사용함으로써 표적 서열을 돌연변이유발하는 데 사용될 수 있다. 일반적으로 관심 표적 부위에서 뉴클레오티드 서열을 삽입, 변화, 또는 결실시키는 이들 방법에 더하여, 프라임 편집제는 또한 고도로 프로그램가능한 라이브러리를 구축하는데, 뿐만 아니라 세포 데이터 기록 및 계통 추적 연구를 수행하는 데 사용될 수 있다. 본 발명자들은 또한 프라임 편집의 효능을 개선시키는 것을 목표로 PEgRNA의 추가의 설계 특색을 고려하였다. 또한 추가로, 본 발명자들은 벡터 전달 시스템을 사용하고, 인테인 도메인을 사용하여 napDNAbp를 분할하는 것을 수반하는 프라임 편집제를 성공적으로 전달하는 방법을 구상하였다.

[0916] 용어 "프라임 편집 시스템" 또는 "프라임 편집제 (PE)"는 napDNAbp, 리버스 트랜스크립타제, 융합 단백질 (예를 들어, napDNAbp 및 리버스 트랜스크립타제를 포함함), 연장된 가이드 RNA, 및 융합 단백질 및 연장된 가이드 RNA를 포함하는 복합체, 뿐만 아니라 프라임 편집 과정을 편집된 생성물 형성쪽으로 구동하는 것을 돕기 위한 보조 요소, 예컨대 제2 가닥 닉킹 성분 및 5' 내인성 DNA 플랩 제거 엔도뉴클레아제 (예를 들어, FEN1)를 포함하나 이에 제한되지는 않는, 본원에 기재된 표적-프라이밍된 역전사 (TPRT)를 사용하는 게놈 편집 방법에 수반되는 조성물을 지칭한다.

[0917] 또 다른 실시양태에서, 도 3f의 개략도는 전형적인 PEgRNA와 이중 가닥 DNA의 표적 부위의 상호작용 및 관심 유전자 변화를 함유하는 3' 단일 가닥 DNA 플랩의 동반 생산을 도시한다. 이중 가닥 DNA는 상단 가닥이 3'에서 5' 배향으로, 하단 가닥이 5'에서 3' 방향으로 제시된다. 상단 가닥은 "프로토스페이서" 및 PAM 서열을 포함하고, "표적 가닥"으로 지칭된다. 상보적 하단 가닥은 "비-표적 가닥"으로 지칭된다. 제시되지 않았지만, 도시

된 PEgRNA는 Cas9 또는 등가물과 복합체화될 것이다. 개략도에 제시된 바와 같이, PEgRNA의 스페이서는 PAM 서열의 바로 하류에 위치하는 프로토스페이서로 지칭되는 표적 가닥 상의 상보적 영역에 어닐링되며, 이는 대략 20개 뉴클레오티드 길이이다. 이러한 상호작용은 스페이서 RNA와 프로토스페이서 DNA 사이에 DNA/RNA 하이브리드로서 형성되고, 프로토스페이서 대향 영역에서 R 루프의 형성을 유도한다. 본원의 다른 곳에 교시된 바와 같이, Cas9 단백질 (제시되지 않음)은 이어서, 제시된 바와 같이 비-표적 가닥에서 닉을 유도한다. 이어서, 이는 3' ssDNA 플랩 영역의 형성으로 이어지며, 이는 *z*에 따르면 프라이머 결합 부위에서 PEgRNA의 3' 단부와 상호작용한다. ssDNA 플랩 (즉, 리버스 트랜스크립타제 프라이머 서열)의 3' 단부는 PEgRNA 상의 프라이머 결합 부위 (A)에 어닐링하여, 리버스 트랜스크립타제를 프라이밍한다. 이어서, 리버스 트랜스크립타제 (예를 들어, Cas9 구축물에 부착된, 융합 단백질로서 트랜스로 제공되거나 또는 시스로 제공됨)는 편집 주형 (B) 및 상동성 아암 (C)에 의해 코딩된 DNA의 단일 가닥을 중합한다. 중합은 연장 아암의 5' 단부쪽으로 계속된다. ssDNA의 중합된 가닥은, 다른 곳에 기재된 바와 같이 (예를 들어, 도 1g에 제시된 바와 같이), 내인성 DNA를 침입하고, 상응하는 내인성 가닥 (이는 내인성 DNA의 5' DNA 플랩으로서 제거됨)을 대체하고, 자연 발생 DNA 복구/복제 라운드를 통해 목적하는 뉴클레오티드 편집 (단일 뉴클레오티드 염기 쌍 변화, 결실, 삽입 (전체 유전자 포함))을 설치하는 ssDNA 3' 단부 플랩을 형성한다.

[0918] 프라임 편집의 이러한 적용은 실시예 1에 추가로 기재될 수 있다.

[0919] 이중 프라임 편집

[0920] 본 명세서는 두 DNA 가닥을 동시에 편집함으로써 플랩 평형 및 비-편집된 상보적 게놈 DNA 가닥 내로의 편집의 후속 혼입과 연관된 도전과제를 다루는 이중 프라임 편집 시스템 (또는 이중-플랩 프라임 편집 시스템)을 기재한다. 이중-플랩 프라임 편집 시스템에서, 2개의 pegRNA는 게놈 부위의 대향 가닥을 표적화하고 편집된 DNA 서열을 함유하는 2개의 상보적 3' 플랩의 합성을 지시하는 데 사용된다 (도 91). 전형적 프라임 편집과는 달리, 상보적인 편집된 가닥이 대신 혼성화에 이용가능하기 때문에, 편집된 DNA 가닥의 쌍 (3' 플랩)이 내인성 게놈 DNA에서 5' 플랩과 직접 경쟁하는 것이 요구되지 않는다. 듀플렉스의 두 가닥이 편집된 DNA로서 합성되기 때문에, 이중-플랩 프라임 편집 시스템은 전형적 프라임 편집에 의해 요구되는 비-편집된 상보적 DNA 가닥의 대체에 대한 필요를 제거한다. 대신, 세포 DNA 복구 기구는 단지 쌍형성된 5' 플랩 (원래 게놈 DNA)만을 절제하고, 쌍형성된 3' 플랩 (편집된 DNA)을 유전자좌 내로 라이게이션시킬 필요가 있다. 따라서, 또한 새로 합성된 DNA 가닥에 게놈 DNA에 상동인 서열을 포함할 필요가 없으므로, 새로운 가닥의 선택적 혼성화를 허용하고 최소 게놈 상동성을 함유하는 편집을 용이하게 한다. DNA의 두 가닥을 절단하는 프라임 편집제의 뉴클레아제-활성 버전은 또한 원래 DNA 서열의 제거를 가속화하는 데 사용될 수 있다. 따라서, 특정 실시양태에서, 이중 플랩 프라임 편집 시스템은 편집될 표적 부위에서 내인성 DNA 서열과의 상동성을 공유하는 새로 합성된 DNA 가닥의 쌍 (예를 들어 3' 플랩)을 수반한다. 일부 실시양태에서, 이중 플랩 프라임 편집 시스템은 새로 합성된 DNA 가닥의 쌍을 수반하며, 여기서 새로 합성된 DNA 가닥 중 적어도 하나는 편집될 표적 부위에서 내인성 DNA 서열과의 상동성을 공유하지 않는다. 일부 실시양태에서, 새로 합성된 DNA 가닥 중 하나는 표적 부위에서 내인성 DNA 서열과의 상동성을 포함하고, 다른 새로 합성된 DNA 가닥과의 상동성은 포함하지 않는다. 일부 실시양태에서, 각각의 새로 합성된 DNA 가닥은 표적 부위에서 내인성 DNA 서열과의 상동성을 포함하고, 다른 새로 합성된 DNA 가닥과의 상동성은 포함하지 않는다. 예를 들어, 이중-플랩 pegRNA 중 하나에 의해 코딩된 새로 합성된 3' 플랩은 다른 이중-플랩 pegRNA의 프로토스페이서 서열에 대한 상보성 영역을 포함할 수 있다. 따라서, 일부 실시양태에서, 각각 다른 pegRNA의 스페이서 서열에 대한 상보성을 갖는 이중-플랩 pegRNA의 쌍은 이중-플랩 pegRNA의 쌍의 프로토스페이서 서열 사이에 위치한 내인성 DNA 서열의 결실을 발생시킬 수 있다. 일부 실시양태에서, 이중 플랩 프라임 편집 시스템은 새로 합성된 DNA 가닥의 쌍을 수반하며, 여기서 새로 합성된 DNA 가닥 중 어느 것도 편집될 표적 부위에서 내인성 DNA 서열과의 상동성을 공유하지 않는다. 오히려, 2개의 새로 합성된 DNA 가닥 (예를 들어 3' 플랩)은 각각 서로에 대한 상보성 영역을 포함하고, 이러한 상보성에 의해 듀플렉스를 형성할 수 있다. 이어서, 듀플렉스에서 편집될 내인성 DNA 서열 표적 부위와 비교하여 목적하는 편집된 부분이 표적 부위에 혼입될 수 있다. 전형적 프라임 편집과 같이, 이중 프라임 편집은 핵산 프로그램가능한 DNA 결합 단백질 ("napDNAbp") (이는 폴리머라제와 회합하여 (즉, 융합 단백질의 형태로 또는 달리 napDNAbp와 트랜스로 제공됨) 작동함)을 사용하여 특정된 DNA 부위 내로 새로운 유전자 정보를 직접 기록하는 다재다능한 정확한 게놈 편집 방법이며, 여기서 프라임 편집 시스템은 표적 부위를 특정하고 또한 가이드 RNA 상에 (예를 들어, 5' 또는 3' 단부에 또는 가이드 RNA의 내부 부분에) 조작된 연장부 (DNA 또는 RNA)에 의해 대체 DNA 가닥의 형태로 목적하는 편집의 합성의 주형이 되는 프라임 편집 (PE) 가이드 RNA ("PEgRNA")로 프로그램화된다. 목적하는 편집 (예를 들어, 단일 핵염기 치환)을 함유하는 대체 가닥은 (목적하는 편집을 포함하는 것을 제외하고는) 편집될 표적 부위의 내인성 가닥과 동일한 서열을 공유한다. DNA 복구 및/또는 복제 기구를 통해, 표적 부위의 내인성 가닥

은 목적하는 편집을 함유하는 새로 합성된 대체 가닥에 의해 대체된다. 일부 경우에, 프라임 편집은 "검색-및-대체" 게놈 편집 기술로서 생각될 수 있으며, 이는 프라임 편집제가, 본원에 기재된 바와 같이, 편집될 목적하는 표적 부위를 검색하고 위치를 찾아낼 뿐만 아니라, 동시에, 상응하는 표적 부위 내인성 DNA 가닥 대신 설치된 목적하는 편집을 함유하는 대체 가닥을 코딩하기 때문이다.

[0921] B. 펩티드 태그부착을 위한 이중 프라임 편집의 사용

[0922] 또 다른 측면에서, 개시내용은 프라임 편집 (예컨대 도 90에 제시된 실시양태)을 사용하여 단백질 상에 1개 이상의 펩티드 태그를 유전자 그라프팅하기 위한, 본원에 기재된 이중 프라임 편집제를 사용하는 방법을 제공한다. 보다 특히, 개시내용은 단백질을 코딩하는 표적 뉴클레오티드 서열을, 1개 이상의 펩티드 태그를 코딩하는 제2 뉴클레오티드 서열을 그 안에 삽입하여 단백질 태그에 융합된 단백질을 포함하는 융합 단백질을 코딩하는 재조합 뉴클레오티드 서열을 생성하도록 구성된 프라임 편집제와 접촉시키는 것을 포함하는, 단백질 상에 1개 이상의 펩티드 태그를 유전자 설치하는 방법을 제공한다.

[0923] 다른 실시양태에서, 개시내용은 단백질을 코딩하는 표적 뉴클레오티드 서열을, 1개 이상의 펩티드 태그를 코딩하는 제2 뉴클레오티드 서열을 그 안에 삽입하여 단백질 태그에 융합된 단백질을 포함하는 융합 단백질을 코딩하는 재조합 뉴클레오티드 서열을 생성하도록 구성된 프라임 편집제와 접촉시키는 것을 포함하는, 관심 펩티드 및 1개 이상의 펩티드 태그를 포함하는 융합 단백질을 제조하는 방법을 제공한다.

[0924] 다양한 실시양태에서, 표적 뉴클레오티드 서열은 게놈 DNA에서의 관심 특이적 유전자이다. 관심 유전자는 관심 단백질 (예를 들어, 수용체, 효소, 치료 단백질, 막 단백질, 수송 단백질, 신호 전달 단백질, 또는 면역학적 단백질 등)을 코딩할 수 있다. 관심 유전자는 또한 메신저 RNA (mRNA), 전달 RNA (tRNA), 리보솜 RNA (rRNA), 소형 핵 RNA (snRNA), 안티센스 RNA, 가이드 RNA, 마이크로RNA (miRNA), 소형 간섭 RNA (siRNA), 및 세포-유리 RNA (cfRNA)를 포함하나 이에 제한되지는 않는 RNA 분자를 코딩할 수 있다.

[0925] 펩티드 태그는 분리, 정제, 가시화, 가용화, 또는 검출과 같은 목적을 위해 단백질에 1종 이상의 기능을 부여하는 임의의 펩티드 태그 또는 그의 변이체일 수 있다. 펩티드 태그는 "친화성 태그" (단백질 정제를 용이하게 하기 위함), "가용화 태그" (단백질의 적절한 폴딩을 보조하기 위함), "크로마토그래피 태그" (단백질의 크로마토그래피 특성을 변경하기 위함), "에피토프 태그" (고친화도 항체에 결합하기 위함), 및 "형광 태그" (세포 또는 시험관내에서 단백질의 가시화를 용이하게 하기 위함)를 포함할 수 있다. 펩티드 태그의 예는 하기 태그를 포함하나 이에 제한되지는 않는다:

명칭	아미노산 서열	SEQ ID NO:
아비태그™	GLNDIFEAQKIEWHE	SEQ ID NO: 245
C-태그	EPEA	SEQ ID NO: 246
칼모듈린-태그	KRRWKKNFIAVSAANRFKKISSSGAL	SEQ ID NO: 247
폴리글루타메이트 태그	EEEEEE	SEQ ID NO: 248
E-태그	GAPVPYPDPLEPR	SEQ ID NO: 249
FLAG-태그	DYKDDDDK	SEQ ID NO: 250
HA-태그	YPYDVPDYA	SEQ ID NO: 251
HIS-태그	H (HIS_1) HH (HIS_2) HHH (HIS_3) HHHH (HIS_4) HHHHH (HIS_5) HHHHHH (HIS_6) HHHHHHH (HIS_7) HHHHHHHH (HIS_8) HHHHHHHHH (HIS_9) HHHHHHHHHH (HIS_{10}) HHHHHHHHHH…H…(HIS_N, WHEREIN N = 1-25)	SEQ ID NO: 252 SEQ ID NO: 253 SEQ ID NO: 254 SEQ ID NO: 255 SEQ ID NO: 256 SEQ ID NO: 257 SEQ ID NO: 258 SEQ ID NO: 259 SEQ ID NO: 260 SEQ ID NO: 261 SEQ ID NO: 262
MYC-태그	EQKLISEEDL	SEQ ID NO: 263
NE-태그	TKENPRSNQEESYDDNES	SEQ ID NO: 264
RHO1D4-태그	TETSQVAPA	SEQ ID NO: 265
S-태그	KETAAAKFERQHMDS	SEQ ID NO: 266
SBP-태그	MDEKTTGWRGGHVVEGLAGELEQLR ARLEHHPQGQREP	SEQ ID NO: 267
소프태그-1	SLAELLNAGLGGS	SEQ ID NO: 268
소프태그-2	TQDPSRVG	SEQ ID NO: 269
스폿-태그	PDRVRAVSHWSS	SEQ ID NO: 270
스트렙-태그	WSHPQFEK	SEQ ID NO: 271
TC 태그	CCPGCC	SEQ ID NO: 272
TY 태그	EVHTNQDPLD	SEQ ID NO: 273
V5 태그	GKPIPNPLLGLDST	SEQ ID NO: 274
VSV-태그	YTDIEMNRLGK	SEQ ID NO: 275
엑스프레스 태그	DLYDDDDK	SEQ ID NO: 276

[0926]

[0927] 퓁티드 태그는 또한 하기 친화성 태그 (단백질의 분리 및/또는 정제를 위함)일 수 있다 (본원에 참조로 포함된 문헌 [Kimple et al., "Overview of Affinity Tags for Protein Purification," Curr Protoc Protein Sci, 2013, 73:Unit-9.9]의 표 9.9.1에 기재된 바와 같음).

명칭	아미노산 서열	
AU1 에피토프	DTYRYI	SEQ ID NO: 278
AU5 에피토프	TDFYLK	SEQ ID NO: 279
박테리오파지 T7 에피토프 (T7-태그)	MASMTGGQQMG	SEQ ID NO: 280
블루텅 바이러스 태그 (B-태그)	QYPALT	SEQ ID NO: 281
E2 에피토프	SSTSSDFRDR	SEQ ID NO: 282
히스티딘 친화성 태그 (HAT)	KDHLIHNVHKEFHAHAHNK	SEQ ID NO: 283
HSV 에피토프	QPELAPED	SEQ ID NO: 284
폴리아르기닌 (ARG-태그)	RRRRR	SEQ ID NO: 285
폴리아스파르테이트 (ASP-태그)	CCCC	SEQ ID NO: 286
폴리페닐알라닌 (PHE-태그)	FFFFFFFFFF	SEQ ID NO: 287
S1-태그	NANNPDWDF	SEQ ID NO: 288
S-태그	KETAAAKFERQHMDS	SEQ ID NO: 289
VSV-G	YTDIEMNRLGK	SEQ ID NO: 290

[0928]

[0929] 특정한 실시양태에서, 퓁티드 태그는 His^6 태그, FLAG-태그, V5-태그, GCN4-태그, HA-태그, Myc-태그, FIAsH/ReAsH-태그, 소르타제 기질, 파이-클램프를 포함할 수 있다.

[0930] 다양한 실시양태에서, 퓁티드 태그는 단백질 형광 표지, 면역침전, 이뮤노블롯팅, 면역조직화학, 단백질 동원, 유도성 단백질 데그론, 및 게놈-전반 스크리닝을 포함하는 적용에 사용될 수 있다.

[0931] 다양한 다른 실시양태에서, 펩티드 태그는 단백질 자기-스플라이싱 기능을 설치하기 위해 인테인 서열을 포함할 수 있다. 본원에 사용된 용어 "인테인"은 모든 생물 역으로부터의 유기체에서 발견되는 자가-프로세싱 폴리펩티드 도메인을 지칭한다. 인테인 (개재 단백질)은 단백질 스플라이싱으로 공지된 고유한 자가-프로세싱 사건을 수행하며, 여기서 이는 2개의 펩티드 결합의 절단을 통해 보다 큰 전구체 폴리펩티드로부터 그 자체를 절제하고, 그 과정에서 새로운 펩티드 결합의 형성을 통해 플랭킹 엑스테인 (외부 단백질) 서열을 라이게이션한다. 이러한 재배열은 인테인 유전자가 다른 단백질-코딩 유전자 내에 인 프레임으로 포매되어 발견되기 때문에 번역후에 (또는 가능하게는 공동-번역으로) 일어난다. 또한, 인테인-매개된 단백질 스플라이싱은 자발적이고; 이는 외부 인자 또는 에너지 공급원을 필요로 하지 않으며, 단지 인테인 도메인의 폴딩만을 필요로 한다. 이 과정은 또한 시스-단백질 스플라이싱으로도 공지되어 있으며, 이는 "분할 인테인"에 의한 트랜스-단백질 스플라이싱의 자연적 과정과 대조된다. 인테인은 엑스테인으로 공지된 플랭킹 단백질 서열의 동반 융합체에 의한 전구체 단백질로부터의 그 자체의 절제를 촉매하는 자기-스플라이싱 RNA 인트론의 단백질 등가물이다 (문헌 [Perler et al., Nucleic Acids Res. 22:1125-1127 (1994)] 참조) (문헌 [Perler et al., Curr. Opin. Chem. Biol. 1:292-299 (1997); Perler, F. B. Cell 92(1):1-4 (1998); Xu et al., EMBO J. 15(19):5146-5153 (1996)]에서 검토됨).

[0932] 단백질 스플라이싱 과정의 메카니즘은 매우 상세하게 연구되었고 (Chong, et al., J. Biol. Chem. 1996, 271, 22159-22168; Xu, M-Q & Perler, F. B. EMBO Journal, 1996, 15, 5146-5153), 보존된 아미노산은 인테인 및 엑스테인 스플라이싱 지점에서 발견되었다 (Xu, et al., EMBO Journal, 1994, 13 5517-522).

[0933] 인테인은 또한 2개의 개별적으로 전사되고 번역되는 유전자에 의해 코딩된 2개의 단편으로서 존재할 수 있다. 이들 소위 분할 인테인은 자기-회합하고, 단백질-스플라이싱 활성을 트랜스로 촉매한다. 분할 인테인은 다양한 시아노박테리아 및 고세균에서 확인되었지만 (Caspi et al, Mol Microbiol. 50: 1569-1577 (2003); Choi J. et al, J Mol Biol. 556: 1093-1106 (2006.); Dassa B. et al, Biochemistry. 46:322-330 (2007.); Liu X. and Yang J., J Biol Chem. 275:26315-26318 (2003); Wu H. et al., Proc Natl Acad Sci USA. £5:9226-9231 (1998.); 및 Zettler J. et al, FEBS Letters. 553:909-914 (2009)), 지금까지 진핵생물에서는 발견되지 않았다. 최근에, 환경 메타게놈 데이터의 생물정보학적 분석은 신규 게놈 배열을 갖는 26개의 상이한 유전자좌를 밝혀내었다. 각각의 유전자좌에서, 보존된 효소 코딩 영역은 분할 인테인에 의해 개재되고, 인테인 서브도메인을 코딩하는 섹션 사이에 자립형 엔도뉴클레아제 유전자가 삽입된다. 이들 중에서, 5개의 유전자좌가 완전히 조립되었다: DNA 헬리카제 (gp41-1, gp41-8); 이노신-5'-모노포스페이트 데히드로게나제 (IMPDH-1); 및 리보뉴클레오티드 리덕타제 촉매 서브유닛 (NrdA-2 및 NrdJ-1). 이러한 파괴된 유전자 조직은 주로 파지에 존재하는 것으로 보인다 (Dassa et al., Nucleic Acids Research. 57:2560-2573 (2009)).

[0934] 특정 실시양태에서, 본원에 기재된 프라임 편집제는 분할-인테인 태그를 2개의 상이한 단백질에 삽입하는 데 사용될 수 있으며, 이는 공동-발현되어 융합 단백질 형성 시 그의 세포내 라이게이션을 유발한다. 단백질 트랜스-스플라이싱에서, 하나의 전구체 단백질은 N-엑스테인 부분에 이어 N-인테인으로 이루어지고, 또 다른 전구체 단백질은 C-인테인 부분에 이어 C-엑스테인으로 이루어지고, 트랜스-스플라이싱 반응 (N- 및 C-인테인에 의해 함께 촉매됨)은 2개의 인테인 서열을 절제하고, 2개의 엑스테인 서열을 펩티드 결합으로 연결한다. 효소 반응인 단백질 트랜스-스플라이싱은 매우 낮은 (예를 들어, 마이크로몰) 농도의 단백질로 작용할 수 있고, 생리학적 조건 하에 수행될 수 있다.

[0935] 분할 인테인 Npu DnaE는 단백질 트랜스-스플라이싱 반응에 대해 보고된 최고 속도를 갖는 것으로 특징화되었다. 또한, Npu DnaE 단백질 스플라이싱 반응은 상이한 엑스테인 서열, 6 내지 37℃의 온도, 및 최대 6M의 우레아의 존재와 관련하여 강건하고 고수율인 것으로 간주된다 (Zettler J. et al., FEBS Letters. 553:909-914 (2009); Iwai I. et al., FEBS Letters 550: 1853-1858 (2006)). 예상된 바와 같이, 이들 인테인의 N-도메인에서 Cysl Ala 돌연변이가 도입되었을 때, 초기 N에서 S- 아실 이동이 차단되었고, 따라서 단백질 스플라이싱이 차단되었다. 불행하게도, C-말단 절단 반응도 또한 거의 완전히 억제되었다. N-말단 절단가능한 펩티드 결합에서의 아실 이동에 대한 C-말단 스플라이스 접합부에서의 아스파라긴 고리화의 의존성은 자연 분할 DnaE 인테인 대립유전자에 공통적인 고유한 특성인 것으로 보인다 (Zettler J. et al., FEBS Letters. 555:909-914 (2009)).

[0936] 분할 인테인에 의해 촉매되는 단백질 트랜스-스플라이싱은 단백질 라이게이션을 위한 전적으로 효소적인 방법을 제공한다. 분할-인테인은 본질적으로 각각 N-인테인 및 C-인테인으로 명명되는 2개의 조각으로 분할된 인접 인테인 (예를 들어 미니-인테인)이다. 분할 인테인의 N-인테인 및 C-인테인은 비-공유적으로 회합되어 활성 인테인을 형성하고, 인접 인테인과 본질적으로 동일한 방식으로 스플라이싱 반응을 촉매할 수 있다. 분할 인테인은 자연에서 발견되었고, 또한 실험실에서 조작되었다. 본원에 사용된 용어 "분할 인테인"은 N-말단 및 C-말단 서

열이 트랜스-스플라이싱 반응에 대해 기능적인 인테인으로 비-공유적으로 재회합 또는 재구성될 수 있는 별개의 분자가 되도록 N-말단 및 C-말단 아미노산 서열 사이에 1개 이상의 펩티드 결합 파괴가 존재하는 임의의 인테인을 지칭한다. 임의의 촉매 활성 인테인 또는 그의 단편은 본 발명의 방법에 사용하기 위한 분할 인테인을 유도하는 데 사용될 수 있다. 예를 들어, 한 측면에서 분할 인테인은 진핵 인테인으로부터 유래될 수 있다. 또 다른 측면에서, 분할 인테인은 박테리아 인테인으로부터 유래될 수 있다. 또 다른 측면에서, 분할 인테인은 고세균 인테인으로부터 유래될 수 있다. 바람직하게는, 이와 같이 유래된 분할 인테인은 트랜스-스플라이싱 반응을 촉매하는데 필수적인 아미노산 서열만을 보유할 것이다.

[0937] 분할 인테인은 미니-인테인의 구조에서 발견되는 ~12개의 보존된 베타-가닥 사이의 비구조화된 루프 또는 개재 아미노산 서열 내의 1개 이상의 분할 부위를 조작함으로써 인접 인테인으로부터 생성될 수 있다. 분할의 생성이 단백질 스플라이싱 활성이 상실되는데 충분한 정도로 인테인, 특히 구조화된 베타-가닥의 구조를 파괴하지 않을 것이라면, 베타-가닥 사이의 영역 내의 분할 부위의 위치에 일부 유연성이 존재할 수 있다.

[0938] 본원에 기재된 프라임 편집제는 펩티드 태그 (인테인 포함)를 관심 단백질의 C-말단 단부 내로 혼입시킬 수 있다. 다른 실시양태에서, 펩티드 태그 (인테인 포함)는 관심 단백질의 N-말단 단부 내로 혼입될 수 있다. 펩티드 태그는 또한 관심 단백질의 내부에 혼입될 수 있다. 본원에 기재된 프라임 편집제에 의해 생성된 생성물 융합 단백질은 하기 구조를 가질 수 있다:

[0939] [관심 단백질] - [펩티드 태그];

[0940] [펩티드 태그] - [관심 단백질]; 또는

[0941] [관심 단백질 - N-말단 영역] - [펩티드 태그] - [관심 단백질 - C-말단 영역].

[0942] 전반에 걸친 펩티드 태그부착에 사용하기 위한 가이드 RNA 설계의 원리는 펩티드 태그부착에 적용될 수 있다. 예를 들어, 한 실시양태에서, 펩티드 태그부착을 위한 PEgRNA 구조는 하기 구조를 가질 수 있다: 5'-[스페이서 서열]-[gRNA 코어 또는 스캐폴드]-[연장 아암]-3', 여기서 연장 아암은 5'에서 3' 방향으로 상동성 아암, 편집 주형 (펩티드 태그를 코딩하는 서열을 포함함), 및 프라이머 결합 부위를 포함한다. 이러한 구성은 도 3d 및 도 24에 도시되어 있다.

[0943] 또 다른 실시양태에서, 펩티드 태그부착을 위한 PEgRNA 구조는 하기 구조를 가질 수 있다: 5'-[연장 아암]-[스페이서 서열]-[gRNA 코어 또는 스캐폴드]-3', 여기서 연장 아암은 5'에서 3' 방향으로 상동성 아암, 편집 주형 (펩티드 태그를 코딩하는 서열을 포함함), 및 프라이머 결합 부위를 포함한다. 이 구성은 도 3e에 도시되어 있다.

[0944] 프라임 편집을 사용한 펩티드 태그부착의 실시양태는 도 25 및 26에 도시되고, 실시예 4에 기재된다.

[0945] C. RNA 태그부착을 위한 이중 프라임 편집의 사용

[0946] 이중 프라임 편집은 또한 RNA 태그부착을 통해 RNA 기능을 코딩하는 DNA의 서열을 조작, 변경, 및 달리 변형시키는데 사용될 수 있고, 이러한 방식으로 RNA의 구조 및 기능을 간접적으로 변형시키는 수단을 제공한다. 예를 들어, 이중 PE는 비-코딩 RNA 또는 mRNA를 태그부착하거나 또는 달리 조작하기 위해 RNA 수준에서 기능적인 모티프 (이하 RNA 모티프)를 삽입하는 데 사용될 수 있다. 이들 모티프는 유전자 발현을 증가시키거나, 유전자 발현을 감소시키거나, 스플라이싱을 변경하거나, 전사후 변형을 변화시키거나, RNA의 세포하 위치에 영향을 미치거나, RNA의 단리 또는 세포내 또는 세포외 위치의 결정을 가능하게 하거나 (예를 들어, 형광 RNA 압타머, 예컨대 스피니치, 스피니치2, 베이비 스피니치, 또는 브로콜리(Broccoli)를 사용함), 내인성 또는 외인성 단백질 또는 RNA 결합제를 동원하거나, sgRNA를 도입하거나, 또는 자기-절단 또는 RNAse에 의한 RNA의 프로세싱을 유도하는 역할을 할 수 있다 (추가의 세부사항에 대해서는 도 28b 및 실시예 6 참조).

[0947] RNA 수송, 발현 수준, 스플라이싱, 및 검출을 포함한 RNA의 다양한 특성에 영향을 미치기 위해 적절한 PEgRNA (본원에 제공된 지침을 사용하여 설계됨)와 함께 이중 프라임 편집을 사용하여 (예를 들어, 도 90의 이중 PE 방법론을 사용하여) 하기 RNA 태그 또는 모티프를 관심 유전자 내로 삽입할 수 있다.

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
폴리오마바이러스 원숭이 바이러스 40 (SV40) 유형1	AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTA (SEQ ID NO: 331)	태그부착된 유전자의 전사의 종결; 시토졸 내로의 MRNA의 수송; RNA 안정성 및 코딩된 단백질 발현의 증가	ATCCTTCCAGTCAGGGCCATGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC*A**CCTGAACCGTATATC***TAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTCTATG*GCCCTGAACTGGAA*(SEQ ID NO: 332)
폴리오마바이러스 원숭이 바이러스 40 (SV40) 유형2	CCATGGCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTC (SEQ ID NO: 333)	태그부착된 유전자의 전사의 종결; 시토졸 내로의 MRNA의 수송; RNA 안정성 및 코딩된 단백질 발현의 증가	ATCCTTCCAGTCAGGGCCATGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC*A**CCTGAACCGTATATC***GAGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTGGGCCATGGCTATG*GCCCTGACTGGAA*(SEQ ID NO: 334)

[0948]

[0949]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
폴리오마바이러스 원숭이 바이러스 40 (SV40) 유형3	TGATCATAATCAAGCCATA TCACATCTGTAGAGGTTTA CTTGCTTTAAAAAACCTC CACACCTCCCCCTGAACC TGAAACATAAAATGAATG CAATTGTTGTTGTTAACTT GTTTATTGCAGCTTATAAT GGTTACAAATAAAGCAAT AGCATCACAAATTTCACA AATAAAGCATTTTTTTCAC TGCATTCTAGTTGTGGTTT GTCCAAACTCATCAATGTA TCTTATCATGTCTGGATCT GC (SEQ ID NO: 335)	태그부착된 유전자의 전사의 종결; 시토졸 내로의 MRNA의 수송; RNA 안정성 및 코딩된 단백질 발현의 증가	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A* *CCTGAACCGTATATC* **GCAGATCCAGACA TGATAAGATACATT GATGAGTTTGGAC AAACCACAACTAG AATGCAGTGAAAA AAATGCTTTATTT GTGAAATTTGTGA TGCTATTGCTTTAT TTGTAACCATTATA AGCTGCAATAAAC AAGTTAACAACAA CAATTGCATTCAT TTTATGTTTCAGG TTCAGGGGGAGG TGTGGAGGTTTTT TAAAGCAAGTAAA CCTCTACAGATGT GATATGGCTTGAT TATGATCA**CTATG*G* *CCCTGACTGGAA* (SEQ ID NO: 336)
인간 성장 호르몬 (HGH)	ACGGGTGGCATCCCTGTG ACCCCTCCCCAGTGCCTC TCCTGGCCCTGGAAGTTG CCACTCCAGTGCCCACCA GCCTTGTCCTAATAAAATT AAGTTGCATCATTTTGTCT GACTAGGTGTCCTTCTATA ATATTATGGGGTGGAGGG GGGTGGTATGGAGCAAGG GGCAAGTTGGGAAGACA ACCTGTAGGGCCTGCGGG GTCTATTGGGAACCAAGC TGGAGTGCAGTGGCACAA TCTTGGCTCACTGCAATCT	세포질 내로의 RNA의 수송; RNA 안정성 및 코딩된 단백질 발현의 증진	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A* *CCTGAACCGTATATC* **AAGGACAGGGAA GGGAGCAGTGGT TCACGCCTGTAAT CCCAGCAATTTGG GAGGCCAAGGTG GGTAGATCACCTG**

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
	CCGCCTCCTGGGTTCAAG CGATTCTCCTGCCTCAGCC TCCCGAGTTGTTGGGATT CCAGGCATGCATGACCAG GCTCAGCTAATTTTTGTTT TTTTGGTAGAGACGGGGT TTCACCATATTGGCCAGGC TGGTCTCCAACTCCTAATC TCAGGTGATCTACCCACCT TGGCCTCCCAAATTGCTG GGATTACAGGCGTGAACC ACTGCTCCCTTCCCTGTCC TT (SEQ ID NO: 337)		**AGATTAGGAGTTG GAGACCAGCCTG GCCAATATGGTGA AACCCCGTCTCTA CCAAAAAAACAAA AATTAGCTGAGCC TGGTCATGCATGC CTGGAATCCCAAC AACTCGGGAGGCT GAGGCAGGAGAAT CGCTTGAACCCAG GAGGCGGAGATTG CAGTGAGCCAAGA TTGTGCCACTGCA CTCCAGCTTGGTT CCCAATAGACCCC GCAGGCCCTACAG GTTGTCTTCCCAA CTTGCCCCTTGCT CCATACCACCCCC CTCCACCCCATAA TATTATAGAAGGA CACCTAGTCAGAC AAAATGATGCAAC TTAATTTTATTAGG ACAAGGCTGGTG GGCACTGGAGTG GCAACTTCCAGGG CCAGGAGAGGCA CTGGGGAGGGGT CACAGGGATGCCA CCCGT**CTATG*GCCC TGACTGGAA* (SEQ ID NO: 338)
소 성장 호르몬 (BGH)	CGACTGTGCCTTCTAGTTG CCAGCCATCTGTTGTTTGC CCCTCCCCCGTGCCTTCCT TGACCCTGGAAGGTGCCA CTCCCACTGTCCTTTCCTA ATAAAATGAGGAAATTGC ATCGCATTGTCTGAGTAGG TGTCATTCTATTCTGGGGG GTGGGGTGGGGCAGGAC	세포질 내로의 RNA의 수송; RNA 안정성 및 코딩된 단백질 발현의 증진	ATCCTTCCAGTCAG GGCCAT<u>GTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC</u>*A CCTGAACCGTATATC* **CCATAGAGCCCAC**

[0950]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
	AGCAAGGGGGAGGATTG GGAAGACAATAGCAGGCA TGCTGGGGATGCGGTGGG CTCTATGG (SEQ ID NO: 339)		**CGCATCCCCAGCA TGCCTGCTATTGT CTTCCCAATCCTC CCCCTTGCTGTCC TGCCCCACCCCAC CCCCCAGAATAGA ATGACACCTACTC AGACAATGCGATG CAATTTCCTCATTT TATTAGGAAAGGA CAGTGGGAGTGG CACCTTCCAGGGT CAAGGAAGGCAC GGGGGAGGGGCA AACAACAGATGGC TGGCAACTAGAAG GCACAGTCGC**TAT G*GCCCTGACTGGAA* (SEQ ID NO: 340)
토끼 베타-글로빈 (RBGLOB)	TTCACTCCTCAGGTGCAG GCTGCCTATCAGAAGGTG GTGGCTGGTGTGGCCAAT GCCCTGGCTCACAAATAC CACTGAGATCTTTTTCCCT CTGCCAAAAATTATGGGG ACATCATGAAGCCCCTTG AGCATCTGACTTCTGGCTA ATAAAGGAAATTTATTTTC ATTGCAATAGTGTGTTGGA ATTTTTTGTGTCTCTCACT CGGAAGGACATATGGGAG GGCAAATCATTTAAAACAT CAGAATGAGTATTTGGTTT AGAGTTTGGCAACATATG CCCATATGCTGGCTGCCAT GAACAAAGGTTGGCTATA AAGAGGTCATCAGTATATG AAACAGCCCCCTGCTGTC CATTCCTTATTCCATAGAA AAGCCTTGACTTGAGGTT AGATTTTTTTTATATTTTGT TTTGTGTTATTTTTTTCTTT AACATCCCTAAAATTTTCC	세포질 내로의 RNA의 수송; RNA 안정성 및 코딩된 단백질 발현의 증진	ATCCTTCCAGTCAG GGCCAT<u>GTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC</u>*A CCTGAACCGTATATC* **GATCTCCATAAGA GAAGAGGGACAG CTATGACTGGGAG TAGTCAGGAGAGG AGGAAAAATCTGG CTAGTAAAACATG TAAGGAAAATTTT AGGGATGTTAAAG AAAAAAATAACAC AAAACAAAATATA AAAAAAATCTAAC CTCAAGTCAAGGC TTTTCTATGGAATA AGGAATGGACAGC AGGGGGCTGTTTC ATATACTGATGAC**

[0951]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
	TTACATGTTTTACTAGCCA GATTTTTCCTCCTCTCCTG ACTACTCCCAGTCATAGCT GTCCCTCTTCTCTTATGGA GATC (SEQ ID NO: 341)		**CTCTTTATAGCCA ACCTTTGTTCATG GCAGCCAGCATAT GGGCATATGTTGC CAAACTCTAAACC AAATACTCATTCT GATGTTTTAAATG ATTTGCCCTCCCA TATGTCCTTCCGA GTGAGAGACACAA AAAATTCCAACAC ACTATTGCAATGA AAATAAATTTCCTT TATTAGCCAGAAG TCAGATGCTCAAG GGGCTTCATGATG TCCCCATAATTTTT GGCAGAGGGAAA AAGATCTCAGTGG TATTTGTGAGCCA GGGCATTGGCCAC ACCAGCCACCACC TTCTGATAGGCAG CCTGCACCTGAGG AGTGAA**CTATG*GCC CTGACTGGAA* (SEQ ID NO: 342)
티미딘 키나제 (TK)	GGGGGAGGCTAACTGAAA CACGGAAGGAGACAATAC CGGAAGGAACCCGCGCTA TGACGGCAATAAAAAGAC AGAATAAAACGCACGGGT GTTGGGTCGTTTGTTCATA AACGCGGGGTTCGGTCCC AGGGCTGGCACTCTGTCG ATACCCCACCGAGACCCC ATTGGGGCCAATACGCCC GCGTTTCTTCCTTTTCCCC ACCCCACCCCCCAAGTTC GGGTGAAGGCCCAGGGCT CGCAGCCAACGTCGGGGC GGCAGGCCCTGCCATAG (SEQ ID NO: 343)	세포질 내로의 RNA의 수송; RNA 안정성 및 코딩된 단백질 발현의 증진	ATCCTTCCAGTCAG GGCCAT<u>GTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC</u>*A CCTGAACCGTATATC* **CTATGGCAGGGCC TGCCGCCCCGACG TTGGCTGCGAGCC CTGGGCCTTCACC CGAACTTGGGGG GTGGGGTGGGGA AAAGGAAGAAAC GCGGGCGTATTGG**

[0952]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
			CCCCAATGGGGTC TCGGTGGGGTATC GACAGAGTGCCA GCCCTGGGACCG AACCCCGCGTTTA TGAACAAACGACC CAACACCCGTGCG TTTTATTCTGTCTT TTTATTGCCGTCA TAGCGCGGGTTCC TTCCGGTATTGTC TCCTTCCGTGTTT CAGTTAGCCTCCC CCCTATG*GCCCTGA CTGGAA* (SEQ ID NO: 344)
MALAT1 ENE-MASCRNA	TAGGGTCATGAAGGTTTTT CTTTTCCTGAGAAAACAA CACGTATTGTTTTCTCAGG TTTTGCTTTTTGGCCTTTT TCTAGCTTAAAAAAAAAA AAAGCAAAAGATGCTGGT GGTTGGCACTCCTGGTTT CCAGGACGGGGTTCAAAT CCCTGCGGCGTCTTTGCTT TGACT (SEQ ID NO: 345)	RNA를 핵 내에 체류시킴, 전사체 종결 및 안정화	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **AGTCAAAGCAAAG ACGCCGCAGGGAT TTGAACCCCGTCC TGGAAACCAGGA GTGCCAACCACCA GCATCTTTTGCTT TTTTTTTTTTTAAG CTAGAAAAAGGCC AAAAAGCAAAACC TGAGAAAACAATA CGTGTTGTTTTCT CAGGAAAAGAAAA ACCTTCATGACCC TA**CTATG*GCCCTGA CTGGAA* (SEQ ID NO: 346)
KSHV PAN ENE	TGTTTTGGCTGGGTTTTTC CTTGTTCGCACCGGACAC CTCCAGTGACCAGACGGC	RNA를 핵 내에 체류시킴,	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA

[0953]

[0954]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
	AAGGTTTTTATCCCAGTGT ATATTGGAAAAACATGTTA TACTTTTGACAATTTAACG TGCCTAGAGCTCAAATTA AACTAATACCATAACGTAA TGCAACTTACAACATAAAT AAAGGTCAATGTTTAATCC ATAAAAAAAAAAAAAAA AAAA (SEQ ID NO: 347)	전사체 종결 및 안정화	GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **TTTTTTTTTTTTTT TTTTTATGGATTAA ACATTGACCTTTA TTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG GTCACTGGAGGTG TCCGGTGCGAACA AGGAAAAACCCAG CCAAAACA**CTATG *GCCCTGACTGGAA* (SEQ ID NO: 348)
짧은, 비보존된 RNA 링커를 갖는 3개의, 순차적 KSHV PAN ENE	TGTTTTGGCTGGGTTTTTC CTTGTTCGCACCGGACAC CTCCAGTGACCAGACGGC AAGGTTTTTATCCCAGTGT ATATTGGAAAAACATGTTA TACTTTTGACAATTTAACG TGCCTAGAGCTCAAATTA AACTAATACCATAACGTAA TGCAACTTACAACATAAAT AAAGGTCAATGTTTAATCC ATAAAAAAAAAAAAAAA AAAAACACACTGTTTTGG CTGGGTTTTTCCTTGTTCG CACCGGACACCTCCAGTG ACCAGACGGCAAGGTTTT TATCCCAGTGTATATTGGA AAAACATGTTATACTTTTG ACAATTTAACGTGCCTAG AGCTCAAATTAAACTAATA CCATAACGTAATGCAACTT	RNA를 핵 내에 체류시킴, 전사체 종결 및 안정화, 단일 PAN ENE보다 더 클 것으로 예측됨	ATCCTTCCAGTCAG GGCCAT<u>GTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC</u>*A CCTGAACCGTATATC* **TTTTTTTTTTTTTT TTTTTATGGATTAA ACATTGACCTTTA TTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG**

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
	ACAACATAAATAAAGGTC AATGTTTAATCCATAAAAA AAAAAAAAAAAAAATCTC TCTGTTTTGGCTGGGTTTT TCCTTGTTCGCACCGGAC ACCTCCAGTGACCAGACG GCAAGGTTTTTATCCCAGT GTATATTGGAAAAACATGT TATACTTTTGACAATTTAA CGTGCCTAGAGCTCAAAT TAAACTAATACCATAACGT AATGCAACTTACAACATA AATAAAGGTCAATGTTTA ATCCATAAAAAAAAAAAA AAAAAAA (SEQ ID NO: 349)		**GTCACTGGAGGTG TCCGGTGCGAACA AGGAAAAACCCAG CCAAAACAGAGAG ATTTTTTTTTTTT TTTTTTATGGATTA AACATTGACCTTT ATTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG GTCACTGGAGGTG TCCGGTGCGAACA AGGAAAAACCCAG CCAAAACAGTGTG TTTTTTTTTTTTT TTTTTTATGGATTA AACATTGACCTTT ATTTATGTTGTAAG TTGCATTACGTTAT GGTATTAGTTTAAT TTGAGCTCTAGGC ACGTTAAATTGTC AAAAGTATAACAT GTTTTTCCAATATA CACTGGGATAAAA ACCTTGCCGTCTG GTCACTGGAGGTG TCCGGTGCGAACA AGGAAAAACCCAG CCAAAACA**CTATG *GCCCTGACTGGAA* (SEQ ID NO: 350)
SM박스/U1 SNRNA 박스	CAGCAAGTTCAGAGAAAT CTGAACTTGCTGGATTTTT GGAGCAGGGAGATGGAAT AGGAGCTTGCTCCGTCCA CTCCACGCATCGACCTGG	RNA의 핵 내 체류 및 전사체 종결을 발생시킴	ATCCTTCCAGTCAG GGCCAT<u>GTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC</u>

[0955]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
	TATTGCAGTACCTCCAGG AACGGTGCACCCACTTTC TGGAGTTTCAAAAGTAGA CTGTACGCTAAGGGTCATA TCTTTTTTTGTTTGGTTTG TGTCTTGGTTGGCGTCTTA AA (SEQ ID NO: 351)		TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **TTTAAGACGCCAA CCAAGACACAAAC CAAACAAAAAAAG ATATGACCCTTAG CGTACAGTCTACT TTTGAAACTCCAG AAAGTGGGTGCAC CGTTCCTGGAGGT ACTGCAATACCAG GTCGATGCGTGGA GTGGACGGAGCA AGCTCCTATTCCA TCTCCCTGCTCCA AAAATCCAGCAAG TTCAGATTTCTCT GAACTTGCTGC**TA TG*GCCCTGACTGGA A* (SEQ ID NO: 352)
U1 SNRNA 3′ 박스	GTTTCAAAAGTAGACTGT ACGCTAAGGGTCATATCTT TTTTTGTTTGGTTTGTGTC TTGGTTGGCGTCTTAAA (SEQ ID NO: 353)	RNA의 핵 내 체류 및 전사체 종결을 발생시킴	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **TTTAAGACGCCAA CCAAGACACAAAC CAAACAAAAAAAG ATATGACCCTTAG CGTACAGTCTACT TTTGAAAC**CTATG*G CCCTGACTGGAA* (SEQ ID NO: 354)
TRNA-리신	GCCCGGCTAGCTCAGTCG GTAGAGCATGAGACTCTT AATCTCAGGGTCGTGGGT TCGAGCCCCACGTTGGGC G (SEQ ID NO: 355)	RNA의 미토콘드리아로의 수송을 가능하게 하는 것으로 보고됨	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC

[0956]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
			TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **CGCCCAACGTGG GGCTCGAACCCAC GACCCTGAGATTA AGAGTCTCATGCT CTACCGACTGAGC TAGCCGGGC**CTAT *GGCCCTGACTGGAA* (SEQ ID NO: 356)
브로콜리 압타머	GAGACGGTCGGGTCCAGA TATTCGTATCTGTCGAGTA GAGTGTGGGCTC (SEQ ID NO: 357)	가시화 (형광)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GAGCCCACACTCT ACTCGACAGATAC GAATATCTGGACC CGACCGTCTC**CTA TG*GCCCTGACTGGAA* (SEQ ID NO: 358)
스피니치 압타머	GACGCAACTGAATGAAAT GGTGAAGGACGGGTCCA GGTGTGGCTGCTTCGGCA GTGCAGCTTGTTGAGTAG AGTGTGAGCTCCGTAACT AGTCGCGTC (SEQ ID NO: 359)	가시화 (형광)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GACGCGACTAGTT ACGGAGCTCACAC TCTACTCAACAAG CTGCACTGCCGAA GCAGCCACACCTG GACCCGTCCTTCA CCATTTCATTCAG TTGCGTC**CTATG*GC CCTGACTGGAA*

[0957]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
			(SEQ ID NO: 360)
스피니치2 압타머	GATGTAACTGAATGAAAT GGTGAAGGACGGGTCCA GTAGGCTGCTTCGGCAGC CTACTTGTTGAGTAGAGT GTGAGCTCCGTAACTAGT TACATC (SEQ ID NO: 361)	가시화 (형광)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GATGTAACTAGTT ACGGAGCTCACAC TCTACTCAACAAG TAGGCTGCCGAAG CAGCCTACTGGAC CCGTCCTTCACCA TTTCATTCAGTTA CATC**CTATG*GCCCT GACTGGAA* (SEQ ID NO: 362)
망고 압타머	GGCACGTACGAAGGGACG GTGCGGAGAGGAGAGTAC GTGC (SEQ ID NO: 363)	가시화 (형광)	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GCACGTACTCTCC TCTCCGCACCGTC CCTTCGTACGTGC C**CTATG*GCCCTGAC TGGAA* (SEQ ID NO: 364)
HDV 리보자임	GGCCGGCATGGTCCCAGC CTCCTCGCTGGCGCCGGC TGGGCAACATGCTTCGGC ATGGCGAATGGGAC (SEQ ID NO: 365)	3′ 단부 RNA 프로세싱	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **GTCCCATTCGCCA TGCCGAAGCATGT**

[0958]

RNA 모티프	RNA 모티프의 서열	기능/ 효과	예시적인 HEXA 유전자 내로의 RNA 모티프의 프라임 편집 삽입을 위한 예시적인 PEGRNA*
			TGCCCAGCCGGC GCCAGCGAGGAG GCTGGGACCATGC CGGCCCTATG*GCCC TGACTGGAA* (SEQ ID NO: 366)
N^6-메틸아데노신 마커 (M^6A)	GGACTCTAGGACTGGACT TTGGACT (SEQ ID NO: 367)	메틸화를 위한 표적화 (밑줄표시된 A가 메틸화됨). M6A 메틸화는 증진된 RNA 안정성 및 발현을 발생시킬 수 있지만, 아직 완전히 이해되어있지 않음	ATCCTTCCAGTCAG GGCCATGTTTGAGA GCTAGAAATAGCAA GTTTAAATAAGGCT AGTCCGTTATCAAC TTGAAAAAGTGGG ACCGAGTCGGTCC*A CCTGAACCGTATATC* **AGTCCAAAGTCCA GTCCTAGAGTCCC** TATG*GCCCTGACTG GAA* (SEQ ID NO: 368)

[0959]

[0960] * 각각의 PEgRNA는 5'에서 3' 방향으로 제시되고, 하기와 같이 글자체 유형에 의해 지정된 바와 같이 도 3f의 하기 구조적 요소를 갖는다: 5' - 스페이서 서열 (통상 글자체) - gRNA 코어 (밑줄표시된 서열) - 상동성 아암 (이탤릭체) - RT 주형 (볼드체) - 프라이머 결합 부위 (이탤릭체) - 3'.

[0961] 상기 표의 PEgRNA는 상기 모티프의 예를 HEXA 유전자 (테이-삭스병에서 결함있음) (예를 들어, 진뱅크 번호 KR710351.1 (서열식별번호: 369)) 내로 부위-특이적으로 삽입하기 위해 설계되지만, 이는 단지 예시의 목적을 위한 것이다. RNA 태그부착에서의 프라임 편집의 사용은 HEXA 유전자에 제한되지 않고, 이는 실제로 임의의 것일 수 있다. HEXA mRNA는 하기 뉴클레오티드 서열을 갖는다:

GTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGT
TGGCATGACAAGTTCCAGGCTTTGGTTTTCGCTGCTGCTGGCGGCAGCGTTCGCAGGACG
GGCGACGGCCCTCTGGCCCTGGCCTCAGAACTTCCAAACCTCCGACCAGCGCTACGTCCT
TTACCCGAACAACTTTCAATTCCAGTACGATGTCAGCTCGGCCGCGCAGCCCGGCTGCTCA
GTCCTCGACGAGGCCTTCCAGCGCTATCGTGACCTGCTTTTCGGTTCCGGGTCTTGGCCCC
GTCCTTACCTCACAGGGAAACGGCATACACTGGAGAAGAATGTGTTGGTTGTCTCTGTAG
TCACACCTGGATGTAACCAGCTTCCTACTTTGGAGTCAGTGGAGAATTATACCCTGACCAT
AAATGATGACCAGTGTTTACTCCTCTCTGAGACTGTCTGGGGAGCTCTCCGAGGTCTGGA
GACTTTTAGCCAGCTTGTTTGGAAATCTGCTGAGGGCACATTCTTTATCAACAAGACTGAG
ATTGAGGACTTTCCCCGCTTTCCTCACCGGGGCTTGCTGTTGGATACATCTCGCCATTACCT
GCCACTCTCTAGCATCCTGGACACTCTGGATGTCATGGCGTACAATAAATTGAACGTGTTC
CACTGGCATCTGGTAGATGATCCTTCCTTCCCATATGAGAGCTTCACTTTTCCAGAGCTCAT
GAGAAAGGGGTCCTACAACCCTGTCACCCACATCTACACAGCACAGGATGTGAAGGAGG
TCATTGAATACGCACGGCTCCGGGGTATCCGTGTGCTTGCAGAGTTTGACACTCCTGGCCA
CACTTTGTCCTGGGGACCAGGTATCCCTGGATTACTGACTCCTTGCTACCCTGGGTCTGAG

[0962]

CCCTCTGGCACCTTTGGACCAGTGAATCCCAGTCTCAATAATACCTATGAGTTCATGAGCA
CATTCTTCTTAGAAGTCAGCTCTGTCTTCCCAGATTTTTATCTTCATCTTGGAGGAGATGAG
GTTGATTTCACCTGCTGGAAGTCCAACCCAGAGATCCAGGACTTTATGAGGAAGAAAGGC
TTCGGTGAGGACTTCAAGCAGCTGGAGTCCTTCTACATCCAGACGCTGCTGGACATCGTCT
CTTCTTATGGCAAGGGCTATGTGGTGTGGCAGGAGGTGTTTGATAATAAAGTAAAGATTCA
GCCAGACACAATCATACAGGTGTGGCGAGAGGATATTCCAGTGAACTATATGAAGGAGCT
GGAACTGGTCACCAAGGCCGGCTTCCGGGCCCTTCTCTCTGCCCCCTGGTACCTGAACCG
TATATCCTATGGCCCTGACTGGAAGGATTTCTACGTAGTGGAACCCCTGGCATTTGAAGGTA
CCCCTGAGCAGAAGGCTCTGGTGATTGGTGGAGAGGCTTGTATGTGGGGAGAATATGTGG
ACAACACAAACCTGGTCCCCAGGCTCTGGCCCAGAGCAGGGGCTGTTGCCGAAAGGCTG
TGGAGCAACAAGTTGACATCTGACCTGACATTTGCCTATGAACGTTTGTCACACTTCCGCT
GTGAGTTGCTGAGGCGAGGTGTCCAGGCCCAACCCCTCAATGTAGGCTTCTGTGAGCAGG
AG
TTTGAACAGACCTGCCCAACTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTAT
CAATTTGTTGCAACGAAC (SEQ ID NO: 369).

[0963]

[0964] 상응하는 HEXA 단백질은 하기 아미노산 서열을 갖는다:

MTSSRLWFSLLLAAAFAGRATALWPWPQNFQTSDQRYVLYPNNFQFQYDVSSAAQPGCSVL
DEAFQRYRDLLFGSGSWPRPYLTGKRHTLEKNVLVVSVVTPGCNQLPTLESVENYTLTINDD
QCLLLSETVWGALRGLETFSQLVWKSAEGTFFINKTEIEDFPRFPHRGLLLDTSRHYLPLSSIL
DTLDVMAYNKLNVFHWHLVDDPSFPYESFTFPELMRKGSYNPVTHIYTAQDVKEVIEYARLR
GIRVLAEFDTPGHTLSWGPGIPGLLTPCYPGSEPSGTFGPVNPSLNNTYEFMSTFFLEVSSVFPD
FYLHLGGDEVDFTCWKSNPEIQDFMRKKGFGEDFKQLESFYIQTLLDIVSSYGKGYVVWQEV
FDNKVKIQPDTIIQVWREDIPVNYMKELELVTKAGFRALLSAPWYLNRISYGPDWKDFYVVE
PLAFEGTPEQKALVIGGEACMWGEYVDNTNLVPRLWPRAGAVAERLWSNKLTSDLTFAYERL
SHFRCELLRRGVQAQPLNVGFCEQEFEQT (SEQ ID NO: 370).

[0965]

[0966] 특히, 생성된 RNA 모티프는 HEXA 유전자의 번역된 영역 내에 포함되어, 단백질 코딩 유전자의 기능을 파괴할 것이다. 삽입된 폴리아데닐화 모티프는 조기 전사체 종결을 초래할 것이다. 이 부위는 전사되어 RNA 생성물을 생성할 게놈 부위 내에 상기 표의 열거된 RNA 모티프의 삽입을 발생시킬 수 있는 잠재적 PEgRNA의 예시일 뿐이다.

[0967] RNA 태그부착을 위해 PE와 함께 사용하기 위한 PEgRNA는 U6 프로모터 (이 경우에 G로 시작하지 않는 프로토스페이서를 포함하는 가이드의 경우 PEgRNA의 5' 단부에 단일 구아노신이 부가될 것이고, 6-7 티민이 3' 단부에 부가될 것임) 또는 pol II 프로모터, 예컨대 pCMV (이 경우에 자기-절단 요소 또는 Csy4 모티프를 통해 RNA의 5' 단부로부터 이 프로모터의 내재적으로 전사된 서열을 제거하는 것이 필요할 수 있고, 핵으로부터 RNA의 유출을 발생시키지 않는 RNA의 3' 단부에 종결 모티프, 예컨대 상기 열거된 3' 박스 모티프가 부가될 필요가 있을 것임. 이 모티프는 어닐링 영역의 3'일 것이기 때문에 PE의 결과로서 게놈 내로 삽입되지 않을 것임에 주목함)로부터 발현될 수 있다. 코어 PEgRNA 스캐폴드는 밑줄표시되고, 상동성 및 어닐링 영역은 이탤릭체로 표시되고, 삽입된 서열은 볼드체로 표시된다. 삽입된 서열은 상기 예의 역 상보체이고 - 하기 기재된 바와 같이, 따라서 이들 PEgRNA는 코딩 가닥에 표적화될 필요가 있을 것임에 주목한다.

[0968] 또한, HDV 외의 자기-절단 리보자임은 일부 실시양태에서 주어진 표적 부위에 맞춰질 필요가 있고; 즉, HDV는 코딩된 전사체를 그 자체에 대해 바로 5'에서 절단하는 반면, 모든 다른 자기-절단 리보자임에 대한 절단 부위는 리보자임 그 자체 내에 있음에 주목한다. 따라서, 제1 및 마지막 대략 5-10개 뉴클레오티드 (및 일부 경우에 잠재적으로 10개 초과)는 실제로 코딩된 서열의 일부일 것이다. 예로서, 서열 5'NNNNNTCATCCTGATAAACTGCAAA3' (서열식별번호: 371)을 5개의 N (여기서 N은 임의의 뉴클레오티드임) 다음에서 해머헤드 자기-절단 리보자임을 사용하여 절단하기 위해, 하기 서열이 삽입될 것이고, 여기서 밑줄표시된 서열은 불완전한 RNA 쌍형성 요소를 형성한다.

5′NNNNN<u>CAGTTTGTACGGA</u>TGACTGATGAGTCCCAAATAGGACGAAACGCGCTTC
GGTGCGTCTCATCCTGATAAACTGCAAA-3′ (SEQ ID NO: 372).

[0969]

[0970] 이러한 쌍형성 요소의 길이 및 성질 면에서는 유의한 유연성이 존재하고, 이는 원래의 제출에 열거된 비-HDV 자기-절단 리보자임 중 임의의 것에 대해 그러할 것이다. 상기 열거된 구축물과 동일한 프로토스페이서를 갖는 PEgRNA를 사용하여 hexA mRNA를 절단하기 위해 해머헤드 리보자임을 설치하기 위해, 하기 PEgRNA 서열을 사용할 수 있었다 (라벨은 상기와 동일함):

[0971] 5′ACCTGAACCGTATATCGACGCACCGAAGCGCGTTTCGTCCTATTTGGGACTCATC
AGGATATACGGTTCAGGTGATATACGGTTCAGGTGACGCACCGAAGCGCGTTTCGTCCTAT
TTGGGACTCATCAGACCTGAACCGTATATCATCCTTCCAGTCAGGGCCATGTTTGAGAGCT
AGAAATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTC
GGTCCACCTGAACCGTATATC**GACGCACCGAAGCGCGTTTCGTCCTATTTGGGACTCA**
TCAGGATATACGGTTCAGGTCTATGGCCCTGACTGGAA-3′

[0972] (여기서 코어 PEgRNA 스캐폴드는 밑줄표시되고, 상동성 및 어닐링 영역은 이탤릭체로 표시되고, 삽입된 서열은 볼드체로 표시됨) (서열식별번호: 373).

[0973] RNA 모티프의 삽입을 위한 다른 PEgRNA의 설계는 본원에 기재된 일반적 원리를 따를 수 있다. 그러나, 많은 RNA 모티프는 잠재적으로 고도로 구조화되어 있으며, 이는 이들이 역전사되고 게놈 내로 삽입되는 것을 어렵게 만들 수 있다는 것에 주목한다. 일부 RNA 서열, 예컨대 단순 헤어핀의 경우, RNA 서열 그 자체 및 그의 상보체 둘 다가 구조화되어 있다. 그러나, 이는 상기 언급된 서열에 대해서는 그러할 가능성이 없다. 따라서, 이들 모티프를 삽입할 경우, PEgRNA가 이들 서열의 역 상보체를 코딩하여 실제로 모티프를 코딩하는 DNA 서열을 게놈 내로 삽입하는 것이 가장 가능성이 높을 것이다. 유사하게, 자기-절단 리보자임을 PEgRNA 주형 영역에 포함시키는 것은 프로세싱 및 비효율적 활성이 발생시킬 것인 반면, 그의 역 상보체를 포함시키는 것은 그렇지 않을 것이다. 따라서, 이들 PEgRNA는 코딩 가닥을 표적화할 가능성이 있을 것인 반면에, 다른 유형의 삽입 (예컨대, 치료 교정)을 코딩하는 PEgRNA는 이론적으로 어느 한 가닥을 표적화할 수 있을 것이다.

[0974] 또한, 많은 삽입된 모티프에 대해, 생성된 PEgRNA는 U6 프로모터로부터 전사될 수 없고, 다른 프로모터, 예컨대 pCMV의 사용을 필요로 할 수 있음에 주목한다. 유사하게, 보다 긴 PEgRNA는 또한 덜 안정할 수 있다. 보다 짧은 모티프, 예컨대 m^6A 마커는 이러한 도전과제를 갖지 않을 것이다.

[0975] D. 이뮤노에피토프의 삽입을 위한 이중 프라임 편집의 사용

[0976] 이중 프라임 편집 (예를 들어, 예컨대 도 90에 제시된 실시양태)는 또한 공지된 면역원성 에피토프를 내인성 또는 외래 게놈 DNA에 삽입하여, 치료 또는 생명공학적 적용을 위한 상응하는 단백질의 변형을 발생시키는 수단으로서 사용될 수 있다 (도 31 및 32 참조). 전형적 프라임 편집 또는 이중 프라임 편집의 발명 전에, 이러한 삽입은 단지 비효율적으로, DSB로부터의 높은 비율의 indel 형성으로 달성될 수 있다. 프라임 편집은 일반적으로 HDR보다 더 높은 효율을 제공하면서 삽입 편집으로부터의 높은 indel 형성의 문제를 해결한다. 이러한 보다 낮은 비율의 indel 형성은, 특히 면역원성 에피토프를 삽입하는 기재된 적용에서 표적화된 DNA 삽입을 위한 방법으로서 HDR에 비해 프라임 편집의 주요 이점을 제공한다. 에피토프의 길이는 수개의 염기 내지 수백개의 염기의 범위이다. 프라임 편집제는 포유동물 세포에서 이러한 표적화된 삽입을 달성하기 위한 효율적인 접근법이다.

[0977] 본 발명의 주요 개념은 이전에 기재된 면역원성 에피토프를 함유하는 뉴클레오티드 서열을 내인성 또는 외래 게놈 DNA 내에 그의 단백질 생성물 및/또는 발현 세포 유형의 하향조절 및/또는 파괴를 위해 삽입하기 위한 이중 프라임 편집제의 사용이다. 면역원성 에피토프 삽입을 위한 뉴클레오티드 서열은 표적화된 유전자의 코딩된 단백질 및 삽입된 면역원성 에피토프의 상응하는 단백질 번역의 융합 단백질을 생산하는 방식으로 유전자에 표적화될 것이다. 환자의 면역계는 예를 들어 파상풍 또는 디프테리아 또는 홍역에 대한 상용 백신접종으로부터의 표준 사전 면역화의 결과로서 이들 에피토프를 인식하도록 이전에 훈련되었을 것이다. 융합된 에피토프의 면역원성 성질의 결과로서, 환자의 면역계는 프라임 편집된 단백질 (삽입된 에피토프만이 아님) 및 잠재적으로 그것이 발현되는 세포를 인식하고 이를 무력화할 것으로 예상될 것이다.

[0978] CRISPR/Cas 시스템을 사용하는 정확한 게놈 표적화 기술은 최근, 조작된 DNA 서열의 표적화된 게놈 유전자좌 내로의 삽입을 포함한 광범위한 적용에서 연구되었다. 이전에는 상동성-지정 복구 (HDR)가 이러한 적용에 사용되어 왔으며, 이는 ssDNA 공여자 주형 및 이중 가닥 DNA 파괴 (DSB)에 의한 복구 개시를 필요로 하였다. 이러한 전략은 세포에서 이루어지는 가장 넓은 범위의 가능한 변화를 제공하고, 대형 DNA 서열을 포유동물 세포 내로

삽입하는데 이용가능한 유일한 방법이다. 그러나, HDR은 그의 개시 DSB로부터 초래되는 바람직하지 않은 세포 부작용, 예컨대 높은 수준의 indel 형성, DNA 전위, 대형 결실, 및 P53 활성화에 의해 방해받는다. 이들 결점에 더하여, HDR은 많은 세포 유형에서의 낮은 효율에 의해 제한된다 (T 세포는 이러한 관찰에 대한 주목할 만한 예외임). 이들 결점을 극복하기 위한 최근의 노력은 인간 Rad51 돌연변이체를 Cas9 D10A 닉카제에 융합시켜 (RDN), 개선된 HDR 생성물:indel 비 및 보다 낮은 오프 타겟 편집을 특색으로 하는 DSB-부재 HDR 시스템을 생성하는 것을 포함하지만, 이는 여전히 세포-유형 의존성 및 단지 보통의 HDR 편집 효율에 의해 방해받는다.

[0979] PEgRNA와 커플링된 리버스 트랜스크립타제에 대한 Cas9의 최근에 개발된 융합체 ("프라임 편집제")는 부위-특이적 방식으로 임의의 단일 뉴클레오티드 치환을 설치하고 임의의 짧은 스트레치의 뉴클레오티드 (최대 적어도 수십개의 염기)를 삽입 또는 결실시키는 능력을 포함하여, 기존의 게놈 편집 방법에 비해 다수의 이점을 제공하는 신규한 게놈 편집 기술을 나타낸다. 특히, PE 편집은 일반적으로 낮은 비율의 비의도된 indel로 달성된다. 따라서, PE는 이전에 불가능하였거나 비실용적이었던 표적화된 삽입-기반 편집 적용을 가능하게 한다.

[0980] 이러한 특정한 측면은 공지된 면역원성 에피토프를 내인성 또는 외래 게놈 DNA에 삽입하여, 치료 또는 생명공학적 적용을 위한 상응하는 단백질의 변형을 발생시키는 수단으로서 프라임 편집을 사용하는 방법을 기재한다 (도 31 및 32 참조). 프라임 편집의 발명 전에, 이러한 삽입은 단지 비효율적으로, DSB로부터의 높은 비율의 indel 형성으로 달성될 수 있다. 프라임 편집은 일반적으로 HDR보다 더 높은 효율을 제공하면서 삽입 편집으로부터의 높은 indel 형성의 문제를 해결한다. 이러한 보다 낮은 비율의 indel 형성은, 특히 면역원성 에피토프를 삽입하는 기재된 적용에서 표적화된 DNA 삽입을 위한 방법으로서 HDR에 비해 프라임 편집의 주요 이점을 제공한다. 에피토프의 길이는 수개의 염기 내지 수백개의 염기의 범위이다. 프라임 편집제는 포유동물 세포에서 이러한 표적화된 삽입을 달성하기 위한 가장 효율적이고 가장 흠없는 기술이다.

[0981] 이러한 측면의 주요 개념은 이전에 기재된 면역원성 에피토프를 함유하는 뉴클레오티드 서열을 내인성 또는 외래 게놈 DNA 내에 그의 단백질 생성물 및/또는 발현 세포 유형의 하향조절 및/또는 파괴를 위해 삽입하기 위한 프라임 편집제의 사용이다. 면역원성 에피토프 삽입을 위한 뉴클레오티드 서열은 표적화된 유전자의 코딩된 단백질 및 삽입된 면역원성 에피토프의 상응하는 단백질 번역의 융합 단백질을 생산하는 방식으로 유전자에 표적화될 것이다. 환자의 면역계는 예를 들어 파상풍 또는 디프테리아 또는 홍역에 대한 상용 백신접종으로부터의 표준 사전 면역화의 결과로서 이들 에피토프를 인식하도록 이전에 훈련되었을 것이다. 융합된 에피토프의 면역원성 성질의 결과로서, 환자의 면역계는 프라임 편집된 단백질 (삽입된 에피토프만이 아님) 및 잠재적으로 그것이 발현되는 세포를 인식하고 이를 무력화할 것으로 예상될 것이다.

[0982] 표적화된 유전자에 대한 융합은 삽입된 에피토프 단백질 번역이 면역계 인식을 위해 노출되는 것을 보장하기 위해 필요에 따라 조작될 것이다. 이는 표적화된 유전자에 대한 면역원성 에피토프의 C-말단 융합체, 표적화된 유전자에 대한 면역원성 에피토프의 N-말단 융합체, 또는 면역원성 에피토프가 단백질 구조의 표면-노출 영역 내에 코딩되도록 하는 유전자 내로의 뉴클레오티드의 삽입을 생성하는 단백질 번역을 발생시키는 표적화된 뉴클레오티드 삽입을 포함할 수 있다.

[0983] 표적 유전자 서열과 삽입된 면역원성 에피토프 뉴클레오티드 서열 사이에 삽입된 뉴클레오티드로서 코딩된 단백질 링커는 표적화된 유전자의 면역계 인식, 세포 트래픽킹, 단백질 기능, 또는 단백질 폴딩을 용이하게 하기 위해 본 발명의 일부로서 조작될 필요가 있을 수 있다. 이들 삽입된 뉴클레오티드-코딩된 단백질 링커는 XTEN 링커의 가변 길이 및 서열 또는 글리신-세린 링커의 가변 길이 및 서열을 포함할 수 있다 (그러나 이에 제한되지는 않음). 이들 조작된 링커는 단백질 융합을 성공적으로 용이하게 하기 위해 이전에 사용되었다. 예시적인 링커는 아미노산 서열 (GGGGS)n (서열식별번호: 165), (G)n (서열식별번호: 166), (EAAAK)n (서열식별번호: 167), (GGS)n (서열식별번호: 168), (SGGS)n (서열식별번호: 169), (XP)n (서열식별번호: 170), 또는 그의 임의의 조합을 포함한, 본원에 기재된 것 중 임의의 것을 포함할 수 있고, 여기서 n은 독립적으로 1 내지 30의 정수이고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 링커는 아미노산 서열 (GGS)n (서열식별번호: 176)을 포함하고, 여기서 n은 1, 3, 또는 7이다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 171)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 172)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSGGSGGS (서열식별번호: 173)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGS (서열식별번호: 174)를 포함한다.

[0984] 이러한 측면의 구별되는 특색은 특정 아미노산 서열에 대해 이전에 획득된 면역 반응을 달리 비-면역원성 단백질에 대한 면역 반응을 유도하는 수단으로서 사용하는 능력을 포함한다. 또 다른 구별되는 특색은 부산물 편집

으로서 원치 않는 indel을 높은 수준으로 유도하지 않고 그의 삽입에 있어서 효율적인 방식으로 이들 면역원성 에피토프의 뉴클레오티드 서열을 표적에 삽입하는 능력이다. 이러한 PE의 특정한 적용은 세포 유형-특이적 전달 방법 (예컨대 AAV 혈청형)을 조합하여 관심 세포 유형 내에 에피토프를 삽입하여 그에 대한 면역 반응을 촉발하는 능력을 갖는다.

[0985] 면역원성 에피토프를 병원성 유전자 내로 삽입하는 수단으로서의 프라임 편집은 환자의 면역계가 매우 다양한 질환 (면역-종양학 전략에 의하는 것과 같이 암으로 제한되지는 않음)과 싸우도록 프로그램화하는 데 사용될 수 있다. 이러한 기술의 즉각적인 관련 사용은 관련 종양유전자 예컨대 HER2 또는 성장 인자 예컨대 EGFR에 대한 면역 반응을 유발함으로써 종양의 면역 회피 메카니즘을 손상시킬 수 있기 때문에, 암 치료제로서의 사용일 것이다. 이러한 접근법은 T-세포 조작과 유사한 것으로 보일 수 있지만, 이러한 접근법의 한가지 신규 진전은 이것이 조작된 T-세포를 생성하여 환자 내로 도입할 필요 없이 많은 세포 유형에서, 그리고 암 이외의 질환에 대해 이용될 수 있다는 것이다.

[0986] PE를 사용하여 대부분의 사람들이 이미 백신접종된 면역원성 에피토프 (파상풍, 백일해, 디프테리아, 홍역, 볼거리, 풍진 등)를 질환을 구동하는 외래 또는 내인성 유전자 내로 삽입하여, 환자의 면역계가 그러한 단백질을 무력화하도록 학습시킨다.

[0987] 상기 언급된 전략으로부터 잠재적 치료 이익을 갖는 것으로 여겨지는 질환은, 예컨대 치명적 가족성 불면증에서 독성 단백질의 응집에 의해 유발되는 것을 포함한다. 유익할 수 있는 다른 질환은 달리 비독성인 내인성 단백질의 병원성 과다발현에 의해 유발되는 것, 및 외래 병원체에 의해 유발되는 것을 포함한다.

[0988] 1차 치료 적응증은 상기 언급된 것, 예컨대 암, 프리온 및 다른 신경변성 질환, 감염성 질환을 위한 치료, 및 예방적 의학을 포함한다. 2차 치료 적응증은 후기-발병 유전 질환을 갖는 환자에 대한 예방적 관리를 포함할 수 있다. 현행 표준 관리 의학은 일부 질환, 예컨대 특히 공격성 암에 대해, 또는 의약이 질환 증상의 완화를 돕는 경우에 질환이 완전히 치유될 때까지 프라임 편집과 함께 사용될 수 있을 것으로 예상된다. 하기는 본원에 개시된 프라임 편집제에 의해 유전자 내로 삽입될 수 있는 이뮤노에피토프의 예이다:

[0989]

백신	질환	에피토프 아미노산 서열	예시적인 핵산 서열 (8)
1	파상풍 톡소이드	QYIKANSKFIGITE L (SEQ ID NO: 396)	CATGATATAAAAGCAAATTCTAAATTTA TAGGTATAACTGAACTA (SEQ ID NO: 397)
2	디프테리아 독소 돌연변이체 CRM197	GADDVVDSSKSF VMENFSSYHGTK PGYVDSIQKGIQK PKSGTQGNYDDD WKEFYSTDNKYD AAGYSVDNENPL SGKAGGVVKVTY PGLTKVLALKVD NAETIKKELGLSL TEPLMEQVGTEEF IKRFGDGASRVVL SLPFAEGSSSVEYI NNWEQAKALSVE LEINFETRGKRGQ DAMYEYMAQAC AGNRVRRSVGSSL SCINLDWDVIRDK TKTKIESLKEHGPI KNKMSESPNKTV SEEKAKQYLEEFH QTALEHPELSELK TVTGTNPVFAGA NYAAWAVNVAQV IDSETADNLEKTT AALSILPGIGSVM GIADGAVHHNTEE IVAQSIALSSLMVA QAIPLVGELVDIGF AAYNFVESIINLFQ VVHNSYNRPAYSP GHKTQPFLHDGY AVSWNTVEDSIIR TGFQGESGHDIKI TAENTPLPIAGVL LPTIPGKLDVNKS KTHISVNGRKIRM RCRAIDGDVTFCR PKSPVYVGNGVH ANLHVAFHRSSSE KIHSNEISSDSIGV	GGCGCCGACGACGTGGTGGACAGCAG CAAGAGCTTCGTGATGGAGAACTTCA GCAGCTACCACGGCACCAAGCCCGGC TACGTGGACAGCATCCAGAAGGGCAT CCAGAAGCCCAAGAGCGGCACCCAGG GCAACTACGACGACGACTGGAAGGAG TTCTACAGCACCGACAACAAGTACGA CGCCGCCGGCTACAGCGTGGACAACG AGAACCCCCTGAGCGGCAAGGCCGGC GGCGTGGTGAAGGTGACCTACCCCGG CCTGACCAAGGTGCTGGCCCTGAAGG TGGACAACGCCGAGACCATCAAGAAG GAGCTGGGCCTGAGCCTGACCGAGCC CCTGATGGAGCAGGTGGGCACCGAGG AGTTCATCAAGAGGTTCGGCGACGGC GCCAGCAGGGTGGTGCTGAGCCTGCC CTTCGCCGAGGGCAGCAGCAGCGTGG AGTACATCAACAACTGGGAGCAGGCC AAGGCCCTGAGCGTGGAGCTGGAGAT CAACTTCGAGACCAGGGGCAAGAGGG GCCAGGACGCCATGTACGAGTACATGG CCCAGGCCTGCGCCGGCAACAGGGTG AGGAGGAGCGTGGGCAGCAGCCTGAG CTGCATCAACCTGGACTGGGACGTGAT CAGGGACAAGACCAAGACCAAGATCG AGAGCCTGAAGGAGCACGGCCCCATC AAGAACAAGATGAGCGAGAGCCCCAA CAAGACCGTGAGCGAGGAGAAGGCC AAGCAGTACCTGGAGGAGTTCCACCA GACCGCCCTGGAGCACCCCGAGCTGA GCGAGCTGAAGACCGTGACCGGCACC AACCCCGTGTTCGCCGGCGCCAACTA CGCCGCCTGGGCCGTGAACGTGGCCC AGGTGATCGACAGCGAGACCGCCGAC AACCTGGAGAAGACCACCGCCGCCCT GAGCATCCTGCCCGGCATCGGCAGCGT GATGGGCATCGCCGACGGCGCCGTGC ACCACAACACCGAGGAGATCGTGGCC CAGAGCATCGCCCTGAGCAGCCTGAT GGTGGCCCAGGCCATCCCCCTGGTGG

백신	질환	에피토프 아미노산 서열	예시적인 핵산 서열 (8)
		LGYQKTVDHTKV NSKLSLFFEIKS (SEQ ID NO: 398)	GCGAGCTGGTGGACATCGGCTTCGCC GCCTACAACTTCGTGGAGAGCATCATC AACCTGTTCCAGGTGGTGCACAACAG CTACAACAGGCCCGCCTACAGCCCCG GCCACAAGACCCAGCCCTTCCTGCAC GACGGCTACGCCGTGAGCTGGAACAC CGTGGAGGACAGCATCATCAGGACCG GCTTCCAGGGCGAGAGCGGCCACGAC ATCAAGATCACCGCCGAGAACACCCC CCTGCCCATCGCCGGCGTGCTGCTGCC CACCATCCCCGGCAAGCTGGACGTGA ACAAGAGCAAGACCCACATCAGCGTG AACGGCAGGAAGATCAGGATGAGGTG CAGGGCCATCGACGGCGACGTGACCT TCTGCAGGCCCAAGAGCCCCGTGTAC GTGGGCAACGGCGTGCACGCCAACCT GCACGTGGCCTTCCACAGGAGCAGCA GCGAGAAGATCCACAGCAACGAGATC AGCAGCGACAGCATCGGCGTGCTGGG CTACCAGAAGACCGTGGACCACACCA AGGTGAACAGCAAGCTGAGCCTGTTC TTCGAGATCAAGAGC (SEQ ID NO: 399)
3	볼거리 이뮤노에피토프 1	GTYRLIPNARANL TA (SEQ ID NO: 400)	GGCACCTACAGGCTGATCCCCAACGC CAGGGCCAACCTGACCGCC (SEQ ID NO: 401)
4	볼거리 이뮤노에피토프 2 볼거리 이뮤노에피토프 1	PSKFFTISDSATFA PGPVSNA (SEQ ID NO: 402) PSKLFIMLDNATF APGPVVNA (SEQ ID NO: 404)	CCGAGCAAATTTTTTACCATTAGCGAT AGCGCGACCTTTGCGCCGGGCCCGGT GAGCAACGCG (SEQ ID NO: 403) CCGAGCAAACTGTTTATTATGCTGGAT AACGCGACCTTTGCGCCGGGCCCGGT GGTGAACGCG (SEQ ID NO: 405) 발생 균주 (표1) 분기 중에서 백신 균주 JL5와 발생 균주 (표2) 사이의 헤마글루티닌-뉴라미니다제 (HN) 다양성으로부터 선택된 예
5	풍진 바이러스 (RV)	TPPPYQVSCGGES DRASARVIDPAAQ S (SEQ ID NO: 406)	ACCCCCCCCCCCTACCAGGTGAGCTGC GGCGGCGAGAGCGACAGGGCCAGCG CCAGGGTGATCGACCCCGCCGCCCAG AGC (SEQ ID NO: 407)

[0990]

백신	질환	에피토프 아미노산 서열	예시적인 핵산 서열 (8)
6	헤마글루티닌	PEYAYKIVKNKK MEDGFLQGMVD GWYGHHSNEQGS GLMENERTLDKA NPNNDLCSWSDH EASSNNTNQEDLL QRESRRKKRIGTS TLNQRGNCNTKC QTEEARLKREEVS LVKSDQCSNGSLQ CRANNSTEQVD (SEQ ID NO: 408)	CCCGAGTACGCCTACAAGATCGTGAA GAACAAGAAGATGGAGGACGGCTTCC TGCAGGGCATGGTGGACGGCTGGTAC GGCCACCACAGCAACGAGCAGGGCAG CGGCCTGATGGAGAACGAGAGGACCC TGGACAAGGCCAACCCCAACAACGAC CTGTGCAGCTGGAGCGACCACGAGGC CAGCAGCAACAACACCAACCAGGAGG ACCTGCTGCAGAGGGAGAGCAGGAGG AAGAAGAGGATCGGCACCAGCACCCT GAACCAGAGGGGCAACTGCAACACCA AGTGCCAGACCGAGGAGGCCAGGCTG AAGAGGGAGGAGGTGAGCCTGGTGA AGAGCGACCAGTGCAGCAACGGCAGC CTGCAGTGCAGGGCCAACAACAGCAC CGAGCAGGTGGAC (SEQ ID NO: 409)
7	뉴라미니다제	TKSTNSRSGGISG PDNEAPVGEAPSP YGDNPRPNDGNN IRIGSKGYNGIITD TIEESCSCYPDAK VVKSVELDSTIWT SGSSPNQKIITIGW DPNGWTGTPMSP NGAYGTDGPSNG QANQHQAESISA GNSSLCPIRDNWH GSNRSWSWPDGA E (SEQ ID NO: 410)	ACCAAGAGCACCAACAGCAGGAGCG GCGGCATCAGCGGCCCCGACAACGAG GCCCCCGTGGGCGAGGCCCCCAGCCC CTACGGCGACAACCCCAGGCCCAACG ACGGCAACAACATCAGGATCGGCAGC AAGGGCTACAACGGCATCATCACCGA CACCATCGAGGAGAGCTGCAGCTGCT ACCCCGACGCCAAGGTGGTGAAGAGC GTGGAGCTGGACAGCACCATCTGGAC CAGCGGCAGCAGCCCCAACCAGAAGA TCATCACCATCGGCTGGGACCCCAACG GCTGGACCGGCACCCCCATGAGCCCC AACGGCGCCTACGGCACCGACGGCCC CAGCAACGGCCAGGCCAACCAGCACC AGGCCGAGAGCATCAGCGCCGGCAAC AGCAGCCTGTGCCCCATCAGGGACAA CTGGCACGGCAGCAACAGGAGCTGGA GCTGGCCCGACGGCGCCGAG (SEQ ID NO: 411)
8	H5N1 바이러스 헤마글루티닌 상의 TAP1 (항원 제시 수송)	EKIVLLLAMMEKI VLLLAKCQTPMG AIKAVDGVTNKCP YLGSPSF (SEQ ID NO: 412)	GAGAAGATCGTGCTGCTGCTGGCCATG ATGGAGAAGATCGTGCTGCTGCTGGCC AAGTGCCAGACCCCCATGGGCGCCAT CAAGGCCGTGGACGGCGTGACCAACA AGTGCCCCTACCTGGGCAGCCCCAGC TTC (SEQ ID NO: 413)

[0991]

백신	질환	에피토프 아미노산 서열	예시적인 핵산 서열 (8)
9	H5N1 바이러스 뉴라미니다제 상의 TAP2 (항원 제시 수송)	IRPCFWVELNPNQKIITIRPCFWVELICYPDAGEIT (SEQ ID NO: 414)	ATCAGGCCCTGCTTCTGGGTGGAGCTGAACCCCAACCAGAAGATCATCACCATCAGGCCCTGCTTCTGGGTGGAGCTGATCTGCTACCCCGACGCCGGCGAGATCACC (SEQ ID NO: 415)
10	부류 I HLA에 대한 헤마글루티닌 에피토프	MEKIVLLLAEKIVLLLAMCPYLGSPSFKCQTPMGAIKAVDGVTNK (SEQ ID NO: 416)	ATGGAGAAGATCGTGCTGCTGCTGGCCGAGAAGATCGTGCTGCTGCTGGCCATGTGCCCCTACCTGGGCAGCCCCAGCTTCAAGTGCCAGACCCCCATGGGCGCCATCAAGGCCGTGGACGGCGTGACCAACAAG (SEQ ID NO: 417)
11	부류 I HLA에 대한 뉴라미니다제 에피토프	NPNQKIITICYPDAGEITIRPCFWVELRPCFWVELI (SEQ ID NO: 418)	AACCCCAACCAGAAGATCATCACCATCTGCTACCCCGACGCCGGCGAGATCACCATCAGGCCCTGCTTCTGGGTGGAGCTGAGGCCCTGCTTCTGGGTGGAGCTGATC (SEQ ID NO: 419)
12	부류 II HLA에 대한 헤마글루티닌 에피토프	MVSLVKSDQIGTSTLNQR (SEQ ID NO: 420)	ATGGTGAGCCTGGTGAAGAGCGACCAGATCGGCACCAGCACCCTGAACCAGAGG (SEQ ID NO: 421)
13	부류 II HLA에 대한 뉴라미니다제 에피토프	YNGIITDTI (SEQ ID NO: 422)	TACAACGGCATCATCACCGACACCATC (SEQ ID NO: 423)
14	헤마글루티닌 에피토프 H5N1-결합된	MEKIVLLLAEKIVLLLAMMVSLVKSDQCPYLGSPSFIGTSTLNQRKCQTPMGAIKAVDGVTNK (SEQ ID NO: 424)	ATGGAGAAGATCGTGCTGCTGCTGGCCGAGAAGATCGTGCTGCTGCTGGCCATGATGGTGAGCCTGGTGAAGAGCGACCAGTGCCCCTACCTGGGCAGCCCCAGCTTCATCGGCACCAGCACCCTGAACCAGAGG (SEQ ID NO: 425)

[0992]

백신	질환	에피토프 아미노산 서열	예시적인 핵산 서열 (8)
	부류 I 및 부류 II HLA		
15	뉴라미니다제 에피토프 H5N1-결합된 부류 I 및 부류 II HLA	NPNQKIITIYNGIITDTICYPDAGEITIRPCFWVELRPCFWVELI (SEQ ID NO: 426)	AACCCCAACCAGAAGATCATCACCATCTACAACGGCATCATCACCGACACCATCTGCTACCCCGACGCCGGCGAGATCACCATCAGGCCCTGCTTCTGGGTGGAGCTGAGGCCCTGCTTCTGGGTGGAGCTGATC (SEQ ID NO: 427)

[0993]

[0994] 관련 기술분야에 공지된 추가의 이뮤노에피토프가 또한 설치될 수 있다. 이뮨 에피토프 데이터베이스 및 분석

자원 (iedb.org/epitopedetails_v3.php) (그의 내용은 본원에 참조로 포함됨)으로부터 입수가능한 이뮤노에피토프 중 임의의 것은 본원에 개시된 프라임 편집제에 의해 설치될 수 있다.

[0995] 일부 실시양태에서, 본원에 개시된 프라임 편집제에 의해 설치될 수 있는 이뮤노에피토프는 하기 에피토프 중 임의의 것을 포함할 수 있다:

[0996]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1	파상풍 톡소이드	QYIKANSKFIGITEL	396	NA
2	디프테리아 독소 돌연변이체 CRM197	GADDVVDSSKSFVMENFSSYHGTKP GYVDSIQKGIQKPKSGTQGNYDDDW KEFYSTDNKYDAAGYSVDNENPLSG KAGGVVKVTYPGLTKVLALKVDNA ETIKKELGLSLTEPLMEQVGTEEFIKR FGDGASRVVLSLPFAEGSSSVEYINN WEQAKALSVELEINFETRGKRGQDA MYEYMAQACAGNRVRRSVGSSLSCI NLDWDVIRDKTKTKIESLKEHGPIKN KMSESPNKTVSEEKAKQYLEEFHQT ALEHPELSELKTVTGTNPVFAGANYA AWAVNVAQVIDSETADNLEKTTAALS ILPGIGSVMGIADGAVHHNTEEIVAQS IALSSLMVAQAIPLVGELVDIGFAAYN FVESIINLFQVVHNSYNRPAYSPGHKT QPFLHDGYAVSWNTVEDSIIRTGFQG ESGHDIKITAENTPLPIAGVLLPTIPGK LDVNKSKTHISVNGRKIRMRCRAIDG DVTFCRPKSPVYVGNGVHANLHVAF HRSSSEKIHSNEISSDSIGVLGYQKTV DHTKVNSKLSLFFEIKS	428	NA
3	볼거리	GTYRLIPNARANLTA	400	NA
4	볼거리	PSKFFTISDSATFAPGPVSNA; PSKLFIMLDNATFAPGPVVNA	402; 404	NA
5	풍진 바이러스 (RV)	TPPPYQVSCGGESDRASARVIDPAAQ S	406	NA
6	헤마글루티닌	PEYAYKIVKNKKMEDGFLQGMVDG WYGHHSNEQGSGLMENERTLDKAN PNNDLCSWSDHEASSNNTNQEDLLQ RESRRKKRIGTSTLNQRGNCNTKCQT EEARLKREEVSLVKSDQCSNGSLQCR ANNSTEQVD	408	NA
7	뉴라미니다제	TKSTNSRSGGISGPDNEAPVGEAPSP YGDNPRPNDGNNIRIGSKGYNGIITD TIEESCSCYPDAKVVKSVELDSTIWTS GSSPNQKIITIGWDPNGWTGTPMSPN GAYGTDGPSNGQANQHQAESISAGN SSLCPIRDNWHGSNRSWSWPDGAE	410	NA
8	H5N1 바이러스 헤마글루티닌	EKIVLLLAMMEKIVLLLAKCQTPMG AIKAVDGVTNKCPYLGSPSF	412	NA

[0997]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
	상의 TAP (항원 제시 수송)			
9	H5N1 바이러스 뉴라미니다제 상의 TAP (항원 제시 수송)	IRPCFWVELNPNQKIITIRPCFWVELIC YPDAGEIT	414	NA
10	부류 I HLA에 대한 헤마글루티닌 에피토프	MEKIVLLLAEKIVLLLAMCPYLGSPS FKCQTPMGAIKAVDGVTNK	416	NA
11	부류 I HLA에 대한 뉴라미니다제 에피토프	NPNQKIITICYPDAGEITIRPCFWVELR PCFWVELI	418	NA
12	부류 II HLA에 대한 헤마글루티닌 에피토프	MVSLVKSDQIGTSTLNQR	420	NA
13	부류 II HLA에 대한 뉴라미니다제 에피토프	YNGIITDTI	422	NA
14	헤마글루티닌 에피토프 H5N1-결합된 부류 I 및 부류 II HLA	MEKIVLLLAEKIVLLLAMMVSLVKSD QCPYLGSPSFIGTSTLNQRKCQTPMG AIKAVDGVTNK	424	NA
15	뉴라미니다제 에피토프 H5N1-결합된 부류 I 및 부류 II HLA	NPNQKIITIYNGIITDTICYPDAGEITIR PCFWVELRPCFWVELI	426	NA
16	코리네박테리움 디프테리아에	AACAGNRVRRSVGSSLKC	899	SRC280292
17	홍역 바이러스 균주 에드몬스톤	AADHCPVVEVNGVTI	900	P69353.1
18	홍역 바이러스 균주 에드몬스톤	AAHLPTGTPLDID	901	P04851.1
19	보르데텔라 페르투시스	AALAVWAGLAVQ	902	Q00879.1
20	홍역 바이러스 균주 에드몬스톤	AALGVATAAQITAGI	903	P69353.1
21	풍진 바이러스 균주 테리엔	AALLNTPPPYQVSCGGESDRATAR	904	P07566.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
22	풍진 바이러스	AAQSFTGVVYGTHTT	905	BAA28178.1
23	풍진 바이러스	ACEVEPAFGHSDAAC	906	BAA28178.1
24	풍진 바이러스	ACTFWAVNAYSSGGY	907	BAA28178.1
25	풍진 바이러스	ADDPLLR	908	BAA19893.1
26	풍진 바이러스	ADDPLLRT	909	CAJ88851.1
27	홍역 바이러스 균주 에드몬스톤	AEMICDIDTYIVEAG	910	P04851.1
28	홍역 바이러스 균주 에드몬스톤	AEMICDIDTYIVEAGLASFI	911	P04851.1
29	홍역 바이러스 균주 에드몬스톤	AEPLLSC	912	P04851.1
30	보르데텔라 페르투시스	AFVSTSSSRRYTEVY	913	CAD44970.1
31	보르데텔라 페르투시스	AGFIYRETFCITTIYKTGQPAADHYYSKVTA	914	P04979.1
32	풍진 바이러스	AGLLACCAKCLYYLR	915	BAA28178.1
33	풍진 바이러스	AHTTSDPWHPPG	916	BAA19893.1
34	홍역 바이러스 균주 에드몬스톤	AIAKLEDAKELLESS	917	P69353.1
35	홍역 바이러스 균주 에드몬스톤	AIDNLRASLETTNQA	918	P69353.1
36	보르데텔라 페르투시스	AKGVEFR	919	ACI16088.1
37	홍역 바이러스 균주 에드몬스톤	AKWAVPTTRTDDKLR	920	P08362.1
38	홍역 바이러스 균주 에드몬스톤	ALAEVLKKPV	921	ABO69699.1
39	홍역 바이러스 균주 에드몬스톤	ALGVINTLEWIPRFK	922	P08362.1
40	홍역 바이러스 균주 에드몬스톤	ALHQSMLNSQAIDNL	923	P69353.1
41	홍역 바이러스 균주 에드몬스톤	ALIGILSLFV	924	ABI54110.1
42	풍진 바이러스	ALLNTPPPYQVSCGGESDRA	925	CAJ88851.1

[0998]

[0999]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
43	풍진 바이러스 균주 M33	ALLNTPPPYQVSCGGESDRASARV	926	CAJ88851.1
44	풍진 바이러스	ALVEGLAPGGGNCHL	927	BAA28178.1
45	보르데텔라 페르투시스	AMAAWSERAGEA	928	P04977.1
46	홍역 바이러스 균주 에드몬스톤	ANCASILCKCYTTGT	929	P69353.1
47	보르데텔라 페르투시스	ANPNPYTSRRSV	930	P04977.1
48	풍진 바이러스	APGPGEVW	931	CAJ88851.1
49	풍진 바이러스	APLPPHTTERIETRSARHPWRIR	932	ABD64214.1
50	풍진 바이러스 백신 균주 RA27/3	APPMPPQPPRAHGQHYGHHHHQLPFLG	933	CAA33016.1
51	풍진 바이러스 균주 테리엔	APPTLPQPPCAHGQHYGHHHHQLPFLG	934	P07566.1
52	풍진 바이러스	APPTLPQPPRAHGQHYGHHHHQLPFLG	935	ABD64214.1
53	보르데텔라 페르투시스	APQPGPQPPQPPQPQPEAPAPQ	936	P14283.3
54	풍진 바이러스	AQLASYFNPGGSYYK	937	BAA28178.1
55	홍역 바이러스 균주 에드몬스톤	ARAAHLPTGTPLD	938	P04851.1
56	홍역 바이러스 균주 에드몬스톤	ARLVSEIAMHTTEDK	939	P04851.1
57	홍역 바이러스 균주 에드몬스톤	ARLVSEIAMHTTEDKISRAV	940	P04851.1
58	홍역 바이러스 균주 에드몬스톤-B	ASDVETAEGGEIHELLR	941	P03422.1
59	홍역 모르빌리바이러스	ASDVETAEGGEIHELLRLQ	942	ABO69699.1
60	홍역 바이러스 균주 에드몬스톤-B	ASDVETAEGGEIHELLRLQSR	943	P03422.1
61	홍역 모르빌리바이러스	ASDVETAEGGEIHKLLRLQ	944	AAA63285.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
62	풍진 바이러스 균주 M33	ASDVLPGHWLQG	945	NP_740663.1
63	홍역 바이러스 균주 에드몬스톤	ASELGITAEDARLVS	946	P04851.1
64	홍역 바이러스 균주 에드몬스톤	ASELGITAEDARLVSEIAMH	947	P04851.1
65	홍역 바이러스 균주 에드몬스톤	ASESSQDPQDSRR	948	P04851.1
66	홍역 바이러스 균주 에드몬스톤	ASILCKCYTTGTIIN	949	P69353.1
67	풍진 바이러스	ASPVCQRHSPDCSRL	950	BAA28178.1
68	보르데텔라 페르투시스	ASQARWTGATRA	951	BAF35031.1
69	풍진 바이러스 균주 테리엔	ASYFNPGGSYYKQYHPTACEVEPAFGHS	952	P07566.1
70	홍역 바이러스 균주 에드몬스톤	ASYKVMTRSSHQSLV	953	P69353.1
71	보르데텔라 페르투시스	ASYVKKPKEDVD	954	ACI16088.1
72	홍역 바이러스 균주 에드몬스톤	ATAAQITAGIALHQS	955	P69353.1
73	풍진 바이러스	ATPERPRL	956	CAJ88851.1
74	홍역 바이러스 균주 에드몬스톤	AVCLGGLIGIPALIC	957	P69353.1
75	홍역 바이러스 균주 에드몬스톤	AVGPRQAQVSF	958	P04851.1
76	풍진 바이러스	AVNAYSSGGYAQLAS	959	BAA28178.1
77	풍진 바이러스	AVSETRQTWAEWAAA	960	BAA28178.1
78	홍역 바이러스 균주 에드몬스톤	AVTAPDTAADSELRR	961	P04851.1
79	홍역 바이러스 균주 에드몬스톤	AVTAPDTAADSELRRWIKYT	962	P04851.1
80	보르데텔라 페르투시스	AYGGIIKDAPPGAGFIYRETFC	963	P04979.1
81	풍진 바이러스	CALPLAGLLACCAKC	964	BAA28178.1

[1000]

[1001]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
82	풍진 바이러스	CARIWNGTQRACTFW	965	BAA28178.1
83	홍역 바이러스 균주 에드몬스톤	CARTLVSGSFGNRFI	966	P69353.1
84	보르데텔라 페르투시스	CASPYEGRYRDMYDALRBRLLY	967	SRC280066
85	보르데텔라 페르투시스	CAVFVRSGQPVIGA	968	AAA83981.1
86	풍진 바이러스	CCAKCLYYLRGAIAPR	969	BAA28178.1
87	홍역 바이러스 균주 에드몬스톤	CCRGRCNKKGEQVGM	970	P69353.1
88	풍진 바이러스	CEIPTDVSCEGLGAW	971	BAA28178.1
89	보르데텔라 페르투시스	CFGKDLKRPGSSPMEV	972	P0A3R5.1
90	홍역 모르빌리바이러스	CFQQACKGKIQALCE	973	P06830.1
91	홍역 모르빌리바이러스	CFQQACKGKIQALCENPEWAPLKDNRIPS	974	AAR89413.1
92	풍진 바이러스	CGGESDRASARVIDP	975	BAA28178.1
93	보르데텔라 페르투시스	CITTIYKTGQPAADHYYSKVTA	976	P04979.1
94	홍역 모르빌리바이러스	CKGKIQALCENPEWA	977	AAR89413.1
95	홍역 바이러스 균주 에드몬스톤	CKPWQESRKNKAQ	978	P04851.1
96	홍역 바이러스 균주 에드몬스톤	CNKKGEQVGMSRPGL	979	P69353.1
97	풍진 바이러스	CNVTTEHPFCNTPHG	980	BAA28178.1
98	보르데텔라 페르투시스	CQVGSSNSAF	981	P04977.1
99	홍역 모르빌리바이러스	CSGPTTIRGQFS	982	P08362.1
100	보르데텔라 페르투시스	CTSPYDGKYWSMYSRL	983	AAA83981.1
101	홍역 바이러스 균주 에드몬스톤	CVLADSESGGHITHS	984	P08362.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
102	홍역 모르빌리바이러스	CYTTGTIINQDPDKILTYIAADHC	985	AAF02706.1
103	풍진 바이러스	DADDPLLR	986	CAJ88851.1
104	홍역 바이러스 균주 에드몬스톤	DARAAHLPTGTPLDI	987	P04851.1
105	홍역 바이러스 균주 에드몬스톤	DARAAHLPTGTPLDIDTASE	988	P04851.1
106	홍역 바이러스 균주 에드몬스톤	DCHAPTYLPAEVDGD	989	P08362.1
107	풍진 바이러스	DCSRLVGATPERPRL	990	BAA28178.1
108	홍역 바이러스 균주 에드몬스톤	DDKLRMETCFQQACK	991	P08362.1
109	풍진 바이러스	DDPLLRTA	992	CAJ88851.1
110	풍진 바이러스	DDPLLRTAPGPGEVW	993	BAA28178.1
111	보르데텔라 페르투시스	DDPPATVYRYD	994	P04977.1
112	보르데텔라 페르투시스	DDPPATVYRYDSRPPED	995	CAD44970.1
113	보르데텔라 페르투시스	DDPPATVYRYDSRPPEDV	996	ACI04548.1
114	홍역 모르빌리바이러스	DEVGLRTPQRFTDLV	997	P06830.1
115	홍역 바이러스 균주 에드몬스톤	DHCPVVEVNGVTIQV	998	P69353.1
116	홍역 바이러스 균주 에드몬스톤	DIDTASESSQDPQ	999	P04851.1
117	홍역 바이러스 균주 에드몬스톤	DINKVLEKLGYSGGD	1000	P69353.1
118	보르데텔라 페르투시스	DLIAYKQ	1001	ACI16088.1
119	홍역 바이러스 균주 에드몬스톤	DLIGQKLGLKLLRYY	1002	P69353.1
120	풍진 바이러스	DLQKALEAQSRALRAELAA	1003	P07566.1
121	홍역 모르빌리바이러스	DLQYVLATYDTSRVE	1004	P06830.1
122	홍역 바이러스 균주 에드몬스톤	DLSLRRFMV	1005	P04851.1

[1002]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
123	홍역 바이러스 균주 에드몬스톤	DLSNCMVALGELKLA	1006	P08362.1
124	풍진 바이러스	DLVEYIMNYTGNQQSRWGLGSPNC	1007	CAJ88851.1
125	홍역 모르빌리바이러스	DLVKFISDKIKFLNP	1008	AAR89413.1
126	홍역 모르빌리바이러스	DLVKFISTKIKFLNP	1009	SRC280117
127	홍역 바이러스 균주 에드몬스톤	DLYKSNHNNV	1010	P08362.1
128	보르데텔라 페르투시스	DNVLDHLTGR	1011	ACI04548.1
129	보르데텔라 페르투시스	DNVLDHLTGRSC	1012	P04977.1
130	보르데텔라 페르투시스	DNVLDHLTGRSCQ	1013	P04977.1
131	홍역 모르빌리바이러스	DPDKILTYIAA	1014	AAF02706.1
132	풍진 바이러스 균주 테리엔	DPGDLVEYIMNYTGNQQSR	1015	P07566.1
133	풍진 바이러스	DPLLRTAP	1016	CAJ88851.1
134	풍진 바이러스	DPLLRTAPGPGEVWVTPVIGSQ	1017	CAJ88851.1
135	홍역 바이러스 균주 에드몬스톤	DPQDSRRSAEPLL	1018	P04851.1
136	홍역 바이러스 균주 에드몬스톤	DPVIDRLYLSSHRGV	1019	P08362.1
137	홍역 바이러스 균주 에드몬스톤	DQILRSMKGLSSTSI	1020	P69353.1
138	홍역 모르빌리바이러스	DQYCADVAAEELMNA	1021	P06830.1
139	홍역 바이러스 균주 에드몬스톤	DSESGGHITH	1022	P08362.1
140	홍역 바이러스 균주 에드몬스톤	DTASESSQDPQDS	1023	P04851.1
141	풍진 바이러스	DTVMSVFALASYVQH	1024	BAA28178.1
142	보르데텔라 페르투시스	DVFQNGFTAWGNND	1025	P04977.1

[1003]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
143	풍진 바이러스	DVGAVPPGKFVTAAL	1026	BAA28178.1
144	코리네박테리움 디프테리아에	DVNKSKTHISVNGRKI	1027	CAE11230.1
145	풍진 바이러스	DVSCEGLGAWVPAAP	1028	BAA28178.1
146	풍진 바이러스 균주 테리엔	DWASPVCQRHSPDCSRLVGATPERPRL	1029	P07566.1
147	홍역 바이러스 균주 에드몬스톤	EARESYRETGPSR	1030	P04851.1
148	보르데텔라 페르투시스	EAVEAERAGRGTG	1031	ACI04548.1
149	홍역 바이러스 균주 에드몬스톤	EDAKELLESSDQILR	1032	P69353.1
150	홍역 바이러스 균주 에드몬스톤	EDRRVKQSRGEAR	1033	P04851.1
151	홍역 바이러스 균주 에드몬스톤	EDSITIPYQGSGKGV	1034	P08362.1
152	풍진 바이러스	EEAFTYLCTAPGCAT	1035	BAA28178.1
153	홍역 바이러스 균주 에드몬스톤	EGFNMILGTILAQIW	1036	P04851.1
154	홍역 바이러스 균주 에드몬스톤	EGFNMILGTILAQIWVLLAK	1037	P04851.1
155	풍진 바이러스	EHPFCNTPHGQLEVQ	1038	BAA28178.1
156	홍역 바이러스 균주 에드몬스톤	EISDIEVQDPEGFNM	1039	P04851.1
157	홍역 바이러스 균주 에드몬스톤	EISDIEVQDPEGFNMILGTI	1040	P04851.1
158	홍역 바이러스 균주 에드몬스톤	EKPNLSSKRSE	1041	P08362.1
159	홍역 바이러스 균주 에드몬스톤	ELKLAALCHGEDSIT	1042	P08362.1
160	홍역 모르빌리바이러스	ELMNALVNSTLLETR	1043	P06830.1
161	홍역 바이러스 균주 에드몬스톤	ELPRL	1044	P04851.1
162	홍역 모르빌리바이러스	ENPEWAPLKDNRIPSYGVLSVDL	1045	AAR89413.1

[1004]

[1005]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
163	홍역 바이러스 균주 에드몬스톤	EPIRDALNAMTQNIR	1046	P69353.1
164	홍역 바이러스 균주 에드몬스톤	EQVGMSRPGLKPDLT	1047	P69353.1
165	풍진 바이러스	ERPRLRLV	1048	CAJ88851.1
166	풍진 바이러스	ERPRLRLVDADDPLL	1049	BAA28178.1
167	홍역 바이러스 CAM/RB	ESPGQLIQRITDDPDVS	1050	P04851.1
168	홍역 바이러스 균주 에드몬스톤	ESRGIKARITHVDTE	1051	P69353.1
169	홍역 바이러스 균주 에드몬스톤	ESSCTFMPEGTVCSQ	1052	P69353.1
170	홍역 바이러스 균주 에드몬스톤	ESSQDPQDSRRSA	1053	P04851.1
171	홍역 바이러스 균주 에드몬스톤	ETRTTNQFLAVSKGN	1054	P08362.1
172	홍역 바이러스 균주 에드몬스톤	EVDGDVKLSSNLVIL	1055	P08362.1
173	홍역 바이러스 균주 에드몬스톤-B	EVNGVTIQV	1056	P26031.1
174	풍진 바이러스	EVWVTPVI	1057	CAJ88851.1
175	풍진 바이러스	EVWVTPVIGSQA	1058	BAA19893.1
176	풍진 바이러스	EWAAAHWWQLTLGAT	1059	BAA28178.1
177	홍역 모르빌리바이러스	EWIPRFKVSPYLFTV	1060	P06830.1
178	보르데텔라 페르투시스	FEYVDTYGDNAG	1061	P04977.1
179	홍역 모르빌리바이러스	FGPLITHGSGMDLYK	1062	P06830.1
180	홍역 바이러스 균주 에드몬스톤	FIFDALAEV	1063	ABK40531.1
181	홍역 바이러스 균주 에드몬스톤	FISDKIKFL	1064	P08362.1
182	홍역 바이러스 균주 에드몬스톤	FKRNKDKPPITSGSG	1065	P04851.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
183	홍역 바이러스 균주 에드몬스톤	FKRNKDKPPITSGSGGAIRG	1066	P04851.1
184	풍진 바이러스	FKTVRPVALPRTLAP	1067	BAA28178.1
185	홍역 바이러스 균주 에드몬스톤	FLMDRHIIV	1068	ABK40531.1
186	홍역 바이러스 균주 에드몬스톤	FMAVLLTLQTPTGQI	1069	P69353.1
187	홍역 바이러스 균주 에드몬스톤	FMPEGTVCSQNALYP	1070	P69353.1
188	홍역 모르빌리바이러스	FMYMSLLGV	1071	AAN09804.1
189	홍역 바이러스 균주 에드몬스톤	FNVPIKEAGEDCHAP	1072	P08362.1
190	홍역 모르빌리바이러스	FRDLTWCINPPERIK	1073	AAC35876.2
191	홍역 바이러스 균주 에드몬스톤	FSHDDPISSDQSRFG	1074	P04851.1
192	홍역 바이러스 균주 에드몬스톤	FSHDDPISSDQSRFGWFENK	1075	P04851.1
193	홍역 모르빌리바이러스	FTDLVKFISDKIKFL	1076	P06830.1
194	홍역 모르빌리바이러스	FTWDQKLWCRHFCVL	1077	P06830.1
195	보르데텔라 페르투시스	FVRDGQSVIGACASPYEGRYRDLYDALRRLLY	1078	SRC280066
196	보르데텔라 페르투시스	FVRSGQPVIGACTSPYDGKYWSILYSRLRKMLY	1079	SRC280066
197	홍역 모르빌리바이러스	FYKDNPHPKGSRIVI	1080	P06830.1
198	보르데텔라 페르투시스	GAASSYFEYVDTYG	1081	ACI04548.1
199	보르데텔라 페르투시스	GAFDLKTTFCIMTTRNTGQPA	1082	AAA83981.1
200	홍역 바이러스 균주 에드몬스톤	GALIGILSLFVESPG	1083	P04851.1
201	홍역 바이러스 균주 에드몬스톤	GALIGILSLFVESPGQLIQR	1084	P04851.1
202	풍진 바이러스	GATPERPR	1085	CAJ88851.1

[1006]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
203	보르데텔라 페르투시스	GAYGRCPNGTRALTVAELRGNAEL	1086	P04979.1
204	풍진 바이러스	GCFAPWDLEATGACI	1087	BAA28178.1
205	홍역 모르빌리바이러스	GDINKVLEKLGYS	1088	BAB60865.1
206	홍역 모르빌리바이러스	GDINKVLEKLGYSGGDLLG	1089	AAL29688.1
207	풍진 바이러스	GDLRAVHHRPVPA	1090	CAA28880.1
208	풍진 바이러스	GDLVEYIMNYTGNQQ	1091	BAA28178.1
209	홍역 바이러스 균주 에드몬스톤	GDSSITTRSRLLDRL	1092	P04851.1
210	홍역 바이러스 균주 에드몬스톤	GDSSITTRSRLLDRLVRLIG	1093	P04851.1
211	홍역 모르빌리바이러스	GEDCHAPTYLPAEVD	1094	P06830.1
212	홍역 바이러스 균주 에드몬스톤	GELSTLESLMNLYQQ	1095	P04851.1
213	홍역 바이러스 균주 에드몬스톤	GELSTLESLMNLYQQMGKPA	1096	P04851.1
214	볼거리 루불라바이러스	GEQARYLALLEA	1097	P21186.1
215	풍진 바이러스	GEVWVT	1098	BAA19893.1
216	풍진 바이러스	GEVWVTPV	1099	CAJ88851.1
217	풍진 바이러스	GEVWVTPVIGSQAR	1100	BAA19893.1
218	보르데텔라 페르투시스	GEYGGVIKDGTPGGA	1101	AAA83981.1
219	풍진 바이러스	GFLSGVGPMRLRHGADT	1102	SRC265968
220	홍역 바이러스 균주 에드몬스톤-B	GFRASDVETAEGGEIHELLRLQ	1103	P03422.1
221	보르데텔라 페르투시스	GGAVPGGAVPGGAVPGGFGPGGFGP	1104	P14283.3
222	보르데텔라 페르투시스	GGAVPGGAVPGGFGPGGFGPGGFGP	1105	CAA09475.1

[1007]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
223	보르데텔라 페르투시스	GGAVPGGAVPGGFGPGGFGPGGFGPGGFGP	1106	CAA09474.1
224	홍역 모르빌리바이러스	GGHITHSGMVGMGVS	1107	P06830.1
225	홍역 모르빌리바이러스	GILESRGIKARITHVDTESY	1108	P26032.1
226	보르데텔라 페르투시스	GITGETTTTEYSNARYV	1109	CAD44970.1
227	홍역 바이러스 균주 에드몬스톤	GKEDRRVKQSRGE	1110	P04851.1
228	보르데텔라 페르투시스	GKVTNGS	1111	ACI16088.1
229	풍진 바이러스	GLGAWVPAAPCARIW	1112	BAA28178.1
230	홍역 바이러스 균주 에드몬스톤	GLIGIPALICCCRGR	1113	P69353.1
231	풍진 바이러스 균주 테리엔	GLLACCAKCLYYLRGAIAPR	1114	P07566.1
232	홍역 모르빌리바이러스	GLLAIAGIRLHRAAI	1115	P06830.1
233	풍진 바이러스	GLQPRADMAAPPTLPQ	1116	NP_740663.1
234	홍역 모르빌리바이러스	GMGVSCTVTREDGTNRR	1117	AAR89413.1
235	홍역 모르빌리바이러스	GMYGGTYLVEKP	1118	AAR89413.1
236	보르데텔라 페르투시스	GNAELQTYLRQITPGWSIYGLYDGTYLG	1119	P04979.1
237	풍진 바이러스	GNCHLTVNGEDVGAV	1120	BAA28178.1
238	보르데텔라 페르투시스	GNNDNVLDHLTGR	1121	P04977.1
239	보르데텔라 페르투시스	GNNDNVLDHLTGRSC	1122	P04977.1
240	홍역 바이러스 균주 에드몬스톤	GNRFILSQGNLIANC	1123	P69353.1
241	풍진 바이러스	GNRGRGQRRDWSRAPPPPEERQETRSQTPAPKPS	1124	P07566.1
242	풍진 바이러스	GPGEVWVT	1125	CAJ88851.1

[1008]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
243	풍진 바이러스 균주 테리엔	GPMRLRHGADTRCGRLI	1126	P07566.1
244	보르데텔라 페르투시스	GPNHTKV	1127	ACI16083.1
245	홍역 바이러스 균주 할레	GPRQAQVSF	1128	P10050.1
246	홍역 바이러스 균주 에드몬스톤	GPRQAQVSFLQGDQS	1129	P04851.1
247	홍역 바이러스 균주 에드몬스톤	GPRQAQVSFLQGDQSENELP	1130	P04851.1
248	홍역 모르빌리바이러스	GRGYNVSSIVTMTSQ	1131	P06830.1
249	보르데텔라 페르투시스	GRTPFII	1132	ACI16083.1
250	풍진 바이러스	GSPNCHGPDWASPVC	1133	BAA28178.1
251	풍진 바이러스	GSQARKCGLHIRAGP	1134	BAA28178.1
252	보르데텔라 페르투시스	GSSNSAFVSTSSSRR	1135	P04977.1
253	홍역 바이러스 균주 에드몬스톤	GSTKSCARTLVSGSF	1136	P69353.1
254	풍진 바이러스	GSYYKQYHPTACEVE	1137	BAA28178.1
255	풍진 바이러스	GTHTTAVSETRQTWA	1138	BAA28178.1
256	홍역 바이러스 균주 에드몬스톤	GTIINQDPDKILTYI	1139	P69353.1
257	보르데텔라 페르투시스	GTLVRMAPVIG	1140	ADA85124.1
258	홍역 바이러스 균주 에드몬스톤	GTPLDIDTASESS	1141	P04851.1
259	보르데텔라 페르투시스	GTYLGQAYGGIIKDAPPGAGFIYRETFC	1142	P04979.1
260	보르데텔라 페르투시스	GVATKGLGVHAKSSDWG	1143	P15318.2
261	코리네박테리움 디프테리아에	GVLLPTIPGKLDVNKSKTHI	1144	AAV70486.1
262	홍역 모르빌리바이러스	GVLSVDLSLTVELKI	1145	P06830.1

[1009]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
263	홍역 모르빌리바이러스	GVPIELQVECFTWDQ	1146	P06830.1
264	홍역 바이러스 균주 에드몬스톤	GVSCTVTREDGTNRR	1147	P08362.1
265	홍역 바이러스 균주 에드몬스톤	GVSYNIGSQEWYTTV	1148	P69353.1
266	홍역 바이러스 균주 에드몬스톤	GYNVSSIVTMTSQGM	1149	P08362.1
267	홍역 모르빌리바이러스	HFCVLADSESGGHIT	1150	P06830.1
268	홍역 모르빌리바이러스	HGEDSITIPYQGSGK	1151	P06830.1
269	풍진 바이러스	HGPDWASP	1152	BAA19893.1
270	풍진 바이러스	HGPDWASPVCQRHSP	1153	BAA28178.1
271	풍진 바이러스	HGPDWASPVCQRHSPDCSRLVG	1154	CAJ88851.1
272	풍진 바이러스 균주 M33	HGPDWASPVCQRHSPDCSRLVGATPERPRLRLV	1155	CAJ88851.1
273	홍역 바이러스 균주 에드몬스톤	HITHSGMEGMGVSCT	1156	P08362.1
274	홍역 모르빌리바이러스	HKSLSTNLDVTNSIE	1157	P06830.1
275	홍역 바이러스 균주 에드몬스톤	HLMIDRPYV	1158	P08362.1
276	홍역 바이러스 균주 에드몬스톤	HLPTGTPLDIDTA	1159	P04851.1
277	홍역 바이러스 균주 에드몬스톤-B	HLPTGTPLDIDTATESSQDPQDSR	1160	Q77M43.1
278	홍역 모르빌리바이러스	HMTNYLEQPVSNDLS	1161	P06830.1
279	홍역 바이러스 균주 에드몬스톤	HQSLVIKLMPNITLL	1162	P69353.1
280	홍역 모르빌리바이러스	HRAAIYTAEIHKSLS	1163	P06830.1
281	보르데텔라 페르투시스	HRMQEAVEAERAGRGTGH	1164	P04977.1
282	홍역 바이러스 균주 에드몬스톤	HVDTESYFIVLSIAY	1165	P69353.1

[1010]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
283	홍역 바이러스 균주 에드몬스톤	HWGNLSKIGVVGIGS	1166	P69353.1
284	풍진 바이러스	HWWQLTLGATCALPL	1167	BAA28178.1
285	풍진 바이러스	HYRNASDVLPGHWLQGGWGCYNL	1168	NP_740663.1
286	홍역 바이러스 균주 에드몬스톤	IDLGPPISLERLDVG	1169	P69353.1
287	홍역 바이러스 균주 에드몬스톤	IEAIRQAGQEMILAV	1170	P69353.1
288	풍진 바이러스 균주 M33	IETRSARHP	1171	CAA28880.1
289	홍역 바이러스 균주 에드몬스톤	IGSQEWYTTVPKYVA	1172	P69353.1
290	홍역 바이러스 균주 에드몬스톤	IKGVIVHRLEGVSYN	1173	P69353.1
291	홍역 바이러스 균주 에드몬스톤	IKHIIIVPIPGDSSI	1174	P04851.1
292	홍역 바이러스 균주 에드몬스톤	IKHIIIVPIPGDSSITTRSR	1175	P04851.1
293	보르데텔라 페르투시스	IKLKDCP	1176	ACI16083.1
294	홍역 바이러스 균주 에드몬스톤	IKLMPNITLLNNCTR	1177	P69353.1
295	홍역 모르빌리바이러스	ILLERLDVGT	1178	AAF85664.1
296	홍역 모르빌리바이러스	ILPGQDLQYV	1179	P08362.1
297	홍역 바이러스 균주 에드몬스톤	ILTYIAADHCPVVEV	1180	P69353.1
298	홍역 모르빌리바이러스	INQDPDKILTY	1181	AAL29688.1
299	홍역 바이러스 균주 에드몬스톤	IPRFKVSPYLFNVPI	1182	P08362.1
300	홍역 바이러스 균주 에드몬스톤	IQALSYALGGDINKV	1183	P69353.1
301	풍진 바이러스	IRAGPYGHATVEMPE	1184	BAA28178.1
302	보르데텔라 페르투시스	IRMGTDK	1185	ACI16088.1

[1011]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
303	홍역 바이러스 균주 에드몬스톤	ISNFDESSCTFMPEG	1186	P69353.1
304	홍역 바이러스 균주 에드몬스톤	ITAGIALHQSMLNSQ	1187	P69353.1
305	홍역 바이러스 균주 에드몬스톤	ITDDPDVSIRLLEVV	1188	P04851.1
306	홍역 바이러스 균주 에드몬스톤	ITDDPDVSIRLLEVVQSDQS	1189	P04851.1
307	보르데텔라 페르투시스	ITTYV	1190	ACI16083.1
308	홍역 바이러스 균주 에드몬스톤	IVEAGLASFILTIKF	1191	P04851.1
309	홍역 바이러스 균주 에드몬스톤	IVEAGLASFILTIKFGIETM	1192	P04851.1
310	홍역 바이러스 균주 에드몬스톤	IVEAGLASFILTIKFGIETMYPALG	1193	P04851.1
311	풍진 바이러스	KALEAQSRALRAELAA	1194	P07566.1
312	홍역 바이러스 균주 에드몬스톤	KARITHVDTESYFIV	1195	P69353.1
313	풍진 바이러스	KCGLHIRAGPYGHAT	1196	BAA28178.1
314	홍역 바이러스 균주 에드몬스톤	KCYTTGTIINQDPDK	1197	P69353.1
315	홍역 바이러스 균주 에드몬스톤	KDNPHPKGSR	1198	P08362.1
316	홍역 바이러스 균주 에드몬스톤	KDNRIPSYGVLSVDL	1199	P08362.1
317	홍역 모르빌리바이러스	KFLNPDREYDFRDLT	1200	AAC35876.2
318	풍진 바이러스	KFVTAALLN	1201	BAA28178.1
319	홍역 바이러스 균주 에드몬스톤	KGNCSGPTTIR	1202	P08362.1
320	홍역 모르빌리바이러스	KIKFLNPDREYDFRD	1203	P06830.1
321	풍진 바이러스	KIVDGGCFAPWDLEA	1204	BAA28178.1
322	홍역 바이러스 균주 에드몬스톤	KLGLKLLRYYTEILS	1205	P69353.1
323	홍역 바이러스 균주 에드몬스톤	KLGVWKSPTDMQSWV	1206	P08362.1

[1012]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
324	보르데텔라 페르투시스	KLKECPQ	1207	ACI16088.1
325	홍역 모르빌리바이러스	KLLRYYTEI	1208	P26031.1
326	홍역 바이러스 균주 에드몬스톤	KLMPFSGDFV	1209	ABK40531.1
327	홍역 모르빌리바이러스	KLMPNITLL	1210	P26031.1
328	홍역 모르빌리바이러스	KLRMETCFQQACKGKIQALCENPEWA	1211	AAR89413.1
329	홍역 모르빌리바이러스	KLWCRHFCV	1212	P08362.1
330	홍역 바이러스 균주 에드몬스톤	KLWCRHFCVL	1213	P08362.1
331	홍역 모르빌리바이러스	KLWESPQEI	1214	BAB60863.1
332	홍역 바이러스 균주 에드몬스톤	KMSSAVGFV	1215	ABO69699.1
333	홍역 바이러스 균주 에드몬스톤-B	KMSSAVGFVPDTGPASR	1216	P03422.1
334	보르데텔라 페르투시스	KMVYATN	1217	ACI16083.1
335	홍역 바이러스 균주 에드몬스톤	KPDLTGTSKSYVRSL	1218	P69353.1
336	홍역 모르빌리바이러스	KPNLSSKRSELSQLS	1219	P08362.1
337	홍역 바이러스 균주 에드몬스톤	KPNLSSKRSELSQLSMYRVF	1220	P08362.1
338	홍역 바이러스 균주 에드몬스톤	KQSRGEARESYRETG	1221	P04851.1
339	홍역 바이러스 균주 에드몬스톤	KQSRGEARESYRETGPSRAS	1222	P04851.1
340	홍역 바이러스 균주 에드몬스톤	KRFAGVVLAGAALGV	1223	P69353.1
341	홍역 바이러스 균주 에드몬스톤	KRTPGNKPRIAEMIC	1224	P04851.1
342	홍역 바이러스 균주 에드몬스톤	KRTPGNKPRIAEMICDIDTY	1225	P04851.1

[1013]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
343	홍역 모르빌리바이러스	KSNHNNVYWLTIPPMKNLALGVINTL	1226	AAR89413.1
344	홍역 모르빌리바이러스	KVSPYLFNV	1227	P08362.1
345	보르데텔라 페르투시스	KVVQLPKISKNALKANG	1228	ACI16083.1
346	보르데텔라 페르투시스	KVVQLPKISKNALRNDG	1229	ACI16087.1
347	보르데텔라 페르투시스	LAHRRIPPENIR	1230	P04977.1
348	보르데텔라 페르투시스	LALALWAGFALS	1231	P11092.1
349	풍진 바이러스	LAPGGGNCHLTVNGE	1232	BAA28178.1
350	홍역 바이러스 균주 에드몬스톤	LAQIWVLLAKAVTAP	1233	P04851.1
351	홍역 바이러스 균주 에드몬스톤	LAQIWVLLAKAVTAPDTAAD	1234	P04851.1
352	풍진 바이러스	LASYFNPGGSYYKQYHPTACEVEPAFGHS	1235	BAA19893.1
353	홍역 모르빌리바이러스	LAVSKGNCSGPTTIR	1236	P06830.1
354	홍역 바이러스 균주 에드몬스톤	LCENPEWAPLKDNRI	1237	P08362.1
355	홍역 모르빌리바이러스	LDRLVRLIG	1238	ABI54110.1
356	코리네박테리움 디프테리아에	LEEEGVTPL	1239	P33120.2
357	보르데텔라 페르투시스	LEHRMQEAVEAERAGRGTGHFI	1240	CAD44970.1
358	홍역 바이러스 균주 에드몬스톤	LEKLGYSGGDLLGIL	1241	P69353.1
359	홍역 바이러스 균주 에드몬스톤	LEQPVSNDLS	1242	P08362.1
360	홍역 바이러스 균주 에드몬스톤	LERKWLDVVRNIIAE	1243	P04851.1
361	홍역 바이러스 균주 에드몬스톤	LERKWLDVVRNIIAEDLSLR	1244	P04851.1
362	홍역 바이러스 균주 에드몬스톤	LFGPSLRDPISAEIS	1245	P69353.1

[1014]

[1015]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
363	홍역 모르빌리바이러스	LGELKLAALCHGEDS	1246	P06830.1
364	홍역 바이러스 균주 에드몬스톤	LGGKEDRRVKQSR	1247	P04851.1
365	풍진 바이러스	LGHDGHHGGTLRVGQHHRNASDVL	1248	ABD64214.1
366	풍진 바이러스 균주 테리엔	LGSPNCHGPDWASPVCQRHS	1249	P07566.1
367	풍진 바이러스 균주 테리엔	LGSPNCHGPDWASPVCQRHSPDCSRLV	1250	P07566.1
368	풍진 바이러스	LHDPDTEAPTEACVTSWL	1251	ABD64214.1
369	홍역 바이러스 균주 에드몬스톤	LIANCASILCKCYTT	1252	P69353.1
370	홍역 모르빌리바이러스	LIGLLAIAGIRLHRAAIYTAEIHK	1253	AAR89413.1
371	홍역 바이러스 균주 에드몬스톤	LIPSMNQLSCDLIGQ	1254	P69353.1
372	홍역 바이러스 균주 에드몬스톤	LKIKIASGFGPLITH	1255	P08362.1
373	홍역 모르빌리바이러스	LKIKIASGFGPLITHGSGMDLYK	1256	AAR89413.1
374	보르데텔라 페르투시스	LKLYFEP	1257	ACI16088.1
375	홍역 바이러스 균주 에드몬스톤	LLAVLFVMFL	1258	P08362.1
376	홍역 바이러스 균주 에드몬스톤	LLDRLVRLIGNPDVS	1259	P04851.1
377	홍역 바이러스 균주 에드몬스톤	LLDRLVRLIGNPDVSGPKLT	1260	P04851.1
378	홍역 바이러스 균주 에드몬스톤	LLESSDQILRSMKGL	1261	P69353.1
379	홍역 모르빌리바이러스	LLETRTTNQFLAVSK	1262	P06830.1
380	홍역 바이러스 균주 에드몬스톤	LLEVVQSDQSQSGLT	1263	P04851.1
381	홍역 바이러스 균주 에드몬스톤	LLEVVQSDQSQSGLTFASR	1264	P04851.1
382	홍역 바이러스 균주 에드몬스톤	LLEVVQSDQSQSGLTFASRG	1265	P04851.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
383	홍역 모르빌리바이러스	LLGILESRGIKARIT	1266	AAL29688.1
384	풍진 바이러스	LLRTAPGP	1267	CAJ88851.1
385	홍역 바이러스 균주 에드몬스톤	LLRYYTEILSLFGPS	1268	P69353.1
386	풍진 바이러스 균주 테리엔	LLVPWVLIFMVCRRACRRRG	1269	P07566.1
387	홍역 바이러스 균주 에드몬스톤	LLWRSRCKIV	1270	ABK40528.1
388	홍역 모르빌리바이러스	LLWSYAMGV	1271	P04851.1
389	홍역 바이러스 균주 에드몬스톤	LLWSYAMGVGVELEN	1272	P04851.1
390	홍역 바이러스 균주 에드몬스톤	LLWSYAMGVGVELENSMGGL	1273	P04851.1
391	홍역 모르빌리바이러스	LMIDRPYVL	1274	P08362.1
392	홍역 바이러스 균주 에드몬스톤	LNAMTQNIRPVQSVA	1275	P69353.1
393	풍진 바이러스	LNTPPPYQVSCGGES	1276	BAA28178.1
394	홍역 바이러스 균주 에드몬스톤	LRDPISAEISIQALS	1277	P69353.1
395	보르데텔라 페르투시스	LRGSGDLQEYLRHVTR	1278	AAA83981.1
396	풍진 바이러스	LRLVDADD	1279	CAJ88851.1
397	풍진 바이러스	LRLVDADDPLLR	1280	BAA19893.1
398	풍진 바이러스	LRLVDADDPLLRTAPGPGEVWVTPVIGSQAR	1281	BAA19893.1
399	보르데텔라 페르투시스	LRRLLYMIYMSGLAVRVHVSKEEQYYDY	1282	P04979.1
400	풍진 바이러스	LRTAPGPG	1283	CAJ88851.1
401	풍진 바이러스	LRVGQHYRNASDVLPGHWLQ	1284	NP_740663.1
402	보르데텔라 페르투시스	LRYLA	1285	ACI16088.1
403	홍역 바이러스 균주 에드몬스톤	LSCKPWQESRKNK	1286	P04851.1

[1016]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
404	홍역 모르빌리바이러스	LSEIKGVIVHRLEGV	1287	AAL29688.1
405	홍역 바이러스 균주 에드몬스톤	LSIAYPTLSEIKGVI	1288	P69353.1
406	홍역 바이러스 균주 에드몬스톤	LSLLDLYLGRGYNVS	1289	P08362.1
407	홍역 바이러스 균주 에드몬스톤	LSQGNLIANCASILC	1290	P69353.1
408	홍역 모르빌리바이러스	LSSHRGVIADNQAKW	1291	P06830.1
409	홍역 바이러스 균주 에드몬스톤	LSVDLSLTVELKIKI	1292	P08362.1
410	보르데텔라 페르투시스	LTGISICNPGSSLC	1293	AAA83981.1
411	홍역 바이러스 균주 에드몬스톤	LTIKFGIETMYPALG	1294	P04851.1
412	홍역 바이러스 균주 에드몬스톤	LTIKFGIETMYPALGLHEFA	1295	P04851.1
413	홍역 바이러스 균주 에드몬스톤	LTLQTPTGQIHWGNL	1296	P69353.1
414	풍진 바이러스	LVDADDPL	1297	CAJ88851.1
415	풍진 바이러스	LVDADDPLLR	1298	BAA19893.1
416	홍역 바이러스 균주 에드몬스톤	LVEKPNLSSKRSELS	1299	P08362.1
417	풍진 바이러스	LVGATPE	1300	BAA19893.1
418	풍진 바이러스	LVGATPER	1301	CAJ88851.1
419	홍역 모르빌리바이러스	LVKLGVWKSPTGMQS	1302	P06830.1
420	홍역 바이러스 균주 에드몬스톤-B	LVSGSFGNRFILSQGNLI	1303	P26031.1
421	홍역 바이러스 균주 에드몬스톤	LYKSNHNNVYWLTIP	1304	P08362.1
422	홍역 바이러스 균주 에드몬스톤	LYPMSPLLQECLRGSTKSCARTLVS	1305	P69353.1
423	풍진 바이러스	MASTTPITMEDLQKALEA	1306	P07566.1

[1017]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
424	풍진 바이러스	MASTTPITMEDLQKALEAQSR	1307	ABD64200.1
425	풍진 바이러스 균주 테리엔	MASTTPITMEDLQKALEAQSRALRAELAA	1308	P07566.1
426	풍진 바이러스	MASTTPITMEDLQKALEAQSRALRAGLAA	1309	ABD64200.1
427	풍진 바이러스 백신 균주 RA27/3	MASTTPITMEDLQKALETQSRVLRAGLTA	1310	CAA33016.1
428	홍역 바이러스 균주 에드몬스톤	MATLLRSLALFKRNK	1311	P04851.1
429	홍역 바이러스 균주 에드몬스톤	MATLLRSLALFKRNKDKPPI	1312	P04851.1
430	홍역 모르빌리바이러스	MDLYKSNHNNVYWLT	1313	P06830.1
431	풍진 바이러스	MEDLQKALEAQSRA	1314	P07566.1
432	풍진 바이러스	MEDLQKALEAQSRALRAELAA	1315	P07566.1
433	홍역 바이러스 균주 에드몬스톤	MGLKVNVSAIFMAVL	1316	P69353.1
434	홍역 모르빌리바이러스	MIDRPYVLLAVLFVM	1317	P06830.1
435	홍역 바이러스 균주 에드몬스톤	MILAVQGVQDYINNE	1318	P69353.1
436	홍역 바이러스 균주 에드몬스톤	MLNSQAIDNLRASLE	1319	P69353.1
437	홍역 바이러스 균주 에드몬스톤	MNALVNSTLLETRTT	1320	P08362.1
438	풍진 바이러스	MNYTGNQQSRWGLGSPNCHGPDWASPVCQRHS	1321	BAA19893.1
439	홍역 바이러스 균주 에드몬스톤	MQSWVPLSTDDPVID	1322	P08362.1
440	홍역 모르빌리바이러스	MSLSLLDLYLGRGYN	1323	P06830.1
441	홍역 바이러스 균주 에드몬스톤	MSPLLQECLRGSTKS	1324	P69353.1
442	홍역 모르빌리바이러스	MSPQRDRINAFYKDN	1325	P06830.1
443	홍역 모르빌리바이러스	MYRVFEVSVIRNPGL	1326	P06830.1
444	홍역 바이러스 균주 에드몬스톤	NALYPMSPLLQECLR	1327	P69353.1

[1018]

[1019]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
445	풍진 바이러스 균주 테리엔	NCHGPDWASPVCQRHSPDCSRLVGAT	1328	P07566.1
446	홍역 바이러스 균주 에드몬스톤	NFGRSYFDPAYFRLG	1329	P04851.1
447	홍역 바이러스 균주 에드몬스톤	NFGRSYFDPAYFRLGQEMVR	1330	P04851.1
448	풍진 바이러스	NGTQRACTFWAVNAY	1331	BAA28178.1
449	홍역 바이러스 균주 에드몬스톤	NGVTIQVGSRRYPDA	1332	P69353.1
450	홍역 바이러스 균주 에드몬스톤	NIIAEDLSLRRFMVA	1333	P04851.1
451	홍역 바이러스 균주 에드몬스톤	NIIAEDLSLRRFMVALILDI	1334	P04851.1
452	홍역 바이러스 균주 에드몬스톤	NITLLNNCTRVEIAE	1335	P69353.1
453	홍역 모르빌리바이러스	NLALGVINTLEWIPR	1336	P06830.1
454	코리네박테리움 디프테리아에	NLFQVVHWSYNRPAYSPG	1337	SRC280292
455	홍역 바이러스 균주 에드몬스톤	NLVILPGQDLQYVLA	1338	P08362.1
456	홍역 바이러스 균주 에드몬스톤	NLYQQMGKPAPYMVN	1339	P04851.1
457	홍역 바이러스 균주 에드몬스톤	NLYQQMGKPAPYMVNLENSI	1340	P04851.1
458	홍역 바이러스 균주 에드몬스톤	NNCTRVEIAEYRRLL	1341	P69353.1
459	홍역 바이러스 균주 에드몬스톤	NPDREYDFRD	1342	P08362.1
460	홍역 바이러스 균주 할레	NPDVSGPKL	1343	P10050.1
461	홍역 바이러스 균주 에드몬스톤	NPDVSGPKLTGALIG	1344	P04851.1
462	홍역 바이러스 균주 에드몬스톤	NPDVSGPKLTGALIGILSLF	1345	P04851.1
463	홍역 모르빌리바이러스	NPPERIKLDYDQYCA	1346	P06830.1
464	홍역 모르빌리바이러스	NQAKWAVPTTRTDDK	1347	P06830.1

[1020]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
465	홍역 모르빌리바이러스	NQDPDKILTYIAADH	1348	AAF02706.1
466	홍역 바이러스 균주 에드몬스톤	NQLSCDLIGQKLGLK	1349	P69353.1
467	풍진 바이러스	NQQSRWGLGSPNCHGPDWASPVCQRHS	1350	ABD64214.1
468	볼거리 루불라바이러스	NSTLGVKSAREF	1351	ABP48111.1
469	풍진 바이러스	NTPHGQLEVQVPPDP	1352	BAA28178.1
470	홍역 바이러스 균주 에드몬스톤	NVSAIFMAVLLTLQT	1353	P69353.1
471	홍역 모르빌리바이러스	PAEVDGDVKLSSNLV	1354	P06830.1
472	풍진 바이러스	PAFGHSDAACWGFPT	1355	BAA28178.1
473	홍역 바이러스 균주 에드몬스톤	PALICCCRGRCNKKG	1356	P69353.1
474	홍역 모르빌리바이러스	PDKILTYIAADHC	1357	AAF02706.1
475	홍역 바이러스 균주 에드몬스톤	PERIKLDYDQYCADV	1358	P08362.1
476	풍진 바이러스	PERPRLRL	1359	CAJ88851.1
477	풍진 바이러스	PGCATQAPVPVRLAG	1360	BAA28178.1
478	풍진 바이러스 백신 균주 RA27/3	PGCATQAPVPVRLAGVRFESKIVDGGCFA	1361	CAJ88851.1
479	풍진 바이러스	PGEVWVTP	1362	CAJ88851.1
480	풍진 바이러스	PGEVWVTPVIGSQAR	1363	BAA28178.1
481	코리네박테리움 디프테리아에	PGKLDVNKSKTHISVN	1364	CAE11230.1
482	홍역 바이러스 균주 에드몬스톤	PGLGAPVFHMTNYLE	1365	P08362.1
483	풍진 바이러스	PGPGEVWV	1366	CAJ88851.1
484	풍진 바이러스	PHKTVRVKFHTETRT	1367	BAA28178.1
485	홍역 모르빌리바이러스	PIELQVECFTWDQKL	1368	AAR89413.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
486	홍역 바이러스 균주 에드몬스톤-B	PISLERLDVG	1369	P26031.1
487	홍역 바이러스 균주 에드몬스톤	PISLERLDVGTNLGN	1370	P69353.1
488	보르데텔라 페르투시스	PKALFTQQGGAYGRC	1371	P04979.1
489	홍역 바이러스 균주 에드몬스톤	PKYVATQGYLISNFD	1372	P69353.1
490	홍역 바이러스 균주 에드몬스톤	PLDIDTASESSQD	1373	P04851.1
491	풍진 바이러스	PLGLKFKTVRPVALP	1374	BAA28178.1
492	홍역 바이러스 균주 에드몬스톤	PLITHGSGMDLYKSN	1375	P08362.1
493	홍역 모르빌리바이러스	PLKDNRIPSYGVLSV	1376	P06830.1
494	풍진 바이러스	PLLRTAPG	1377	CAJ88851.1
495	홍역 바이러스 균주 에드몬스톤	PLLSCKPWQESRK	1378	P04851.1
496	보르데텔라 페르투시스	PPATVYRYDSRPPE	1379	P04977.1
497	홍역 모르빌리바이러스	PPISLERLDVGT	1380	AAL29688.1
498	풍진 바이러스	PPPPEERQETRSQTPAPKPS	1381	P07566.1
499	보르데텔라 페르투시스	PQEQITQHGSPYGRC	1382	AAA83981.1
500	보르데텔라 페르투시스	PQEQITQHGSPYGRCANK	1383	AAA83981.1
501	보르데텔라 페르투시스	PQPGPQPPQPPQPQPEAPAPQPPAG	1384	P14283.3
502	홍역 바이러스 균주 에드몬스톤	PRLGGKEDRRVKQ	1385	P04851.1
503	풍진 바이러스	PRLRLVDA	1386	CAJ88851.1
504	풍진 바이러스	PRNVRVTGCYQCGTP	1387	BAA28178.1
505	홍역 바이러스 균주 에드몬스톤	PSRASDARAAHLP	1388	P04851.1

[1021]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
506	홍역 바이러스 균주 에드몬스톤	PTGQIHWGNLSKIGV	1389	P69353.1
507	홍역 바이러스 균주 에드몬스톤	PTGTPLDIDTASE	1390	P04851.1
508	홍역 바이러스 균주 에드몬스톤	PTLSEIKGVIVHRLE	1391	P69353.1
509	홍역 모르빌리바이러스	PTTIRGQFSNMSLSL	1392	P06830.1
510	홍역 모르빌리바이러스	PTTRTDDKLR	1393	AAR89413.1
511	홍역 모르빌리바이러스	PTTRTDDKLRMETCFQQACKG	1394	AAR89413.1
512	풍진 바이러스	PVALPRTLAPPRNVR	1395	BAA28178.1
513	코리네박테리움 디프테리아에	PVFAGANYAAWAVNVAQVI	1396	AAV70486.1
514	풍진 바이러스	PVIGSQAR	1397	CAJ88851.1
515	홍역 모르빌리바이러스	PVVEVNGVTIQVGSR	1398	AAL29688.1
516	풍진 바이러스 균주 테리엔	PWELVVLTARPEDGWTCRGV	1399	P07566.1
517	홍역 바이러스 균주 에드몬스톤	PWQESRKNKAQTR	1400	P04851.1
518	홍역 바이러스 균주 에드몬스톤	PYMVNLENSIQNKFS	1401	P04851.1
519	홍역 바이러스 균주 에드몬스톤	PYMVNLENSIQNKFSAGSYP	1402	P04851.1
520	홍역 바이러스 균주 에드몬스톤	PYQGSGKGVS	1403	P08362.1
521	풍진 바이러스	PYQVSCGGESDRASA	1404	BAA28178.1
522	홍역 모르빌리바이러스	QACKGKIQALCEN	1405	P08362.1
523	홍역 바이러스 균주 에드몬스톤	QAGQEMILAVQGVQD	1406	P69353.1
524	홍역 모르빌리바이러스	QALCENPECVPLKDN	1407	P06830.1
525	보르데텔라 페르투시스	QALGALK	1408	ACI16088.1

[1022]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
526	풍진 바이러스	QAPVPVRLAGVRFES	1409	BAA28178.1
527	풍진 바이러스	QCGTPALVEGLAPGG	1410	BAA28178.1
528	홍역 바이러스 균주 에드몬스톤	QDPDKILTYIAADHC	1411	P69353.1
529	홍역 바이러스 균주 에드몬스톤	QECLRGSTKSCARTL	1412	P69353.1
530	보르데텔라 페르투시스	QEQITQHGSPYGRC	1413	AAA83981.1
531	홍역 바이러스 균주 에드몬스톤	QESRKNKAQTRTP	1414	P04851.1
532	홍역 바이러스 균주 에드몬스톤	QGDQSENELPRLGGK	1415	P04851.1
533	홍역 바이러스 균주 에드몬스톤	QGDQSENELPRLGGKEDRRV	1416	P04851.1
534	코리네박테리움 디프테리아에	QGESGHDIKITAENTPLPIA	1417	AAV70486.1
535	홍역 모르빌리바이러스	QGSGKGVSFQLVKLG	1418	P06830.1
536	홍역 바이러스 균주 에드몬스톤	QGVQDYINNELIPSM	1419	P69353.1
537	풍진 바이러스	QLEVQVPPDPGDLVE	1420	BAA28178.1
538	홍역 모르빌리바이러스	QLPEATFMV	1421	ABK40528.1
539	풍진 바이러스	QLPFLGHDGHHGGTLRVGQHYRNAS	1422	NP_740663.1
540	홍역 바이러스 균주 에드몬스톤	QLSMYRVFEV	1423	P08362.1
541	보르데텔라 페르투시스	QLSNIT	1424	ACI16083.1
542	홍역 바이러스 균주 에드몬스톤	QNIRPVQSVASSRRH	1425	P69353.1
543	홍역 바이러스 균주 에드몬스톤	QNKFSAGSYPLLWSY	1426	P04851.1
544	홍역 바이러스 균주 에드몬스톤	QNKFSAGSYPLLWSYAMGVG	1427	P04851.1
545	홍역 바이러스 균주 에드몬스톤	QNKFSAGSYPLLWSYAMGVGVELEN	1428	P04851.1

[1023]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
546	홍역 바이러스 균주 에드몬스톤	QQACKGKIQALCENP	1429	P08362.1
547	홍역 바이러스 균주 에드몬스톤	QQRRVVGEFRLERKW	1430	P04851.1
548	홍역 바이러스 균주 에드몬스톤	QQRRVVGEFRLERKWLDVVR	1431	P04851.1
549	보르데텔라 페르투시스	QQTRANPNPYTSRRSVAS	1432	P04977.1
550	풍진 바이러스	QRHSPDCSRLVGATP	1433	BAA28178.1
551	홍역 바이러스 균주 에드몬스톤	QSGLTFASRGTNMED	1434	P04851.1
552	홍역 바이러스 균주 에드몬스톤	QSGLTFASRGTNMEDEADQY	1435	P04851.1
553	코리네박테리움 디프테리아에	QSIALSSLMVAQAIPLVGEL	1436	AAV70486.1
554	홍역 바이러스 균주 에드몬스톤	QSRFGWFENKEISDI	1437	P04851.1
555	홍역 바이러스 균주 에드몬스톤	QSRFGWFENKEISDIEVQDP	1438	P04851.1
556	홍역 바이러스 균주 에드몬스톤	QSRGEAR	1439	P04851.1
557	홍역 바이러스 균주 에드몬스톤	QSRGEARESYRETGPSRA	1440	P04851.1
558	풍진 바이러스 균주 테리엔	QTGRGGSAPRPELGPPTN	1441	P07566.1
559	풍진 바이러스 균주 테리엔	QTPAPKPSRAPPQQPQPPRMQTGRG	1442	P07566.1
560	홍역 바이러스 균주 에드몬스톤	QVGSRRYPDAVYLHR	1443	P69353.1
561	홍역 바이러스 균주 에드몬스톤	QVSFLQGDQSENE	1444	P04851.1
562	풍진 바이러스	QYHPTACEVEPAFGH	1445	BAA28178.1
563	홍역 바이러스 균주 에드몬스톤	QYVLATYDTSRVEHA	1446	P08362.1
564	보르데텔라 페르투시스	RANPNPYTSRRSV	1447	ACI04548.1
565	홍역 바이러스 균주 에드몬스톤	RASDARAAHLPTG	1448	P04851.1

[1024]

[1025]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
566	홍역 바이러스 균주 에드몬스톤	RASLETTNQAIEAIR	1449	P69353.1
567	풍진 바이러스 균주 테리엔	RCGRLICGLSTTAQYPPTRF	1450	P07566.1
568	보르데텔라 페르투시스	RDGQSVIGACASPYEGRYR	1451	P04979.1
569	홍역 바이러스 균주 에드몬스톤	RESYRETGPSRAS	1452	P04851.1
570	홍역 바이러스 균주 에드몬스톤	RETGPSRASDARA	1453	P04851.1
571	볼거리 루불라바이러스	RFAKYQQQGRLEAR	1454	P21186.1
572	풍진 바이러스 균주 테리엔	RFGAPQAFLAGLLLATVAVGTARA	1455	P07566.1
573	홍역 바이러스 균주 에드몬스톤	RFMVALILDIKRTPG	1456	P04851.1
574	홍역 바이러스 균주 에드몬스톤	RFMVALILDIKRTPGNKPRI	1457	P04851.1
575	홍역 바이러스 균주 에드몬스톤	RGEARESYRETGP	1458	P04851.1
576	풍진 바이러스	RGTTPPAYG	1459	CAA28880.1
577	풍진 바이러스 균주 M33	RIETRSARH	1460	ABD64214.1
578	보르데텔라 페르투시스	RILAGALATYQ	1461	P04977.1
579	보르데텔라 페르투시스	RIPPENIRRVT	1462	ACI04548.1
580	보르데텔라 페르투시스	RISNLND	1463	ACI16083.1
581	보르데텔라 페르투시스	RLANLNG	1464	ACI16088.1
582	홍역 바이러스 균주 에드몬스톤	RLDVGTNLGNAIAKL	1465	P69353.1
583	홍역 바이러스 균주 에드몬스톤	RLERKWLDV	1466	P04851.1
584	홍역 바이러스 균주 에드몬스톤	RLGGKEDRRVKQSRG	1467	P04851.1
585	홍역 바이러스 균주 에드몬스톤	RLGGKEDRRVKQSRGEARES	1468	P04851.1

[1026]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
586	홍역 바이러스 균주 에드몬스톤	RLLDRLVRL	1469	ABI54110.1
587	풍진 바이러스	RLRLVDAD	1470	CAJ88851.1
588	풍진 바이러스	RLRLVDADDPLLR	1471	BAA19893.1
589	풍진 바이러스	RLRLVDADDPLLRTAPGPGEVWVTPVIGSQA	1472	BAA19893.1
590	풍진 바이러스	RLRLVQDADDPLLRIAPGPGEVWVTPVIGSQA	1473	SRC265968
591	홍역 바이러스 균주 에드몬스톤	RLSDNGYYTV	1474	ABK40528.1
592	풍진 바이러스	RLVDADDP	1475	CAJ88851.1
593	풍진 바이러스	RLVDADDPLLRTAPG	1476	BAA28178.1
594	풍진 바이러스	RLVGATPE	1477	CAJ88851.1
595	풍진 바이러스	RMQTGRGGSAPRPELGPPTNPFQAAVA	1478	ABD64216.1
596	홍역 모르빌리바이러스	RMSKGVFKV	1479	ABY21184.1
597	홍역 모르빌리바이러스	RNPGLGAPVFHMTNY	1480	P06830.1
598	풍진 바이러스	RPRLRLVD	1481	CAJ88851.1
599	보르데텔라 페르투시스	RQAESSEAMAAWSERAGEA	1482	P04977.1
600	풍진 바이러스	RQTWAEWAAAHWWQL	1483	BAA28178.1
601	홍역 바이러스 균주 에드몬스톤	RRVKQSRGEARES	1484	P04851.1
602	홍역 모르빌리바이러스	RRYPDAVYL	1485	ACA09725.1
603	홍역 바이러스 균주 에드몬스톤	RRYPDAVYLHRIDLG	1486	P69353.1
604	홍역 바이러스 균주 에드몬스톤	RSAGKVSSTLASELG	1487	P04851.1
605	홍역 바이러스 균주 에드몬스톤	RSAGKVSSTLASELGITAED	1488	P04851.1
606	홍역 바이러스 균주 에드몬스톤	RSELSQLS	1489	P08362.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
607	홍역 바이러스 균주 에드몬스톤	RSELSQLSMYRVFEV	1490	P08362.1
608	풍진 바이러스	RSQTPAPKPSRAPPQQPQPPRMQT	1491	ABD64214.1
609	풍진 바이러스	RTAPGPGE	1492	CAJ88851.1
610	풍진 바이러스	RTAPGPGEVWVTPVI	1493	BAA28178.1
611	홍역 모르빌리바이러스	RTDDKLRMETCFQQA	1494	P06830.1
612	풍진 바이러스	RTLAPPRNVRVTGCY	1495	BAA28178.1
613	홍역 바이러스 균주 에드몬스톤	RTVLEPIRDALNAMT	1496	P69353.1
614	홍역 바이러스 균주 에드몬스톤	RVEHAVVYYVYSPSR	1497	P08362.1
615	홍역 바이러스 균주 에드몬스톤	RVFEVGVIRNPGLGA	1498	P08362.1
616	보르데텔라 페르투시스	RVHVSKEEQYYDYEDATFE	1499	P04978.2
617	풍진 바이러스	RVIDPAAQ	1500	BAA28178.1
618	풍진 바이러스	RVKFHTETRTVWQLS	1501	BAA28178.1
619	보르데텔라 페르투시스	RVYHNGITGET	1502	ACI04548.1
620	보르데텔라 페르투시스	RYDSRPPEDVF	1503	ACI04548.1
621	홍역 바이러스 균주 에드몬스톤	RYPDAVYLHRIDLGP	1504	P69353.1
622	홍역 바이러스 균주 에드몬스톤	SAEISIQALSYALGG	1505	P69353.1
623	홍역 바이러스 균주 에드몬스톤	SAEPLLSCKPWQESR	1506	P04851.1
624	홍역 바이러스 균주 에드몬스톤	SAEPLLSCKPWQESRKNKAQ	1507	P04851.1
625	홍역 모르빌리바이러스	SAGKVSSTLASELG	1508	P04851.1
626	홍역 바이러스 균주 에드몬스톤	SAGKVSSTLASELGITAEDARLVS	1509	P04851.1

[1027]

[1028]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
627	홍역 바이러스 균주 에드몬스톤	SCTVTREDGT	1510	P08362.1
628	풍진 바이러스	SDAACWGFPTDTVMS	1511	BAA28178.1
629	홍역 바이러스 균주 에드몬스톤	SDARAAHLPTGTP	1512	P04851.1
630	풍진 바이러스 균주 테리엔	SDWHQGTHVCHTKHMDFWCVEHD	1513	P07566.1
631	홍역 바이러스 균주 에드몬스톤	SELRRWIKYTQQRRV	1514	P04851.1
632	홍역 바이러스 균주 에드몬스톤	SELRRWIKYTQQRRVVGEFR	1515	P04851.1
633	홍역 바이러스 균주 에드몬스톤	SELSQL	1516	P08362.1
634	홍역 바이러스 균주 에드몬스톤	SELSQLS	1517	P08362.1
635	보르데텔라 페르투시스	SEYLAHRRIPPENIRRVTRV	1518	CAD44970.1
636	홍역 바이러스 균주 에드몬스톤	SFLQGDQSENELP	1519	P04851.1
637	홍역 바이러스 균주 에드몬스톤	SGKGVSFQLVKLGVW	1520	P08362.1
638	홍역 바이러스 균주 에드몬스톤	SHRGVIADNQAKWAV	1521	P08362.1
639	홍역 바이러스 균주 에드몬스톤	SIEHQVKDVLTPLFK	1522	P08362.1
640	홍역 바이러스 균주 에드몬스톤	SKIGVVGIGSASYKV	1523	P69353.1
641	홍역 모르빌리바이러스	SKRSELSQLSMYRVF	1524	P06830.1
642	홍역 모르빌리바이러스	SLFVESPGQLIQRITDDPDVS	1525	ABI54110.1
643	홍역 바이러스 균주 에드몬스톤	SLSTNLDVTNSIEHQ	1526	P08362.1
644	홍역 바이러스 균주 에드몬스톤	SLSTNLDVTNSIEHQVKDVLTPLFK	1527	P08362.1
645	홍역 바이러스 균주 에드몬스톤	SLWGSGLLML	1528	BAE98296.1
646	홍역 바이러스 균주 에드몬스톤	SMKGLSSTSIVYILI	1529	P69353.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
647	홍역 모르빌리바이러스	SMYRVFEVGV	1530	P08362.1
648	홍역 모르빌리바이러스	SNDLSNCMVALGELK	1531	P06830.1
649	홍역 바이러스 균주 할레	SPGQLIQR	1532	P10050.1
650	홍역 바이러스 균주 에드몬스톤	SQDPQDSRRSAEP	1533	P04851.1
651	홍역 모르빌리바이러스	SRIVINREHLMIDRP	1534	P06830.1
652	홍역 바이러스 균주 에드몬스톤	SRKNKAQTRTPLQ	1535	P04851.1
653	풍진 바이러스	SRLVGATP	1536	CAJ88851.1
654	풍진 바이러스	SRLVGATPERPRLRLVDADDPLLR	1537	CAJ88851.1
655	홍역 바이러스 균주 에드몬스톤	SRPGLKPDLTGTSKS	1538	P69353.1
656	보르데텔라 페르투시스	SRRSVASIVGTLVRM	1539	CAD44970.1
657	풍진 바이러스	SRWGLGSPNCHGPDW	1540	BAA28178.1
658	보르데텔라 페르투시스	SSATK	1541	ACI16088.1
659	풍진 바이러스	SSGGYAQLASYFNPG	1542	BAA28178.1
660	보르데텔라 페르투시스	SSLGNGV	1543	ACI16083.1
661	홍역 모르빌리바이러스	SSNLVILPGQDLQYV	1544	P06830.1
662	홍역 바이러스 균주 에드몬스톤	SSQDPQDSRRSAEPL	1545	P04851.1
663	홍역 바이러스 균주 에드몬스톤	SSQDPQDSRRSAEPLLSCKP	1546	P04851.1
664	홍역 바이러스 균주 에드몬스톤	SSRRHKRFAGVVLAG	1547	P69353.1
665	홍역 바이러스 균주 에드몬스톤	SSTSIVYILIAVCLG	1548	P69353.1
666	보르데텔라 페르투시스	STPGIVI	1549	AAA83981.1

[1029]

[1030]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
667	보르데텔라 페르투시스	STPGIVIPPQEQITQHGSPYGRC	1550	AAA83981.1
668	보르데텔라 페르투시스	STSSSRRYTEVY	1551	P04977.1
669	홍역 바이러스 균주 에드몬스톤	SYFIVLSIAYPTLSE	1552	P69353.1
670	홍역 바이러스 균주 에드몬스톤	SYRETGPSRASDA	1553	P04851.1
671	보르데텔라 페르투시스	SYVK	1554	ACI16083.1
672	풍진 바이러스	SYVQHPHKTVRVKFH	1555	BAA28178.1
673	풍진 바이러스	TAPGPGEV	1556	CAJ88851.1
674	보르데텔라 페르투시스	TATRLLSSTNSRLC	1557	AAA83981.1
675	홍역 모르빌리바이러스	TDDPVIDRLYLSSHR	1558	P06830.1
676	홍역 바이러스 균주 에드몬스톤	TEILSLFGPSLRDPI	1559	P69353.1
677	풍진 바이러스	TETRTVWQLSVAGVS	1560	BAA28178.1
678	보르데텔라 페르투시스	TEVYLEHRMQEAVE	1561	P04977.1
679	홍역 바이러스 균주 에드몬스톤	TFMPEGTVCSQNALY	1562	P69353.1
680	풍진 바이러스	TGACICEIPTDVSCE	1563	BAA28178.1
681	보르데텔라 페르투시스	TGDLRAY	1564	ACI16083.1
682	홍역 모르빌리바이러스	TGMQSWVPLSTDDPV	1565	P06830.1
683	풍진 바이러스	TGNQQSRWGLGSPNC	1566	BAA28178.1
684	홍역 바이러스 균주 에드몬스톤	TGPSRASDARAAH	1567	P04851.1
685	홍역 모르빌리바이러스	TGTIINQDPDKILTY	1568	AAF02706.1
686	풍진 바이러스	TGVVYGTHTTAVSET	1569	BAA28178.1

[1031]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
687	홍역 바이러스 균주 에드몬스톤	TIRGQFSNMSLSLLD	1570	P08362.1
688	풍진 바이러스	TLGATCALPLAGLLA	1571	BAA28178.1
689	홍역 모르빌리바이러스	TLLNNCTRV	1572	P26031.1
690	홍역 바이러스 균주 에드몬스톤-B	TLNVPPPPDPGR	1573	P03422.1
691	홍역 바이러스 균주 에드몬스톤-B	TLNVPPPPDPGRASTSGTPIKK	1574	P03422.1
692	홍역 모르빌리바이러스	TMTSQGMYGGTYPVE	1575	P06830.1
693	홍역 바이러스 균주 에드몬스톤	TNLGNAIAKLEDAKE	1576	P69353.1
694	홍역 바이러스 균주 에드몬스톤	TNMEDEADQYFSHDD	1577	P04851.1
695	홍역 바이러스 균주 에드몬스톤	TNMEDEADQYFSHDDPISSD	1578	P04851.1
696	풍진 바이러스 균주 M33	TNPFQAAVARGLRPP	1579	CAA28880.1
697	홍역 모르빌리바이러스	TNSIEHQVKDVLTPL	1580	P06830.1
698	홍역 바이러스 균주 에드몬스톤	TNYLEQPVSNDLSNC	1581	P08362.1
699	홍역 모르빌리바이러스	TNYLEQPVSNDLSNCMVALGELKLAAL	1582	AAR89413.1
700	풍진 바이러스	TPERPRLR	1583	CAJ88851.1
701	풍진 바이러스 균주 테리엔	TPERPRLRLVDADDPLLRTA	1584	P07566.1
702	홍역 바이러스 균주 할레	TPGNKPRIA	1585	P10050.1
703	홍역 바이러스 균주 에드몬스톤	TPLDIDTASESSQDP	1586	P04851.1
704	홍역 바이러스 균주 에드몬스톤	TPLDIDTASESSQDPQDSRR	1587	P04851.1
705	홍역 바이러스 균주 에드몬스톤	TPLFKIIGDEVGLRT	1588	P08362.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
706	홍역 바이러스 균주 에드몬스톤	TPLQCTM	1589	P04851.1
707	풍진 바이러스	TPVIGSQA	1590	CAJ88851.1
708	풍진 바이러스	TPVIGSQARK	1591	BAA19893.1
709	홍역 바이러스 균주 에드몬스톤	TQGYLISNFDESSCT	1592	P69353.1
710	보르데텔라 페르투시스	TRANPNPYTSRRSVASIVGTLVRM	1593	P04977.1
711	풍진 바이러스	TRFGCAMRWGLPP	1594	NP_740663.1
712	보르데텔라 페르투시스	TRNTGQPATDHYYSNV	1595	AAA83981.1
713	홍역 바이러스 균주 에드몬스톤	TRTPLQCTMTEIF	1596	P04851.1
714	풍진 바이러스	TRWHRLLRMPVR	1597	ABD64216.1
715	홍역 바이러스 균주 에드몬스톤	TSGSGGAIRGIKHII	1598	P04851.1
716	홍역 바이러스 균주 에드몬스톤	TSGSGGAIRGIKHIIIVPIP	1599	P04851.1
717	홍역 바이러스 균주 에드몬스톤	TSQGMYGGTYLVEKP	1600	P08362.1
718	홍역 모르빌리바이러스	TSRVEHAVVYYVYSP	1601	P06830.1
719	보르데텔라 페르투시스	TSSSRRYTEVYL	1602	ACI04548.1
720	보르데텔라 페르투시스	TSYVG	1603	ACI16088.1
721	홍역 바이러스 균주 에드몬스톤	TTEDKISRAVGPRQA	1604	P04851.1
722	홍역 바이러스 균주 에드몬스톤	TTEDKISRAVGPRQAQVSFL	1605	P04851.1
723	풍진 바이러스 균주 M33	TTERIETRSARHP	1606	ABD64214.1
724	홍역 바이러스 균주 에드몬스톤	TTNQAIEAIRQAGQE	1607	P69353.1
725	풍진 바이러스	TTSDPWHPPGPLGLK	1608	BAA28178.1

[1032]

[1033]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
726	홍역 바이러스 균주 에드몬스톤	TVCSQNALYPMSPLL	1609	P69353.1
727	풍진 바이러스	TVNGEDVGAVPPGKF	1610	BAA28178.1
728	홍역 모르빌리바이러스	TYPVEKPNLSSKRSE	1611	P06830.1
729	풍진 바이러스	VAGVSCNVTTEHPFC	1612	BAA28178.1
730	보르데텔라 페르투시스	VAPGIVIPPKALFTQQGGAYGRC	1613	P04979.1
731	풍진 바이러스	VCHTKHMDFWCVEHDRPPPATPTPL	1614	NP_740663.1
732	풍진 바이러스	VCQRHSPDCSRLVGATPER	1615	BAA19893.1
733	풍진 바이러스	VDADDPLL	1616	CAJ88851.1
734	풍진 바이러스	VDADDPLLRTAPGPGEVWVT	1617	BAA19893.1
735	코리네박테리움 디프테리아에	VDIGFAAYNFVESIINLFQV	1618	AAV70486.1
736	홍역 바이러스 균주 에드몬스톤	VEIAEYRRLLRTVLE	1619	P69353.1
737	홍역 바이러스 균주 에드몬스톤	VELENSMGGLNFGRS	1620	P04851.1
738	홍역 바이러스 균주 에드몬스톤	VELENSMGGLNFGRSYFDPA	1621	P04851.1
739	홍역 모르빌리바이러스	VELKIKIASGFGPLI	1622	P06830.1
740	풍진 바이러스	VEMDEWIHAHTTSD	1623	SRC265968
741	풍진 바이러스	VEMPEWIHAHTTSDP	1624	BAA28178.1
742	코리네박테리움 디프테리아에	VERRLVKVL	1625	P33120.2
743	홍역 모르빌리바이러스	VESPGQLI	1626	ABI54110.1
744	홍역 바이러스 균주 에드몬스톤	VESPGQLIQRITDDP	1627	P04851.1
745	홍역 바이러스 균주 에드몬스톤	VESPGQLIQRITDDPDVSIR	1628	P04851.1
746	풍진 바이러스	VFALASYVQHPHKTV	1629	BAA28178.1

[1034]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
747	풍진 바이러스	VGATPERP	1630	CAJ88851.1
748	풍진 바이러스	VGATPERPRL	1631	BAA19893.1
749	풍진 바이러스	VGATPERPRLRLVDA	1632	BAA28178.1
750	홍역 바이러스 균주 에드몬스톤	VGIGSASYKVMTRSS	1633	P69353.1
751	홍역 바이러스 균주 에드몬스톤	VGLRTPQRFTDLVKF	1634	P08362.1
752	코리네박테리움 디프테리아에	VHHNTEEIVAQSIALSSLMV	1635	AAV70486.1
753	홍역 바이러스 균주 에드몬스톤	VHRLEGVSYNIGSQE	1636	P69353.1
754	풍진 바이러스	VIGSQARK	1637	CAJ88851.1
755	보르데텔라 페르투시스	VITGSI	1638	ACI16088.1
756	보르데텔라 페르투시스	VITGTI	1639	ACI16083.1
757	홍역 바이러스 균주 에드몬스톤	VKQSRGEA	1640	P04851.1
758	홍역 바이러스 균주 에드몬스톤	VLFVMFLSLI	1641	P08362.1
759	홍역 모르빌리바이러스	VLFVMFLSLIGLLAI	1642	P06830.1
760	홍역 모르빌리바이러스	VLTPLFKIIGDEVGL	1643	P06830.1
761	홍역 바이러스 균주 에드몬스톤	VMTRSSHQSLVIKLM	1644	P69353.1
762	풍진 바이러스	VPAAPCARIWNGTQR	1645	BAA28178.1
763	풍진 바이러스	VPPDPGDLVEYIMNY	1646	BAA28178.1
764	홍역 바이러스 균주 에드몬스톤	VQSVASSRRHKRFAG	1647	P69353.1
765	보르데텔라 페르투시스	VQTGGTSRTVTMRYLAS	1648	ACI16083.1
766	보르데텔라 페르투시스	VQVRI	1649	ACI16083.1
767	풍진 바이러스	VRAYNQPAGDV	1650	NP_740662.1

[1035]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
768	풍진 바이러스	VRFESKIVDGGCFAP	1651	BAA28178.1
769	풍진 바이러스	VRLAGVRFESKIVDG	1652	BAA28178.1
770	홍역 바이러스 균주 에드몬스톤	VSGSFGNRFILSQGN	1653	P69353.1
771	보르데텔라 페르투시스	VSKEEQYYDYEDAT	1654	AAA83981.1
772	홍역 바이러스 균주 에드몬스톤	VSKGNCSGPTTIRGQ	1655	P08362.1
773	풍진 바이러스	VTAALLNTPPPYQVS	1656	BAA28178.1
774	풍진 바이러스	VTGCYQCGTPALVEG	1657	BAA28178.1
775	풍진 바이러스	VTPVIGSQ	1658	CAJ88851.1
776	풍진 바이러스 균주 테리엔	VTTEHPFCNTPHGQLEVQVPPD	1659	P07566.1
777	홍역 바이러스 균주 에드몬스톤	VVLAGAALGVATAAQ	1660	P69353.1
778	풍진 바이러스	VWQLSVAGVSCNVTT	1661	BAA28178.1
779	풍진 바이러스	VWVTPVIG	1662	CAJ88851.1
780	풍진 바이러스	VWVTPVIGSQAR	1663	BAA19893.1
781	홍역 바이러스 균주 에드몬스톤	VYILIAVCLGGLIGI	1664	P69353.1
782	홍역 바이러스 균주 에드몬스톤	VYLHRIDLGPPISLE	1665	P69353.1
783	보르데텔라 페르투시스	VYRYDSRP	1666	P04977.1
784	보르데텔라 페르투시스	VYRYDSRPPEDV	1667	P04977.1
785	홍역 모르빌리바이러스	VYWLTIPPMKNLALG	1668	P06830.1
786	풍진 바이러스	WDLEATGACICEIPT	1669	BAA28178.1
787	홍역 모르빌리바이러스	WDQKLWCRHFCVLAD	1670	AAR89413.1
788	풍진 바이러스	WGFPTDTVMSVFALA	1671	BAA28178.1

[1036]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
789	풍진 바이러스	WHPPGPLGLKFKTVR	1672	BAA28178.1
790	풍진 바이러스	WIHAHTTSDPWHPPG	1673	BAA28178.1
791	홍역 바이러스 균주 에드몬스톤	WLTIPPMKNLALGVI	1674	P08362.1
792	홍역 바이러스 균주 에드몬스톤	WQESRKNKAQTRTPLQCTMT	1675	P04851.1
793	풍진 바이러스	WVCIFMVCRRACR	1676	SRC265968
794	풍진 바이러스	WVTPVIGS	1677	CAJ88851.1
795	홍역 바이러스 균주 에드몬스톤	WYTTVPKYVATQGYL	1678	P69353.1
796	홍역 바이러스 균주 에드몬스톤	YALGGDINKVLEKLG	1679	P69353.1
797	홍역 바이러스 균주 에드몬스톤	YAMGVGVELE	1680	P04851.1
798	홍역 모르빌리바이러스	YAMGVGVELEN	1681	ABI54110.1
799	홍역 모르빌리바이러스	YCADVAAEELMNALV	1682	AAR89413.1
800	홍역 모르빌리바이러스	YDFRDLTWCINPPER	1683	P06830.1
801	보르데텔라 페르투시스	YFEPGPT	1684	ACI16083.1
802	풍진 바이러스	YFNPGGSYYKQYHPT	1685	BAA28178.1
803	홍역 바이러스 균주 에드몬스톤	YFRLGQEMVRRSAGK	1686	P04851.1
804	홍역 바이러스 균주 에드몬스톤	YFRLGQEMVRRSAGKVSSTL	1687	P04851.1
805	홍역 모르빌리바이러스	YFYPFRLPIKGVPIE	1688	P06830.1
806	보르데텔라 페르투시스	YGDNAGRILAGALAT	1689	P04977.1
807	풍진 바이러스	YGHATVEMPEWIHAH	1690	BAA28178.1
808	풍진 바이러스	YIMNYTGNQQSRWGL	1691	BAA28178.1
809	홍역 바이러스 균주 에드몬스톤	YINNELIPSMNQLSC	1692	P69353.1

[1037]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
810	풍진 바이러스	YLCTAPGCATQAPVP	1693	BAA28178.1
811	홍역 모르빌리바이러스	YLFTVPIKEAGEDCH	1694	P06830.1
812	홍역 바이러스 균주 에드몬스톤	YLHDPEFNL	1695	ABK40531.1
813	홍역 바이러스 균주 에드몬스톤	YLNMSRLFV	1696	ABK40531.1
814	홍역 바이러스 균주 할레	YPALGLHEF	1697	P10050.1
815	홍역 모르빌리바이러스	YPALGLHEFA	1698	ABI54110.1
816	홍역 바이러스 균주 에드몬스톤	YPALGLHEFAGELST	1699	P04851.1
817	홍역 바이러스 균주 에드몬스톤	YPALGLHEFAGELSTLESLM	1700	P04851.1
818	홍역 바이러스 균주 에드몬스톤	YPFRLPIKGVPIELQ	1701	P08362.1
819	홍역 바이러스 균주 할레	YPLLWSYAM	1702	P10050.1
820	볼거리 루불라바이러스	YQQQGRL	1703	P21186.1
821	보르데텔라 페르투시스	YQSEYLAHRR	1704	P04977.1
822	홍역 바이러스 균주 에드몬스톤	YRETGPSRASDARAA	1705	P04851.1
823	홍역 바이러스 균주 에드몬스톤	YRETGPSRASDARAAHLPTG	1706	P04851.1
824	풍진 바이러스	YRNASDVLPGHWLQGGWGCYNLSDW	1707	NP_740663.1
825	홍역 바이러스 균주 에드몬스톤	YRRLLRTVLEPIRDA	1708	P69353.1
826	보르데텔라 페르투시스	YRYDSRPP	1709	P04977.1
827	홍역 바이러스 균주 에드몬스톤	YSGGDLLGILESRGI	1710	P69353.1
828	보르데텔라 페르투시스	YSKVTATBLLASTNSRLCAVFVRDG	1711	SRC280066
829	홍역 바이러스 균주 에드몬스톤	YSPSRSFSYFYPFRL	1712	P08362.1

[1038]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
830	풍진 바이러스 균주 테리엔	YTGNQQSRWGLGSPNCHGPDWASPV	1713	P07566.1
831	홍역 모르빌리바이러스	YVLLAVLFV	1714	P08362.1
832	홍역 모르빌리바이러스	YVYSPGRSFSYFYPF	1715	P06830.1
833	보르데텔라 페르투시스	YYDYEDATFQTYALTGISLCNPAASIC	1716	P04979.1
834	홍역 모르빌리바이러스	GDLLGILESRGIKAR	1717	AAF02706.1
835	홍역 모르빌리바이러스	TVPKYVATQGYLISN	1718	AAL29688.1
836	홍역 모르빌리바이러스	KPWDSPQEI	1719	P26035.1
837	홍역 모르빌리바이러스	KPWESPQEI	1720	CAA34579.1
838	보르데텔라 페르투시스	ATYQSEYLAHRRIPP	1721	ACI04548.1
839	보르데텔라 페르투시스	CMARQAESSEAMAAWSERAGEAMVLVYYESIAYSF	1722	ACI04548.1
840	보르데텔라 페르투시스	CQVGSSNSAFVSTSSSRRYTEVYL	1723	ACI04548.1
841	보르데텔라 페르투시스	DDPPATVYRYDSRPP	1724	ACI04548.1
842	보르데텔라 페르투시스	GALATYQSEYLAHRRIPP	1725	ACI04548.1
843	보르데텔라 페르투시스	MAAWSERAGEAMVLVYYESIAYSF	1726	ACI04548.1
844	보르데텔라 페르투시스	MVLVYYESIAYSF	1727	ACI04548.1
845	보르데텔라 페르투시스	PATVYRYDSRPPEDV	1728	ACI04548.1
846	보르데텔라 페르투시스	YDSRPPEDV	1729	ACI04548.1
847	보르데텔라 페르투시스	EPGITTNYDT	1730	ACI16087.1
848	보르데텔라 페르투시스	GDLRAYKMVYATNPQTQLSN	1731	ACI16083.1
849	보르데텔라 페르투시스	KNGDVEASAITTYVGFSVVYP	1732	ACI16083.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
850	보르데텔라 페르투시스	KVTNGSKSYTLRYLASYVK	1733	ACI16088.1
851	보르데텔라 페르투시스	QALGALKLYFEPGITTNYDTGDLIAYKQTYNASGN	1734	ACI16088.1
852	보르데텔라 페르투시스	YATNPQTQLS	1735	ACI16083.1
853	코리네박테리움 디프테리아에	DNENPLSGKAGGVVKVTYPGLTKV	1736	AAV70486.1
854	코리네박테리움 디프테리아에	ENFSSYHGTKPGYVDSI	1737	AAV70486.1
855	코리네박테리움 디프테리아에	KVDNAETIKKELGLSLTEP	1738	AAV70486.1
856	풍진 바이러스 균주 M33	MEDLQKALEAQSRALRAGLAA	1739	CAA28880.1
857	코리네박테리움 디프테리아에	QKGIQKPKSGTQGNYDDDWKGFY	1740	AAV70486.1
858	풍진 바이러스 균주 M33	RTGAWQRKDWSRAPPPPEERQESRSQTPAPKPSR	1741	CAA28880.1
859	풍진 바이러스	AAGASQSRRPRPPRHARAQHLPEMTPAVT	1742	SRC265968
860	풍진 바이러스	CVTSWLWSEGEGAVFYRVDLHFINLGTP	1743	CAA28880.1
861	풍진 바이러스	FRVGGTRWHRLLRMPVRGLDGDTAPLP	1744	CAA28880.1
862	코리네박테리움 디프테리아에	GRKIRMRCRAIDGDVTFCRPKSPVYVGN	1745	1007216A
863	풍진 바이러스	GTPPLDEDGRWDPALMYNPCGPEPPAHV	1746	CAA28880.1
864	코리네박테리움 디프테리아에	GVHANLHVAFHRSSSEKIHSNEISSDSIGVLGYQKTVDHTKVNSKLSLFFEIKS	1747	AAV70486.1
865	풍진 바이러스	MASTTPITMEDLQKALEAQSRALRAGLAAG	1748	ABD64200.1
866	풍진 바이러스	PELGPPTNPFQAAVARGLRPPLHDPDTEAPTEAC	1749	CAA28880.1
867	풍진 바이러스	PLPPHTTERIETRSARHPWRIRFGAP	1750	CAA28880.1
868	풍진 바이러스	SRAPPPPEERQESRSQTPAPKPSRAPP	1751	CAA28880.1
869	풍진 바이러스	SRAPPQQPQPPRMQTGRGGSAPRPELGP	1752	CAA28880.1
870	풍진 바이러스	TPAVTPEGPAPPRTGAWQRKDWSRAPP	1753	CAA28880.1

[1039]

[1040]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
871	풍진 바이러스	VRAYNQPAGDVRGVWGKGERTYAEQDFRV	1754	CAA28880.1
872	풍진 바이러스	AFGHSDAACWGFPTDTVMSV	1755	CAA28880.1
873	풍진 바이러스	CARIWNGTQRACTFWAVNAYS	1756	CAA28880.1
874	풍진 바이러스	EEAFTYLCTAPGCATQTPVPVR	1757	CAA28880.1
875	풍진 바이러스	FAPWDLEATGACICEIPTDV	1758	CAA28880.1
876	풍진 바이러스	GEDVGAFPPGKFVTAAL	1759	CAA28880.1
877	풍진 바이러스	GEVWVTPVIGSQARKCGLHI	1760	CAA28880.1
878	풍진 바이러스	GQLEVQVPPDPGDLVEYIMN	1761	CAA28880.1
879	풍진 바이러스	GSYYKQYHPTACEVEPAFGH	1762	CAA28880.1
880	풍진 바이러스	IHAHTTSDPWHPPGPLGLKF	1763	CAA28880.1
881	풍진 바이러스	IMNYTGNQQSRWGLGSPNCH	1764	CAA28880.1
882	풍진 바이러스	LHIRAGPYGHATVEMPEWIH	1765	CAA28880.1
883	풍진 바이러스	LKFKTVRPVALPRALAPPRN	1766	CAA28880.1
884	풍진 바이러스	LNTPPPYQVSCGGESDRASAGH	1767	CAA28880.1
885	풍진 바이러스	NCHGPDWASPVCQRHSPDCS	1768	CAA28880.1
886	풍진 바이러스	PDCSRLVGATPERPRLRLVD	1769	CAA28880.1
887	풍진 바이러스	PRNVRVTGCYQCGTPALVEG	1770	CAA28880.1
888	풍진 바이러스	PTDVSCEGLGAWVPTAPCARI	1771	CAA28880.1
889	풍진 바이러스	RLVDADDPLLRTAPGPGEVW	1772	CAA28880.1
890	풍진 바이러스	SVFALASYVQHPHKTVRVKF	1773	CAA28880.1
891	풍진 바이러스	VEGLAPGGGNCHLTVNGEDV	1774	CAA28880.1
892	풍진 바이러스	VKFHTETRTVWQLSVAGVSC	1775	CAA28880.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
893	풍진 바이러스	VPVRLAGVGFESKIVDGGCF	1776	CAA28880.1
894	풍진 바이러스	VSCNVTTEHPFCNTPHGQLE	1777	CAA28880.1
895	보르데텔라 페르투시스	AAASSPDAHVPF	1778	AAA22983.1
896	보르데텔라 페르투시스	AASSPDA	1779	AAA22983.1
897	보르데텔라 페르투시스	AKLGAAASSPDA	1780	AAA22983.1
898	보르데텔라 페르투시스	AMKPYEVTPTRM	1781	AAA22983.1
899	보르데텔라 페르투시스	AMTHLSPALADVPYVLVKTNMVVTS	1782	AAA22983.1
900	보르데텔라 페르투시스	ASSPDAHVPFCF	1783	AAA22983.1
901	보르데텔라 페르투시스	ASSPDAHVPFCFGKDLKRPGSSPME	1784	AAA22983.1
902	보르데텔라 페르투시스	CFGKDLKRPGSS	1785	AAA22983.1
903	보르데텔라 페르투시스	CFGKDLKRPGSSPMEVMLRAVFMQQ	1786	AAA22983.1
904	보르데텔라 페르투시스	CGIAAKLGAAAS	1787	AAA22983.1
905	보르데텔라 페르투시스	CGIAAKLGAAASSPDAHVPFCFGKD	1788	AAA22983.1
906	보르데텔라 페르투시스	DAHVPFCFGKDL	1789	AAA22983.1
907	보르데텔라 페르투시스	DLKRPGSSPMEV	1790	AAA22983.1
908	보르데텔라 페르투시스	DVPYVLVKTNMV	1791	AAA22983.1
909	보르데텔라 페르투시스	DVPYVLVKTNMVVTSVAMKPYEVTPT	1792	AAA22983.1
910	보르데텔라 페르투시스	EVMLRAVFMQQR	1793	AAA22983.1
911	보르데텔라 페르투시스	FEGKPALELIRM	1794	AAA22983.1
912	보르데텔라 페르투시스	FLGPKQLTFEGK	1795	AAA22983.1

[1041]

[1042]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
913	보르데텔라 페르투시스	FLGPKQLTFEGKPALELIRMVECSG	1796	AAA22983.1
914	보르데텔라 페르투시스	FMQQRPLRMFLGPKQLT	1797	AAA22983.1
915	보르데텔라 페르투시스	GKDLKRPGSSPM	1798	AAA22983.1
916	보르데텔라 페르투시스	GKDLKRPGSSPME	1799	AAA22983.1
917	보르데텔라 페르투시스	GKPALELIRMVE	1800	AAA22983.1
918	보르데텔라 페르투시스	GPKQLTFEGKPA	1801	AAA22983.1
919	보르데텔라 페르투시스	HVPFCFGKDLKR	1802	AAA22983.1
920	보르데텔라 페르투시스	IAAKLGAAASSP	1803	AAA22983.1
921	보르데텔라 페르투시스	KPYEVTPTRMLV	1804	AAA22983.1
922	보르데텔라 페르투시스	KQLTFEGKPALE	1805	AAA22983.1
923	보르데텔라 페르투시스	KRPGSSPMEVML	1806	AAA22983.1
924	보르데텔라 페르투시스	LELIRMVECSGK	1807	AAA22983.1
925	보르데텔라 페르투시스	LGAAASSPDAHV	1808	AAA22983.1
926	보르데텔라 페르투시스	LIRMVECSGKQD	1809	AAA22983.1
927	보르데텔라 페르투시스	LVCGIAAKLGAA	1810	AAA22983.1
928	보르데텔라 페르투시스	MKPYEVTPTRM	1811	AAA22983.1
929	보르데텔라 페르투시스	MLRAVFMQQRPL	1812	AAA22983.1
930	보르데텔라 페르투시스	MQQRPLRM	1813	AAA22983.1
931	보르데텔라 페르투시스	MQQRPLRMFLGP	1814	AAA22983.1
932	보르데텔라 페르투시스	MVVTSVAMKPYE	1815	AAA22983.1

[1043]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
933	보르데텔라 페르투시스	MVVTSVAMKPYEVTPTRMLVCGIAA	1816	AAA22983.1
934	보르데텔라 페르투시스	PALELIRMVECS	1817	AAA22983.1
935	보르데텔라 페르투시스	PALELIRMVECSGK	1818	AAA22983.1
936	보르데텔라 페르투시스	PFCFGKDLKRPG	1819	AAA22983.1
937	보르데텔라 페르투시스	PGSSPMEVMLRA	1820	AAA22983.1
938	보르데텔라 페르투시스	PGSSPMEVMLRAVF	1821	AAA22983.1
939	보르데텔라 페르투시스	PKQLTFEGK	1822	AAA22983.1
940	보르데텔라 페르투시스	PLRMFLGPKQLT	1823	AAA22983.1
941	보르데텔라 페르투시스	PTRMLVCGIAAK	1824	AAA22983.1
942	보르데텔라 페르투시스	PYVLVKTNMVVT	1825	AAA22983.1
943	보르데텔라 페르투시스	QLTFEGKPALELIRMVECSGKQDCP	1826	AAA22983.1
944	보르데텔라 페르투시스	QRPLRMFLGPKQ	1827	AAA22983.1
945	보르데텔라 페르투시스	RAVFMQQRPLRM	1828	AAA22983.1
946	보르데텔라 페르투시스	RMFLGPKQLTFE	1829	AAA22983.1
947	보르데텔라 페르투시스	RMLVCGIAAKLG	1830	AAA22983.1
948	보르데텔라 페르투시스	RMVECSGKQDCP	1831	AAA22983.1
949	보르데텔라 페르투시스	SPDAHVPFCFGK	1832	AAA22983.1
950	보르데텔라 페르투시스	SSPMEVMLRAVF	1833	AAA22983.1
951	보르데텔라 페르투시스	SSPMEVMLRAVFMQQRPLRMFLGPK	1834	AAA22983.1
952	보르데텔라 페르투시스	SVAMKPYEVTPT	1835	AAA22983.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
953	보르데텔라 페르투시스	VECSGKQDCP	1836	AAA22983.1
954	보르데텔라 페르투시스	VFMQQRPLRMFL	1837	AAA22983.1
955	보르데텔라 페르투시스	VFMQQRPLRMFLGPKQLTFEGKPAL	1838	AAA22983.1
956	보르데텔라 페르투시스	VKTNMVVTSVAM	1839	AAA22983.1
957	보르데텔라 페르투시스	VLVKTNMVVTSV	1840	AAA22983.1
958	보르데텔라 페르투시스	VTPTRMLVCGIA	1841	AAA22983.1
959	보르데텔라 페르투시스	VTSVAMKPYEVT	1842	AAA22983.1
960	보르데텔라 페르투시스	YEVTPTRMLVCG	1843	AAA22983.1
961	보르데텔라 페르투시스	YEVTPTRMLVCGIAAKLGAAASSPD	1844	AAA22983.1
962	보르데텔라 페르투시스	CASPYEGRYRDMYDALR	1845	P04979.1
963	보르데텔라 페르투시스	CAVFVRDGQSV	1846	P04979.1
964	보르데텔라 페르투시스	CITTIYKTG	1847	P04979.1
965	보르데텔라 페르투시스	CPNGTRALTV	1848	P04979.1
966	보르데텔라 페르투시스	DALRRLLYMIYMSG	1849	P04979.1
967	풍진 바이러스 균주 테리엔	GNRGRGQRRDWSRAPPPPEERQETRS	1850	P07566.1
968	보르데텔라 페르투시스	GQPAADHYYSKVT	1851	P04979.1
969	풍진 바이러스	GSPNCHGPDWASPVCQRHS	1852	ABD64214.1
970	홍역 바이러스 균주 에드몬스톤	HKSLSTNLDVTNSIEHQ	1853	P08362.1
971	보르데텔라 페르투시스	LFTQQGGAYGRC	1854	P04979.1
972	홍역 바이러스 균주 에드몬스톤	LIGLLAIAGIRLHRAAIYTAEI	1855	P08362.1

[1044]

[1045]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
973	홍역 모르빌리바이러스	PDTAADSELRRWIKY	1856	ABI54110.1
974	풍진 바이러스 균주 테리엔	PNCHGPDWASPVCQRHS	1857	P07566.1
975	풍진 바이러스	QTPAPKPSRAPPQQPQPPRMQTGR	1858	ABD64216.1
976	풍진 바이러스 백신 균주 RA27/3	RAGLTAGASQSRRPRPPR	1859	CAA33016.1
977	풍진 바이러스 백신 균주 RA27/3	RFGAPQAFLAGLLLAAVAVGTARA	1860	ABD64214.1
978	보르데텔라 페르투시스	RGNAELQTYLRQITPG	1861	P04979.1
979	보르데텔라 페르투시스	RVHVSKEEQYYDYED	1862	P04979.1
980	보르데텔라 페르투시스	SIYGLYDGTYL	1863	P04979.1
981	보르데텔라 페르투시스	SKVTATRLLASTNS	1864	P04979.1
982	보르데텔라 페르투시스	TQQGGAYGRCPNGTRA	1865	P04979.1
983	보르데텔라 페르투시스	VAPGIVIPPKAL	1866	P04979.1
984	보르데텔라 페르투시스	DSRPPEDVFQNGFTAWG	1867	ACI04548.1
985	보르데텔라 페르투시스	EHRMQEAVEAERAGR	1868	ACI04548.1
986	홍역 바이러스 균주 에드몬스톤	ETCFQQACKGKIQALCENPEWA	1869	P08362.1
987	보르데텔라 페르투시스	EYVDTYGDNAGRILAGALATYQ	1870	ACI04548.1
988	보르데텔라 페르투시스	HRRIPPENIRRVTR	1871	ACI04548.1
989	보르데텔라 페르투시스	MARQAESSE	1872	ACI04548.1
990	보르데텔라 페르투시스	MQEAVEAERAGR	1873	ACI04548.1
991	보르데텔라 페르투시스	SQQTRANPNPYTSRR	1874	ACI04548.1
992	보르데텔라 페르투시스	TRANPNPYTSRRSVASIVGTLVHG	1875	SRC280066

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
993	보르데텔라 페르투시스	TVYRYDSRPPED	1876	ACI04548.1
994	홍역 모르빌리바이러스	NDRNLLD	1877	P10050.1
995	홍역 모르빌리바이러스	NMEDEADQYFSHDDPISSDQSRFGWFENK	1878	P04851.1
996	홍역 모르빌리바이러스	SRASDARAAHLPTGTPLDID	1879	P04851.1
997	보르데텔라 페르투시스	EDVFQNGFTAW	1880	ACI04548.1
998	코리네박테리움 디프테리아에	AEGSSSVEYINNWEQAK	1881	AAV70486.1
999	코리네박테리움 디프테리아에	GPIKNKMSESPNKT	1882	AAV70486.1
1000	홍역 바이러스 균주 할레	GPKLTGALIGILSLFVESPGQLIQRITDDPDV	1883	P10050.1
1001	코리네박테리움 디프테리아에	GYQKTVDHTKVNSK	1884	AAV70486.1
1002	코리네박테리움 디프테리아에	KTVDH	1885	AAV70486.1
1003	코리네박테리움 디프테리아에	SESPNK	1886	AAV70486.1
1004	코리네박테리움 디프테리아에	AEGSSSVEYINNWEQAKALS	1887	AAV70486.1
1005	코리네박테리움 디프테리아에	AQAIPLVGELVDIGFAAYNF	1888	AAV70486.1
1006	코리네박테리움 디프테리아에	ASRVVLSLPFAEGSSSVEYI	1889	AAV70486.1
1007	코리네박테리움 디프테리아에	CINLDWDVIRDKTKTKIESL	1890	AAV70486.1
1008	코리네박테리움 디프테리아에	CRAIDGDVTFCRPKSPVYVG	1891	AAV70486.1
1009	코리네박테리움 디프테리아에	CRPKSPVYVGNGVHANLHVA	1892	AAV70486.1
1010	코리네박테리움 디프테리아에	DAAGYSVDNENPLSGKAGGV	1893	AAV70486.1
1011	코리네박테리움 디프테리아에	DKTKTKIESLKEHGPIKNKM	1894	AAV70486.1
1012	코리네박테리움 디프테리아에	EEFIKRFGDGASRVVLSLPF	1895	AAV70486.1

[1046]

[1047]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1013	코리네박테리움 디프테리아에	EKAKQYLEEFHQTALEHPEL	1896	AAV70486.1
1014	코리네박테리움 디프테리아에	FHRSSSEKIHSNEISSDSIG	1897	AAV70486.1
1015	코리네박테리움 디프테리아에	GADDVVDSSKSFVMENFSSY	1898	CAE11230.1
1016	코리네박테리움 디프테리아에	GKRGQDAMYEYMAQACAGNR	1899	AAV70486.1
1017	코리네박테리움 디프테리아에	GSVMGIADGAVHHNTEEIVA	1900	AAV70486.1
1018	코리네박테리움 디프테리아에	GTQGNYDDDWKGFYSTDNKY	1901	AAV70486.1
1019	코리네박테리움 디프테리아에	HDGYAVSWNTVEDSIIRTGF	1902	AAV70486.1
1020	코리네박테리움 디프테리아에	HGTKPGYVDSIQKGIQKPKS	1903	AAV70486.1
1021	코리네박테리움 디프테리아에	HQTALEHPELSELKTVTGTN	1904	AAV70486.1
1022	코리네박테리움 디프테리아에	IQKGIQKPKSGTQGNYDDDW	1905	AAV70486.1
1023	코리네박테리움 디프테리아에	KEHGPIKNKMSESPNKTVSE	1906	AAV70486.1
1024	코리네박테리움 디프테리아에	KGFYSTDNKYDAAGYSVDNE	1907	AAV70486.1
1025	코리네박테리움 디프테리아에	LDVNKSKTHISVNGRKIRMR	1908	AAV70486.1
1026	풍진 바이러스	MASTIPITMEDLQKALEA	1909	SRC265968
1027	코리네박테리움 디프테리아에	NGVHANLHVAFHRSSSEKIH	1910	AAV70486.1
1028	코리네박테리움 디프테리아에	NNWEQAKALSVELEINFETR	1911	AAV70486.1
1029	코리네박테리움 디프테리아에	NPLSGKAGGVVKVTYPGLTK	1912	AAV70486.1
1030	코리네박테리움 디프테리아에	PVFAGANYAAWAVNVAQVID	1913	AAV70486.1
1031	코리네박테리움 디프테리아에	SELKTVTGTNPVFAGANYAA	1914	AAV70486.1
1032	코리네박테리움 디프테리아에	SESPNKTVSEEKAKQYLEEF	1915	AAV70486.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1033	코리네박테리움 디프테리아에	SETADNLEKTTAALSILPGI	1916	AAV70486.1
1034	코리네박테리움 디프테리아에	SFVMENFSSYHGTKPGYVDS	1917	AAV70486.1
1035	코리네박테리움 디프테리아에	SNEISSDSIGVLGYQKTVDH	1918	AAV70486.1
1036	코리네박테리움 디프테리아에	SPGHKTQPFLHDGYAVSWNT	1919	AAV70486.1
1037	코리네박테리움 디프테리아에	SVNGRKIRMRCRAIDGDVTF	1920	AAV70486.1
1038	코리네박테리움 디프테리아에	TAALSILPGIGSVMGIADGA	1921	AAV70486.1
1039	코리네박테리움 디프테리아에	TAENTPLPIAGVLLPTIPGK	1922	AAV70486.1
1040	코리네박테리움 디프테리아에	TEPLMEQVGTEEFIKRFGDG	1923	AAV70486.1
1041	코리네박테리움 디프테리아에	TIKKELGLSLTEPLMEQVGT	1924	AAV70486.1
1042	코리네박테리움 디프테리아에	TKVNSKLSLFFEIKS	1925	AAV70486.1
1043	코리네박테리움 디프테리아에	VEDSIIRTGFQGESGHDIKI	1926	AAV70486.1
1044	코리네박테리움 디프테리아에	VELEINFETRGKRGQDAMYE	1927	AAV70486.1
1045	코리네박테리움 디프테리아에	VESIINLFQVVHNSYNRPAY	1928	AAV70486.1
1046	코리네박테리움 디프테리아에	VHNSYNRPAYSPGHKTQPFL	1929	AAV70486.1
1047	코리네박테리움 디프테리아에	VKVTYPGLTKVLALKVDNAE	1930	AAV70486.1
1048	코리네박테리움 디프테리아에	VLALKVDNAETIKKELGLSL	1931	AAV70486.1
1049	코리네박테리움 디프테리아에	VLGYQKTVDHTKVNSKLSLF	1932	AAV70486.1
1050	코리네박테리움 디프테리아에	VRRSVGSSLSCINLDWDVIR	1933	AAV70486.1
1051	코리네박테리움 디프테리아에	WAVNVAQVIDSETADNLEKT	1934	AAV70486.1
1052	코리네박테리움 디프테리아에	YMAQACAGNRVRRSVGSSLS	1935	AAV70486.1

[1048]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1053	보르데텔라 페르투시스	AKAPPAPKPAPQPGP	1936	ABO77783.1
1054	보르데텔라 페르투시스	APKPAPQPGP	1937	ABO77783.1
1055	보르데텔라 페르투시스	APKPAPQPGPQPPQP	1938	ABO77783.1
1056	홍역 바이러스 균주 에드몬스톤	AQTRTPLQCTMTEIF	1939	P04851.1
1057	홍역 바이러스 균주 에드몬스톤	ASRGTNMEDEADQYFSHDD	1940	P04851.1
1058	보르데텔라 페르투시스	ATIRR	1941	ABO77783.1
1059	보르데텔라 페르투시스	DNRAG	1942	ABO77783.1
1060	보르데텔라 페르투시스	EAPAPQPPAGRELSA	1943	ABO77783.1
1061	홍역 바이러스 균주 에드몬스톤	EMVRRSAGKVSSTLASELGI	1944	P04851.1
1062	보르데텔라 페르투시스	GASEL	1945	ABO77783.1
1063	보르데텔라 페르투시스	GDALAGGAVP	1946	AAA22980.1
1064	보르데텔라 페르투시스	GDAPAGGAVP	1947	ABO77783.1
1065	보르데텔라 페르투시스	GDTWDDD	1948	ABO77783.1
1066	보르데텔라 페르투시스	GERQH	1949	ABO77783.1
1067	보르데텔라 페르투시스	GGAVP	1950	ABO77783.1
1068	보르데텔라 페르투시스	GGFGP	1951	P14283.3
1069	보르데텔라 페르투시스	GGFGPGGFGP	1952	BAF35031.1
1070	보르데텔라 페르투시스	GGFGPVLDGW	1953	ABO77783.1
1071	홍역 바이러스 균주 에드몬스톤	GGKEDRRVKQSRGEARESYR	1954	P04851.1
1072	보르데텔라 페르투시스	GILLEN	1955	ABO77783.1

[1049]

[1050]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1073	보르데텔라 페르투시스	GIRRFL	1956	ABO77783.1
1074	홍역 바이러스 균주 에드몬스톤	HDDPISSDQSRFGWFENKEI	1957	P04851.1
1075	홍역 바이러스 균주 에드몬스톤	HEFAGELSTLESLMNLY	1958	P04851.1
1076	보르데텔라 페르투시스	HLGGLAGY	1959	ABO77783.1
1077	홍역 바이러스 균주 에드몬스톤	HTTEDKISRAVGPRQAQVSFL	1960	P04851.1
1078	홍역 바이러스 균주 에드몬스톤	ICDIDTYIVEAGLASFILTI	1961	P04851.1
1079	홍역 바이러스 균주 에드몬스톤	IIIVPIPGDSSITTRSRLLD	1962	P04851.1
1080	홍역 바이러스 균주 에드몬스톤	ILDIKRTPGNKPRIAEMICD	1963	P04851.1
1081	보르데텔라 페르투시스	KALLYR	1964	ABO77783.1
1082	홍역 바이러스 균주 에드몬스톤	KPPITSGSGGAIRGIKHIII	1965	P04851.1
1083	보르데텔라 페르투시스	LAGSGL	1966	ABO77783.1
1084	홍역 바이러스 균주 에드몬스톤	LGITAEDARLVSEIAMHTTE	1967	P04851.1
1085	홍역 바이러스 균주 에드몬스톤	LGTILAQIWVLLAKAVTA	1968	P04851.1
1086	홍역 바이러스 균주 에드몬스톤	LPTGTPLDIDTASESSQD	1969	P04851.1
1087	홍역 바이러스 균주 에드몬스톤	MATLLRSLALFKRNKDK	1970	P04851.1
1088	보르데텔라 페르투시스	PAPQPP	1971	ABO77783.1
1089	홍역 바이러스 균주 에드몬스톤	PDTAADSELRRWIKYTQQRR	1972	P04851.1
1090	홍역 바이러스 균주 에드몬스톤	PKLTGALIGILSLFVESPGQ	1973	P04851.1
1091	보르데텔라 페르투시스	PQP	1974	ABO77783.1
1092	보르데텔라 페르투시스	PQPGP	1975	ABO77783.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1093	보르데텔라 페르투시스	PQPGPQPPQPPQPQP	1976	ABO77783.1
1094	홍역 바이러스 균주 에드몬스톤	QDPQDSRRSAEPLLSCKPWQ	1977	P04851.1
1095	홍역 바이러스 균주 에드몬스톤	QRRVVGEFRLERKWLDVVR	1978	P04851.1
1096	보르데텔라 페르투시스	RELSA	1979	ABO77783.1
1097	보르데텔라 페르투시스	RFAPQ	1980	ABO77783.1
1098	홍역 바이러스 균주 에드몬스톤	SIQNKFSAGSYPLLWSYAMG	1981	P04851.1
1099	보르데텔라 페르투시스	SITLQAGAH	1982	ABO77783.1
1100	보르데텔라 페르투시스	SLQPED	1983	ABO77783.1
1101	보르데텔라 페르투시스	SNALSKRL	1984	ABO77783.1
1102	홍역 바이러스 균주 에드몬스톤	SPGQLIQRITDDPDVSIRLL	1985	P04851.1
1103	보르데텔라 페르투시스	TELPSIPG	1986	ABO77783.1
1104	보르데텔라 페르투시스	TFTLANK	1987	ABO77783.1
1105	보르데텔라 페르투시스	TWDDD	1988	ABO77783.1
1106	홍역 바이러스 균주 에드몬스톤	VSFLQGDQSENELPRLGGKE	1989	P04851.1
1107	홍역 바이러스 균주 에드몬스톤	WQESRKNKAQTRTPLQC	1990	P04851.1
1108	홍역 바이러스 균주 에드몬스톤	YAMGVGVELENSMGGLNFGR	1991	P04851.1
1109	홍역 바이러스 균주 에드몬스톤	YQQMGKPAPYMVNLENSI	1992	P04851.1
1110	홍역 바이러스 균주 할레	MTRSSHQSLVIKLMP	1993	P69355.1
1111	홍역 바이러스 균주 할레	PIRDALNAMTQNIRP	1994	P69355.1
1112	풍진 바이러스 균주 M33	ALLNTPPPYQVSCGGESDRASAGH	1995	CAA28880.1

[1051]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1113	풍진 바이러스 균주 테리엔	GLGSPNCHGPDWASPVCQRHS	1996	P07566.1
1114	풍진 바이러스 균주 테리엔	GLGSPNCHGPDWASPVCQRHSPDCSRLV	1997	P07566.1
1115	풍진 바이러스 균주 테리엔	NYTGNQQSRWGLGSPNCHGPDWASPV	1998	P07566.1
1116	풍진 바이러스 균주 테리엔	TLPQPPCAHGQHYGHHHHQL	1999	P07566.1
1117	풍진 바이러스 균주 테리엔	TVRVKFHTETRTVWQLSVAGVSCNVT	2000	P07566.1
1118	풍진 바이러스	ASYFNPGGSYYKQYH	2001	BAA28178.1
1119	풍진 바이러스	FALASYVQHPHKTVR	2002	BAA28178.1
1120	풍진 바이러스	GGESDRASARVIDPAAQSFTG	2003	BAA28178.1
1121	풍진 바이러스	GPGEVWVTPVIGSQARKC	2004	BAA28178.1
1122	풍진 바이러스	GSQARKCGLHIRAG	2005	BAA28178.1
1123	풍진 바이러스	LVGATPERPRLRLVDADDPLLRTAP	2006	BAA28178.1
1124	풍진 바이러스	TAPGPGEVWVTPVI	2007	BAA28178.1
1125	홍역 모르빌리바이러스	FIVLSIAYPTLSEIK	2008	AAL29688.1
1126	홍역 바이러스 균주 에드몬스톤-자그레브	ETCFQQACKGKIQALCENPEWAPLKDNRIPSY	2009	P08362.1
1127	홍역 모르빌리바이러스	LPRLGGKEDR	2010	P04851.1
1128	홍역 모르빌리바이러스	MSKTEWNASQ	2011	SRC280080
1129	홍역 모르빌리바이러스	SRFGWFENKE	2012	P04851.1
1130	풍진 바이러스	EACVTSWLWSEGEGAVFYRVDLHFINLGT	2013	CAA28880.1
1131	풍진 바이러스	MDFWCVEHDRPPPATPTSLTT	2014	CAA33016.1
1132	풍진 바이러스	PFLGHDGHHGGTLRVGQHHRNASDV	2015	CAA33016.1

[1052]

[1053]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1133	풍진 바이러스	RVKFHTETRTVWQLSVAGVSC	2016	BAA19893.1
1134	홍역 모르빌리바이러스	AEEQARHVKNGLE	2017	ABO69699.1
1135	홍역 모르빌리바이러스	ESPQEISKHQALG	2018	SRC280080
1136	홍역 모르빌리바이러스	GVGVELENSMGGLNF	2019	ABI54110.1
1137	홍역 모르빌리바이러스	IKGANDLAKFHQMLMKIIMK	2020	ABO69699.1
1138	홍역 모르빌리바이러스	MSKTLHAQLGFKKT	2021	ABK40528.1
1139	홍역 모르빌리바이러스	NASGLSRPSPSAH	2022	BAE98296.1
1140	홍역 모르빌리바이러스	VRVIDPSLGDRKDE	2023	SRC280148
1141	보르데텔라 페르투시스	DLSDGDLLV	2024	AAC31207.1
1142	보르데텔라 페르투시스	EAERAGRGTG	2025	ACI04548.1
1143	보르데텔라 페르투시스	YRYDSRPPEDV	2026	ACI04548.1
1144	보르데텔라 페르투시스	YVDTYGDNAG	2027	ACI04548.1
1145	홍역 모르빌리바이러스	SFSYFYPFR	2028	CAB43772.1
1146	볼거리 루불라바이러스	DIFIVSPR	2029	ADF49557.1
1147	홍역 모르빌리바이러스	QDSRRSADALLRLQAMAGISEEQGSDTDTPIVYNDRN	2030	CAA59302.1
1148	홍역 모르빌리바이러스	SAEALLRLQA	2031	BAH22350.1
1149	홍역 모르빌리바이러스	RIVINREHL	2032	BAB39835.1
1150	홍역 바이러스 유전자형 A	IPRFK	2033	BAB39848.1
1151	보르데텔라 페르투시스	AAALSPMEI	2034	P15318.2
1152	보르데텔라 페르투시스	AAASVVGAPV	2035	P15318.2

[1054]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1153	보르데텔라 페르투시스	AALGRQDSI	2036	P15318.2
1154	보르데텔라 페르투시스	AAQRLVHAIA	2037	P15318.2
1155	보르데텔라 페르투시스	AAVEAAEL	2038	P15318.2
1156	보르데텔라 페르투시스	AGANVLNGL	2039	P15318.2
1157	보르데텔라 페르투시스	AGYANAAD	2040	P15318.2
1158	보르데텔라 페르투시스	AGYEQFEFRV	2041	P15318.2
1159	보르데텔라 페르투시스	AITGNADNL	2042	P15318.2
1160	보르데텔라 페르투시스	AKEKNATLM	2043	P15318.2
1161	보르데텔라 페르투시스	AKGVFLSL	2044	P15318.2
1162	보르데텔라 페르투시스	APHEYGFGI	2045	P15318.2
1163	보르데텔라 페르투시스	ARQGNDLEI	2046	P15318.2
1164	보르데텔라 페르투시스	ASVVGAPV	2047	P15318.2
1165	보르데텔라 페르투시스	ATLMFRLV	2048	P15318.2
1166	보르데텔라 페르투시스	AVAAAQRL	2049	P15318.2
1167	보르데텔라 페르투시스	DAGANVLNGL	2050	P15318.2
1168	보르데텔라 페르투시스	DALLAQLYR	2051	P15318.2
1169	보르데텔라 페르투시스	DANGVLKHSI	2052	P15318.2
1170	보르데텔라 페르투시스	DGDMNIGVI	2053	P15318.2
1171	보르데텔라 페르투시스	DHVKNIENL	2054	P15318.2
1172	보르데텔라 페르투시스	DIDMFAIM	2055	P15318.2

[1055]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1173	보르데텔라 페르투시스	DMFAIMPHL	2056	P15318.2
1174	보르데텔라 페르투시스	DNVRNVENV	2057	P15318.2
1175	보르데텔라 페르투시스	DNVRNVENVI	2058	P15318.2
1176	보르데텔라 페르투시스	DTVDYSAM	2059	P15318.2
1177	보르데텔라 페르투시스	DTVDYSAMI	2060	P15318.2
1178	보르데텔라 페르투시스	DYLRQAGL	2061	P15318.2
1179	보르데텔라 페르투시스	DYYDNVRNV	2062	P15318.2
1180	보르데텔라 페르투시스	EFTTFVEI	2063	P15318.2
1181	보르데텔라 페르투시스	EFTTFVEIV	2064	P15318.2
1182	보르데텔라 페르투시스	EGYVFYEN	2065	P15318.2
1183	보르데텔라 페르투시스	ENVQYRHV	2066	P15318.2
1184	보르데텔라 페르투시스	EQLANSDGL	2067	P15318.2
1185	보르데텔라 페르투시스	FGVGYGHDTI	2068	P15318.2
1186	보르데텔라 페르투시스	FSPDVLETVP	2069	P15318.2
1187	보르데텔라 페르투시스	FSVDHVKNI	2070	P15318.2
1188	보르데텔라 페르투시스	GDDTYLFGV	2071	P15318.2
1189	보르데텔라 페르투시스	GDDVFLQDL	2072	P15318.2
1190	보르데텔라 페르투시스	GEDGNDIFL	2073	P15318.2
1191	보르데텔라 페르투시스	GERFNVRKQL	2074	P15318.2
1192	보르데텔라 페르투시스	GGAGNDTLV	2075	P15318.2

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1193	보르데텔라 페르투시스	GGDDFEAV	2076	P15318.2
1194	보르데텔라 페르투시스	GKSEFTTFV	2077	P15318.2
1195	보르데텔라 페르투시스	GKSLFDDGL	2078	P15318.2
1196	보르데텔라 페르투시스	GNADNLKSV	2079	P15318.2
1197	보르데텔라 페르투시스	GQLVEVDTL	2080	P15318.2
1198	보르데텔라 페르투시스	GRSKFSPDV	2081	P15318.2
1199	보르데텔라 페르투시스	GSSAYDTV	2082	P15318.2
1200	보르데텔라 페르투시스	GTVEKWPAL	2083	P15318.2
1201	보르데텔라 페르투시스	GVDYYDNV	2084	P15318.2
1202	보르데텔라 페르투시스	GYEQFEFRV	2085	P15318.2
1203	보르데텔라 페르투시스	HAVGAQDVV	2086	P15318.2
1204	보르데텔라 페르투시스	IAAGRIGLGI	2087	P15318.2
1205	보르데텔라 페르투시스	IGDAQANTL	2088	P15318.2
1206	보르데텔라 페르투시스	IGLGILADL	2089	P15318.2
1207	보르데텔라 페르투시스	IGNAAGIPL	2090	P15318.2
1208	보르데텔라 페르투시스	IGTSMKDVL	2091	P15318.2
1209	보르데텔라 페르투시스	IGVITDFEL	2092	P15318.2
1210	보르데텔라 페르투시스	IPLTADIDM	2093	P15318.2
1211	보르데텔라 페르투시스	ISKSALEL	2094	P15318.2
1212	보르데텔라 페르투시스	ITGNADNL	2095	P15318.2

[1056]

[1057]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1213	보르데텔라 페르투시스	KIFVVSAT	2096	P15318.2
1214	보르데텔라 페르투시스	KQLNNANVYR	2097	P15318.2
1215	보르데텔라 페르투시스	KVIGNAAGI	2098	P15318.2
1216	보르데텔라 페르투시스	LAKVVSQL	2099	P15318.2
1217	보르데텔라 페르투시스	LANDYARKI	2100	P15318.2
1218	보르데텔라 페르투시스	LDYLRQAGL	2101	P15318.2
1219	보르데텔라 페르투시스	LGKGFASL	2102	P15318.2
1220	보르데텔라 페르투시스	LGKGFASLM	2103	P15318.2
1221	보르데텔라 페르투시스	LGVDYYDN	2104	P15318.2
1222	보르데텔라 페르투시스	LGVDYYDNV	2105	P15318.2
1223	보르데텔라 페르투시스	LKHSIKLDVI	2106	P15318.2
1224	보르데텔라 페르투시스	LQAGYIPV	2107	P15318.2
1225	보르데텔라 페르투시스	LQLTGGTVE	2108	P15318.2
1226	보르데텔라 페르투시스	LSAAVFGL	2109	P15318.2
1227	보르데텔라 페르투시스	LSLGKGFASL	2110	P15318.2
1228	보르데텔라 페르투시스	LSPMEIYGL	2111	P15318.2
1229	보르데텔라 페르투시스	NAHDNFLAGG	2112	P15318.2
1230	보르데텔라 페르투시스	NANVYREGV	2113	P15318.2
1231	보르데텔라 페르투시스	NDTLYGGL	2114	P15318.2
1232	보르데텔라 페르투시스	NGLAGNDVL	2115	P15318.2

[1058]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1233	보르데텔라 페르투시스	NNANVYREGV	2116	P15318.2
1234	보르데텔라 페르투시스	NTVSYAAL	2117	P15318.2
1235	보르데텔라 페르투시스	NVLRNIENAV	2118	P15318.2
1236	보르데텔라 페르투시스	PALTFITPL	2119	P15318.2
1237	보르데텔라 페르투시스	PETSNVLRNI	2120	P15318.2
1238	보르데텔라 페르투시스	PMEIYGLV	2121	P15318.2
1239	보르데텔라 페르투시스	PQAYFEKNL	2122	P15318.2
1240	보르데텔라 페르투시스	PVNPNLSKL	2123	P15318.2
1241	보르데텔라 페르투시스	QAGWNASSV	2124	P15318.2
1242	보르데텔라 페르투시스	QAGWNASSVI	2125	P15318.2
1243	보르데텔라 페르투시스	QDAANAGNL	2126	P15318.2
1244	보르데텔라 페르투시스	QDAANAGNLL	2127	P15318.2
1245	보르데텔라 페르투시스	QDSGYDSL	2128	P15318.2
1246	보르데텔라 페르투시스	QQSHYADQL	2129	P15318.2
1247	보르데텔라 페르투시스	RALQGAQAV	2130	P15318.2
1248	보르데텔라 페르투시스	RGGLGLDTL	2131	P15318.2
1249	보르데텔라 페르투시스	RKQLNNANV	2132	P15318.2
1250	보르데텔라 페르투시스	RQDSGYDSL	2133	P15318.2
1251	보르데텔라 페르투시스	RQFRYDGDM	2134	P15318.2
1252	보르데텔라 페르투시스	RSKFSPDVL	2135	P15318.2

[1059]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1253	보르데텔라 페르투시스	SAGAAAGAL	2136	P15318.2
1254	보르데텔라 페르투시스	SAHWGQRAL	2137	P15318.2
1255	보르데텔라 페르투시스	SAMIHPGRI	2138	P15318.2
1256	보르데텔라 페르투시스	SAMIHPGRIV	2139	P15318.2
1257	보르데텔라 페르투시스	SAYDTVSGI	2140	P15318.2
1258	보르데텔라 페르투시스	SAYGYEGD	2141	P15318.2
1259	보르데텔라 페르투시스	SGGAGDDVL	2142	P15318.2
1260	보르데텔라 페르투시스	SGLQVAGA	2143	P15318.2
1261	보르데텔라 페르투시스	SGYDSLDGV	2144	P15318.2
1262	보르데텔라 페르투시스	SLLTGALNGI	2145	P15318.2
1263	보르데텔라 페르투시스	SPMEIYGL	2146	P15318.2
1264	보르데텔라 페르투시스	SQMLTRGQL	2147	P15318.2
1265	보르데텔라 페르투시스	SSAYDTVSGI	2148	P15318.2
1266	보르데텔라 페르투시스	SSLAHGHTA	2149	P15318.2
1267	보르데텔라 페르투시스	SSVTSGDSV	2150	P15318.2
1268	보르데텔라 페르투시스	SVIGVQTTEI	2151	P15318.2
1269	보르데텔라 페르투시스	TNTVSYAAL	2152	P15318.2
1270	보르데텔라 페르투시스	TSLIAEGV	2153	P15318.2
1271	보르데텔라 페르투시스	TSLLTGAL	2154	P15318.2
1272	보르데텔라 페르투시스	TVPASPGL	2155	P15318.2

[1060]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1273	보르데텔라 페르투시스	VAKEKNATL	2156	P15318.2
1274	보르데텔라 페르투시스	VAPHEYGFGI	2157	P15318.2
1275	보르데텔라 페르투시스	VAVVTSLL	2158	P15318.2
1276	보르데텔라 페르투시스	VFYENRAYG	2159	P15318.2
1277	보르데텔라 페르투시스	VFYENRAYGV	2160	P15318.2
1278	보르데텔라 페르투시스	VITDFELEV	2161	P15318.2
1279	보르데텔라 페르투시스	VNPHSTSL	2162	P15318.2
1280	보르데텔라 페르투시스	VNPHSTSLI	2163	P15318.2
1281	보르데텔라 페르투시스	VNPNLSKL	2164	P15318.2
1282	보르데텔라 페르투시스	VNPNLSKLF	2165	P15318.2
1283	보르데텔라 페르투시스	VQQPIIEKL	2166	P15318.2
1284	보르데텔라 페르투시스	VQYRHVEL	2167	P15318.2
1285	보르데텔라 페르투시스	VSIAAAASV	2168	P15318.2
1286	보르데텔라 페르투시스	VSIAAAASVV	2169	P15318.2
1287	보르데텔라 페르투시스	VTSLLTGAL	2170	P15318.2
1288	보르데텔라 페르투시스	VVLANASRI	2171	P15318.2
1289	보르데텔라 페르투시스	WPALNLFSV	2172	P15318.2
1290	보르데텔라 페르투시스	WVRKASAL	2173	P15318.2
1291	보르데텔라 페르투시스	YAVQYRRKGG	2174	P15318.2
1292	보르데텔라 페르투시스	YGGLGDDTL	2175	P15318.2

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1293	보르데텔라 페르투시스	YGLVQQSHYA	2176	P15318.2
1294	보르데텔라 페르투시스	YGYEGDALL	2177	P15318.2
1295	보르데텔라 페르투시스	YIPVNPNL	2178	P15318.2
1296	보르데텔라 페르투시스	YSAMIHPGRI	2179	P15318.2
1297	보르데텔라 페르투시스	YSQTGAHAGI	2180	P15318.2
1298	코리네박테리움 디프테리아에	AYNFVESIINLFQVVHNSYN	2181	CAE11230.1
1299	보르데텔라 페르투시스	SGTTIK	2182	BAF35031.1
1300	보르데텔라 페르투시스	RGHTLESAEGRKIFG	2183	AAA22974.1
1301	보르데텔라 페르투시스	AGAMTVRDVAAAADLALQAGDA	2184	AAA22974.1
1302	보르데텔라 페르투시스	AGAMTVRDVAAAADLALQAGDAL	2185	AAA22974.1
1303	보르데텔라 페르투시스	ALAAVLVNPHIFTRIGAAQTSLADGA AGPA	2186	AAA22974.1
1304	보르데텔라 페르투시스	ALSIDSMTALGA	2187	AAA22974.1
1305	보르데텔라 페르투시스	DLSAARGADISGEGR	2188	AAA22974.1
1306	보르데텔라 페르투시스	DQNRYEYIWGLY	2189	AAA22974.1
1307	보르데텔라 페르투시스	DYTVSADAIALA	2190	AAA22974.1
1308	보르데텔라 페르투시스	GPIVVEAGELVSHAGG	2191	AAA22974.1
1309	보르데텔라 페르투시스	GRPEGLKIGAHSATSVSGSFDAL	2192	AAA22974.1
1310	보르데텔라 페르투시스	ITVTSRGGFDNEGKMESNK	2193	AAA22974.1
1311	보르데텔라 페르투시스	LDQNRYEYIWGLYP	2194	AAA22974.1
1312	보르데텔라 페르투시스	LSAARGADISG	2195	AAA22974.1

[1061]

[1062]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1313	보르데텔라 페르투시스	NKIRLMGPLQ	2196	AAA22974.1
1314	보르데텔라 페르투시스	NKLGRIRAGEDM	2197	AAA22974.1
1315	보르데텔라 페르투시스	NKLGRIRAGEDMHLDAPRIE	2198	AAA22974.1
1316	보르데텔라 페르투시스	PHLRNTGQVVAG	2199	AAA22974.1
1317	보르데텔라 페르투시스	QVDLHDLSAARGADISG	2200	AAA22974.1
1318	보르데텔라 페르투시스	RDVAAAADLALQ	2201	AAA22974.1
1319	보르데텔라 페르투시스	SAARGADISGEG	2202	AAA22974.1
1320	보르데텔라 페르투시스	TKGEMQIAGKGGGSP	2203	AAA22974.1
1321	보르데텔라 페르투시스	TVSADAIALAAQ	2204	AAA22974.1
1322	보르데텔라 페르투시스	VVAGHDIHI	2205	AAA22974.1
1323	보르데텔라 페르투시스	PSGPNHTKVVQLPKISKNALKANG	2206	CAD12823.1
1324	풍진 바이러스	LVGATPERPRLRLVDADDPLLRTAPG PGEVWVTPVIGSQAR	2207	BAA19902.1
1325	풍진 바이러스	QQSRWGLGSPNCHGPDWASPVCQRH SP	2208	BAA19902.1
1326	보르데텔라 페르투시스	AGEAMVLVYYESIAYSF	2209	ACI04548.1
1327	보르데텔라 페르투시스	GGVGLASTLWYAESNALSKRLGEL	2210	AAZ74322.1
1328	보르데텔라 페르투시스	GTLVRIAPVIGACMARQA	2211	ACI04548.1
1329	보르데텔라 페르투시스	IRRVTRVYHNGITGETTT	2212	ACI04548.1
1330	보르데텔라 페르투시스	IVKTGERQHGIHIQGSDP	2213	AAZ74322.1
1331	보르데텔라 페르투시스	IVKTGERQHGIHIQGSDPGGVRTA	2214	AAZ74338.1
1332	보르데텔라 페르투시스	LRDTNVTAVPASGAPAAVSVLGAS	2215	AAZ74338.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1333	보르데텔라 페르투시스	PEAPAPQPPAGRELSAAANAAVNT	2216	AAZ74322.1
1334	보르데텔라 페르투시스	AAADFAHAE	2217	WP_019247158.1
1335	보르데텔라 페르투시스	AAAEVAGAL	2218	WP_019249248.1
1336	보르데텔라 페르투시스	AAESTFESY	2219	WP_019247158.1
1337	보르데텔라 페르투시스	AAGFDPEVQ	2220	WP_019248145.1
1338	보르데텔라 페르투시스	AALGRGHSL	2221	AGS56996.1
1339	보르데텔라 페르투시스	AAMQGAVVH	2222	AGT50936.1
1340	보르데텔라 페르투시스	AAPAAHADW	2223	AGS56996.1
1341	보르데텔라 페르투시스	AAQATVVQR	2224	WP_019247158.1
1342	보르데텔라 페르투시스	AARVAGDNY	2225	WP_019249248.1
1343	보르데텔라 페르투시스	AAVALLNKL	2226	WP_019249248.1
1344	보르데텔라 페르투시스	ADDPPATVY	2227	AAW72734.1
1345	보르데텔라 페르투시스	AEAGRFKVL	2228	AGS56996.1
1346	보르데텔라 페르투시스	AEATQLVTA	2229	WP_019247158.1
1347	보르데텔라 페르투시스	AEGGATLGA	2230	WP_019249248.1
1348	보르데텔라 페르투시스	AEHGEVSIQ	2231	WP_019249248.1
1349	보르데텔라 페르투시스	AEIAFYPKE	2232	WP_019249248.1
1350	보르데텔라 페르투시스	AEKVTTPAV	2233	WP_019247158.1
1351	보르데텔라 페르투시스	AELQTYLRQ	2234	1BCP_C
1352	보르데텔라 페르투시스	AEQSLIEVG	2235	WP_019249248.1

[1063]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1353	보르데텔라 페르투시스	AESNALSKR	2236	AGS56996.1
1354	보르데텔라 페르투시스	AESSEAMAA	2237	AFK26302.1
1355	보르데텔라 페르투시스	AEVKVGYRA	2238	WP_019247158.1
1356	보르데텔라 페르투시스	AEVTDTSPS	2239	WP_019249248.1
1357	보르데텔라 페르투시스	AGKSLKKKN	2240	WP_019247158.1
1358	보르데텔라 페르투시스	AGLAGPSAV	2241	WP_019249248.1
1359	보르데텔라 페르투시스	AIRVGRGAR	2242	AGT50936.1
1360	보르데텔라 페르투시스	ALAAIASAA	2243	WP_019248145.1
1361	보르데텔라 페르투시스	ALADVPYVL	2244	AAA22983.1
1362	보르데텔라 페르투시스	ALANDGTIV	2245	WP_019248658.1
1363	보르데텔라 페르투시스	ALGRGHSLY	2246	AGS56996.1
1364	보르데텔라 페르투시스	ALILAASPV	2247	WP_019248658.1
1365	보르데텔라 페르투시스	ALMLACTGL	2248	AAA22974.1
1366	보르데텔라 페르투시스	AMQGAVVHL	2249	AGT50936.1
1367	보르데텔라 페르투시스	AMTHLSPAL	2250	AAA22983.1
1368	보르데텔라 페르투시스	AMYGKHITL	2251	WP_019249248.1
1369	보르데텔라 페르투시스	ANEANALLW	2252	WP_019249248.1
1370	보르데텔라 페르투시스	APLSITLQA	2253	AGT50936.1
1371	보르데텔라 페르투시스	APNALAWAL	2254	AAA22974.1
1372	보르데텔라 페르투시스	APPAPKPAP	2255	AGS56996.1

[1064]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1373	보르데텔라 페르투시스	APPGAGFIY	2256	1BCP_C
1374	보르데텔라 페르투시스	APQAAPLSI	2257	AGT50936.1
1375	보르데텔라 페르투시스	APRIENTAK	2258	WP_019249248.1
1376	보르데텔라 페르투시스	AQGKALLYR	2259	AGT50936.1
1377	보르데텔라 페르투시스	AQITSYVGF	2260	WP_019248658.1
1378	보르데텔라 페르투시스	AQLEVRGQR	2261	WP_019249248.1
1379	보르데텔라 페르투시스	AQQLKQADR	2262	WP_019247699.1
1380	보르데텔라 페르투시스	AQVTVAGRY	2263	WP_019249248.1
1381	보르데텔라 페르투시스	ARRSRVRAL	2264	NP_882284.1
1382	보르데텔라 페르투시스	ASPRRARRA	2265	WP_019249248.1
1383	보르데텔라 페르투시스	ASSPDAHVP	2266	AAA22983.1
1384	보르데텔라 페르투시스	ASVSNPGTF	2267	WP_019249248.1
1385	보르데텔라 페르투시스	ATWNFQSTY	2268	WP_019249248.1
1386	보르데텔라 페르투시스	ATYIADSGF	2269	AGS56996.1
1387	보르데텔라 페르투시스	AVAAPAVGA	2270	WP_019249248.1
1388	보르데텔라 페르투시스	AVFMQQRPL	2271	AAA22983.1
1389	보르데텔라 페르투시스	AVLVNPHIF	2272	WP_019249248.1
1390	보르데텔라 페르투시스	CFGKDLKRP	2273	AAA22983.1
1391	보르데텔라 페르투시스	CPSSLGNGV	2274	WP_019248145.1
1392	보르데텔라 페르투시스	DAGHEHDTW	2275	AAA22984.1

[1065]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1393	보르데텔라 페르투시스	DAKHDLTVT	2276	WP_019249248.1
1394	보르데텔라 페르투시스	DASGQHRLW	2277	AGS56996.1
1395	보르데텔라 페르투시스	DATFETYAL	2278	YP_006628018.1
1396	보르데텔라 페르투시스	DATFQTYAL	2279	1BCP_C
1397	보르데텔라 페르투시스	DATLVGAKF	2280	WP_019247158.1
1398	보르데텔라 페르투시스	DDEVDVSGR	2281	AAA22974.1
1399	보르데텔라 페르투시스	DENGKPQTY	2282	WP_019247158.1
1400	보르데텔라 페르투시스	DGPPSRPTT	2283	WP_019247158.1
1401	보르데텔라 페르투시스	DHLTGRSCQ	2284	AAW72734.1
1402	보르데텔라 페르투시스	DNEGKMESN	2285	WP_019249248.1
1403	보르데텔라 페르투시스	DPPATVYRY	2286	AAW72734.1
1404	보르데텔라 페르투시스	EATEGDATL	2287	WP_019247158.1
1405	보르데텔라 페르투시스	EATQQAAGF	2288	WP_019248145.1
1406	보르데텔라 페르투시스	ECSGKQDCP	2289	AAA22983.1
1407	보르데텔라 페르투시스	EGGKLRGKD	2290	AAA22974.1
1408	보르데텔라 페르투시스	EGKMESNKD	2291	WP_019249248.1
1409	보르데텔라 페르투시스	EHRMQEAVE	2292	AAW72734.1
1410	보르데텔라 페르투시스	EKRLDIDDA	2293	WP_019249248.1
1411	보르데텔라 페르투시스	EPASANTLL	2294	AGS56996.1
1412	보르데텔라 페르투시스	EPQAELAVF	2295	AGS56996.1

[1066]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1413	보르데텔라 페르투시스	EPVKLTLTG	2296	AGT50936.1
1414	보르데텔라 페르투시스	ESAEGRKIF	2297	WP_019249248.1
1415	보르데텔라 페르투시스	ESYSESHNF	2298	WP_019247158.1
1416	보르데텔라 페르투시스	ETFCITTIY	2299	1BCP_C
1417	보르데텔라 페르투시스	EVAGALELS	2300	WP_019249248.1
1418	보르데텔라 페르투시스	EVAKVEVVP	2301	WP_019247158.1
1419	보르데텔라 페르투시스	EVDGIIQEF	2302	WP_019249248.1
1420	보르데텔라 페르투시스	EVRADNNFY	2303	WP_019248344.1
1421	보르데텔라 페르투시스	FAILSSTTE	2304	WP_019247158.1
1422	보르데텔라 페르투시스	FAISAYALK	2305	AAA22984.1
1423	보르데텔라 페르투시스	FALYDGTYL	2306	AFK26303.1
1424	보르데텔라 페르투시스	FDTMLGFAI	2307	AAA22984.1
1425	보르데텔라 페르투시스	FEGKPALEL	2308	AAA22983.1
1426	보르데텔라 페르투시스	FELGADHAV	2309	AGS56996.1
1427	보르데텔라 페르투시스	FEPGITTNY	2310	WP_019248658.1
1428	보르데텔라 페르투시스	FETYALTGI	2311	YP_006628018.1
1429	보르데텔라 페르투시스	FIYRETFCI	2312	1BCP_C
1430	보르데텔라 페르투시스	FPTRTTAPG	2313	NP_882284.1
1431	보르데텔라 페르투시스	FQTYALTGI	2314	1BCP_C
1432	보르데텔라 페르투시스	FTHADGWFL	2315	AGS56996.1

[1067]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1433	보르데텔라 페르투시스	FVRDGQSVI	2316	1BCP_C
1434	보르데텔라 페르투시스	FVRSGQPVI	2317	YP_006628018.1
1435	보르데텔라 페르투시스	FVWYVDTVI	2318	WP_019248866.1
1436	보르데텔라 페르투시스	GAASSRQAL	2319	WP_019249248.1
1437	보르데텔라 페르투시스	GAASSYFEY	2320	AAW72734.1
1438	보르데텔라 페르투시스	GAFDLKTTF	2321	AFK26303.1
1439	보르데텔라 페르투시스	GAPAAVSVL	2322	AGS56996.1
1440	보르데텔라 페르투시스	GATRAVDSL	2323	AGS56996.1
1441	보르데텔라 페르투시스	GAVPGGAVP	2324	AGT50936.1
1442	보르데텔라 페르투시스	GEAMVLVYY	2325	AFK26302.1
1443	보르데텔라 페르투시스	GEIALGDAT	2326	WP_019249248.1
1444	보르데텔라 페르투시스	GELMAAQVA	2327	WP_019247158.1
1445	보르데텔라 페르투시스	GGVPGGAVP	2328	AAZ74338.1
1446	보르데텔라 페르투시스	GHEHDTWFD	2329	AAA22984.1
1447	보르데텔라 페르투시스	GIGALKAGA	2330	WP_019249248.1
1448	보르데텔라 페르투시스	GIVIPPKAL	2331	1BCP_C
1449	보르데텔라 페르투시스	GKDLKRPGS	2332	AAA22983.1
1450	보르데텔라 페르투시스	GKLPKPVTV	2333	WP_019247158.1
1451	보르데텔라 페르투시스	GKSLKKKNQ	2334	WP_019247158.1
1452	보르데텔라 페르투시스	GLDVQQGTV	2335	WP_019249248.1

[1068]

[1069]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1453	보르데텔라 페르투시스	GLTDGVSRI	2336	WP_019249248.1
1454	보르데텔라 페르투시스	GLYPTYTEW	2337	WP_019249248.1
1455	보르데텔라 페르투시스	GLYQTYTEW	2338	YP_006626470.1
1456	보르데텔라 페르투시스	GPPSRPTTP	2339	WP_019247158.1
1457	보르데텔라 페르투시스	GPSAVAAPA	2340	WP_019249248.1
1458	보르데텔라 페르투시스	GVAPTAQQL	2341	WP_019248866.1
1459	보르데텔라 페르투시스	GVGLASTLW	2342	AGS56996.1
1460	보르데텔라 페르투시스	GYEAGFSLG	2343	WP_019247158.1
1461	보르데텔라 페르투시스	HADDGTIVI	2344	WP_019248145.1
1462	보르데텔라 페르투시스	HADWNNQSI	2345	AGS56996.1
1463	보르데텔라 페르투시스	HAEHEKDVR	2346	WP_019247158.1
1464	보르데텔라 페르투시스	HANHYGTRI	2347	WP_019247158.1
1465	보르데텔라 페르투시스	HAQGKALLY	2348	AGT50936.1
1466	보르데텔라 페르투시스	HFIGYIYEV	2349	AAW72734.1
1467	보르데텔라 페르투시스	HLSPALADV	2350	AAA22983.1
1468	보르데텔라 페르투시스	HSLYASYEY	2351	AGS56996.1
1469	보르데텔라 페르투시스	HVRGMLVPV	2352	AAA22974.1
1470	보르데텔라 페르투시스	HVSKEEQYY	2353	YP_006628018.1
1471	보르데텔라 페르투시스	HVTRGWSIF	2354	AFK26303.1
1472	보르데텔라 페르투시스	IADSGFYLD	2355	AGS56996.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1473	보르데텔라 페르투시스	IAHRTELRG	2356	AGS56996.1
1474	보르데텔라 페르투시스	IENTAKLSG	2357	WP_019249248.1
1475	보르데텔라 페르투시스	IESKISQSV	2358	WP_019249248.1
1476	보르데텔라 페르투시스	IETGGARRF	2359	AGT50936.1
1477	보르데텔라 페르투시스	IIKDAPPGA	2360	1BCP_C
1478	보르데텔라 페르투시스	IIQEFAADL	2361	WP_019249248.1
1479	보르데텔라 페르투시스	ILAGALATY	2362	AAW72734.1
1480	보르데텔라 페르투시스	ILLENPAAE	2363	AGS56996.1
1481	보르데텔라 페르투시스	ILPILVLAL	2364	NP_882286.1
1482	보르데텔라 페르투시스	IPFQRALRL	2365	WP_019248145.1
1483	보르데텔라 페르투시스	ISVRVHVSK	2366	YP_006628018.1
1484	보르데텔라 페르투시스	ITNETGKTY	2367	WP_019247158.1
1485	보르데텔라 페르투시스	ITNKRAALI	2368	WP_019249248.1
1486	보르데텔라 페르투시스	ITSYVGFSV	2369	WP_019248658.1
1487	보르데텔라 페르투시스	ITVTSRGGF	2370	WP_019249248.1
1488	보르데텔라 페르투시스	IVIPPKALF	2371	1BCP_C
1489	보르데텔라 페르투시스	IVVEAGELV	2372	WP_019249248.1
1490	보르데텔라 페르투시스	KAAKSVNLM	2373	WP_019247158.1
1491	보르데텔라 페르투시스	KAAPLRRTT	2374	AGS56996.1
1492	보르데텔라 페르투시스	KAGKLSATG	2375	WP_019249248.1

[1070]

[1071]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1493	보르데텔라 페르투시스	KAGTIAAPW	2376	WP_019249248.1
1494	보르데텔라 페르투시스	KAKSLTTEI	2377	WP_019249248.1
1495	보르데텔라 페르투시스	KATVTTVQV	2378	WP_019247158.1
1496	보르데텔라 페르투시스	KDYRDKDGG	2379	WP_019247158.1
1497	보르데텔라 페르투시스	KEAATIVAA	2380	WP_019249248.1
1498	보르데텔라 페르투시스	KEDVDAAQI	2381	WP_019248658.1
1499	보르데텔라 페르투시스	KEVDGIIQE	2382	WP_019249248.1
1500	보르데텔라 페르투시스	KGPKLAMPW	2383	AGS56996.1
1501	보르데텔라 페르투시스	KLASGGGAV	2384	WP_019249248.1
1502	보르데텔라 페르투시스	KLKGKNQEF	2385	AAA22984.1
1503	보르데텔라 페르투시스	KLLHHILPI	2386	NP_882286.1
1504	보르데텔라 페르투시스	KPAPQPGPQ	2387	AGS56996.1
1505	보르데텔라 페르투시스	KPAPTAPPM	2388	WP_019249248.1
1506	보르데텔라 페르투시스	KPAVSVKVA	2389	WP_019249248.1
1507	보르데텔라 페르투시스	KPDRAARVA	2390	WP_019249248.1
1508	보르데텔라 페르투시스	KPLADIAVI	2391	YP_006626470.1
1509	보르데텔라 페르투시스	KPLADIAVV	2392	WP_019249248.1
1510	보르데텔라 페르투시스	KPLPKPLPV	2393	WP_019247158.1
1511	보르데텔라 페르투시스	KQADRDFVW	2394	WP_019247699.1
1512	보르데텔라 페르투시스	KSLPGGKLP	2395	WP_019247158.1

[1072]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1513	보르데텔라 페르투시스	KSYTLRYLA	2396	WP_019248658.1
1514	보르데텔라 페르투시스	KTNMVVTSV	2397	AAA22983.1
1515	보르데텔라 페르투시스	KVLAPRLYL	2398	AAA22974.1
1516	보르데텔라 페르투시스	KVLSTKTTL	2399	WP_019247158.1
1517	보르데텔라 페르투시스	LAAGAGLTL	2400	WP_019249248.1
1518	보르데텔라 페르투시스	LAAIASAAH	2401	WP_019248145.1
1519	보르데텔라 페르투시스	LAANGNGQW	2402	AGS56996.1
1520	보르데텔라 페르투시스	LAAQVTQRG	2403	WP_019249248.1
1521	보르데텔라 페르투시스	LAARGDGAL	2404	AAA22974.1
1522	보르데텔라 페르투시스	LAAVLVNPH	2405	WP_019249248.1
1523	보르데텔라 페르투시스	LAGSGLFRM	2406	AGS56996.1
1524	보르데텔라 페르투시스	LAKALSAAL	2407	WP_019249248.1
1525	보르데텔라 페르투시스	LALQAGDAL	2408	WP_019249248.1
1526	보르데텔라 페르투시스	LAMPWTFHA	2409	AGS56996.1
1527	보르데텔라 페르투시스	LANDGTIVI	2410	WP_019248658.1
1528	보르데텔라 페르투시스	LAPTVGVAF	2411	WP_019247158.1
1529	보르데텔라 페르투시스	LASDGSVDF	2412	AGS56996.1
1530	보르데텔라 페르투시스	LATYQSEYL	2413	AAW72734.1
1531	보르데텔라 페르투시스	LAWALMLAC	2414	AAA22974.1
1532	보르데텔라 페르투시스	LEAGRRFTH	2415	AGS56996.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1533	보르데텔라 페르투시스	LFTQQGGAY	2416	1BCP_C
1534	보르데텔라 페르투시스	LKRPGSSPM	2417	AAA22983.1
1535	보르데텔라 페르투시스	LLGSHVARA	2418	AFK26303.1
1536	보르데텔라 페르투시스	LLHHILPIL	2419	NP_882286.1
1537	보르데텔라 페르투시스	LLNAGGTLL	2420	WP_019249248.1
1538	보르데텔라 페르투시스	LMGPLQVNA	2421	WP_019249248.1
1539	보르데텔라 페르투시스	LNDSKITMG	2422	WP_019248145.1
1540	보르데텔라 페르투시스	LPEPVKLTL	2423	AGT50936.1
1541	보르데텔라 페르투시스	LPILVLALL	2424	NP_882286.1
1542	보르데텔라 페르투시스	LPKISKNAL	2425	WP_019248145.1
1543	보르데텔라 페르투시스	LPKPVTVKL	2426	WP_019247158.1
1544	보르데텔라 페르투시스	LPLKANPMH	2427	NP_882285.1
1545	보르데텔라 페르투시스	LPPRPVVAE	2428	WP_019247158.1
1546	보르데텔라 페르투시스	LPSIPGTSI	2429	AGS56996.1
1547	보르데텔라 페르투시스	LPTHLYKNF	2430	AAA22984.1
1548	보르데텔라 페르투시스	LPVRGVALV	2431	WP_019249248.1
1549	보르데텔라 페르투시스	LPVSLTALD	2432	WP_019249248.1
1550	보르데텔라 페르투시스	LQGGNKVPV	2433	WP_019249248.1
1551	보르데텔라 페르투시스	LSAALGADW	2434	WP_019249248.1
1552	보르데텔라 페르투시스	LSDAGHEHD	2435	AAA22984.1

[1073]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1553	보르데텔라 페르투시스	LSGEVQRKG	2436	WP_019249248.1
1554	보르데텔라 페르투시스	LSSPSAITV	2437	WP_019249248.1
1555	보르데텔라 페르투시스	LTWLAILAV	2438	AAW72734.1
1556	보르데텔라 페르투시스	LVFSHVRGM	2439	AAA22974.1
1557	보르데텔라 페르투시스	LVSDAGADL	2440	WP_019249248.1
1558	보르데텔라 페르투시스	LVYYESIAY	2441	AFK26302.1
1559	보르데텔라 페르투시스	MAAESTFES	2442	WP_019247158.1
1560	보르데텔라 페르투시스	MAAGHDATL	2443	WP_019249248.1
1561	보르데텔라 페르투시스	MAAWSERAG	2444	AFK26302.1
1562	보르데텔라 페르투시스	MALGALGAA	2445	AGS56996.1
1563	보르데텔라 페르투시스	MAPVMGACM	2446	ADA85123.1
1564	보르데텔라 페르투시스	MATKGEMQI	2447	WP_019249248.1
1565	보르데텔라 페르투시스	MDAKGGTLL	2448	WP_019249248.1
1566	보르데텔라 페르투시스	MESNKDIVI	2449	WP_019249248.1
1567	보르데텔라 페르투시스	MEVMLRAVF	2450	AAA22983.1
1568	보르데텔라 페르투시스	MEYFKTPLP	2451	WP_019249248.1
1569	보르데텔라 페르투시스	MHTIASILL	2452	AAA22984.1
1570	보르데텔라 페르투시스	MIYMSGLAV	2453	1BCP_C
1571	보르데텔라 페르투시스	MLACTGLPL	2454	AAA22974.1
1572	보르데텔라 페르투시스	MPIDRKTLC	2455	AFK26303.1

[1074]

[1075]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1573	보르데텔라 페르투시스	MPKAPELDL	2456	WP_019249248.1
1574	보르데텔라 페르투시스	MQRQAGLPL	2457	NP_882285.1
1575	보르데텔라 페르투시스	NALAWALML	2458	AAA22974.1
1576	보르데텔라 페르투시스	NITNKRAAL	2459	WP_019249248.1
1577	보르데텔라 페르투시스	NLMAAESTF	2460	WP_019247158.1
1578	보르데텔라 페르투시스	NNETMSGRQ	2461	WP_019249248.1
1579	보르데텔라 페르투시스	NPGSLIAEV	2462	WP_019249248.1
1580	보르데텔라 페르투시스	NPMHTIASI	2463	NP_882285.1
1581	보르데텔라 페르투시스	NPQTQLSNI	2464	WP_019248145.1
1582	보르데텔라 페르투시스	NPYTSRRSV	2465	AFK26302.1
1583	보르데텔라 페르투시스	PAPTAPPMP	2466	WP_019249248.1
1584	보르데텔라 페르투시스	PASANTLLL	2467	AGS56996.1
1585	보르데텔라 페르투시스	PAVALPRPL	2468	WP_019249248.1
1586	보르데텔라 페르투시스	PDAHVPFCF	2469	AAA22983.1
1587	보르데텔라 페르투시스	PELGAAIRV	2470	AGT50936.1
1588	보르데텔라 페르투시스	PFIIKLKDC	2471	WP_019248145.1
1589	보르데텔라 페르투시스	PGPQPPQPP	2472	AGS56996.1
1590	보르데텔라 페르투시스	PGPQPPQPQ	2473	AAZ74338.1
1591	보르데텔라 페르투시스	PGPTTDYST	2474	WP_019248145.1
1592	보르데텔라 페르투시스	PGTFTAGKD	2475	WP_019249248.1

[1076]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1593	보르데텔라 페르투시스	PGTPGDLLE	2476	AAA22984.1
1594	보르데텔라 페르투시스	PKPKPKAER	2477	WP_019247158.1
1595	보르데텔라 페르투시스	PKPKPKPKA	2478	WP_019247158.1
1596	보르데텔라 페르투시스	PKPKPKPKP	2479	WP_019247158.1
1597	보르데텔라 페르투시스	PLPPRPVVA	2480	WP_019247158.1
1598	보르데텔라 페르투시스	PPAPKPAPQ	2481	AGS56996.1
1599	보르데텔라 페르투시스	PPKPAPVAK	2482	WP_019247158.1
1600	보르데텔라 페르투시스	PPRPVAAQV	2483	WP_019247158.1
1601	보르데텔라 페르투시스	PPRPVVAEK	2484	WP_019247158.1
1602	보르데텔라 페르투시스	PPSRPTTPP	2485	WP_019247158.1
1603	보르데텔라 페르투시스	PRRARRALR	2486	WP_019249248.1
1604	보르데텔라 페르투시스	QAAPLSITL	2487	AGT50936.1
1605	보르데텔라 페르투시스	QADRDFVWY	2488	WP_019247699.1
1606	보르데텔라 페르투시스	QAIVVGKDL	2489	WP_019249248.1
1607	보르데텔라 페르투시스	QALGALKLY	2490	ACI16088.1
1608	보르데텔라 페르투시스	QELALKLKG	2491	AAA22984.1
1609	보르데텔라 페르투시스	QITQHGGPY	2492	NP_882283.1
1610	보르데텔라 페르투시스	QITQHGSPY	2493	AFK26303.1
1611	보르데텔라 페르투시스	QPLPPRPVA	2494	WP_019247158.1
1612	보르데텔라 페르투시스	QPPAGRELS	2495	AGS56996.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1613	보르데텔라 페르투시스	QQLKQADRD	2496	WP_019247699.1
1614	보르데텔라 페르투시스	QQVQVLQRQ	2497	WP_019247158.1
1615	보르데텔라 페르투시스	QSIVEAPEL	2498	AGT50936.1
1616	보르데텔라 페르투시스	QVGSSNSAF	2499	AAW72734.1
1617	보르데텔라 페르투시스	RAGLSPATW	2500	WP_019249248.1
1618	보르데텔라 페르투시스	RARRALRQD	2501	WP_019249248.1
1619	보르데텔라 페르투시스	RASASRARI	2502	WP_019249248.1
1620	보르데텔라 페르투시스	RELSAAANA	2503	AGS56996.1
1621	보르데텔라 페르투시스	RETFCITTI	2504	1BCP_C
1622	보르데텔라 페르투시스	RGFAQRQQL	2505	AGS56996.1
1623	보르데텔라 페르투시스	RGSAATFTL	2506	WP_019248295.1
1624	보르데텔라 페르투시스	RGWSIFALY	2507	AFK26303.1
1625	보르데텔라 페르투시스	RKMLYLIYV	2508	YP_006628018.1
1626	보르데텔라 페르투시스	RLRKMLYLI	2509	YP_006628018.1
1627	보르데텔라 페르투시스	RPQITDAVT	2510	WP_019249248.1
1628	보르데텔라 페르투시스	RPSVNGGRI	2511	WP_019249248.1
1629	보르데텔라 페르투시스	RRFTHADGW	2512	AGS56996.1
1630	보르데텔라 페르투시스	RSGARATSL	2513	AAA22974.1
1631	보르데텔라 페르투시스	RSRVRALAW	2514	NP_882284.1
1632	보르데텔라 페르투시스	RSRVRALSW	2515	YP_006628019.1

[1077]

[1078]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1633	보르데텔라 페르투시스	RTHGVGASL	2516	AGS56996.1
1634	보르데텔라 페르투시스	RTRGQARSG	2517	AAA22974.1
1635	보르데텔라 페르투시스	RVAPPAVAL	2518	WP_019249248.1
1636	보르데텔라 페르투시스	RVLPEPVKL	2519	AGT50936.1
1637	보르데텔라 페르투시스	RVRALAWLL	2520	NP_882284.1
1638	보르데텔라 페르투시스	RVRALSWLL	2521	YP_006628019.1
1639	보르데텔라 페르투시스	RVTVSGGSL	2522	AGT50936.1
1640	보르데텔라 페르투시스	SEAMAAWSE	2523	AFK26302.1
1641	보르데텔라 페르투시스	SESHNFHAS	2524	WP_019247158.1
1642	보르데텔라 페르투시스	SGEGRVNIG	2525	WP_019249248.1
1643	보르데텔라 페르투시스	SGLAVRVHV	2526	1BCP_C
1644	보르데텔라 페르투시스	SLADISLGA	2527	WP_019249248.1
1645	보르데텔라 페르투시스	SLFAILSST	2528	WP_019247158.1
1646	보르데텔라 페르투시스	SLFAPHGNV	2529	AAZ74338.1
1647	보르데텔라 페르투시스	SLSIDNATW	2530	AGS56996.1
1648	보르데텔라 페르투시스	SPMEVMLRA	2531	AAA22983.1
1649	보르데텔라 페르투시스	SPQPIRATV	2532	WP_019247158.1
1650	보르데텔라 페르투시스	SPRRARRAL	2533	WP_019249248.1
1651	보르데텔라 페르투시스	SPSRLAGTL	2534	WP_019249248.1
1652	보르데텔라 페르투시스	SSTPLGSLF	2535	WP_019247158.1

[1079]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1653	보르데텔라 페르투시스	STYELLDYL	2536	WP_019249248.1
1654	보르데텔라 페르투시스	SVAMKPYEV	2537	AAA22983.1
1655	보르데텔라 페르투시스	SVAPNALAW	2538	AAA22974.1
1656	보르데텔라 페르투시스	SVKVAKKLF	2539	WP_019249248.1
1657	보르데텔라 페르투시스	TAFMSGRSL	2540	AAA22984.1
1658	보르데텔라 페르투시스	TAGATPFDI	2541	WP_019248658.1
1659	보르데텔라 페르투시스	TAPVTSPAW	2542	AAW72734.1
1660	보르데텔라 페르투시스	TARTGWLTW	2543	AAW72734.1
1661	보르데텔라 페르투시스	TEAQGVQVR	2544	WP_019248145.1
1662	보르데텔라 페르투시스	TEVYLEHRM	2545	AAW72734.1
1663	보르데텔라 페르투시스	TFEGKPALE	2546	AAA22983.1
1664	보르데텔라 페르투시스	TFTGKVTNG	2547	WP_019248658.1
1665	보르데텔라 페르투시스	TGDGGGHTD	2548	AGS56996.1
1666	보르데텔라 페르투시스	TLAKALSAA	2549	WP_019249248.1
1667	보르데텔라 페르투시스	TLANVGDTW	2550	AGS56996.1
1668	보르데텔라 페르투시스	TLNASNLTL	2551	WP_019249248.1
1669	보르데텔라 페르투시스	TLSSAHGNV	2552	WP_019249248.1
1670	보르데텔라 페르투시스	TPFDIKLKE	2553	WP_019248658.1
1671	보르데텔라 페르투시스	TPFIIKLKD	2554	WP_019248145.1
1672	보르데텔라 페르투시스	TPGWSIYGL	2555	1BCP_C

[1080]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1673	보르데텔라 페르투시스	TPLGSAATF	2556	AGS56996.1
1674	보르데텔라 페르투시스	TPLGSLFAI	2557	WP_019247158.1
1675	보르데텔라 페르투시스	TPLPVSLTA	2558	WP_019249248.1
1676	보르데텔라 페르투시스	TRQGIMDQY	2559	YP_006626873.1
1677	보르데텔라 페르투시스	TSKQDERNY	2560	WP_019247158.1
1678	보르데텔라 페르투시스	TSPYDGKYW	2561	YP_006628018.1
1679	보르데텔라 페르투시스	TSRRSVASI	2562	AFK26302.1
1680	보르데텔라 페르투시스	TSRTVTMRY	2563	NP_879898.1
1681	보르데텔라 페르투시스	TTEYPNARY	2564	ADA85123.1
1682	보르데텔라 페르투시스	TTEYSNARY	2565	AFK26302.1
1683	보르데텔라 페르투시스	TTLGLEQTF	2566	WP_019247158.1
1684	보르데텔라 페르투시스	TVLAAGAGL	2567	WP_019249248.1
1685	보르데텔라 페르투시스	TVQELALKL	2568	AAA22984.1
1686	보르데텔라 페르투시스	TVVQRNKHW	2569	WP_019247158.1
1687	보르데텔라 페르투시스	VAAAADLAL	2570	WP_019249248.1
1688	보르데텔라 페르투시스	VALASQARW	2571	AGS56996.1
1689	보르데텔라 페르투시스	VAMKPYEVT	2572	AAA22983.1
1690	보르데텔라 페르투시스	VARLVKLQG	2573	WP_019249248.1
1691	보르데텔라 페르투시스	VAVAGGRWH	2574	AGS56996.1
1692	보르데텔라 페르투시스	VEASAITTY	2575	WP_019248145.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1693	보르데텔라 페르투시스	VEDIGGKNY	2576	WP_019247158.1
1694	보르데텔라 페르투시스	VEVSSPPPV	2577	WP_019247158.1
1695	보르데텔라 페르투시스	VGAGGYEAG	2578	WP_019247158.1
1696	보르데텔라 페르투시스	VGGGEHGRW	2579	WP_019249248.1
1697	보르데텔라 페르투시스	VHVSKEEQY	2580	YP_006628018.1
1698	보르데텔라 페르투시스	VIDGQKVLA	2581	AAA22974.1
1699	보르데텔라 페르투시스	VIGACTSPY	2582	YP_006628018.1
1700	보르데텔라 페르투시스	VKLGGVYEA	2583	AAA22974.1
1701	보르데텔라 페르투시스	VLAPRLYLT	2584	AAA22974.1
1702	보르데텔라 페르투시스	VLVKTNMVV	2585	AAA22983.1
1703	보르데텔라 페르투시스	VPASGAPAA	2586	AGS56996.1
1704	보르데텔라 페르투시스	VPFCFGKDL	2587	AAA22983.1
1705	보르데텔라 페르투시스	VPVSEHCTV	2588	AAA22974.1
1706	보르데텔라 페르투시스	VPVTPPKVE	2589	WP_019247158.1
1707	보르데텔라 페르투시스	VRTVSAMEY	2590	WP_019249248.1
1708	보르데텔라 페르투시스	VSGGSLSAP	2591	AGT50936.1
1709	보르데텔라 페르투시스	VSSATKAKG	2592	WP_019248658.1
1710	보르데텔라 페르투시스	VSSPPPVSV	2593	WP_019247158.1
1711	보르데텔라 페르투시스	VSVKVAKKL	2594	WP_019249248.1
1712	보르데텔라 페르투시스	VTMRYLASY	2595	WP_019248145.1

[1081]

[1082]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1713	보르데텔라 페르투시스	VTSVAMKPY	2596	AAA22983.1
1714	보르데텔라 페르투시스	VVAEKVTTP	2597	WP_019247158.1
1715	보르데텔라 페르투시스	VVDGPPSRP	2598	WP_019247158.1
1716	보르데텔라 페르투시스	VVETAQPLP	2599	WP_019247158.1
1717	보르데텔라 페르투시스	WLTWLAILA	2600	AAW72734.1
1718	보르데텔라 페르투시스	WTFHAGYRY	2601	AGS56996.1
1719	보르데텔라 페르투시스	WVMTDNSNV	2602	AGS56996.1
1720	보르데텔라 페르투시스	YAEHGEVSI	2603	WP_019249248.1
1721	보르데텔라 페르투시스	YAIDGTAAG	2604	WP_019249248.1
1722	보르데텔라 페르투시스	YALKSRIAL	2605	AAA22984.1
1723	보르데텔라 페르투시스	YATNPQTQL	2606	WP_019248145.1
1724	보르데텔라 페르투시스	YDTGDLIAY	2607	WP_019248658.1
1725	보르데텔라 페르투시스	YEAGFSLGS	2608	WP_019247158.1
1726	보르데텔라 페르투시스	YEDATFETY	2609	YP_006628018.1
1727	보르데텔라 페르투시스	YENKSSTPL	2610	WP_019247158.1
1728	보르데텔라 페르투시스	YEVTPTRML	2611	AAA22983.1
1729	보르데텔라 페르투시스	YEYIWGLYP	2612	WP_019249248.1
1730	보르데텔라 페르투시스	YEYIWGLYQ	2613	YP_006626470.1
1731	보르데텔라 페르투시스	YEYSKGPKL	2614	AGS56996.1
1732	보르데텔라 페르투시스	YFEPGPTTD	2615	WP_019248145.1

[1083]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1733	보르데텔라 페르투시스	YIWGLYPTY	2616	WP_019249248.1
1734	보르데텔라 페르투시스	YIWGLYQTY	2617	YP_006626470.1
1735	보르데텔라 페르투시스	YLRQITPGW	2618	1BCP_C
1736	보르데텔라 페르투시스	YMIYMSGLA	2619	1BCP_C
1737	보르데텔라 페르투시스	YPALRAALI	2620	WP_019248658.1
1738	보르데텔라 페르투시스	YPGTPGDLL	2621	AAA22984.1
1739	보르데텔라 페르투시스	YPTYTEWSV	2622	WP_019249248.1
1740	보르데텔라 페르투시스	YQTYTEWSV	2623	YP_006626470.1
1741	보르데텔라 페르투시스	YSTGDLRAY	2624	WP_019248145.1
1742	보르데텔라 페르투시스	YTLRYLASY	2625	WP_019248658.1
1743	보르데텔라 페르투시스	YVLVKTNMV	2626	AAA22983.1
1744	보르데텔라 페르투시스	YYDYEDATF	2627	YP_006628018.1
1745	보르데텔라 페르투시스 509	AAFIALYPNSQLAPT	2628	Q7VU05
1746	보르데텔라 페르투시스 509	GGAEYNLALGQRRA	2629	Q7VU04
1747	보르데텔라 페르투시스 509	GGAEYNLALGQRRADA	2630	Q7VU04
1748	보르데텔라 페르투시스 509	IALYPNSQLAPT	2631	Q7VU05
1749	보르데텔라 페르투시스 509	KPDQGEVVAVGPGKKTED	2632	P0A339.1
1750	보르데텔라 페르투시스 509	KPDQGEVVAVGPGKKTEDG	2633	P0A339.1
1751	보르데텔라 페르투시스 509	LAEVLDYHNFVLTQ	2634	Q7VWM1.1
1752	코리네박테리움 디프테리아에	QSIALSSLMVAQAIP	2635	AAV70486.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1753	코리네박테리움 디프테리아에	SIGVLGYQKTVDHTKVNSKLSLF	2636	AAV70486.1
1754	보르데텔라 페르투시스	AAHADWNNQSIVKT	2637	ABO77783.1
1755	보르데텔라 페르투시스	AALGRG	2638	ABO77783.1
1756	보르데텔라 페르투시스	AALGRGHSLYASYE	2639	ABO77783.1
1757	보르데텔라 페르투시스	AAPLRRTTLAMALG	2640	ABO77783.1
1758	보르데텔라 페르투시스	AAPLSITLQAGAHA	2641	ABO77783.1
1759	보르데텔라 페르투시스	ADAQGDIVATELPS	2642	ABO77783.1
1760	보르데텔라 페르투시스	ADSGFYLDATLRAS	2643	ABO77783.1
1761	보르데텔라 페르투시스	AELA	2644	ABO77783.1
1762	보르데텔라 페르투시스	AELAVFRAGGGAYR	2645	ABO77783.1
1763	보르데텔라 페르투시스	AELQFRNGSVTSSG	2646	ABO77783.1
1764	보르데텔라 페르투시스	AGGRWHLGGLAGYT	2647	ABO77783.1
1765	보르데텔라 페르투시스	AGVAAMQGAVVHLQ	2648	ABO77783.1
1766	보르데텔라 페르투시스	AGYTRGDRGFTGDG	2649	ABO77783.1
1767	보르데텔라 페르투시스	ALASQARWTGATRA	2650	ABO77783.1
1768	보르데텔라 페르투시스	AMPWTFHAGYRYSW	2651	ABO77783.1
1769	보르데텔라 페르투시스	AMQGAVVHLQRATIRRGDAP	2652	ABO77783.1
1770	보르데텔라 페르투시스	ANGLRVRDE	2653	ABO77783.1
1771	보르데텔라 페르투시스	ANGLRVRDEGGSSV	2654	ABO77783.1
1772	보르데텔라 페르투시스	ANKDGKVDIGTYRY	2655	ABO77783.1

[1084]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1773	보르데텔라 페르투시스	APAAVSVLGASELT	2656	ABO77783.1
1774	보르데텔라 페르투시스	APPAPKPAPQPGPQ	2657	ABO77783.1
1775	보르데텔라 페르투시스	AQGILLENPAAELQ	2658	ABO77783.1
1776	보르데텔라 페르투시스	ARWTGATRAVDSLS	2659	ABO77783.1
1777	보르데텔라 페르투시스	ASLEAGRRFTHADG	2660	ABO77783.1
1778	보르데텔라 페르투시스	ASYEYSKGPKLAMP	2661	ABO77783.1
1779	보르데텔라 페르투시스	ATFTLANKD	2662	ABO77783.1
1780	보르데텔라 페르투시스	ATFTLANKDGKVDI	2663	ABO77783.1
1781	보르데텔라 페르투시스	ATRAVDSLSIDNAT	2664	ABO77783.1
1782	보르데텔라 페르투시스	DDDGIALYVAGEQAQ	2665	ABO77783.1
1783	보르데텔라 페르투시스	DDGIALYVAGEQAQ	2666	ABO77783.1
1784	보르데텔라 페르투시스	DGGHITGGRAAGVA	2667	ABO77783.1
1785	보르데텔라 페르투시스	DGIRRFLGTVTVKAGK	2668	ABO77783.1
1786	보르데텔라 페르투시스	DGSVDFQQPAEAGR	2669	ABO77783.1
1787	보르데텔라 페르투시스	DGYAVKGKYRTHGV	2670	ABO77783.1
1788	보르데텔라 페르투시스	DIVATELPSIPGTS	2671	ABO77783.1
1789	보르데텔라 페르투시스	DKLVVMQDASGQHR	2672	ABO77783.1
1790	보르데텔라 페르투시스	DLGLSDKLVVMQDA	2673	ABO77783.1
1791	보르데텔라 페르투시스	DNATWVMTDNSNVGA	2674	ABO77783.1
1792	보르데텔라 페르투시스	DNATWVMTDNSNVGALRLA	2675	ABO77783.1

[1085]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1793	보르데텔라 페르투시스	DNRAGRRFDQKVAG	2676	ABO77783.1
1794	보르데텔라 페르투시스	EAGRFKVLTVNTLA	2677	ABO77783.1
1795	보르데텔라 페르투시스	ELAQSIVEAPELGA	2678	ABO77783.1
1796	보르데텔라 페르투시스	ELGAAIRVGRGARV	2679	ABO77783.1
1797	보르데텔라 페르투시스	ELGADHAVAVAGGR	2680	ABO77783.1
1798	보르데텔라 페르투시스	ELPSIPGTSIGPLD	2681	ABO77783.1
1799	보르데텔라 페르투시스	EPVKLTLTGGADAQ	2682	ABO77783.1
1800	보르데텔라 페르투시스	EQAQASIADSTLQG	2683	ABO77783.1
1801	보르데텔라 페르투시스	ERGANVTVQRSAIV	2684	ABO77783.1
1802	보르데텔라 페르투시스	ERQHGIHIQGSDPG	2685	ABO77783.1
1803	보르데텔라 페르투시스	EVGKRIELAGGRQV	2686	ABO77783.1
1804	보르데텔라 페르투시스	FDGAGTVHTNGIAH	2687	ABO77783.1
1805	보르데텔라 페르투시스	FQQPAEAGRFKVLT	2688	ABO77783.1
1806	보르데텔라 페르투시스	FRAGGGAYRAANGL	2689	ABO77783.1
1807	보르데텔라 페르투시스	GAHAQGKALLYRVL	2690	ABO77783.1
1808	보르데텔라 페르투시스	GARVTVSGGSLSAP	2691	ABO77783.1
1809	보르데텔라 페르투시스	GAYRAANGLRVRDE	2692	ABO77783.1
1810	보르데텔라 페르투시스	GDAPAGGAVPGGAV	2693	ABO77783.1
1811	보르데텔라 페르투시스	GGAVPGGAVPGGFG	2694	ABO77783.1
1812	보르데텔라 페르투시스	GGAVPGGFGPVLDG	2695	ABO77783.1

[1086]

[1087]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1813	보르데텔라 페르투시스	GGFGPVLDGWYGVD	2696	ABO77783.1
1814	보르데텔라 페르투시스	GGLHIGALQSLQPE	2697	ABO77783.1
1815	보르데텔라 페르투시스	GGVQIERGANVTVQ	2698	ABO77783.1
1816	보르데텔라 페르투시스	GHSLYASYEYSKGP	2699	ABO77783.1
1817	보르데텔라 페르투시스	GHTDSVHVGGYATY	2700	ABO77783.1
1818	보르데텔라 페르투시스	GIAHRTELRGTRAE	2701	ABO77783.1
1819	보르데텔라 페르투시스	GKALLYRVLPEPVK	2702	ABO77783.1
1820	보르데텔라 페르투시스	GLGMAAALGRGHSL	2703	ABO77783.1
1821	보르데텔라 페르투시스	GNVIETGGARRFAP	2704	ABO77783.1
1822	보르데텔라 페르투시스	GPLDVALASQARWT	2705	ABO77783.1
1823	보르데텔라 페르투시스	GQHRLWVRN	2706	ABO77783.1
1824	보르데텔라 페르투시스	GRGFAQRQQLDNRA	2707	ABO77783.1
1825	보르데텔라 페르투시스	GRLGLEVGKRIELA	2708	ABO77783.1
1826	보르데텔라 페르투시스	GRQVQPYIKASVLQ	2709	ABO77783.1
1827	보르데텔라 페르투시스	GRRFTHADGWFLEPQAELA	2710	ABO77783.1
1828	보르데텔라 페르투시스	GSEPASANTLLLVQ	2711	ABO77783.1
1829	보르데텔라 페르투시스	GSSVLGRLGLEVGK	2712	ABO77783.1
1830	보르데텔라 페르투시스	GTTIKVSGRQAQGI	2713	ABO77783.1
1831	보르데텔라 페르투시스	GTVTVKAGKLVADH	2714	ABO77783.1
1832	보르데텔라 페르투시스	HAVAVAGGRWHLGG	2715	ABO77783.1

[1088]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1833	보르데텔라 페르투시스	IELAGGRQVQPYIK	2716	ABO77783.1
1834	보르데텔라 페르투시스	IHIQGSDPGGVRTA	2717	ABO77783.1
1835	보르데텔라 페르투시스	IRRFLGTVTVKAGK	2718	ABO77783.1
1836	보르데텔라 페르투시스	IRVGRGARVTVSGG	2719	ABO77783.1
1837	보르데텔라 페르투시스	ITLQAGAHA	2720	ABO77783.1
1838	보르데텔라 페르투시스	ITLQAGAHAQGKAL	2721	ABO77783.1
1839	보르데텔라 페르투시스	IVEAPELGAAIRVG	2722	ABO77783.1
1840	보르데텔라 페르투시스	IVKTGERQHGIHIQ	2723	ABO77783.1
1841	보르데텔라 페르투시스	KAGKLVADHATLAN	2724	ABO77783.1
1842	보르데텔라 페르투시스	KGKYRTHGVGASLE	2725	ABO77783.1
1843	보르데텔라 페르투시스	KPAPQPGPQPPQPP	2726	ABO77783.1
1844	보르데텔라 페르투시스	KVAGFELGADHAVA	2727	ABO77783.1
1845	보르데텔라 페르투시스	KVAGSDGYAVKGKY	2728	ABO77783.1
1846	보르데텔라 페르투시스	KVDIGTYRYRLAAN	2729	ABO77783.1
1847	보르데텔라 페르투시스	KVLTVNTLAGSGLF	2730	ABO77783.1
1848	보르데텔라 페르투시스	LAANGNGQWSLVGA	2731	ABO77783.1
1849	보르데텔라 페르투시스	LAMPWTFHAGYRYS	2732	ABO77783.1
1850	보르데텔라 페르투시스	LASTLWYAESNALS	2733	ABO77783.1
1851	보르데텔라 페르투시스	LENDFKVAGSDGYA	2734	ABO77783.1
1852	보르데텔라 페르투시스	LENPAAELQFRNGS	2735	ABO77783.1

[1089]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1853	보르데텔라 페르투시스	LGAAPAAHADWNNQ	2736	ABO77783.1
1854	보르데텔라 페르투시스	LGGLAGYTRGDRGFTGDG	2737	ABO77783.1
1855	보르데텔라 페르투시스	LLENP	2738	ABO77783.1
1856	보르데텔라 페르투시스	LLVQTPLGSAATFT	2739	ABO77783.1
1857	보르데텔라 페르투시스	LPPSRVVLRDTNVT	2740	ABO77783.1
1858	보르데텔라 페르투시스	LQPEDLPPS	2741	ABO77783.1
1859	보르데텔라 페르투시스	LQPEDLPPSRVVLR	2742	ABO77783.1
1860	보르데텔라 페르투시스	LRASRLENDFKVAG	2743	ABO77783.1
1861	보르데텔라 페르투시스	LRLASDGSVDFQQP	2744	ABO77783.1
1862	보르데텔라 페르투시스	LSAAANAAVNTGGV	2745	ABO77783.1
1863	보르데텔라 페르투시스	LSAPHGNVIETGGA	2746	ABO77783.1
1864	보르데텔라 페르투시스	LSDDGIRRFLGTVT	2747	ABO77783.1
1865	보르데텔라 페르투시스	LVGAKAPPAPKPAP	2748	ABO77783.1
1866	보르데텔라 페르투시스	LYVAGEQAQASIAD	2749	ABO77783.1
1867	보르데텔라 페르투시스	MALGALGAAPAAHA	2750	ABO77783.1
1868	보르데텔라 페르투시스	MNMSLSRIVKAAPL	2751	ABO77783.1
1869	보르데텔라 페르투시스	MQDASGQHR	2752	ABO77783.1
1870	보르데텔라 페르투시스	MQGAVVHLQRATIR	2753	ABO77783.1
1871	보르데텔라 페르투시스	NAAVNTGGVGLAST	2754	ABO77783.1
1872	보르데텔라 페르투시스	NALSKRLGELRLNP	2755	ABO77783.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1873	보르데텔라 페르투시스	NGQWSLVGAKAPPA	2756	ABO77783.1
1874	보르데텔라 페르투시스	NTLAGSGLFRMNVF	2757	ABO77783.1
1875	보르데텔라 페르투시스	PAGRELSAAANAAV	2758	ABO77783.1
1876	보르데텔라 페르투시스	PAPQPPAGRELSAA	2759	ABO77783.1
1877	보르데텔라 페르투시스	PGPQPPQPPQPQPE	2760	ABO77783.1
1878	보르데텔라 페르투시스	PGTSIGPLDVALAS	2761	ABO77783.1
1879	보르데텔라 페르투시스	PLGSAATFTLANKD	2762	ABO77783.1
1880	보르데텔라 페르투시스	PQPEAPAPQPPAGR	2763	ABO77783.1
1881	보르데텔라 페르투시스	PQPPQPQPEAPAPQ	2764	ABO77783.1
1882	보르데텔라 페르투시스	PYIKASVLQEFDGA	2765	ABO77783.1
1883	보르데텔라 페르투시스	RFAPQAAPLSITLQ	2766	ABO77783.1
1884	보르데텔라 페르투시스	RLGELRLNPDAGGA	2767	ABO77783.1
1885	보르데텔라 페르투시스	RLNPDAGGAWGRGF	2768	ABO77783.1
1886	보르데텔라 페르투시스	RNGSVTSSGQLSDD	2769	ABO77783.1
1887	보르데텔라 페르투시스	RRFDQKVAGFELGA	2770	ABO77783.1
1888	보르데텔라 페르투시스	RTTLAMALGALGAA	2771	ABO77783.1
1889	보르데텔라 페르투시스	SAIVDGGLHIGALQ	2772	ABO77783.1
1890	보르데텔라 페르투시스	SANTLLLVQTPLGS	2773	ABO77783.1
1891	보르데텔라 페르투시스	SDPGGVRTASGTTI	2774	ABO77783.1
1892	보르데텔라 페르투시스	SELTLDGGHITGGR	2775	ABO77783.1

[1090]

[1091]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1893	보르데텔라 페르투시스	SGLFRMNVF	2776	ABO77783.1
1894	보르데텔라 페르투시스	SGLFRMNVFADLGL	2777	ABO77783.1
1895	보르데텔라 페르투시스	SGSSVELAQSIVEA	2778	ABO77783.1
1896	보르데텔라 페르투시스	SIADSTLQGAGGVQ	2779	ABO77783.1
1897	보르데텔라 페르투시스	SKGPKLAMPWTFHA	2780	ABO77783.1
1898	보르데텔라 페르투시스	SNVGALRLASDGSV	2781	ABO77783.1
1899	보르데텔라 페르투시스	SRIVKAAPLRRTTL	2782	ABO77783.1
1900	보르데텔라 페르투시스	SVLGASELTLDGGH	2783	ABO77783.1
1901	보르데텔라 페르투시스	SVLQEFDGA	2784	ABO77783.1
1902	보르데텔라 페르투시스	SVLQEFDGAGTVHT	2785	ABO77783.1
1903	보르데텔라 페르투시스	TELR	2786	ABO77783.1
1904	보르데텔라 페르투시스	TELRGTRAELGLGM	2787	ABO77783.1
1905	보르데텔라 페르투시스	TGDGGGHTDSVHVG	2788	ABO77783.1
1906	보르데텔라 페르투시스	TGGARRFAPQAAPL	2789	ABO77783.1
1907	보르데텔라 페르투시스	TGGRAAGVAAMQGA	2790	ABO77783.1
1908	보르데텔라 페르투시스	TGGVGLASTLWYAE	2791	ABO77783.1
1909	보르데텔라 페르투시스	THGVGASLEAGRRF	2792	ABO77783.1
1910	보르데텔라 페르투시스	TIRRGDAPA	2793	ABO77783.1
1911	보르데텔라 페르투시스	TLANVGDTWDDDGI	2794	ABO77783.1
1912	보르데텔라 페르투시스	TLQGAGGVQIERGA	2795	ABO77783.1

[1092]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1913	보르데텔라 페르투시스	TLTGGADAQGDIVA	2796	ABO77783.1
1914	보르데텔라 페르투시스	TNVTAVPASGAPAA	2797	ABO77783.1
1915	보르데텔라 페르투시스	TRAELGLGMAAALG	2798	ABO77783.1
1916	보르데텔라 페르투시스	TSSGQLSDDGIRRF	2799	ABO77783.1
1917	보르데텔라 페르투시스	TVHTNGIAHRTELR	2800	ABO77783.1
1918	보르데텔라 페르투시스	TYRYRLAANGNGQW	2801	ABO77783.1
1919	보르데텔라 페르투시스	VADHATLANVGDTW	2802	ABO77783.1
1920	보르데텔라 페르투시스	VHVGGYATYIADSG	2803	ABO77783.1
1921	보르데텔라 페르투시스	VLDGWYGVD	2804	ABO77783.1
1922	보르데텔라 페르투시스	VPASGAPAAVSVLG	2805	ABO77783.1
1923	보르데텔라 페르투시스	VRDEGGSSVLGRLG	2806	ABO77783.1
1924	보르데텔라 페르투시스	VRTASGTTIKVSGR	2807	ABO77783.1
1925	보르데텔라 페르투시스	VSGGSLSAPHGNVI	2808	ABO77783.1
1926	보르데텔라 페르투시스	VSGRQAQGILLENP	2809	ABO77783.1
1927	보르데텔라 페르투시스	VTVQRSAIVDGGLH	2810	ABO77783.1
1928	보르데텔라 페르투시스	VVLRDTNVTAVPAS	2811	ABO77783.1
1929	보르데텔라 페르투시스	WNNQSIVKTGERQH	2812	ABO77783.1
1930	보르데텔라 페르투시스	WVRNSGSEPASANT	2813	ABO77783.1
1931	보르데텔라 페르투시스	WYAESNALSKRLGE	2814	ABO77783.1
1932	보르데텔라 페르투시스	YATYIADSGFYLDA	2815	ABO77783.1

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1933	보르데텔라 페르투시스	YGVDVSGSS	2816	ABO77783.1
1934	보르데텔라 페르투시스	YGVDVSGSSVELAQ	2817	ABO77783.1
1935	보르데텔라 페르투시스	YLDATLRASRLEND	2818	ABO77783.1
1936	보르데텔라 페르투시스	YRVLPEPVKLTLTG	2819	ABO77783.1
1937	보르데텔라 페르투시스	VKAQNITNKRAALIEA	2820	AAA22974.1
1938	보르데텔라 페르투시스	YYSNVTATRLLSSTNS	2821	AAA83981.1
1939	보르데텔라 페르투시스	SPNLTDERAAQAGVT	2822	CPP72976.1
1940	홍역 모르빌리바이러스	SSRASDERAAHLPTS	2823	BAA33867.1
1941	코리네박테리움 디프테리아에	QVVHNSYNRPAYSPG	2824	1007216A
1942	홍역 바이러스 균주 에드몬스톤-B	AEGGEIHEL	2825	AAF85692.1
1943	홍역 바이러스 균주 에드몬스톤-B	AENLISNGIGKY	2826	AAF85698.1
1944	홍역 바이러스 균주 에드몬스톤-B	AEVDGDVKL	2827	CAB43772.1
1945	홍역 바이러스 균주 에드몬스톤-B	AIYTAEIHK	2828	AAF85697.1
1946	홍역 바이러스 균주 에드몬스톤-B	APVFHMTNY	2829	CAB43772.1
1947	홍역 바이러스 균주 에드몬스톤-B	APVFHMTNYLEQPVSN	2830	AAR89413.1
1948	홍역 바이러스 균주 에드몬스톤-B	AQRLNEIY	2831	AAF85698.1

[1093]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1949	홍역 바이러스 균주 에드몬스톤-B	ARVPHAYSL	2832	AAF85698.1
1950	홍역 바이러스 균주 에드몬스톤-B	AVRDLERAM	2833	P03424.1
1951	홍역 바이러스 균주 에드몬스톤-B	AVRDLERAMTTLK	2834	P03424.1
1952	홍역 바이러스 균주 에드몬스톤-B	DALLRLQAM	2835	Q89933.1
1953	홍역 바이러스 균주 에드몬스톤-B	DIKEKVINL	2836	AAF85698.1
1954	홍역 바이러스 균주 에드몬스톤-B	DQGLFKVL	2837	AAF85695.1
1955	홍역 바이러스 균주 에드몬스톤-B	DTGVDTRIW	2838	Q9EMA9.1
1956	홍역 바이러스 균주 에드몬스톤-B	EPIGSLAIEEAM	2839	AAF85692.1
1957	홍역 바이러스 균주 에드몬스톤-B	EPIRDALNAM	2840	P69354.1
1958	홍역 바이러스 균주 에드몬스톤-B	FPKLGKTL	2841	AAF85692.1
1959	홍역 바이러스 균주 에드몬스톤-B	FRSVNAVAF	2842	AAF85695.1
1960	홍역 바이러스 균주 에드몬스톤-B	GKIIDNTEQL	2843	AAF85695.1
1961	홍역 바이러스 균주 에드몬스톤-B	GLNEKLVFY	2844	AAF85695.1
1962	홍역 바이러스 균주 에드몬스톤-B	GMYGGTYLVEK	2845	AAC35876.2

[1094]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1963	홍역 바이러스 균주 에드몬스톤-B	GPPISLERLDVGTN	2846	P69354.1
1964	홍역 바이러스 균주 에드몬스톤-B	GPRQAQVSFL	2847	Q89933.1
1965	홍역 바이러스 균주 에드몬스톤-B	GRLVPQVRVID	2848	AAF85695.1
1966	홍역 바이러스 균주 에드몬스톤-B	GSAPISMGFR	2849	AAF85692.1
1967	홍역 바이러스 균주 에드몬스톤-B	HILAKSTAL	2850	AAF85698.1
1968	홍역 바이러스 균주 에드몬스톤-B	HYREVNLVY	2851	AAF85698.1
1969	홍역 바이러스 균주 에드몬스톤-B	IPPMKNLAL	2852	AAC35876.2
1970	홍역 바이러스 균주 에드몬스톤-B	IPYQGSGKGVSF	2853	CAB43772.1
1971	홍역 바이러스 균주 에드몬스톤-B	ISKESQHVY	2854	AAF85698.1
1972	홍역 바이러스 균주 에드몬스톤-B	IVSSHFFVY	2855	AAF85698.1
1973	홍역 바이러스 균주 에드몬스톤-B	KEIKETGRLF	2856	AAF85698.1
1974	홍역 바이러스 균주 에드몬스톤-B	KESQHVYYL	2857	AAF85698.1
1975	홍역 바이러스 균주 에드몬스톤-B	KIIDNTEQL	2858	AAF85695.1
1976	홍역 바이러스 균주 에드몬스톤-B	KKQINRQN	2859	AAA63285.1

[1095]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1977	홍역 바이러스 균주 에드몬스톤-B	KKVDTNFIYQ	2860	AAF85698.1
1978	홍역 바이러스 균주 에드몬스톤-B	KLIDGFFPA	2861	AAF85698.1
1979	홍역 바이러스 균주 에드몬스톤-B	KPNLSSKRSEL	2862	BAB39848.1
1980	홍역 바이러스 균주 에드몬스톤-B	KVSPYLFTV	2863	AAR89413.1
1981	홍역 바이러스 균주 에드몬스톤-B	LETRTTNQFL	2864	CAB43772.1
1982	홍역 바이러스 균주 에드몬스톤-B	LLKEATEL	2865	AAF85695.1
1983	홍역 바이러스 균주 에드몬스톤-B	LLKKGNSLY	2866	AAF85698.1
1984	홍역 바이러스 균주 에드몬스톤-B	LPAPIGGMNY	2867	AAF85698.1
1985	홍역 바이러스 균주 에드몬스톤-B	MPEETLHQVM	2868	AAF85698.1
1986	홍역 바이러스 균주 에드몬스톤-B	PTTIRGQFS	2869	CAB43772.1
1987	홍역 바이러스 균주 에드몬스톤-B	QEISRHQALGY	2870	P03424.1
1988	홍역 바이러스 균주 에드몬스톤-B	RITHVDTESY	2871	P69354.1
1989	홍역 바이러스 균주 에드몬스톤-B	RPGLKPDL	2872	P69354.1
1990	홍역 바이러스 균주 에드몬스톤-B	RPIYGLEV	2873	AAF85698.1

[1096]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
1991	홍역 바이러스 균주 에드몬스톤-B	RQAGQEMILAV	2874	P69354.1
1992	홍역 바이러스 균주 에드몬스톤-B	SAVRIATVY	2875	AAF85698.1
1993	홍역 바이러스 균주 에드몬스톤-B	SLMPEETLHQV	2876	AAF85698.1
1994	홍역 바이러스 균주 에드몬스톤-B	SMIDLVTKF	2877	AAF85698.1
1995	홍역 바이러스 균주 에드몬스톤-B	SMLNSQAIDNLRA	2878	P69354.1
1996	홍역 바이러스 균주 에드몬스톤-B	SMYRVFEV	2879	CAB43772.1
1997	홍역 바이러스 균주 에드몬스톤-B	SQQGMFHAY	2880	AAF85698.1
1998	홍역 바이러스 균주 에드몬스톤-B	TDTPIVYNDRNLLD	2881	Q89933.1
1999	홍역 바이러스 균주 에드몬스톤-B	VIINDDQGLFKV	2882	AAF85695.1
2000	홍역 바이러스 균주 에드몬스톤-B	YESGVRIASL	2883	AAF85698.1
2001	홍역 바이러스 균주 에드몬스톤-B	YLKDKALA	2884	AAF85698.1
2002	홍역 바이러스 균주 에드몬스톤-B	YVYDHSGEAVK	2885	AAF85692.1
2003	풍진 바이러스	ARVIDPAAQSFTGVV	2886	BAA28178.1
2004	풍진 바이러스	SDRASARVIDPAAQS	2887	BAA28178.1
2005	풍진 바이러스	VPPGKFVTAALLNTP	2888	BAA28178.1

[1097]

NO.	질환	에피토프 아미노산 서열	SEQ ID NO	ACCESS. NO.
2006	풍진 바이러스	WVTPVIGSQARKCGL	2889	BAA28178.1
2007	볼거리 루불라바이러스	GTYRLIPNARANLTA	400	AGC97176.1

[1098]

[1099] E. 다중-플랩 프라임 편집제의 전달

[1100] 또 다른 측면에서, 본 개시내용은 개별 벡터 상에서의 분할 인테인의 사용 및 뿐만 아니라 전기천공, 양이온성 지질-매개된 제제의 사용, 및 리보뉴클레오단백질 복합체에 융합된 수용체 리간드를 사용한 유도된 세포내이입 방법과 같은 기술을 사용한 리보핵단백질 복합체 (즉, PEgRNA 및/또는 제2-부위 gRNA에 복합체화된 프라임 편집제)의 직접 전달 전략을 포함한 다양한 전략을 사용한, 다중-플랩 프라임 편집제의 시험관내 및 생체내 전달을 제공한다. 임의의 이러한 방법이 본원에서 고려된다.

[1101] 전달 옵션의 개관

[1102] 일부 측면에서, 본 발명은 1개 이상의 다중-플랩 프라임 편집제-코딩 폴리뉴클레오티드, 예컨대 또는 본원에 기재된 다중-플랩 프라임 편집 시스템의 1개 이상의 성분을 코딩하는 본원에 기재된 바와 같은 1개 이상의 벡터, 그의 1개 이상의 전사체, 및/또는 그로부터 전사된 1개 이상의 단백질을 숙주 세포에 전달하는 것을 포함하는 방법을 제공한다. 일부 측면에서, 본 발명은 이러한 방법에 의해 생산된 세포, 및 이러한 세포를 포함하거나 또는 그로부터 생산된 유기체 (예컨대 동물, 식물, 또는 진균)를 추가로 제공한다. 일부 실시양태에서, 가이드 서열과 조합된 (및 임의로 그와 복합체화된) 본원에 기재된 바와 같은 다중-플랩 프라임 편집제가 세포에 전달된다. 통상적인 바이러스 및 비-바이러스 기반 유전자 전달 방법이 핵산을 포유동물 세포 또는 표적 조직에 도입하는 데 사용될 수 있다. 이러한 방법은 다중-플랩 프라임 편집제 시스템의 성분을 코딩하는 핵산을 배양물 또는 숙주 유기체 내의 세포에 투여하는 데 사용될 수 있다. 비-바이러스 벡터 전달 시스템은 DNA 플라스미드, RNA (예를 들어, 본원에 기재된 벡터의 전사체), 네이키드 핵산, 및 전달 비히클, 예컨대 리포솜과 복합체화된 핵산을 포함한다. 바이러스 벡터 전달 시스템은 DNA 및 RNA 바이러스를 포함하며, 이는 세포로 전달된 후에 에피솜 또는 통합된 게놈을 갖는다. 유전자 요법 절차의 검토를 위해, 문헌 [Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Bihm (eds) (1995); 및 Yu et al., Gene Therapy 1:13-26 (1994)]을 참조한다.

[1103] 핵산의 비-바이러스 전달 방법은 리포펙션, 뉴클레오펙션, 미세주사, 바이오리스틱, 비로솜, 리포솜, 이뮤노리포솜, 다가양이온 또는 지질:핵산 접합체, 네이키드 DNA, 인공 비리온, 및 DNA의 작용제-증진된 흡수를 포함한다. 리포펙션은 예를 들어 미국 특허 번호 5,049,386, 4,946,787; 및 4,897,355에 기재되어 있고, 리포펙션 시약은 상업적으로 판매된다 (예를 들어, 트랜스펙탐(Transfectam)™ 및 리포펙틴(Lipofectin)™). 폴리뉴클레오티드의 효율적인 수용체-인식 리포펙션에 적합한 양이온성 및 중성 지질은 파이너(Feigner), WO 91/17424; WO 91/16024의 것을 포함한다. 전달은 세포 (예를 들어, 시험관내 또는 생체외 투여) 또는 표적 조직 (예를 들어, 생체내 투여)에 대한 것일 수 있다.

[1104] 표적화된 리포솜, 예컨대 면역지질 복합체를 포함한 지질:핵산 복합체의 제조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 [Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992)]; 미국 특허 번호 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 및 4,946,787 참조).

[1105] 핵산의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 바이러스를 신체 내의 특이적 세포에 표적화하고 바이러스 페이로드를 핵으로 트래픽킹하기 위한 고도로 진화된 과정을 이용한다. 바이러스 벡터는 환자에게 직접 투여될 수 있거나 (생체내), 또는 시험관내에서 세포를 처리하는 데 사용될 수 있고, 변형된 세포는 임의로 환자에게 투여될 수 있다 (생체외). 통상적인 바이러스 기반 시스템은 유전자 전달을 위한 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 및 단순 포진 바이러스 벡터를 포함할 수 있다. 숙주 게놈에서의 통합은 레트로바이러스, 렌티바이러스, 및 아데노-연관 바이러스 유전자 전달 방법에 의해 가능하며, 이는 종종 삽입된 트랜스진의 장기간 발현을 발생시킨다. 추가적으로, 많은 상이한 세포 유형 및 표적 조직에서 높은 형질도입 효율이 관찰되었다.

[1106] 바이러스의 향성은 외래 외피 단백질을 혼입하여 표적 세포의 잠재적 표적 집단을 확장시킴으로써 변경될 수 있다. 렌티바이러스 벡터는 비-분열 세포를 형질도입 또는 감염시킬 수 있고 전형적으로 높은 바이러스 역가를 생산할 수 있는 레트로바이러스 벡터이다. 따라서, 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에 좌우될 것이다. 레트로바이러스 벡터는 최대 6-10 kb의 외래 서열에 대한 패키징 능력을 갖는 시스-작용성 긴 말단 반복부로 구성된다. 최소 시스-작용성 LTR은 벡터의 복제 및 패키징에 충분하며, 이는 이어서 치료 유전자를 표적 세포 내로 통합시켜 영구적 트랜스진 발현을 제공하는 데 사용된다. 널리 사용되는 레트로바이러스 벡터는 뮤린 백혈병 바이러스 (MuLV), 긴팔원숭이 유인원 백혈병 바이러스 (GaLV), 원숭이 면역 결핍 바이러스 (SIV), 인간 면역 결핍 바이러스 (HIV), 및 그의 조합을 기반으로 하는 것을 포함한다 (예를 들어, 문헌 [Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992);

Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991)]; PCT/US94/05700 참조). 일시적 발현이 바람직한 적용에서는, 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 많은 세포 유형에서 매우 높은 형질도입 효율을 가질 수 있고, 세포 분열을 필요로 하지 않는다. 이러한 벡터에 의해, 높은 역가 및 발현 수준이 수득되었다. 이 벡터는 비교적 간단한 시스템에서 대량으로 생산될 수 있다. 아데노-연관 바이러스 ("AAV") 벡터는 또한, 예를 들어 핵산 및 펩티드의 시험관내 생산에서, 그리고 생체내 및 생체외 유전자 요법 절차를 위해, 세포를 표적 핵산으로 형질도입하는 데 사용될 수 있다 (예를 들어, 문헌 [West et al., Virology 160:38-47 (1987)]; 미국 특허 번호 4,797,368; WO 93/24641; [Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994)] 참조). 재조합 AAV 벡터의 구축은 미국 특허 번호 5,173,414; 문헌 [Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); 및 Samulski et al., J. Virol. 63:03822-3828 (1989)]을 포함한 다수의 간행물에 기재되어 있다.

[1107] 패키징 세포는 전형적으로 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성하는 데 사용된다. 이러한 세포는 아데노바이러스를 패키징하는 293 세포, 및 레트로바이러스를 패키징하는 ψ2 세포 또는 PA317 세포를 포함한다. 유전자 요법에 사용되는 바이러스 벡터는 통상적으로, 핵산 벡터를 바이러스 입자 내로 패키징하는 세포주를 생산함으로써 생성된다. 벡터는 전형적으로 패키징 및 숙주 내로의 후속 통합에 요구되는 최소 바이러스 서열을 함유하고, 다른 바이러스 서열은 발현될 폴리뉴클레오티드(들)에 대한 발현 카세트에 의해 대체된다. 누락된 바이러스 기능은 전형적으로 패키징 세포주에 의해 트랜스로 공급된다. 예를 들어, 유전자 요법에 사용되는 AAV 벡터는 전형적으로 숙주 게놈 내로의 패키징 및 통합에 요구되는 AAV 게놈으로부터의 ITR 서열만을 보유한다. 바이러스 DNA는 다른 AAV 유전자, 즉 rep 및 cap를 코딩하지만 ITR 서열은 결여되어 있는 헬퍼 플라스미드를 함유하는 세포주 내에 패키징된다. 세포주는 또한 헬퍼로서 아데노바이러스로 감염될 수 있다. 헬퍼 바이러스는 헬퍼 플라스미드로부터 AAV 벡터의 복제 및 AAV 유전자의 발현을 촉진한다. 헬퍼 플라스미드는 ITR 서열의 결여로 인해 유의한 양으로 패키징되지 않는다. 아데노바이러스에 의한 오염은, 예를 들어 아데노바이러스가 AAV보다 더 감수성인 열 처리에 의해 감소될 수 있다. 핵산을 세포에 전달하기 위한 추가의 방법은 관련 기술분야의 통상의 기술자에게 공지되어 있다. 예를 들어, 본원에 참조로 포함된 US20030087817을 참조한다.

[1108] 다양한 실시양태에서, 이중 PE 구축물 (분할-구축물 포함)은 1개 이상의 rAAV 벡터에서의 전달을 위해 조작될 수 있다. 본원에 제공된 방법 및 조성물 중 임의의 것과 관련된 rAAV는 임의의 유도체 또는 유사형 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 2/1, 2/5, 2/8, 2/9, 3/1, 3/5, 3/8, 또는 3/9)을 포함한 임의의 혈청형의 것일 수 있다. rAAV는 세포에 전달될 유전자 로드 (즉, 관심 유전자, 예컨대 rAAV에 의해 세포 내로 운반되는 전체 또는 분할 PE 융합 단백질을 발현하는 재조합 핵산 벡터)를 포함할 수 있다. rAAV는 키메라일 수 있다.

[1109] 본원에 사용된 rAAV의 혈청형은 재조합 바이러스의 캡시드 단백질의 혈청형을 지칭한다. 유도체 및 유사형의 비제한적 예는 rAAV2/1, rAAV2/5, rAAV2/8, rAAV2/9, AAV2-AAV3 하이브리드, AAVrh.10, AAVhu.14, AAV3a/3b, AAVrh32.33, AAV-HSC15, AAV-HSC17, AAVhu.37, AAVrh.8, CHt-P6, AAV2.5, AAV6.2, AAV2i8, AAV-HSC15/17, AAVM41, AAV9.45, AAV6(Y445F/Y731F), AAV2.5T, AAV-HAE1/2, AAV 클론 32/83, AAVShH10, AAV2 (Y->F), AAV8 (Y733F), AAV2.15, AAV2.4, AAVM41, 및 AAVr3.45를 포함한다. 키메라 VP1 단백질을 갖는 유도체 및 유사형의 비제한적 예는 rAAV2/5-1VP1u로, 이는 AAV2의 게놈, AAV5의 캡시드 백본 및 AAV1의 VP1u를 갖는다. 키메라 VP1 단백질을 갖는 유도체 및 유사형의 다른 비제한적 예는 rAAV2/5-8VP1u, rAAV2/9-1VP1u, 및 rAAV2/9-8VP1u이다.

[1110] AAV 유도체/유사형, 및 이러한 유도체/유사형을 생산하는 방법은 관련 기술분야에 공지되어 있다 (예를 들어, 문헌 [Mol Ther. 2012 Apr;20(4):699-708. doi: 10.1038/mt.2011.287. Epub 2012 Jan 24. The AAV vector toolkit: poised at the clinical crossroads. Asokan A1, Schaffer DV, Samulski RJ.] 참조). 유사형화 rAAV 벡터를 생산하고 사용하는 방법은 관련 기술분야에 공지되어 있다 (예를 들어, 문헌 [Duan et al., J. Virol., 75:7662-7671, 2001; Halbert et al., J. Virol., 74:1524-1532, 2000; Zolotukhin et al., Methods, 28:158-167, 2002; 및 Auricchio et al., Hum. Molec. Genet., 10:3075-3081, 2001] 참조).

[1111] rAAV 입자를 제조하거나 패키징 방법은 관련 기술분야에 공지되어 있고, 시약은 상업적으로 입수가능하다 (예를 들어, 문헌 [Zolotukhin et al., Production and purification of serotype 1, 2, and 5 recombinant adeno-associated viral vectors. Methods 28 (2002) 158-167]; 및 미국 특허 공개 번호 US20070015238 및

US20120322861 참조, 이는 본원에 참조로 포함됨; 및 플라스미드 및 키트는 ATCC 및 셀 바이오랩스, 인크.(Cell Biolabs, Inc.)로부터 입수가능함). 예를 들어, 관심 유전자를 포함하는 플라스미드는, 예를 들어 rep 유전자 (예를 들어, Rep78, Rep68, Rep52 및 Rep40을 코딩함) 및 cap 유전자 (VP1, VP2, 및 VP3을 코딩함, 본원에 기재된 바와 같은 변형된 VP2 영역을 포함함)를 함유하는 1개 이상의 헬퍼 플라스미드와 조합될 수 있고, rAAV 입자가 패키징되고 후속적으로 정제될 수 있도록 재조합 세포 내로 형질감염될 수 있다.

[1112] 재조합 AAV는 (a) 관심 단백질 또는 폴리펩티드를 코딩하는 서열 또는 관심 RNA (예를 들어, siRNA 또는 마이크로RNA)를 포함하는 1개 이상의 이종 핵산 영역, 및 (b) 1개 이상의 핵산 영역 (예를 들어, 이종 핵산 영역)에 플랭킹된 역전된 말단 반복부 (ITR) 서열 (예를 들어, 야생형 ITR 서열 또는 조작된 ITR 서열)을 포함하는 1개 이상의 영역을 최소한으로 포함할 수 있는 핵산 벡터를 포함할 수 있다. 본원에서, 관심 단백질을 코딩하는 서열 또는 관심 RNA를 포함하는 이종 핵산 영역은 관심 유전자로 지칭된다.

[1113] 본원에 제공된 rAAV 입자 중 어느 하나는 VP1u 영역 외부에 상이한 혈청형의 아미노산을 갖는 캡시드 단백질을 가질 수 있다. 일부 실시양태에서, VP1 단백질의 백본의 혈청형은 ITR 및/또는 Rep 유전자의 혈청형과 상이하다. 일부 실시양태에서, 입자의 VP1 캡시드 단백질의 백본의 혈청형은 ITR의 혈청형과 동일하다. 일부 실시양태에서, 입자의 VP1 캡시드 단백질의 백본의 혈청형은 Rep 유전자의 혈청형과 동일하다. 일부 실시양태에서, rAAV 입자의 캡시드 단백질은 개선된 형질도입 효율을 발생시키는 아미노산 돌연변이를 포함한다.

[1114] 일부 실시양태에서, 핵산 벡터는 핵산 (예를 들어, 이종 핵산)의 발현을 용이하게 하는 서열, 예를 들어 핵산에 작동가능하게 연결된 발현 제어 서열을 포함하는 1개 이상의 영역을 포함한다. 수많은 이러한 서열이 관련 기술분야에 공지되어 있다. 발현 제어 서열의 비제한적 예는 프로모터, 인슐레이터, 사일렌서, 반응 요소, 인트론, 인핸서, 개시 부위, 종결 신호, 및 폴리(A) 테일을 포함한다. 이러한 제어 서열의 임의의 조합이 본원에서 고려된다 (예를 들어, 프로모터 및 인핸서).

[1115] 최종 AAV 구축물은 PEgRNA를 코딩하는 서열을 포함할 수 있다. 다른 실시양태에서, AAV 구축물은 제2-부위 닉킹 가이드 RNA를 코딩하는 서열을 포함할 수 있다. 또 다른 실시양태에서, AAV 구축물은 제2-부위 닉킹 가이드 RNA를 코딩하는 서열 및 PEgRNA를 코딩하는 서열을 포함할 수 있다.

[1116] 다양한 실시양태에서, PEgRNA 및 제2-부위 닉킹 가이드 RNA는 적절한 프로모터, 예컨대 인간 U6 (hU6) 프로모터, 마우스 U6 (mU6) 프로모터, 또는 다른 적절한 프로모터로부터 발현될 수 있다. PEgRNA 및 제2-부위 닉킹 가이드 RNA는 동일한 프로모터 또는 상이한 프로모터에 의해 구동될 수 있다.

[1117] 일부 실시양태에서, rAAV 구축물 또는 본원의 조성물은 대상체에게 경장으로 투여된다. 일부 실시양태에서, rAAV 구축물 또는 본원의 조성물은 대상체에게 비경구로 투여된다. 일부 실시양태에서, rAAV 입자 또는 본원의 조성물은 대상체에게 피하로, 안구내로, 유리체내로, 망막하로, 정맥내로 (IV), 뇌실내로, 근육내로, 척수강내로 (IT), 수조내로, 복강내로, 흡입을 통해, 국소로, 또는 1개 이상의 세포, 조직, 또는 기관으로의 직접 주사에 의해 투여된다. 일부 실시양태에서, rAAV 입자 또는 본원의 조성물은 간 동맥 또는 문맥 내로의 주사에 의해 대상체에게 투여된다.

[1118] 분할 다중-플랩 PE 벡터-기반 전략

[1119] 이러한 측면에서, 다중-플랩 프라임 편집제는 분할 부위에서 나뉘어질 수 있고, 전체/완전 프라임 편집제의 2개의 절반으로서 제공될 수 있다. 2개의 절반은 세포에 (예를 들어, 발현된 단백질로서 또는 개별 발현 벡터 상에서) 전달될 수 있고, 일단 세포 내부에서 접촉하면, 2개의 절반은 각각의 프라임 편집제 절반 상의 인테인의 자기-스플라이싱 작용을 통해 완전한 프라임 편집제를 형성한다. 분할 인테인 서열은 코딩된 프라임 편집제의 각각의 절반 내로 조작되어 세포 내부에서의 그의 트랜스스플라이싱 및 완전한, 기능성 PE의 동반 복구를 용이하게 할 수 있다.

[1120] 이들 분할 인테인-기반 방법은 생체내 전달에 대한 여러 장벽을 극복한다. 예를 들어, 프라임 편집제를 코딩하는 DNA는 rAAV 패키징 제한보다 더 크고, 따라서 특별한 해결책을 필요로 한다. 하나의 이러한 해결책은 세포에 공동-전달되는 경우에 기능적 편집제 단백질을 재구성하는 2개의 개별 rAAV 입자 내로 패키징된 분할 인테인 쌍에 융합된 편집제를 제제화하는 것이다. 제2-부위 닉킹 표적의 최적화 및 렌티바이러스 및 rAAV를 포함한 바이러스 벡터 내로의 다중-플랩 프라임 편집제의 적절한 패키징을 포함한, 다중-플랩 프라임 편집의 고유한 특색을 설명하기 위한 여러 다른 특별한 고려사항이 기재된다.

[1121] 이러한 측면에서, 다중-플랩 프라임 편집제는 분할 부위에서 나뉘어질 수 있고, 전체/완전 프라임 편집제의 2개의 절반으로서 제공될 수 있다. 2개의 절반은 세포에 (예를 들어, 발현된 단백질로서 또는 개별 발현 벡터 상

에서) 전달될 수 있고, 일단 세포 내부에서 접촉하면, 2개의 절반은 각각의 프라임 편집제 절반 상의 인테인의 자기-스플라이싱 작용을 통해 완전한 프라임 편집제를 형성한다. 분할 인테인 서열은 코딩된 프라임 편집제의 각각의 절반 내로 조작되어 세포 내부에서의 그의 트랜스스플라이싱 및 완전한, 기능성 PE의 동반 복구를 용이하게 할 수 있다.

[1122] 도 66은 각각의 프라임 편집제 절반 단백질의 단부 또는 시작부에 위치한 분할-인테인 절반의 자기-스플라이싱 작용을 통해 전체 프라임 편집제로서 재생되는, 2개의 PE 절반 단백질로서 제공된 프라임 편집제의 한 실시양태를 도시한다. 본원에 기재된 다중-플랩 프라임 편집제는 또한 동일한 방식으로 제공될 수 있다. 본원에 사용된 용어 "PE N-말단 절반"은 완전 프라임 편집제의 N-말단 절반을 지칭하고, 이는 완전 프라임 편집제의 PE N-말단 절반 (즉, N-말단 엑스테인)의 C-말단 단부에 "N 인테인"을 포함한다. "N 인테인"은 완전한, 전체-형성된 분할-인테인 모이어티의 N-말단 절반을 지칭한다. 본원에 사용된 용어 "PE C-말단 절반"은 완전 프라임 편집제의 C-말단 절반을 지칭하고, 이는 완전 프라임 편집제의 C-말단 절반 (즉, C-말단 엑스테인)의 N-말단 단부에 "C 인테인"을 포함한다. 2개의 절반 단백질, 즉 PE N-말단 절반 및 PE C-말단 절반이, 예를 들어 세포 내에서 서로 접촉하게 되는 경우에, N 인테인 및 C 인테인은 자기-절제 및 PE N-말단 절반의 C-말단 단부와 PE C-말단 절반의 N-말단 단부 사이의 펩티드 결합의 형성의 동시 과정을 거쳐 완전한 napDNAbp 도메인 (예를 들어, Cas9 닉카제) 및 RT 도메인을 포함하는 완전한 프라임 편집제 단백질을 재형성한다. 도면에 제시되지는 않았지만, 프라임 편집제는 또한 N-말단 및/또는 C-말단의 NLS, 뿐만 아니라 각각의 도메인을 연결하는 아미노산 링커 서열을 포함한 추가의 서열을 포함할 수 있다.

[1123] 다양한 실시양태에서, 다중-플랩 프라임 편집제는 전체 프라임 편집제를 "분할 부위"로서 "분할"함으로써 2개의 절반 단백질 (즉, PE N-말단 절반 및 PE C-말단 절반)로서 조작될 수 있다. "분할 부위"는 프라임 편집제에서 2개의 인접한 아미노산 잔기 사이의 분할 인테인 서열 (즉, N 인테인 및 C 인테인)의 삽입 위치를 지칭한다. 보다 구체적으로, "분할 부위"는 전체 프라임 편집제를 2개의 별개의 절반으로 나누는 위치를 지칭하며, 여기서 각각의 절반은 분할 부위에서 N 인테인 또는 C 인테인 모티프에 융합된다. 분할 부위는 프라임 편집제 융합 단백질의 임의의 적합한 위치에 있을 수 있지만, 바람직하게는 분할 부위는 (예를 들어, 발현 벡터에 의한) 전달을 위해 적절하게 사이징된 2개의 절반 단백질의 형성을 가능하게 하는 위치에 위치하고, 여기서 분할 부위 말단에서 각각의 절반 단백질에 융합된 인테인은 하나의 절반 단백질이 세포 내부에서 다른 절반 단백질과 접촉할 때 서로 충분히 상호작용하는데 이용가능하다.

[1124] 일부 실시양태에서, 분할 부위는 napDNAbp 도메인에 위치한다. 다른 실시양태에서, 분할 부위는 RT 도메인에 위치한다. 다른 실시양태에서, 분할 부위는 napDNAbp 도메인과 RT 도메인을 연결하는 링커에 위치한다.

[1125] 다양한 실시양태에서, 분할 부위 설계는 2개의 절반 프라임 편집제 도메인을 2개의 상이한 AAV 게놈 내로 패키징할 목적으로, 둘 다 구조적으로 허용되는 N- 및 C-말단 인테인을 분할 및 삽입하기 위한 부위를 발견하는 것을 필요로 한다. 추가적으로, 트랜스 스플라이싱에 필요한 인테인 잔기는 C 말단 엑스테인의 N 말단에서 잔기를 돌연변이시키거나 또는 인테인 "자국"을 남길 잔기를 삽입함으로써 혼입될 수 있다.

[1126] SpCas9 닉카제 또는 SaCas9 닉카제를 포함하는 분할 다중-플랩 (예를 들어, 이중 플랩 또는 사중 플랩) 프라임 편집제의 예시적인 분할 구성은 하기와 같다.

에스. 피오게네스 PE, 링커에서 분할, N 말단 부분
구조: [N 엑스테인]-[N 인테인]

MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLG**
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYF
TVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
DSGGSSGGSCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWH
DRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPNSG
GSKRTADGSEFEPKKKRKV(SEQ ID NO: 443)

설명:
NLS (SEQ ID NO: 124, 125)
CAS9 (SEQ ID NO: 445)
링커 (SEQ ID NO: 446)
NPUN 인테인 (SEQ ID NO: 447)

[1127]

에스. 피오게네스 PE, 링커에서 분할, C 말단 부분
구조: [C 인테인]-[C 엑스테인]

MKRTADGSEFESPKKKRKV*IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN***CFNS**
GSETPGTSESATPESSGGSSGGSS*TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWA*
ETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW
NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKD
AFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRI
QHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGY
LLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKT
GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW
RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQP
PDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTR
PDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIA
LTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFL
*PKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP***SGGS**KRTA
DGSEFEPKKKRKV(SEQ ID NO: 450)

설명:
NLS (SEQ ID NO: 124, 125)
링커 1 (SEQ ID NO: 453)
링커 2 (SEQ ID NO: 174)

NPUC 인테인 (SEQ ID NO: 452)
RT (SEQ ID NO: 454)

에스. 아우레우스 PE, 잔기 740/741 사이에서 분할, N 말단 부분
구조: [N 엑스테인]-[N 인테인]

MKRTADGSEFESPKKKRKV**GKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLF**
KEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRN
SKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQL
DQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELR
SVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQ
IAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKI
LTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWH
TNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAII
KKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAK
YLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKV
LVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLL
EERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGF
TSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQ
MFEEKQAECLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWH
DRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN**SG**
GSKRTADGSEFEPKKKRKV(SEQ ID NO: 458)

[1128]

설명: NLS (SEQ ID NO: 124, 125) **CAS9** (SEQ ID NO: 460) **링커** (SEQ ID NO: 174) NPUN 인테인 (SEQ ID NO: 462)
에스. 아우레우스 PE, 잔기 740/741 사이에서 분할, C 말단 부분 구조: [C 인테인]-[C 엑스테인]
MKRTADGSEFESPKKKRKV*IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN***CFN*****E** **IETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKG** **NTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGD** **EKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRN** **KVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLK** **KISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMN** **DKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG*SGGSSGGSSGS*** ***ETPGTSESATPESSGGSSGGSS****TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAET* *GGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNT* *PLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAF* *FCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQ* *HPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYL* *LKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKT* *GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW* *RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQP* *PDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTR* *PDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIA* *LTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFL* *PKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP****SGGS***KRTA DGSEFEPKKKRKV(SEQ ID NO: 465) 설명: NLS (SEQ ID NO: 124, 125) **CAS9** (SEQ ID NO: 467) **링커 1** (SEQ ID NO: 127) **링커 2** (SEQ ID NO: 174) *NPUC 인테인* (SEQ ID NO: 452) *RT* (SEQ ID NO: 471)

[1129]

[1130] 예로서 SpCas9 닉카제 (서열식별번호: 18, 1368개 아미노산)를 사용하는 다양한 실시양태에서, 분할은 1 내지 1368의 임의의 2개의 아미노산 사이에서 이루어질 수 있다. 그러나, 바람직한 분할은 단백질의 중심 영역 사이에, 예를 들어 서열식별번호: 18의 아미노산 50-1250, 또는 100-1200, 또는 150-1150, 또는 200-1100, 또는 250-1050, 또는 300-1000, 또는 350-950, 또는 400-900, 또는 450-850, 또는 500-800, 또는 550-750, 또는 600-700 사이에 위치할 것이다. 구체적인 예시적인 실시양태에서, 분할 부위는 740/741, 또는 801/802, 또는 1010/1011, 또는 1041/1042 사이일 수 있다. 다른 실시양태에서, 분할 부위는 서열식별번호: 18의 SpCas9 대비 1/2, 2/3, 3/4, 4/5, 5/6, 6/7, 7/8, 8/9, 9/10, 10/11, 12/13, 14/15, 15/16, 17/18, 19/20, 20/21, 21/22, 22/23, 23/24, 24/25, 25/26, 26/27, 27/28, 28/29, 29/30, 30/31, 31/32, 32/33, 33/34, 34/35, 35/36, 36/37, 38/39, 39/40, 41/42, 42/43, 43/44, 44/45, 45/46, 46/47, 47/48, 48/49, 49/50, 51/52, 52/53, 53/54, 54/55, 55/56, 56/57, 57/58, 58/59, 59/60, 61/62, 62/63, 63/64, 64/65, 65/66, 66/67, 67/68, 68/69, 69/70, 71/72, 72/73, 73/74, 74/75, 75/76, 76/77, 77/78, 78/79, 79/80, 81/82, 82/83, 83/84, 84/85, 85/86, 86/87, 87/88, 88/89, 89/90 사이, 또는 90-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1050, 1050-1100, 1100-1150, 1150-1200, 1200-1250, 1250-1300, 1300-1350, 및 1350-1368 사이의 인접 잔기의 임의의 2개의 쌍 사이, 서열식별번호: 18과 적어도 80%, 85%, 90%, 95%, 98%, 99%, 또는 99.9% 서열 동일성을 갖는 아미노산 서열에서의 임의의 2개의 상응하는 잔기 사이, 또는 아미노산 서열 서열식별번호: 19-88 중 임의의 것, 또는 서열식별번호: 19-88 중 임의의 것과 적어도 80%, 85%, 90%, 95%, 98%, 99%, 또는 99.9% 서열 동일성을 갖는 아미노산 서열의 SpCas9의 변이체 또는 등가물에서의 임의의 2개의 상응하는 잔기 사이일 수 있다.

[1131] 다양한 실시양태에서, 분할 인테인 서열은 하기 인테인 서열로부터 조작될 수 있다.

명칭	리간드-의존성 인테인의 서열
2-4 인테인:	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 472)
3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDR VAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASM MGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEIL MIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATS SRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRAL DKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHL YSMKYTNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDK FLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHN C (SEQ ID NO: 473)
30R3-1 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 474)
30R3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 475)
30R3-3 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH

[1132]

	RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 476)
37R3-1 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYNPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC ((SEQ ID NO: 477)
37R3-2 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 478)
37R3-3 인테인	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 479)

[1133]

[1134] 다양한 다른 실시양태에서, 분할 인테인 서열은 하기와 같이 사용될 수 있다:

인테인-N	인테인-C
NPU-N CLSYETEILTVEYGLLPIGKIVEKRIEC TVYSVDNNGNIYTQPVAQWHDRGEQ EVFEYCLEDGSLIRATKDHKFMTVDG QMLPIDEIFERELDLMRVDNLPNSGG S (SEQ ID NO: 447)	NPU-C IKIATRKYLGKQNVYDIGVERDHNFALKNG FIASN (SEQ ID NO: 452)

[1135]

[1136] 다양한 실시양태에서, 분할 인테인은 완전한 PE 융합 단백질의 개별 부분을 세포에 개별적으로 전달하는 데 사용될 수 있으며, 이는 세포에서 발현 시 트랜스 스플라이싱을 통해 완전한 PE 융합 단백질로서 재구성된다.

[1137] 일부 실시양태에서, 개시내용은

[1138] (a) 제1 분할 인테인 서열에 융합된 PE 융합 단백질의 N-말단 단편을 코딩하는 제1 발현 벡터를 구축하는 단계;

[1139] (b) 제2 분할 인테인 서열에 융합된 PE 융합 단백질의 C-말단 단편을 코딩하는 제2 발현 벡터를 구축하는 단계;

[1140] (c) 제1 및 제2 발현 벡터를 세포에 전달하는 단계

[1141] 를 포함하는, PE 융합 단백질을 세포에 전달하는 방법을 제공하며,

[1142] 여기서 N-말단 및 C-말단 단편은 제1 및 제2 분할 인테인 서열의 자기-절제를 유발하는 트랜스 스플라이싱 활성의 결과로서 세포에서 PE 융합 단백질로서 재구성된다.

[1143] 일부 실시양태에서 분할 부위는 napDNAbp 도메인, 링커, 또는 리버스 트랜스크립타제 도메인을 포함한, 프라임 편집제 융합체 내의 임의의 곳일 수 있다.

[1144] 다른 실시양태에서, 분할 부위는 napDNAbp 도메인 내에 있다.

[1145] 또 다른 실시양태에서, 분할 부위는 리버스 트랜스크립타제 또는 폴리머라제 도메인 내에 있다.

[1146] 또 다른 실시양태에서, 분할 부위는 링커 내에 있다.

[1147] 다양한 실시양태에서, 본 개시내용은 napDNAbp (예를 들어, Cas9 도메인) 및 리버스 트랜스크립타제를 포함하는 다중-플랩 프라임 편집제를 제공하며, 여기서 napDNAbp 및/또는 리버스 트랜스크립타제 중 하나 또는 둘 다는 인테인, 예를 들어 리간드-의존성 인테인을 포함한다. 전형적으로, 인테인은 리간드 (예를 들어, 소분자, 예컨대 4-히드록시타목시펜, 펩티드, 단백질, 폴리뉴클레오티드, 아미노산, 및 뉴클레오티드)의 부재 하에 단백질 스플라이싱 활성을 나타내지 않거나 최소로 나타내는 리간드-의존성 인테인이다. 리간드-의존성 인테인은 공지되어 있고, 그의 전체 내용이 본원에 참조로 포함되는, U.S. 2014/0065711 A1로서 공개된 미국 특허 출원 U.S.S.N. 14/004,280에 기재된 것을 포함한다. 또한, 분할-Cas9 아키텍처의 사용에서, 일부 실시양태에서, 인테인은 서열식별번호: 8-15, 447, 452, 462, 및 472-479로 이루어진 군으로부터 선택된 아미노산 서열을 포함한다.

[1148] 다양한 실시양태에서, napDNAbp 도메인은 서열식별번호: 18의 정규 SpCas9 도메인과 비교하여 보다 작은 크기의 napDNAbp 도메인이다.

[1149] 정규 SpCas9 단백질은 1368개 아미노산 길이이고, 158 킬로달톤의 예측 분자량을 갖는다. 본원에 사용된 용어 "작은-크기의 Cas9 변이체"는 적어도 1300개 미만의 아미노산, 또는 적어도 1290개 미만의 아미노산, 또는 1280개 미만의 아미노산, 또는 1270개 미만의 아미노산, 또는 1260개 미만의 아미노산, 또는 1250개 미만의 아미노산, 또는 1240개 미만의 아미노산, 또는 1230개 미만의 아미노산, 또는 1220개 미만의 아미노산, 또는 1210개 미만의 아미노산, 또는 1200개 미만의 아미노산, 또는 1190개 미만의 아미노산, 또는 1180개 미만의 아미노산, 또는 1170개 미만의 아미노산, 또는 1160개 미만의 아미노산, 또는 1150개 미만의 아미노산, 또는 1140개 미만의 아미노산, 또는 1130개 미만의 아미노산, 또는 1120개 미만의 아미노산, 또는 1110개 미만의 아미노산, 또는 1100개 미만의 아미노산, 또는 1050개 미만의 아미노산, 또는 1000개 미만의 아미노산, 또는 950개 미만의 아미노산, 또는 900개 미만의 아미노산, 또는 850개 미만의 아미노산, 또는 800개 미만의 아미노산, 또는 750개 미만의 아미노산, 또는 700개 미만의 아미노산, 또는 650개 미만의 아미노산, 또는 600개 미만의 아미노산, 또는 550개 미만의 아미노산, 또는 500개 미만의 아미노산이지만, 적어도 약 400개 초과의 아미노산이고, Cas9 단백질의 요구되는 기능을 보유하는 임의의 Cas9 변이체-자연 발생, 조작, 또는 그 밖의 것-를 지칭한다.

[1150] 한 실시양태에서, 실시예 20에 도시된 바와 같이, 본 명세서는 정규 SpCas9 (서열식별번호: 18)의 잔기 1024와 1025 (또는 Met-마이너스 서열식별번호: 18 대비 각각 잔기 1023 및 1024로 지칭될 수 있음) 사이에 분할된 하기 분할-인테인 PE 구축물을 포괄한다.

[1151] 먼저, 1024 ("K") 및 1025 ("S") 잔기 사이의 분할 부위의 위치를 나타낸 서열식별번호: 18의 아미노산 서열은 하기와 같이 제시된다:

설명	서열	SEQ ID NO:
SpCas9 *스트렙토코쿠스 피오게네스* M1 스위스프롯 수탁 번호 Q99ZW2 야생형	**M**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA **KS**EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SEQ ID NO: 18, 분할 부위 1024/1025를 볼드체로 표시함 위치 1의 M은 특정 실시양태에서 PE 융합 단백질에 반드시 존재하는 것은 아니다. 따라서, 아미노산 서열에서 위치 1의 Met이 제외되어 있는 경우, 분할 부위의 넘버링은 1023/1024이다.

[1152]

[1153] 이러한 구성에서, N-말단 절반의 아미노산 서열 (아미노산 1-1024)은 하기와 같다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV
KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF
ANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIA**K** (SEQ ID NO: 3877).

[1154]

[1155] 이러한 구성에서, N-말단 절반의 아미노산 서열 (아미노산 1-1023) (여기서 단백질은 위치 1에서 Met-마이너스임)은 하기와 같다:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV
KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF
ANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIA**K** (SEQ ID NO: 3878).

[1156]

[1157] 이러한 구성에서, C-말단 절반의 아미노산 서열 (아미노산 1024-1368 (또는 Met-마이너스 Cas9에서 아미노산 1023-1367로서 카운팅됨))은 하기와 같다:

SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 3879).

[1158]

[1159] 실시예 20에 제시된 바와 같이, PE2 (서열식별번호: 18의 SpCas9를 기초로 함) 구축물을 하기와 같이 위치 1023/1024 (Met-마이너스 서열식별번호: 18 대비)에서 2개의 별개의 구축물로 분할하였다:

[1160] 1023/1024 N 말단 절반에서의 SpPE2 분할

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF
RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE
QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK*CLSYETEILTVEYGLL*
PIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQM
*LPIDEIFERELDLMRVDNLPN***SGGS***KRTADGSEFEPKKKRKV* (SEQ ID NO: 3875)

[1161]

[1162] 설명: NLS, Cas9, 돌연변이된 잔기, **링커**, *NpuN 인테인*, *NpuC 인테인*, **RT**

[1163] 1023/1024 C 말단 절반에서의 SpPE2 분할

MKRTADGSEFESPKKKRKV*IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN*CFNEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**TLNIEDEYRLHE
**TSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAR
LGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP
NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL
PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTL
GNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFL
GKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDL
TKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVL
TKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV
ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQR
KAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAH
IHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRM
ADQAARKAAITETPDTSTLLIENSSP**SGGS***KRTADGSEFEPKKKRK* (SEQ ID NO: 3876)

[1164]

[1165] 설명: NLS, Cas9, 돌연변이된 잔기, **링커**, *NpuN 인테인*, *NpuC 인테인*, **RT**

[1166] 본 개시내용은 또한 분할-인테인 다중-플랩 프라임 편집제를 세포에 전달하고/거나 세포를 분할-인테인 다중-플랩 프라임 편집제로 처리하는 방법을 고려한다.

[1167] 일부 실시양태에서, 개시내용은

[1168] (a) 제1 분할 인테인 서열에 융합된 PE 융합 단백질의 N-말단 단편을 코딩하는 제1 발현 벡터를 구축하는 단계;

[1169] (b) 제2 분할 인테인 서열에 융합된 PE 융합 단백질의 C-말단 단편을 코딩하는 제2 발현 벡터를 구축하는 단계;

[1170] (c) 제1 및 제2 발현 벡터를 세포에 전달하는 단계

[1171] 를 포함하는, PE 융합 단백질을 세포에 전달하는 방법을 제공하며,

[1172] 여기서 N-말단 및 C-말단 단편은 제1 및 제2 분할 인테인 서열의 자기-절제를 유발하는 트랜스 스플라이싱 활성의 결과로서 세포에서 PE 융합 단백질로서 재구성된다.

[1173] 특정 실시양태에서, 제1 분할 인테인 서열에 융합된 PE 융합 단백질의 N-말단 단편은 서열식별번호: 3875, 또는 서열식별번호: 3875와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 또는 적어도 99.9% 서열 동일성을 갖는 아미노산 서열이다.

[1174] 다른 실시양태에서, 제1 분할 인테인 서열에 융합된 PE 융합 단백질의 C-말단 단편은 서열식별번호: 3876, 또는 서열식별번호: 3876과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 또는 적어도 99.9% 서열 동일성을 갖는 아미노산 서열이다.

[1175] 다른 실시양태에서, 개시내용은

[1176] (a) 제1 분할 인테인 서열에 융합된 PE 융합 단백질의 N-말단 단편을 코딩하는 제1 발현 벡터를 구축하는 단계;

[1177] (b) 제2 분할 인테인 서열에 융합된 PE 융합 단백질의 C-말단 단편을 코딩하는 제2 발현 벡터를 구축하는 단계;

[1178] (c) 제1 및 제2 발현 벡터를 세포에 전달하는 단계

[1179] 를 포함하는, 세포 내의 표적 DNA 서열을 편집하는 방법을 제공하며,

[1180] 여기서 N-말단 및 C-말단 단편은 제1 및 제2 분할 인테인 서열의 자기-절제를 유발하는 트랜스 스플라이싱 활성의 결과로서 세포에서 PE 융합 단백질로서 재구성된다.

[1181] 특정 실시양태에서, 제1 분할 인테인 서열에 융합된 PE 융합 단백질의 N-말단 단편은 서열식별번호: 3875, 또는 서열식별번호: 3875와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 또는 적어도 99.9% 서열 동일성을 갖는 아미노산 서열이다.

[1182] 다른 실시양태에서, 제1 분할 인테인 서열에 융합된 PE 융합 단백질의 C-말단 단편은 서열식별번호: 3876, 또는 서열식별번호: 3876과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 또는 적어도 99.9% 서열 동일성을 갖는 아미노산 서열이다.

[1183] PE 리보핵단백질 복합체의 전달

[1184] 이러한 측면에서, 다중-플랩 프라임 편집제는 전기천공 및 지질 나노입자를 포함한 다양한 방법에 의해 PEgRNA와 복합체화된 다중-플랩 프라임 편집제 (즉, PE 리보핵단백질 복합체)의 전달을 수반하는 비-바이러스 전달 전략에 의해 전달될 수 있다. 핵산의 비-바이러스 전달 방법은 리포펙션, 뉴클레오펙션, 미세주사, 바이오리스틱, 비로솜, 리포솜, 이뮤노리포솜, 다가양이온 또는 지질:핵산 접합체, 네이키드 DNA, 인공 비리온, 및 DNA의 작용제-증진된 흡수를 포함한다. 리포펙션은 예를 들어 미국 특허 번호 5,049,386, 4,946,787; 및 4,897,355에 기재되어 있고, 리포펙션 시약은 상업적으로 판매된다 (예를 들어, 트랜스펙탐(Transfectam)™ 및 리포펙틴(Lipofectin)™). 폴리뉴클레오티드의 효율적인 수용체-인식 리포펙션에 적합한 양이온성 및 중성 지질은 파이너(Feigner), WO 91/17424; WO 91/16024의 것을 포함한다. 전달은 세포 (예를 들어, 시험관내 또는 생체외 투여) 또는 표적 조직 (예를 들어, 생체내 투여)에 대한 것일 수 있다.

[1185] 표적화된 리포솜, 예컨대 면역지질 복합체를 포함한 지질:핵산 복합체의 제조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 [Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992)]; 미국 특허 번호 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 및 4,946,787 참조).

[1186] 리보핵단백질 복합체의 비-바이러스 전달을 위한 접근법을 논의하는 하기 참고문헌을 추가로 참조할 수 있고, 이들 각각은 본원에 참조로 포함된다.

Chen, Sean, et al. "Highly efficient mouse genome editing by CRISPR ribonucleoprotein electroporation of zygotes." *Journal of Biological Chemistry* (2016): jbc-M116. PubMed

Zuris, John A., et al. "Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo." *Nature biotechnology* 33.1 (2015): 73. PubMed

Rouet, Romain, et al. "Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing." *Journal of the American Chemical Society* 140.21 (2018): 6596-6603. PubMed.

[1187]

[1188] 도 68c는 다양한 개시된 PE 리보핵단백질 복합체 (고농도의 PE2, 고농도의 PE3 및 저농도의 PE3)가 이러한 방식으로 전달될 수 있음을 보여주는 데이터를 제공한다.

[1189] mRNA에 의한 PE의 전달

[1190] 다중-플랩 프라임 편집-기반 게놈 편집이 바람직한 세포에 다중-플랩 프라임 편집제 및/또는 PEgRNA를 전달하는 데 활용될 수 있는 또 다른 방법은 메신저 RNA (mRNA) 전달 방법 및 기술의 사용을 활용하는 것에 의한다. 예를 들어, 각각의 전문이 본원에 참조로 포함되는 PCT/US2014/028330, US8822663B2, NZ700688A, ES2740248T3, EP2755693A4, EP2755986A4, WO2014152940A1, EP3450553B1, BR112016030852A2, 및 EP3362461A1을 포함한, 본 개시내용에서 이용될 수 있는 mRNA 전달 방법 및 조성물의 예. 본원에 참조로 포함되는 추가의 개시내용은 문헌 [Kowalski et al., "Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery," Mol Therap., 2019; 27(4): 710-728]에서 찾아볼 수 있다.

[1191] 다중-플랩 프라임 편집제를 코딩하는 DNA 벡터와 대조적으로, 다중-플랩 프라임 편집제를 위한 전달 작용제로서의 RNA의 사용은 유전 물질이 그의 기능을 수행하기 위해 핵에 진입할 필요가 없다는 이점을 갖는다. 전달된 mRNA는 세포질에서 목적하는 단백질 (예를 들어, 프라임 편집제 융합 단백질) 및 핵산 생성물 (예를 들어, PEgRNA)로 직접 번역될 수 있다. 그러나, 보다 안정하기 위해 (예를 들어, 세포질에서 RNA-분해 효소에 저항하기 위해), 일부 실시양태에서 mRNA를 안정화시켜 전달 효율을 개선시키는 것이 필요하다. 특정 전달 담체, 예컨대 양이온성 지질 또는 중합체성 전달 담체도 또한, 목적하는 프라임 편집제 융합 단백질을 코딩하는 치료

mRNA를 달리 분해할 수 있는 내인성 RNase 효소로부터 형질감염된 mRNA를 보호하는 것을 도울 수 있다. 또한, 변형된 mRNA의 증가된 안정성에도 불구하고, mRNA, 특히 전장 단백질을 코딩하는 mRNA를 세포에 치료 수준의 단백질 생산을 가능하게 하는 방식으로 생체내 전달하는 것은 도전과제로 남아있다.

[1192] 일부 예외로, mRNA의 세포내 전달은 일반적으로 작은 올리고뉴클레오티드의 세포내 전달보다 더 큰 도전과제이며, 부분적으로 다른 유형의 RNA (소형 간섭 RNA [siRNA], ~14 kDa; 안티센스 올리고뉴클레오티드 [ASO], 4-10 kDa)와 비교하여 mRNA 분자의 유의하게 더 큰 크기 (300-5,000 kDa, ~1-15 kb)로 인해 전달 나노입자 내로 캡슐화하는 것을 필요로 한다.

[1193] mRNA는 세포질에 도달하기 위해 세포막을 통과해야 한다. 세포막은 세포내 전달에 대한 동적이고 강력한 장벽이다. 이는 주로 쯔비터이온성 및 음으로 하전된 인지질의 지질 이중층으로 구성되며, 여기서 인지질의 극성 머리는 수성 환경을 향하고, 소수성 테일은 소수성 코어를 형성한다.

[1194] 일부 실시양태에서, 개시내용의 mRNA 조성물은 mRNA (프라임 편집제 및/또는 PEgRNA를 코딩함), 수송 비히클, 및 임의로 표적 세포와의 접촉 및 후속 형질감염을 용이하게 하는 작용제를 포함한다.

[1195] 일부 실시양태에서, mRNA는 (예를 들어, mRNA의 야생형 또는 천연 버전과 비교하여) mRNA에 안정성을 부여하고 단백질의 연관된 비정상적 발현에 수반되는 1개 이상의 변형을 포함할 수 있다. 결함을 교정하는 야생형에 대한 1개 이상의 변형이 또한 포함될 수 있다. 예를 들어, 본 발명의 핵산은 5' 비번역 영역 또는 3' 비번역 영역 중 하나 또는 둘 다의 변형을 포함할 수 있다. 이러한 변형은 시토메갈로바이러스 (CMV) 극초기 1 (IE1) 유전자, 폴리 A 테일, Cap1 구조, 또는 인간 성장 호르몬 (hGH)의 부분 서열을 코딩하는 서열의 포함을 포함할 수 있다. 일부 실시양태에서, mRNA는 mRNA 면역원성을 감소시키도록 변형된다.

[1196] 한 실시양태에서, 본 발명의 조성물 중 "다중-플랩 프라임 편집제" mRNA는 표적 세포로의 전달을 용이하게 하기 위해 리포솜 전달 비히클 중에 제제화될 수 있다. 고려되는 전달 비히클은 1종 이상의 양이온성 지질, 비-양이온성 지질, 및/또는 PEG-변형된 지질을 포함할 수 있다. 예를 들어, 전달 비히클은 하기 양이온성 지질 중 적어도 1종을 포함할 수 있다: C12-200, DLin-KC2-DMA, DODAP, HGT4003, ICE, HGT5000, 또는 HGT5001. 실시양태에서, 전달 비히클은 콜레스테롤 (chol) 및/또는 PEG 변형된 지질을 포함한다. 일부 실시양태에서, 전달 비히클은 DMG-PEG2K를 포함한다. 특정 실시양태에서, 전달 비히클은 하기 지질 제제를 갖는다: C12-200, DOPE, chol, DMG-PEG2K; DODAP, DOPE, 콜레스테롤, DMG-PEG2K; HGT5000, DOPE, chol, DMG-PEG2K, HGT5001, DOPE, chol, DMG-PEG2K 중 1종.

[1197] 본 개시내용은 1개 이상의 PE-코딩 mRNA 분자로의 표적 세포의 형질감염을 용이하게 하는데 유용한 조성물 및 방법을 또한 제공한다. 예를 들어, 본 발명의 조성물 및 방법은 1개 이상의 표적 세포에 대한 조성물의 친화도를 증가시킬 수 있는 표적화 리간드의 사용을 고려한다. 한 실시양태에서, 표적화 리간드는 아포지단백질 B 또는 아포지단백질 E이고, 상응하는 표적 세포는 저밀도 지단백질 수용체를 발현하고, 따라서 표적화 리간드의 인식을 촉진한다. 매우 많은 표적 세포가 본 개시내용의 방법 및 조성물을 사용하여 우선적으로 표적화될 수 있다. 예를 들어, 고려되는 표적 세포는 간세포, 상피 세포, 조혈 세포, 상피 세포, 내피 세포, 폐 세포, 골 세포, 줄기 세포, 중간엽 세포, 신경 세포, 심장 세포, 지방세포, 혈관 평활근 세포, 심근세포, 골격근 세포, 베타 세포, 뇌하수체 세포, 활막 내층 세포, 난소 세포, 고환 세포, 섬유모세포, B 세포, T 세포, 망상적혈구, 백혈구, 과립구, 및 종양 세포를 포함한다. 그러나, 이들에 제한되지는 않는다.

[1198] 일부 실시양태에서, PE-코딩 mRNA는 임의로, 예를 들어 이러한 mRNA의 안정성 및/또는 반감기를 개선시키거나 또는 단백질 생산을 개선시키거나 또는 달리 용이하게 하는 화학적 또는 생물학적 변형을 가질 수 있다. 형질감염 시, 본 발명의 조성물 중 천연 mRNA는 30분 내지 수일의 반감기로 붕괴될 수 있다. 개시내용의 조성물 중 mRNA는 적어도 일부의 번역되는 능력을 보유함으로써 기능적 단백질 또는 효소를 생산할 수 있다. 따라서, 본 발명은 안정화된 mRNA를 포함하는 조성물 및 그의 투여 방법을 제공한다. 일부 실시양태에서, mRNA의 활성은 연장된 기간에 걸쳐 지속된다. 예를 들어, mRNA의 활성은 본 개시내용의 조성물이 대상체에게 주 2회 또는 격주 기준으로, 또는 보다 바람직하게는 매월, 격월, 분기마다 또는 연간 기준으로 투여되도록 지속될 수 있다. 본 발명의 mRNA의 연장된 또는 지속된 활성은 이러한 mRNA로부터 생산된 단백질 또는 효소의 양에 직접적으로 관련된다. 유사하게, 본 개시내용의 조성물의 활성은 mRNA의 번역을 개선 또는 증진시키기 위해 이루어진 변형에 의해 추가로 연장 또는 지속될 수 있다. 또한, 표적 세포에 의해 생산되는 기능적 단백질 또는 효소의 양은 표적 세포에 전달되는 mRNA의 양 및 이러한 mRNA의 안정성의 함수이다. 본 발명의 mRNA의 안정성이 개선되거나 증진될 수 있는 정도로, 생산된 단백질 또는 효소의 반감기, 활성 및 조성물의 투여 빈도가 추가로 연장될 수 있다.

[1199] 따라서, 일부 실시양태에서, 개시내용의 조성물 중의 mRNA는, 예를 들어 생체내 뉴클레아제 소화에 대한 개선된 저항성을 포함한, 핵산에 증가된 또는 증진된 안정성을 부여하는 적어도 1개의 변형을 포함한다. 본원에 사용된 용어 "변형" 및 "변형된"은, 이러한 용어가 본원에 제공된 핵산에 관한 것일 때, 바람직하게는 안정성을 증진시키고 mRNA를 야생형 또는 자연 발생 버전의 mRNA보다 더 안정하게 (예를 들어, 뉴클레아제 소화에 대해 저항성이게) 하는 적어도 1개의 변경을 포함한다. 본원에 사용된 용어 "안정한" 및 "안정성"은, 이러한 용어가 본 발명의 핵산에 관한 것일 때, 특히 mRNA와 관련한 것일 때, 예를 들어 이러한 mRNA를 정상적으로 분해할 수 있는 뉴클레아제 (즉, 엔도뉴클레아제 또는 엑소뉴클레아제)에 의한 분해에 대한 증가된 또는 증진된 저항성을 지칭한다. 증가된 안정성은, 예를 들어, 가수분해 또는 표적 세포 또는 조직 내의 내인성 효소 (예를 들어, 엔도뉴클레아제 또는 엑소뉴클레아제) 또는 조건에 의한 다른 파괴에 대한 보다 낮은 감수성을 포함할 수 있으며, 그에 의해 표적 세포, 조직, 대상체 및/또는 세포질 내에서 이러한 mRNA의 체류가 증가 또는 증진된다. 본원에 제공된 안정화된 mRNA 분자는 그의 자연 발생, 비변형된 대응물 (예를 들어 mRNA의 야생형 버전)에 비해 더 긴 반감기를 입증한다. 또한, 용어 "변형" 및 "변형된"은, 이러한 용어가 본 발명의 mRNA에 관한 것일 때, 예를 들어 단백질 번역의 개시 시 기능하는 서열 (예를 들어, 코작(Kozak) 컨센서스 서열)의 포함을 포함한, mRNA 핵산의 번역을 개선 또는 증진시키는 변경이 고려된다. (Kozak, M., Nucleic Acids Res 15 (20): 8125-48 (1987)).

[1200] 일부 실시양태에서, 개시내용의 조성물에 사용된 mRNA는 이를 보다 안정하게 하는 화학적 또는 생물학적 변형을 거쳤다. mRNA에 대한 예시적인 변형은 염기의 고갈 (예를 들어, 결실에 의하거나 또는 하나의 뉴클레오티드로 또 다른 뉴클레오티드를 치환하는 것에 의함) 또는 염기의 변형, 예를 들어 염기의 화학적 변형을 포함한다. 본원에 사용된 어구 "화학적 변형"은 자연 발생 mRNA에서 관찰되는 것과 상이한 화학을 도입하는 변형, 예를 들어 공유 변형, 예컨대 변형된 뉴클레오티드의 도입 (예를 들어, 뉴클레오티드 유사체, 또는 이러한 mRNA 분자에서 자연적으로 발견되지 않는 펜던트 기의 포함)을 포함한다.

[1201] 개시내용의 조성물에 사용되는 PE-코딩 mRNA 내로 혼입될 수 있는 다른 적합한 폴리뉴클레오티드 변형은 4'-티오-변형된 염기: 4'-티오-아데노신, 4'-티오-구아노신, 4'-티오-시티딘, 4'-티오-우리딘, 4'-티오-5-메틸-시티딘, 4'-티오-슈도우리딘, 및 4'-티오-2-티오우리딘, 피리딘-4-온 리보뉴클레오시드, 5-아자-우리딘, 2-티오-5-아자-우리딘, 2-티오우리딘, 4-티오-슈도우리딘, 2-티오-슈도우리딘, 5-히드록시우리딘, 3-메틸우리딘, 5-카르복시메틸-우리딘, 1-카르복시메틸-슈도우리딘, 5-프로피닐-우리딘, 1-프로피닐-슈도우리딘, 5-타우리노메틸우리딘, 1-타우리노메틸-슈도우리딘, 5-타우리노메틸-2-티오-우리딘, 1-타우리노메틸-4-티오-우리딘, 5-메틸-우리딘, 1-메틸-슈도우리딘, 4-티오-1-메틸-슈도우리딘, 2-티오-1-메틸-슈도우리딘, 1-메틸-1-데아자-슈도우리딘, 2-티오-1-메틸-1-데아자-슈도우리딘, 디히드로우리딘, 디히드로슈도우리딘, 2-티오-디히드로우리딘, 2-티오-디히드로슈도우리딘, 2-메톡시우리딘, 2-메톡시-4-티오-우리딘, 4-메톡시-슈도우리딘, 4-메톡시-2-티오-슈도우리딘, 5-아자-시티딘, 슈도이소시티딘, 3-메틸-시티딘, N4-아세틸시티딘, 5-포르밀시티딘, N4-메틸시티딘, 5-히드록시메틸시티딘, 1-메틸-슈도이소시티딘, 피롤로-시티딘, 피롤로-슈도이소시티딘, 2-티오-시티딘, 2-티오-5-메틸-시티딘, 4-티오-슈도이소시티딘, 4-티오-1-메틸-슈도이소시티딘, 4-티오-1-메틸-1-데아자-슈도이소시티딘, 1-메틸-1-데아자-슈도이소시티딘, 제뷸라린, 5-아자-제뷸라린, 5-메틸-제뷸라린, 5-아자-2-티오-제뷸라린, 2-티오-제뷸라린, 2-메톡시-시티딘, 2-메톡시-5-메틸-시티딘, 4-메톡시-슈도이소시티딘, 4-메톡시-1-메틸-슈도이소시티딘, 2-아미노퓨린, 2,6-디아미노퓨린, 7-데아자-아데닌, 7-데아자-8-아자-아데닌, 7-데아자-2-아미노퓨린, 7-데아자-8-아자-2-아미노퓨린, 7-데아자-2,6-디아미노퓨린, 7-데아자-8-아자-2,6-디아미노퓨린, 1-메틸아데노신, N6-메틸아데노신, N6-이소펜테닐아데노신, N6-(시스-히드록시이소펜테닐)아데노신, 2-메틸티오-N6-(시스-히드록시이소펜테닐)아데노신, N6-글리시닐카르바모일아데노신, N6-트레오닐카르바모일아데노신, 2-메틸티오-N6-트레오닐 카르바모일아데노신, N6,N6-디메틸아데노신, 7-메틸아데닌, 2-메틸티오-아데닌, 및 2-메톡시-아데닌, 이노신, 1-메틸-이노신, 와이오신, 와이부토신, 7-데아자-구아노신, 7-데아자-8-아자-구아노신, 6-티오-구아노신, 6-티오-7-데아자-구아노신, 6-티오-7-데아자-8-아자-구아노신, 7-메틸-구아노신, 6-티오-7-메틸-구아노신, 7-메틸이노신, 6-메톡시-구아노신, 1-메틸구아노신, N2-메틸구아노신, N2,N2-디메틸구아노신, 8-옥소-구아노신, 7-메틸-8-옥소-구아노신, 1-메틸-6-티오-구아노신, N2-메틸-6-티오-구아노신, 및 N2,N2-디메틸-6-티오-구아노신, 및 그의 조합을 포함하나 이에 제한되지는 않는다. 용어 변형은 또한, 예를 들어 비-뉴클레오티드 연결 또는 변형된 뉴클레오티드의 본 발명의 mRNA 서열 내로의 혼입 (예를 들어, 기능적 단백질 또는 효소를 코딩하는 mRNA 분자의 3' 및 5' 단부 중 하나 또는 둘 다에 대한 변형)을 포함한다. 이러한 변형은 mRNA 서열에 대한 염기의 부가 (예를 들어, 폴리 A 테일 또는 보다 긴 폴리 A 테일의 포함), 3' UTR 또는 5' UTR의 변경, mRNA와 작용제 (예를 들어, 단백질 또는 상보적 핵산 분자)의 복합체화, 및 mRNA 분자의 구조를 변화시키는 (예를 들어, 2차 구조를 형성하는) 요소의 포함을 포함한다.

[1202] 일부 실시양태에서, PE-코딩 mRNA는 5' 캡 구조를 포함한다. 5' 캡은 전형적으로 하기와 같이 부가된다: 먼저, RNA 말단 포스파타제는 5' 뉴클레오티드로부터 말단 포스페이트 기 중 하나를 제거하여, 2개의 말단 포스페이트를 남기고; 구아노신 트리포스페이트 (GTP)가 이어서 구아닐릴 트랜스퍼라제를 통해 말단 포스페이트에 부가되어, 5'5'5 트리포스페이트 연결을 생성하고; 구아닌의 7-질소는 이어서 메틸트랜스퍼라제에 의해 메틸화된다. 캡 구조의 예는 m7G(5')ppp (5'(A,G(5')ppp(5')A 및 G(5')ppp(5')G를 포함하나 이에 제한되지는 않는다. 자연 발생 캡 구조는 제1 전사된 뉴클레오티드의 5'-단부에 트리포스페이트 가교를 통해 연결되어 m7G(5')ppp(5')N의 디뉴클레오티드 캡을 생성하는 7-메틸 구아노신을 포함하며, 여기서 N은 임의의 뉴클레오시드이다. 생체내에서, 캡은 효소적으로 부가된다. 캡은 핵에 부가되고, 효소 구아닐릴 트랜스퍼라제에 의해 촉매된다. RNA의 5' 말단 단부에 대한 캡의 부가는 전사 개시 직후에 일어난다. 말단 뉴클레오시드는 전형적으로 구아노신이고, 모든 다른 뉴클레오티드, 즉 G(5')ppp(5')GpNpNp에 대해 역 배향이다.

[1203] 추가의 캡 유사체는 m7GpppG, m7GpppA, m7GpppC; 비메틸화 캡 유사체 (예를 들어, GpppG); 디메틸화 캡 유사체 (예를 들어, m2,7GpppG), 트리메틸화 캡 유사체 (예를 들어, m2,2,7GpppG), 디메틸화 대칭 캡 유사체 (예를 들어, m7Gpppm7G), 또는 항 역 캡 유사체 (예를 들어, ARCA; m7,2'OmeGpppG, m72'dGpppG, m7,3'OmeGpppG, m7,3'dGpppG 및 그의 테트라포스페이트 유도체)로 이루어진 군으로부터 선택된 화학 구조를 포함하나 이에 제한되지는 않는다 (예를 들어, 문헌 [Jemielity, J. et al., "Novel 'anti-reverse' cap analogs with superior translational properties", RNA, 9: 1108-1122 (2003)] 참조).

[1204] 전형적으로, "테일"의 존재는 mRNA를 엑소뉴클레아제 분해로부터 보호하는 역할을 한다. 폴리 A 또는 폴리 U 테일은 천연 메신저 및 합성 센스 RNA를 안정화시키는 것으로 생각된다. 따라서, 특정 실시양태에서, 긴 폴리 A 또는 폴리 U 테일을 mRNA 분자에 부가하여 RNA를 보다 안정하게 할 수 있다. 폴리 A 또는 폴리 U 테일은 다양한 관련 기술분야-인식 기술을 사용하여 부가될 수 있다. 예를 들어, 긴 폴리 A 테일은 폴리 A 폴리머라제를 사용하여 합성 또는 시험관내 전사된 RNA에 부가될 수 있다 (Yokoe, et al., Nature Biotechnology. 1996; 14: 1252-1256). 전사 벡터는 또한 긴 폴리 A 테일을 코딩할 수 있다. 또한, 폴리 A 테일은 PCR 생성물로부터 직접 전사에 의해 부가될 수 있다. 폴리 A는 또한 RNA 리가제에 의해 센스 RNA의 3' 단부에 라이게이션될 수 있다 (예를 들어, 문헌 [Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1991 edition)] 참조).

[1205] 전형적으로, 폴리 A 또는 폴리 U 테일의 길이는 적어도 약 10, 50, 100, 200, 300, 400, 적어도 500개 뉴클레오티드일 수 있다. 일부 실시양태에서, mRNA의 3' 말단 상의 폴리-A 테일은 전형적으로 약 10 내지 300개의 아데노신 뉴클레오티드 (예를 들어, 약 10 내지 200개의 아데노신 뉴클레오티드, 약 10 내지 150개의 아데노신 뉴클레오티드, 약 10 내지 100개의 아데노신 뉴클레오티드, 약 20 내지 70개의 아데노신 뉴클레오티드, 또는 약 20 내지 60개의 아데노신 뉴클레오티드)를 포함한다. 일부 실시양태에서, mRNA는 3' 폴리(C) 테일 구조를 포함한다. mRNA의 3' 말단 상의 적합한 폴리-C 테일은 전형적으로 약 10 내지 200개의 시토신 뉴클레오티드 (예를 들어, 약 10 내지 150개의 시토신 뉴클레오티드, 약 10 내지 100개의 시토신 뉴클레오티드, 약 20 내지 70개의 시토신 뉴클레오티드, 약 20 내지 60개의 시토신 뉴클레오티드, 또는 약 10 내지 40개의 시토신 뉴클레오티드)를 포함한다. 폴리-C 테일은 폴리-A 또는 폴리 U 테일에 부가될 수 있거나, 또는 폴리-A 또는 폴리 U 테일을 치환할 수 있다.

[1206] 본 개시내용에 따른 PE-코딩 mRNA는 임의의 다양한 공지된 방법에 따라 합성될 수 있다. 예를 들어, 본 발명에 따른 mRNA는 시험관내 전사 (IVT)를 통해 합성될 수 있다. 간략하게, IVT는 전형적으로 프로모터, 리보뉴클레오티드 트리포스페이트의 풀, DTT 및 마그네슘 이온을 포함할 수 있는 완충제 시스템, 및 적절한 RNA 폴리머라제 (예를 들어, T3, T7 또는 SP6 RNA 폴리머라제), DNAse I, 피로포스파타제, 및/또는 RNAse 억제제를 함유하는 선형 또는 원형 DNA 주형으로 수행된다. 정확한 조건은 특정한 적용에 따라 달라질 것이다.

[1207] mRNA 전달을 수반하는 실시양태에서, PE 융합 단백질을 코딩하는 mRNA 대 PEgRNA의 비는 효율적인 편집에 중요할 수 있다. 특정 실시양태에서, mRNA (PE 융합 단백질을 코딩함) 대 PEgRNA의 중량비는 1:1이다. 특정의 다른 실시양태에서, mRNA (PE 융합 단백질을 코딩함) 대 PEgRNA의 중량비는 2:1이다. 또 다른 실시양태에서, mRNA (PE 융합 단백질을 코딩함) 대 PEgRNA의 중량비는 1:2이다. 추가 실시양태에서, mRNA (PE 융합 단백질을 코딩함) 대 PEgRNA의 중량비는 약 1:1000, 1:900; 1:800; 1:700; 1:600; 1:500; 1:400; 1:300; 1:200; 1:100; 1:90; 1:80; 1:70; 1:60; 1:50; 1:40; 1:30; 1:20; 1:10; 및 1:1로 이루어진 군으로부터 선택된다. 다른 실시양태에서, mRNA (PE 융합 단백질을 코딩함) 대 PEgRNA의 중량비는 약 1:1000, 1:900; 800:1; 700:1; 600:1; 500:1; 400:1; 300:1; 200:1; 100:1; 90:1; 80:1; 70:1; 60:1; 50:1; 40:1; 30:1; 20:1; 10:1; 및 1:1로 이루

어진 군으로부터 선택된다.

[1208] F. 유도성 이량체화 도메인의 삽입을 위한 다중-플랩 프라임 편집의 사용

[1209] 본원에 기재된 다중-플랩 프라임 편집제 (도 90, 95, 97a, 98a에 의해 나타내어진 바와 같음)는 또한 이량체화 도메인을 1개 이상의 단백질 표적에 설치하는 데 사용될 수 있다. 이량체화 도메인은 이중특이적 방식으로 결합하는 연결 모이어티 (예를 들어, 소분자, 펩티드, 또는 단백질)를 통해 1개 이상의 단백질 표적의 이량체화와 연관된 활성의 유도성 조절을 용이하게 할 수 있다. 다양한 측면에서, 이량체화 도메인은 상이한 단백질 (예를 들어, 동일한 유형 또는 상이한 단백질) 상에 설치될 경우, 각각 동일한 이중특이적 모이어티 (예를 들어, 이량체화 도메인에 개별적으로 결합하는 적어도 2개의 영역을 갖는 이중특이적 소분자, 펩티드, 또는 폴리펩티드)에 결합하여, 이중특이적 리간드에 대한 그의 공통 상호작용을 통해 단백질의 이량체화를 유발한다. 이러한 방식으로, 이중특이적 리간드는 2개의 단백질의 이량체화의 "유도제"로서 기능한다. 일부 경우에, 이중특이적 리간드 또는 화합물은 동일한 2개의 표적화 모이어티를 가질 것이다. 다른 실시양태에서, 이중특이적 리간드 또는 화합물은 각각 서로 상이한 표적화 모이어티를 가질 것이다. 동일한 2개의 표적화 모이어티를 갖는 이중특이적 리간드 또는 화합물은 상이한 단백질 표적 상에 설치된 동일한 이량체화 도메인을 표적화할 수 있을 것이다. 상이한 표적화 모이어티를 갖는 이중특이적 리간드 또는 화합물은 상이한 단백질 표적 상에 설치된 상이한 이량체화 도메인을 표적화할 수 있을 것이다.

[1210] 본원에 사용된 용어 "이량체화 도메인"은 이중특이적 리간드의 결합 모이어티에 결합하는 리간드-결합 도메인을 지칭한다. "제1" 이량체화 도메인은 이중특이적 리간드의 제1 결합 모이어티에 결합하고, "제2" 이량체화 도메인은 동일한 이중특이적 리간드의 제2 결합 모이어티에 결합한다. 제1 이량체화 도메인이 (예를 들어, 본원에 논의된 바와 같은 PE를 통해) 제1 단백질에 융합되고, 제2 이량체화 도메인이 (예를 들어, 본원에 논의된 바와 같은 PE를 통해) 제2 단백질에 융합되는 경우에, 제1 및 제2 단백질은 이중특이적 리간드의 존재 하에 이량체화되고, 여기서 이중특이적 리간드는 제1 이량체화 도메인에 결합하는 적어도 1개의 모이어티 및 제2 이량체화 도메인에 결합하는 적어도 또 다른 모이어티를 갖는다.

[1211] 본원에 사용된 용어 "이중특이적 리간드" 또는 "이중특이적 모이어티"는 2개의 상이한 리간드-결합 도메인에 결합하는 리간드를 지칭한다. 다양한 실시양태에서, 이중특이적 모이어티 자체는 2개의 동일한 또는 2개의 상이한 화학적 모이어티의 이량체이고, 여기서 각각의 모이어티는 이량체화 도메인에 특이적으로 단단히 결합한다. 특정 실시양태에서, 리간드는 소분자 화합물, 또는 펩티드, 또는 폴리펩티드이다. 다른 실시양태에서, 리간드-결합 도메인은 펩티드 태그로서 단백질 상에 설치될 수 있는 "이량체화 도메인"이다. 다양한 실시양태에서, 각각 동일하거나 상이한 이량체화 도메인을 포함하는 2개의 단백질은 이중특이적 리간드에 대한 각각의 이량체화 도메인의 결합을 통해 이량체화되도록 유도될 수 있다. 이들 분자는 또한 "이량체화의 화학적 유도제" 또는 CID로 지칭될 수 있다. 또한, 이중특이적 리간드는 2개의 동일한 모이어티를 함께, 또는 2개의 상이한 모이어티를 함께 커플링시킴으로써 (예를 들어, 표준화 화학적 연결을 통함) 제조될 수 있으며, 여기서 각각의 모이어티는 이량체화 도메인에 단단히 특이적으로 결합한다.

[1212] 다양한 측면에서, 다중-플랩 PE에 의해 설치된 이량체화 도메인은 동일하거나 상이할 수 있다.

[1213] 예를 들어, 이량체화 도메인은 하기 아미노산 서열을 갖는 FKBP12일 수 있다:

[1214]

FKBP12	MGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE (SEQ ID NO: 488)

[1215] 또 다른 예에서, 이량체화 도메인은 합성 범핑된 FK506 모방체 (2, 도 3)에 결합하는 조작된 홀을 갖는 FKBP12의 돌연변이체인, FKBP12-F36V로 지칭되는 FKBP12의 돌연변이체일 수 있다[107]:

[1216]

FKBP12-F36V	MGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKVDSSRDRNKPFKFMLGKQEVIRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE (SEQ ID NO: 489)

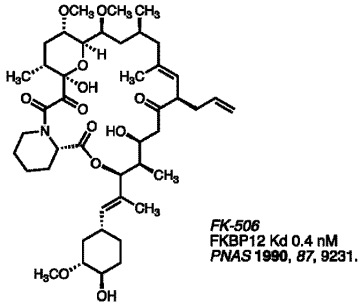
[1217] 또 다른 예에서, 이량체화 도메인은 하기와 같은 시클로필린일 수 있다:

시클로필린	MVNPTVFFDIAVDGEPLGRVSFELFADKVP KTAENFRALSTGEKGFGYKGSCFHRIIPGF MCQGGDFTRHNGTGGKSIYGEKFEDENFI LKHTGPGILSMANAGPNTNGSQFFICTAKT EWLDGKHVVFGKVKEGMNIVEAMERFGS RNGKTSKKITIADCGQLE (SEQ ID NO: 490)

[1218]

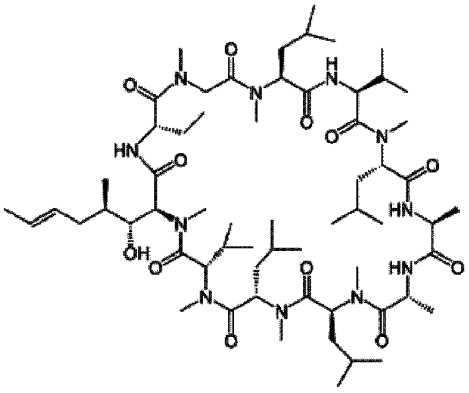
[1219] 다양한 실시양태에서, 이들 이량체화 도메인의 아미노산 서열은 결합을 최적화하거나 또는 천연 표적에 대한 결합 직교성을 개선시키기 위해 변경될 수 있다. 소분자 결합 단백질을 코딩하는 유전자의 핵산 서열은 PE 과정의 효율을 최적화하기 위해, 예컨대 PEgRNA 2차 구조를 감소시킴으로써 변경될 수 있다.

[1220] 적합한 이량체화 도메인 및 그에 결합하는 동족 소분자 화합물의 다른 예가 하기와 같이 제공된다. 동족 소분자 화합물은 2개의 이량체화 도메인에 결합할 수 있는 이중특이적 리간드를 형성하기 위해 제2 소분자 화합물(동일한 화합물 또는 상이한 화합물)에 (예를 들어, 화학적 링커를 통해) 커플링될 수 있음에 주목한다. 일부 경우, 예컨대 FK506 및 시클로스포린 A에서, 각각의 이량체화 (예를 들어, FK506-FK506 또는 시클로스포린 A-시클로스포린 A)는 단량체 화합물의 면역억제 활성을 감소시키거나 제거한다.

<table>
<tr><th>소분자 - 이량체화 도메인에 결합함 - 이들 분자의 이량체는 2개의 이량체화 도메인에 결합할 이중특이적 리간드를 구성할 것이다</th><th>이량체화 도메인(들)</th></tr>
<tr><td>FK506


표적: FKBP12 + 칼시뉴린</td><td>FKBP12의 아미노산 서열:

MGVQVETISPGDGRTFPKRGQTCVVH YTGMLEDGKKFDSSRDRNKPFKFMLG KQEVIRGWEEGVAQMSVGQRAKLTISP DYAYGATGHPGIIPPHATLVFDVELLKL E (SEQ ID NO: 491)

칼시뉴린:</td></tr>
<tr><td>시클로스포린 A

표적: 시클로필린 + 칼시뉴린</td><td>인간 시클로필린 A의 아미노산 서열:

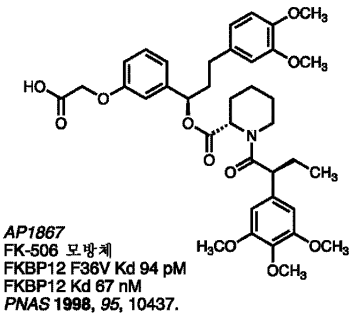
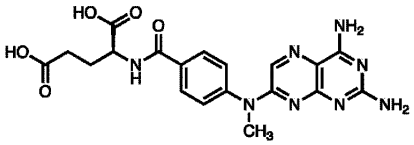
MVNPTVFFDIAVDGEPLGRVSFELF ADKVPKTAENFRALSTGEKGFGYK GSCFHRIIPGFMCQGGDFTRHNGTG GKSIYGEKFEDENFILKHTGPGILSM ANAGPNTNGSQFFICTAKTEWLDG KHVVFGKVKEGMNIVEAMERFGS RNGKTSKKITIADCGQLE (SEQ ID NO: 490)

인간 시클로필린 B의 아미노산 서열:

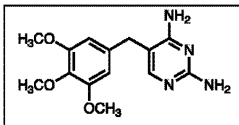
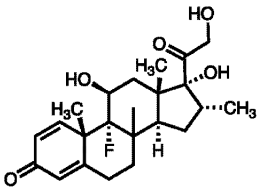
MLRLSERNMKVLLAAALIAGSVFF LLLPGPSAADEKKKGPKVTVKVYF DLRIGDEDVGRVIFGLFGKTVPKTV DNFVALATGEKGFGYKNSKFHRVIK DFMIQGGDFTRGDGTGGKSIYGER FPDENFKLKHYGPGWVSMANAGK DTNGSQFFITTVKTAWLDGKHVVF GKVLEGMEVVRKVESTKTDSRDKP LKDVIIADCGKIEVEKPFAIAKE (SEQ ID NO: 493)

뮤린 시클로필린 C의 아미노산 서열:</td></tr>
</table>

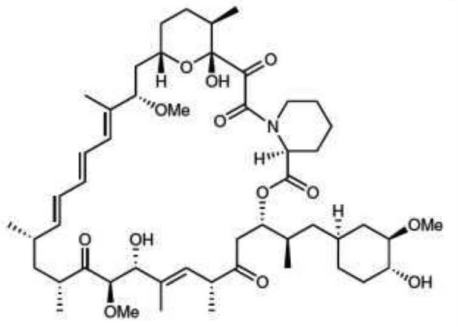
[1221]

	MSPGPRLLLPAVLCLGLGALVSSSG SSGVRKRGPSVTDKVFFDVRIGDK DVGRIVIGLFGNVVPKTVENFVALA TGEKGYGYKGSIFHRVIKDFMIQGG DFTARDGTGGMSIYGETFPDENFKL KHYGIGWVSMANAGPDTNGSQFFI TLTKPTWLDGKHVVFGKVLDGMT VVHSIELQATDGHDRPLTDCTIVNS GKIDVKTPFVVEVPDW (SEQ ID NO: 494)
AP1867  *AP1867* FK-506 모방체 FKBP12 F36V Kd 94 pM FKBP12 Kd 67 nM *PNAS* **1998**, *95*, 10437. 표적(들): FKBP12	FKBP12-F36V의 아미노산 서열 MGVQVETISPGDGRTFPKRGQTCVVH YTGMLEDGKKVDSSRDRNKPFKFMLG KQEVIRGWEEGVAQMSVGQRAKLT ISPDYAYGATGHPGIIPPHATLVFDVELL KLE (SEQ ID NO: 495)
메토트렉세이트  메토트렉세이트 인간 DHFR Kd < 10 nM *J. Biol. Chem.* **1988**, *263*, 10304. 이.콜라이DHFR Kd 9.5 nM *PNAS* **2002**, *99*, 13481. 표적(들): 디히드로폴레이트 리덕타제	인간 디히드로폴레이트 리덕타제의 아미노산 서열 MVGSLNCIVAVSQNMGIGKNGDLPWP PLRNEFRYFQRMTTTSSVEGKQNLVIM GKKTWFSIPEKNRPLKGRINLVLSRELK EPPQGAHFLSRSLDDALKLTEQPELAN KVDMVWIVGGSSVYKEAMNHPGHLK LFVTRIMQDFESDTFFPEIDLEKYKLLP EYPGVLSDVQEEKGIKYKFEVYEKND (SEQ ID NO: 496)
트리메토프림	이. 콜라이 디히드로폴레이트 리덕타제에 대한 아미노산 서열

[1222]

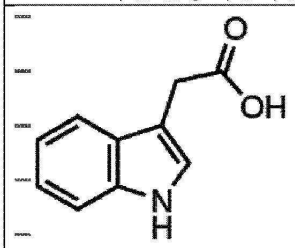
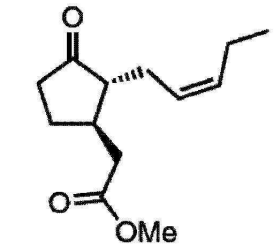
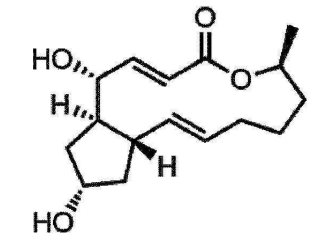
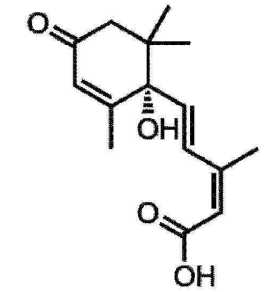
 트리메토프림 이.콜라이DHFR K_i 1.3 nM *Biochemistry* **1982**, *21*, 5068. 표적(들): 디히드로폴레이트 리덕타제	MISLIAALAVDRVIGMENAMPWNLPAD LAWFKRNTLNKPVIMGRHTWESIGRPL PGRKNIILSSQPGTDDRVTWVKSVDEAI AACGDVPEIMVIGGGRVYEQFLPKAQ KLYLTHIDAEVEGDTHFPDYEPDDWES VFSEFHDADAQNSHSYCFEILERR (SEQ ID NO: 497)
덱사메타손  덱사메타손 인간 GR Kd 4.6 nM *Mol. Endocrin.* **1999**, *13*, 1855.	인간 글루코코르티코이드 수용체에 대한 아미노산 서열 MDSKESLTPGREENPSSVLAQERGDV MDFYKTLRGGATVKVSASSPSLAVASQ SDSKQRRLLVDFPKGSVSNAQQPDLSK AVSLSMGLYMGETETKVMGNDLGFPQ QGQISLSSGETDLKLLEESIANLNRSTS VPENPKSSASTAVSAAPTEKEF PKTHSDVSSEQQHLKGQTGTNGGNVK LYTTDQSTFDILQDLEFSSGSPGKETNE SPWRSDLLIDENCLLSPLAGEDDSFLLE GNSNEDCKPLILPDTKPKIKDNGDLVLS SPSNVTLPQVKTEKEDFIELCTPGVIKQ EKLGTVYCQASFPGANIIGNKMSAISV HGVSTSGGQMYHYDMNTASLSQQQD QKPIFNVIPPIPVGSENWNRCQGSGDD NLTSLGTLNFPGRTVFSNGYSSPSMRPD VSSPPSSSSTATTGPPPKLCLVCSDEASG CHYGVLTCGSCKVFFKRAVEGQHNYL CAGRNDCIIDKIRRKNCPACRYRKCLQ AGMNLEARKTKKKIKGIQQATTGVSQ ETSENPGNKTIVPATLPQLTPTLVSLLEV IEPEVLYAGYDSSVPDSTWRIMTTLNM LGGRQVIAAVKWAKAIPGFRNLHLDD QMTLLQYSWMFLMAFALGWRSYRQS SANLLCFAPDLIINEQRMTLPCMYDQC KHMLYVSSELHRLQVSYEEYLCMKTL LLLSSVPKDGLKSQELFDEIRMTYIKEL GKAIVKREGNSSQNWQRFYQLTKLLD SMHEVVENLLNYCFQTFLDKTMSIEFP EMLAEIITNQIPKYSNGNIKKLLFHQK (SEQ ID NO: 498)

[1223]

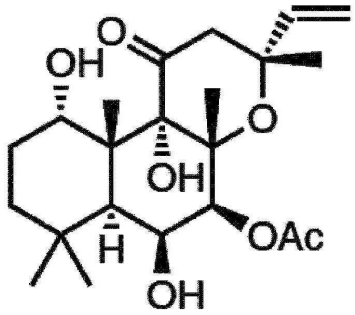
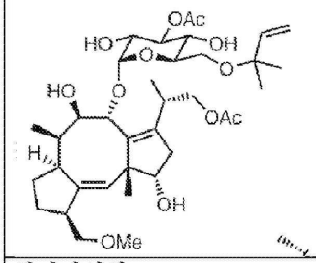
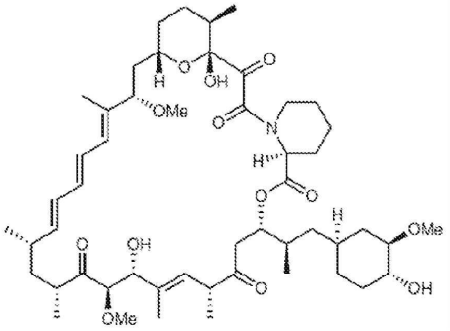
 라파마이신	FKBP12의 아미노산 서열: MGVQVETISPGDGRTFPKRGQTCVVH YTGMLEDGKKFDSSRDRNKPFKFMLG KQEVIRGWEEGVAQMSVGQRAKLTISP DYAYGATGHPGIIPPHATLVFDVELLKL E (SEQ ID NO: 491)

[1224]

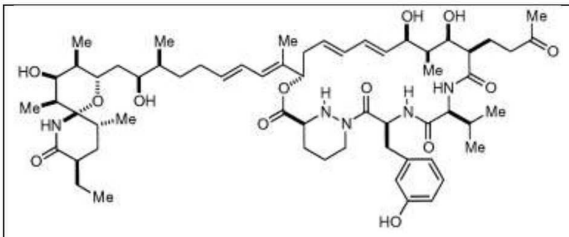
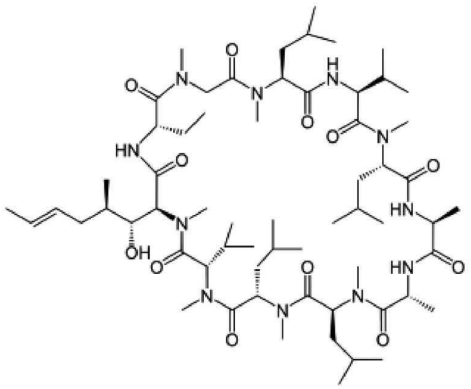
[1225] 자연 발생 이중기능적 분자 및 그의 이중 표적 수용체의 다른 예는 하기와 같다. 프라임 편집은 이중 표적 수용체를 상이한 단백질에 설치하는 데 사용될 수 있다. 일단 상이한 단백질이 이중기능적 분자 수용체를 함유하도록 PE에 의해 변형되면, 이중기능적 분자가 도입되어, 상이한 이량체화 도메인을 포함하도록 변형된 단백질의 이량체화가 유발될 수 있다. (1) 이중기능적 분자 및 (2) 그의 이중 표적 수용체의 쌍 형성의 예는 하기와 같다:

자연 발생 이중기능적 분자	이중기능적 분자의 표적 수용체
 옥신	표적 수용체 1: 옥신 수용체 표적 수용체 2: TIR1 E3 리가제
 메틸 자스모네이트	표적 수용체 1: JAZ 수용체 표적 수용체 2: Col1 E3 리가제
브레펠딘 A 	표적 수용체 1: GBF1 표적 수용체 2: GTPase Arf1p
아브시스산 	표적 수용체 1: PYR 수용체 표적 수용체 2: 인단백질 포스파타제 2C

[1226]

포르스콜린 	표적 수용체 1: 아데닐릴 시클라제 단량체 표적 수용체 2: 아데닐릴 시클라제 단량체
푸시콕신 A 	표적 수용체 1: 14-3-3 단백질 표적 수용체 2: H^+-ATPase
라파마이신 	표적 수용체 1: FKBP12 표적 수용체 2: mTOR
상글리페린 A	표적 수용체 1: 시클로필린 표적 수용체 2: IMP 데히드로게나제 2

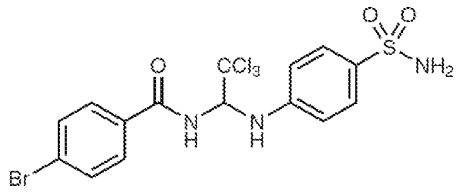
[1227]

 상글리페린 A (SFA)	
시클로스포린 A 	표적 수용체 1: 시클로필린 표적 수용체 2: 칼시뉴린

[1228]

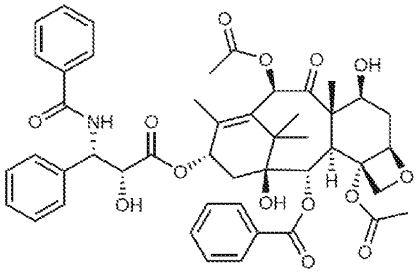
[1229] 프라임 편집의 이러한 측면에서 사용될 수 있는 다른 이중기능적 분자의 예는 하기와 같다:

[1230] 신스탑 A:



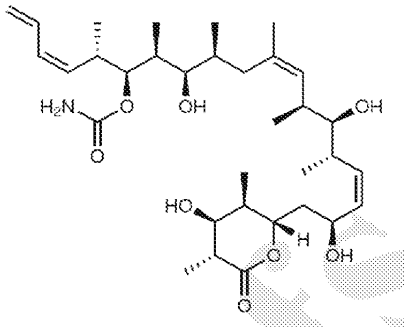
[1231]

[1232] 파클리탁셀:



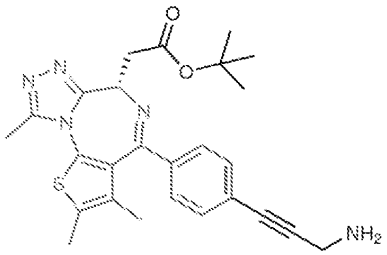
[1233]

[1234] 디스코데르몰리드:



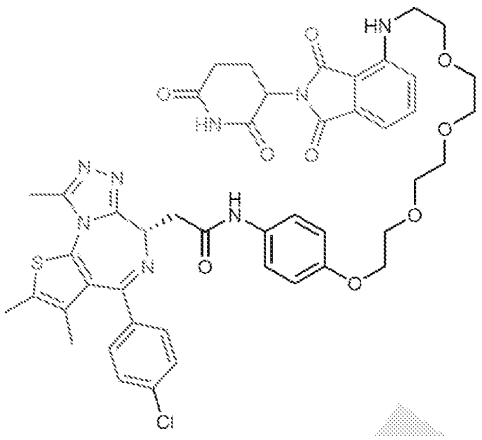
[1235]

[1236] GNE-0011



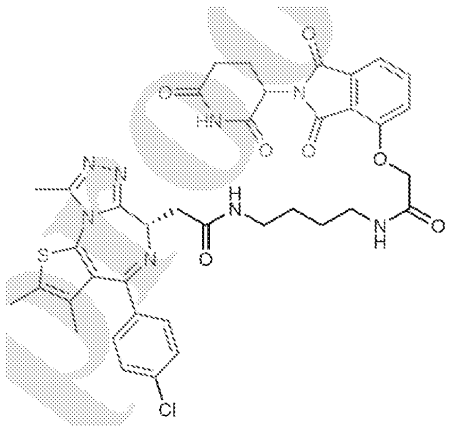
[1237]

[1238] ARV-825, 및



[1239]

[1240] dBET1



[1241]

[1242] 신스탑 A, 파클리탁셀, 및 디스코데르몰리드는 미세관 안정화제이다. 따라서, 이들 화합물은 미세관 단백질을

포함하도록 PE에 의해 변형된 단백질을 이량체화하는 데 사용될 수 있다. GNE-0011, ARV-825, 및 dBET1은 BRD4 결합 모티프 및 CRBN 결합 모티프를 포함한다. 따라서, 이들 화합물은 이들 표적화 도메인을 포함하도록 PE에 의해 변형된 단백질을 이량체화하는 데 사용될 수 있다.

[1243] 이량체화 도메인을 설치하기 위한 PEgRNA는 하기 구조를 포함할 수 있다 (도 3d 참조):

[1244] 5'-[스페이서]-[gRNA 코어]-[연장 아암]-3', 여기서 연장 아암은 5'-[상동성 아암]-[편집 주형]-[프라이머 결합 부위]-3'를 포함함; 또는

[1245] 5'-[연장 아암]-[스페이서]-[gRNA 코어]-3', 여기서 연장 아암은 5'-[상동성 아암]-[편집 주형]-[프라이머 결합 부위]-3'를 포함하고, 여기서 어느 하나의 구성을 갖는 "편집 주형"은 이량체화 도메인의 뉴클레오티드 서열을 포함함.

[1246] 한 예에서, 인간 인슐린 수용체의 C-말단 단부에의 FKBP12 이량체화 도메인의 삽입을 위한 PEgRNA (스페이서는 밑줄표시되고, gRNA 코어는 특별한 표시가 없고, 플랩 상동성은 볼드체이고, FKBP12 삽입은 이탤릭체이고, 어닐링 영역은 볼드체 이탤릭체임):

인간 인슐린 수용체 내에 FKBP12를 설치하기 위한 PEGRNA	CACGGUAGGCACUGUUAGGAGUUUUAG AGCUAGAAAUAGCAAGUUAAAAUAAG GCUAGUCCGUUAUCAACUUGAAAAAGU GGCACCGAGUCGGUGC**UUGCCUCGGU CCAAUCCUUCC***GGAGUGCAGGUGGAAA CCAUCUCCCCAGGAGACGGGCGCACCUU CCCCAAGCGCGGCCAGACCUGCGUGGUG CACUACACCGGGAUGCUUGAAGAUGGAA AGAAAUUUGAUUCCUCCCGGGACAGAAA CAAGCCCUUUAAGUUUAUGCUAGGCAAG CAGGAGGUGAUCCGAGGCUGGGAAGAAG GGGUUGCCCAGAUGAGUGUGGGUCAGA GAGCCAAACUGACUAUAUCUCCAGAUUA UGCCUAUGGUGCCACUGGGCACCCAGGC AUCAUCCCACCACAUGCCACUCUCGUCU UCGAUGUGGAGCUUCUAAAACUGGAA**UA ACAGUGCCUACC*** (SEQ ID NO: 499)

[1247]

[1248] 또 다른 예에서, HEK3 유전자좌에의 FKBP12 이량체화 도메인의 삽입을 위한 PEgRNA (최적화를 위함):

HEK3 내에 FKBP12를 설치하기 위한 PEGRNA	GGCCCAGACTGAGCACGTGAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGCTGGAGGAAGCAGGG CTTCCTTTCCTCTGCCATCATTCCAGTTT TAGAAGCTCCACATCGAAGACGAGAGT GGCATGTGGTGGGATGATGCCTGGGTGC CCAGTGGCACCATAGGCATAATCTGGAG ATATAGTCAGTTTGGCTCTCTGACCCAC ACTCATCTGGGCAACCCCTTCTTCCCAG CCTCGGATCACCTCCTGCTTGCCTAGCA TAAACTTAAAGGGCTTGTTTCTGTCCCG GGAGGAATCAAATTTCTTTCCATCTTCA AGCATCCCGGTGTAGTGCACCACGCAGG TCTGGCCGCGCTTGGGGAAGGTGCGCCC GTCTCCTGGGGAGATGGTTTCCACCTGC ACTCCCGTGCTCAGTCTG (SEQ ID NO: 500)

[1249]

[1250] 이량체화 도메인을 설치하기 위한 표적 단백질은 특히 제한되지는 않지만; 이중특이적 리간드의 존재 하에서의 그의 이량체화가 (일단 PE에 의해 변형되면) 일부 유리한 생물학적 효과, 예를 들어 신호전달 경로, 감소된 면역반응성 등을 생성한다는 점에서 유리하다. 다양한 측면에서, 이량체화 도메인의 PE-의존성 설치를 통해 이량체화될 표적 단백질은 동일한 단백질 또는 상이한 단백질일 수 있다. 바람직하게는, 단백질은 이량체화되는 경우 1종 이상의 하류 생물학적 캐스케이드, 예를 들어, 신호 전달 캐스케이드, 인산화 등을 촉발한다. PE가 이

량체화 도메인을 설치하는 데 사용될 수 있는 예시적인 표적 단백질은 하기를 포함하나 이에 제한되지는 않는다:

막-결합된 수용체	하기에 융합된 키나제 도메인	사용된 CID	신호전달 캐스케이드	참조
T-세포 수용체	FKBP12	FK1012 (FK506 이량체)	T-세포 수용체 신호전달	*SCIENCE* 262, 1019-24 (1993) *CHEM. BIOL.* 1, 163-172 (1994).
FAS 수용체	뮤린 시클로필린 C	시클로스포린 A 이량체	아폽토시스를 위한 FAS 경로	*CHEM. BIOL.* 3, 731-738 (1996).
인슐린 수용체	FKBP12	FK1012 (FK506 이량체)	인슐린 신호전달	*CURR. BIOL.* 8, 11-18 (1998).
혈소판-유래 성장 인자 (PDGF) 베타	FKBP12	FK1012 (FK506 이량체)	PDGF 중배엽 형성 신호전달	*CURR. BIOL.* 8, 11-18 (1998).
에리트로포이에틴 수용체 (EPOR)	FKBP12	FK1012 (FK506 이량체)	EPOR-매개된 증식성 신호전달	*PROC. NATL. ACAD. SCI.* 94, 3076-3081 (2002).

[1251]

[1252] 한 측면에서, 본원에 기재된 다중-플랩 프라임 편집제 시스템은 살아있는 세포 또는 환자에서 관심 표적 단백질을 코딩하는 1개 이상의 유전자 내로 이량체화 도메인을 코딩하는 서열을 설치하는 데 사용될 수 있다. 이는 "다중-플랩 프라임 편집-CID 시스템"으로 지칭될 수 있으며, 여기서 CID는 각각 다중-플랩 PE에 의해 설치된 이량체화 도메인에 융합된 표적 단백질의 이량체화를 유도하는 이중특이적 리간드이다. 이러한 편집 단독은 생리학적 효과를 갖지 않아야 한다. 전형적으로 각각이 표적 단백질의 카피에 융합된 2개의 이량체화 도메인에 동시에 결합할 수 있는 이량체 소분자인 이중특이적 리간드의 투여 시, 이중특이적 리간드는 표적화된 단백질의 이량체화를 유발한다. 이어서, 이러한 표적 단백질 이량체화 사건은 생물학적 신호전달 사건, 예컨대 적혈구생성 또는 인슐린 신호전달을 유도한다. 다중-플랩 프라임 편집을 사용한 소분자 결합 단백질 (즉, 이량체화 도메인)을 코딩하는 유전자의 게놈 통합에 의해 편리한 소분자 약물 (즉, 이중특이적 리간드)의 제어 하에 이량체화-유도된 생물학적 과정, 예컨대 수용체 신호전달을 배치하는 신규 방법이 본원에 기재된다.

[1253] 단백질 이량체화는 편재하는 생물학적 과정이다. 특히, 많은 막-결합 수용체의 동종이량체화는 종종 막대한 생물학적 결과를 갖는 신호전달 캐스케이드를 개시하는 것으로 공지되어 있다. 약물로서 사용하도록 승인된 다수의 소분자 천연 생성물은 그의 작용 메카니즘의 일부로서 단백질 이량체화의 화학적 유도제로서 작용한다.[92] 예를 들어, FK506은 FKBP12에 단단히 결합하고, 생성된 소분자-단백질 복합체는 이어서 포스파타제 칼시뉴린에 결합하여, T 세포 수용체 신호전달의 단계를 억제한다.[93] 마찬가지로, 시클로스포린 A는 시클로필린 및 칼시뉴린의 이량체화를 유도하고, 라파마이신은 FKBP 및 mTOR의 이량체화를 유도한다.[93,94]

[1254] 한 실시양태에서, FK506:FKBP12 및 시클로스포린 A:시클로필린 소분자:단백질 결합 상호작용의 선택적 고친화도 결합을 사용하여, 이량체화의 합성 화학적 유도제를 또한 개발하였다. 한 예에서, FK1012로 명명된 FK506의 2개의 유닛으로 구성된 소분자는 신호전달 수용체의 세포질 도메인을 FKBP12로 태그부착한 경우 신호 전달에 영향을 미치는 것으로 제시되었다.[95] 이량체화의 화학적 유도제 (CID)는 이후에 다수의 신호전달 경로를 제어하는 데 사용되어 왔다.[96-103]

[1255] 생물학적 과정을 연구하기 위한 유용한 도구이지만, 치료 용도를 위한 합성 CID가 직면한 하나의 도전과제는 환자 내로 FKBP12- 또는 시클로필린-표적 단백질 키메라를 도입하는 것이 도전과제라는 것이다.

[1256] 본 개시내용은 이전에 접근불가능한 치료 방법을 생성하기 위해 다수의 개념을 합한다. 제1 개념은 살아있는 세포에서 표적화된 삽입을 포함한 정확한 게놈 편집을 가능하게 하는, 본원에 기재된 다중-플랩 프라임 편집 (예를 들어, 이중 플랩 또는 사중 플랩 프라임 편집)이다. 제2 개념은 세포 배양에서 신호전달 및 올리고머화 과정에 대한 소분자 제어를 가능하게 하는 강력한 도구인 화학적-유도된 이량체화이다.

[1257] 단백질 이량체화에 대한 화학적 제어가 유익한 치료 효과를 가질 수 있는 특정 사례가 확인되었다.

[1258] 인슐린 수용체는 세포질 키나제 도메인의 인산화에 의해 세포외 도메인에 대한 인슐린 결합에 반응하는 이종사량체 막횡단 단백질이다.[104] 막-국재화 성분, 인슐린 수용체의 C-말단 키나제 도메인, 및 FKBP12의 3개의 카피로 구성된 조작된 키메라 단백질은 FK1012에 반응하고, 세포 배양에서 인슐린 반응을 개시한다.[99] 유사하게, 환자 세포에서 천연 인슐린 수용체의 키나제 도메인의 C-말단 단부에 대한 FKBP12의 융합은 FK1012-의존성 인산화 및 인슐린 신호전달 캐스케이드의 개시를 가능하게 할 것으로 예상된다. 이 시스템은 인슐린을 만들 수 없거나 (예를 들어, 제1형 당뇨병), 또는 인슐린에 약하게 반응하는 (예를 들어, 제2형 당뇨병) 환자에서 인슐린 사용을 대체하거나 보완할 수 있다.

[1259] 추가적으로, 에리트로포이에틴은 에리트로포이에틴 수용체 (EpoR)에 결합하여 Jak/STAT 신호전달 캐스케이드의 활성화를 발생시키는 이량체화 또는 사전형성된 수용체 이량체의 입체형태적 변화를 유도함으로써 적혈구 증식을 자극한다.[105] FKBP12로 태그부착된 EpoR의 막-고정된 세포질 도메인의 FK1012-유도된 올리고머화는 Jak/STAT 신호전달 캐스케이드를 개시하고 세포 증식을 촉진하기에 충분하다는 것이 입증되었다.[106] 환자 세포에서의 프라임 편집에 의해 FKBP12를 천연 EpoR에 융합시키는 것은 적혈구 증식 (적혈구생성)에 대한 FK1012-유도된 제어를 가능하게 할 것으로 예상된다. 이 시스템은 빈혈 환자에서 적혈구 성장을 촉발하는 데 사용될 수 있다. FK1012-유도성 EpoR은 또한 생체외 조작을 거친 혈액 세포에 대한 생체내 선택 마커로서 사용될 수 있다.

[1260] 원칙적으로, 임의의 수용체 티로신 키나제는 다중-플랩 프라임 편집-CID 치료를 위한 실행가능한 표적일 수 있다. 하기 표는 인간 게놈 내의 모든 수용체 티로신 키나제의 목록을 포함한다.[110]

패밀리	수용체	동의어	NT 수탁번호	프롯 수탁번호	염색체
ALK 패밀리	ALK	Ki1	NM_004304	NP_004295	2p23
	LTK	TYK1	NM_002344	NP_002335	15q15.1-q21.1
AXL 패밀리	AXL	UFO, Tyro7(r) Ark(m)	NM_001699	NP_001690	19q13.1
	MER	MERTK, NYK, Eyk(ch)	NM_006343	NP_006334	2q14.1
	TYRO3	RSE, SKY, BRT, DTK, TIF	NM_006293	NP_006284	15q15.1-q21.1
DDR 패밀리	DDR1	CAK, TRKE, NEP NTRK4, EDDR1, PTK3	NM_013993	NP_001945	6p21.3
	DDR2	TKT, TYRO10, NTRKR3	NM_006182	NP_006173	1q21-q22
EGFR 패밀리	EGFR	ERBB, ERBB1	NM_005228	NP_005219	7p12
	ERBB2	HER2, Neu(r), NGL	NM_004448	NP_004439	17q11.2-q12
	ERBB3	HER3	NM_001982	NP_001973	12q13
	ERBB4	HER4	NM_005235	NP_005226	2q33.3-q34
EPH 패밀리	EPHA1	EPH, EPHT	NM_005232	NP_005223	7q32-q36
	EPHA2	ECK, Sek(m), Myk2(m)	NM_004431	NP_004422	1p34
	EPHA3	HEK, ETK1, Tyro4(r), Mek4(m), Cek4(ch)	NM_005233	NP_005224	3p11.2
	EPHA4	HEK8, Tyro1(r), Sek1(m), Cek8(ch)	NM_004438	NP_004429	2q36qter
	EPHA5	HEK7, Ehk(r), Bsk(r), Cek7(ch)	L36644	P54756	
	EPHA6	DKFZp434C1418, Ehk2(r)	AL133666		
	EPHA7	HEK11, Mdk1(m), Ebk(m), Ehk3(r), Cek11(ch)	NM_004440	NP_004431	6q21
	EPHA8	HEK3, KIAA1459, Eek(r), Cek10(ch)	AB040892	CAB81612	1q23-q24
	EPHB1	NET, EPHT2, HEK6, Elk(r), Cek6(ch)	NM_004441	NP_004432	3q21-q23
	EPHB2	HEK5, ERK, DRT, EPHT3, Tyro5(r), Nuk(m), Sek3(m), Cek5(ch)	AF025304	AAB94602	1p36.1-p35
	EPHB3	HEK2, Tyro6, Mdk5(m), Sek4(m)	NM_004443	NP_004434	3q21-qter
	EPHB4	HTK, Tyro11(r), Mdk2(m), Myk1(m)	NM_004444	NP_004435	
	EPHB6	HEP, Mep(m), Cek1(ch)	NM_004445	NP_004436	7q33-q35
FGFR 패밀리	FGFR1	FLT2, bFGFR, FLG, N-SAM	M34641	AAA35835	8p11.2
	FGFR2	KGFR, K-SAM, Bek(m), CFD1, JWS, Cek3(ch)	NM_000141	NP_000132	10q26
	FGFR3	HBGFR, ACH, Cek2(ch)	NM_000142	NP_000133	4p16.3
	FGFR4		NM_002011	NP_002002	5q35.1-qter
	IGF1R	JTK13	NM_000875	NP_000866	15q25-q26

[1261]

INSR 패밀리	INSR	IR	NM_000208	NP_000199	19p13.3-p13.2
	INSRR	IRR	J05046	AAC31759	1q21-q23
MET 패밀리	MET	HGFR	NM_000245	NP_000236	7q31
	RON	MST1R, CDw136, Fv2(m), STK(m), SEA(ch)	NM_002447	NP_002438	3p21.3
MUSK 패밀리	MUSK	Nsk2(m), Mlk1(m), Mk2(m)	NM_005592	NP_005583	9q31.3-q32
PDGFR 패밀리	CSF1R	FMS, C-FMS, CD115	NM_005211	NP_005202	5q31-q32
	FLT3	FLK2, STK1, CD135	NM_004119	NP_0041110	13q12
	KIT	Sfr(m), CKIT	NM_000222	NP_000213	4q11-q12
	PDGFRA		NM_006206	NP_006197	4q11-q13
	PDGFRB	PDGFR, JTK12	NM_002609	NP_002600	5q31-q32
PTK7 패밀리	PTK7	CCK4, KLG(ch)	NM_002821	NP_002812	6p21.2-p12.2
RET 패밀리	RET	MEN2A/B, HSCR1, MTC1	X12949	P07949	10q11.2
ROR 패밀리	ROR1	NTRKR1	NM_005012	NP_005003	1p32-p31
	ROR2	NTRKR2	NM_004560	NP_004551	
ROS 패밀리	ROS1	MCF3	NM_002944	NP_002935	6q22
RYK 패밀리	RYK	Vlk(m), Mrk(m)	S59184	AAB263411	3q22
TIE 패밀리	TEK	TIE2	NM_000459	NP_000450	9p21
	TIE	TIE1, JTK14	NM_005424	NP_005415	1p34-p33
TRK 패밀리	NTRK1	TRK, TRKA	NM_002529	NP_002520	1q21-q22
	NTRK2	TRKB	NM_006180	NP_006171	9q22.1
	NTRK3	TRKC	NM_002530	NP_002521	15q25
VEGFR 패밀리	VEGFR1	FLT1	NM_002019	NP_002010	13q12
	VEGFR2	KDR, FLK1		AAB88005	4q11-q12
	VEGFR3	FLT4, PCL	NM_002020	NP_002011	5q34-q35
AATYK 패밀리	AATYK	AATK, KIAA0641	NM_004920	NP_004911	17q25.3
	AATYK2	KIAA1079	NM_014916	NP_055731	7q21-q22
	AATYK3				19q13.2-q13.3

[1262]

[1263] 다중-플랩 프라임 편집-CID 시스템에는 수많은 이점이 존재한다. 하나의 이러한 이점은, 전형적으로 제조, 비용, 전달, 생산, 또는 저장에 있어서 문제를 야기하는 단백질인 내인성 리간드를, IV 또는 주사에 의해 투여되는 대신 경구로 투여될 수 있는 약물-유사 소분자로 대체할 수 있고, FDA-승인된 약물로부터 용이하게 제조되고 (또는 그 자체가 이미 약물이고), 전형적으로 단백질 약물과 연관된 특별한 생산 또는 저장 비용을 초래하지 않는다는 것이다. 또 다른 이점은 편집 단독이 생리학적 효과를 갖지 않는다는 것이다. 표적 단백질 이량체화의 양은 소분자 CID를 투여하는 것에 의해 제어될 수 있다. 또한, 표적 단백질 이량체화는 단량체 형태의 CID를 부가하는 것에 의해 용이하고 빠르게 가역적이다. 또 다른 이점은 단일 리간드가 다중 수용체를 표적화하는 경우에, 단지 1개의 수용체를 프라임-편집하는 것에 의해 선택성이 달성될 수 있다는 것이다. 마지막으로, 프라임 편집에 사용되는 전달 방법에 의존하여, 편집을 국부 조직 또는 기관으로 제한하여, 특정 영역에서만 유도성 수용체 활성화를 가능하게 하는 것이 또한 가능할 수 있다.

[1264] 2개의 개별 편집 사건이 높은 수준으로 발생할 수 있을 정도로 다중-플랩 프라임 편집에 의한 편집 효율이 충분히 높은 경우에, 2개의 관심 단백질을 상이한 소분자 결합 도메인 (예컨대 FKBP 및 시클로필린)으로 태그부착하고, 소분자 이종이량체 (예컨대 FK506-시클로스포린 A 이량체)와의 이종이량체화를 유도하는 것이 또한 가능할 것이다.

[1265] 천연 단백질에 대한 FKBP12 또는 다른 소분자 결합 단백질의 융합은 일반적으로 조직 배양에서 플라스미드로부터의 과다발현에 의해 달성되었다. 후속 화학적-유도된 이량체화는 융합 단백질을 생산하는 세포로의 표현형 변화를 유도하는 것으로 입증되었다.

[1266] 하기 참고문헌은 상기 섹션에서 인용되며, 본원에 참조로 포함된다.

1. Crabtree, G. R. & Schreiber, S. L. Three-part inventions: intracellular signaling and induced proximity. *Trends Biochem. Sci.* 21, 418-22 (1996).

2. Liu, J. *et al.* Calcineurin Is a Common Target of A and FKBP-FK506 Complexes. *Cell* 66, 807-815 (1991).

3. Keith, C. T. *et al.* A mammalian protein targeted by Gl-arresting rapamycin-receptor complex. *Nature* 369, 756-758 (2003).

4. Spencer, D. M., Wandless, T. J., Schreiber, S. L. S. & Crabtree, G. R. Controlling signal transduction with synthetic ligands. *Science* 262, 1019-24 (1993).

5. Pruschy, M. N. *et al.* Mechanistic studies of a signaling pathway activated by the organic dimerizer FK1012. *Chem. Biol.* 1, 163-172 (1994).

6. Spencer, D. M. *et al.* Functional analysis of Fas signaling in vivo using synthetic inducers of dimerization. *Curr. Biol.* 6, 839-847 (1996).

7. Belshaw, P. J., Spencer, D. M., Crabtree, G. R. & Schreiber, S. L. Controlling programmed cell death with a cyclophilin-cyclosporin-based chemical inducer of dimerization. *Chem. Biol.* 3, 731-738 (1996).

[1267]

8. Yang, J. X., Symes, K., Mercola, M. & Schreiber, S. L. Small-molecule control of insulin and PDGF receptor signaling and the role of membrane attachment. *Curr. Biol.* 8, 11-18 (1998).

9. Belshaw, P. J., Ho, S. N., Crabtree, G. R. & Schreiber, S. L. Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins. *Proc.Natl. Acad. Sci.* 93, 4604-4607 (2002).

10. Stockwell, B. R. & Schreiber, S. L. Probing the role of homomeric and heteromeric receptor interactions in TGF-β signaling using small molecule dimerizers. *Curr. Biol.* 8, 761—773 (2004).

11. Spencer, D. M., Graef, L, Austin, D. J., Schreiber, S. L. & Crabtree, G. R. A general strategy for producing conditional alleles of Src-like tyrosine kinases. *Proc. Natl. Acad. Sci.* 92, 9805-9809 (2006).

12. Holsinger, L. J., Spencer, D. M., Austin, D. J., Schreiber, S. L. & Crabtree, G. R. Signal transduction in T lymphocytes using a conditional allele of Sos. *Proc. Natl. Acad. Sci.* 92, 9810-9814 (2006).

13. Myers, M. G. Insulin Signal Transduction and the IRS Proteins. *Annu. Rev. Pharmacol. Toxicol.* 36, 615-658 (1996).

14. Watowich, S. S. The erythropoietin receptor: Molecular structure and hematopoietic signaling pathways. *J. Investig. Med.* 59, 1067-1072 (2011).

15. Blau, C. A., Peterson, K. R., Drachman, J. G. & Spencer, D. M. A proliferation switch for genetically modified cells. *Proc. Natl. Acad. Sci.* 94, 3076-3081 (2002).

16. Clackson, T. *et al.* Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci.* 95, 10437-10442 (1998).

17. Diver, S. T. & Schreiber, S. L. Single-step synthesis of cell-permeable protein dimerizers that activate signal transduction and gene expression. *J. Am. Chem. Soc.* 119, 5106-5109 (1997).

18. Guo, Z. F., Zhang, R. & Liang, F. Sen. Facile functionalization of FK506 for biological studies by the thiol-ene 'click' reaction. *RSC Adv.* 4, 11400-11403 (2014).

19. Robinson, D. R., Wu, Y.-M. & Lin, S.-F. The protein tyrosine kinase family of the human genome. *Oncogene* 19, 5548-5557 (2000).

[1268]

[1269] G. 레콤비나제 표적 부위를 삽입하기 위한 다중-플랩 프라임 편집의 사용

[1270] 또 다른 측면에서, 다중-플랩 프라임 편집 (예를 들어, 도 90, 95, 97a, 98a에 나타내어진 바와 같음)은 레콤비나제 부위 (또는 "레콤비나제 인식 서열")를 목적하는 게놈 부위에 삽입하는 데 사용될 수 있다. 레콤비나제 부위의 삽입은 게놈에서 부위-특이적 유전자 변화를 달성하기 위한 프로그램화된 위치를 제공한다. 이러한 유전자 변화는 다른 유전자 변화 중에서도, 예를 들어 플라스미드의 게놈 통합, 게놈 결실 또는 삽입, 염색체 전위, 및 카세트 교환을 포함할 수 있다. 이들 예시적인 유형의 유전자 변화가 도 64(b)-(f)에 예시된다. 이어서, 설치된 레콤비나제 인식 서열을 사용하여 그 부위에서 부위-특이적 재조합을 수행하여 다양한 재조합 결과, 예컨대 DNA 단편의 절제, 통합, 역위, 또는 교환을 발생시킬 수 있다. 예를 들어, 도 65는 GFP 발현 마커를 포함하는 DNA 공여자 주형을 통합하는 데 사용될 수 있는 레콤비나제 부위의 설치를 예시한다. 레콤비나제 부위

내로 통합된 GFP 발현 시스템을 함유하는 세포는 형광을 낼 것이다.

[1271] 레콤비나제 부위를 게놈 내로 설치하는 메카니즘은 다른 서열, 예컨대 펩티드/단백질 및 RNA 태그를 게놈 내로 설치하는 것과 유사하다. 레콤비나제 표적 서열의 설치를 예시하는 개략도가 도 64(a)에 제시된다. 과정은 레콤비나제 표적 서열을 도입할 목적하는 표적 유전자좌를 선택하는 것으로 시작된다. 다음으로, 프라임 편집제 융합체가 제공된다 ("RT-Cas9:gRNA"). 여기서, "gRNA"는 본원에 기재된 원리를 사용하여 설계될 수 있는 PEgRNA를 지칭한다. 다양한 실시양태에서 PEgRNA는 도 3d에 상응하는 아키텍처를 포함할 것이다 (5'-[~20-nt 스페이서]-[gRNA 코어]-[연장 아암]-3', 여기서 연장 아암은 3'에서 5' 방향으로 프라이머 결합 부위 ("A"), 편집 주형 ("B"), 및 상동성 아암 ("C")을 포함함). 편집 주형 ("B")은 레콤비나제 부위에 상응하는 서열, 즉 레콤비나제 부위의 센스 또는 안티센스 가닥이고 다중-플랩 프라임 편집 과정을 통해 게놈 DNA 표적 유전자좌 내로 혼입되는 상보적 단일 가닥 DNA를 코딩하는 PEgRNA의 단일 가닥 RNA를 포함할 것이다.

[1272] 다양한 측면에서, 본 개시내용은 부위-특이적 레콤비나제(들)에 대한 노출 후에 정확하고 효율적인 게놈 변형을 지시할, 인간 또는 다른 게놈 내의 고가치 유전자좌에 레콤비나제 인식 서열을 도입하기 위한 다중-플랩 PE의 사용을 제공한다 (도 64). 도 64에 제시된 다양한 실시양태에서, PE는 (b) DNA 공여자 주형의 게놈 통합을 위한 부위로서 사용하기 위한 단일 SSR 표적을 삽입하는 데 사용될 수 있다. (c)는 게놈의 한 부분을 결실시키기 위해 SSR 표적 부위의 탠덤 삽입이 어떻게 사용될 수 있는지 보여준다. (d)는 게놈의 한 부분을 역전시키기 위해 SSR 표적 부위의 탠덤 삽입이 어떻게 사용될 수 있는지 보여준다. (e)는 2개의 원위 염색체 영역에서의 2개의 SSR 표적 부위의 삽입이 염색체 전위를 어떻게 발생시킬 수 있는지 보여준다. (f)는 게놈 내의 2개의 상이한 SSR 표적 부위의 삽입이 DNA 공여자 주형으로부터 카세트를 교환하는데 어떻게 사용될 수 있는지 보여준다. SSR 표적을 삽입하기 위해 PE를 사용하는 것에 의한 각각의 유형의 게놈 변형이 고려되지만, 이 목록은 또한 제한적인 것으로 의도되지 않는다.

[1273] 레콤비나제 인식 서열의 다중-플랩 PE-매개된 도입은 대규모 게놈 결함, 예컨대 유전자 상실, 역위, 또는 중복, 또는 염색체 전위에 의해 유발되는 유전 질환의 치료에 특히 유용할 수 있다[1-7] (표 6). 예를 들어, 윌리암스-보이렌 증후군은 염색체 721에서의 24의 결실에 의해 유발되는 발달 장애이다. 살아있는 세포에의 다중 전체 유전자의 효율적이고 표적화된 삽입을 위한 기술은 현재 존재하지 않지만 (이러한 전장 유전자 삽입을 행하는 PE의 잠재력이 현재 연구되고 있지만, 아직 확립되지 않았음); PE에 의해 삽입된 표적에서의 레콤비나제-매개된 통합은 이러한 질환 및 다른 질환에 대한 영구적인 치유를 위한 하나의 접근법을 제공한다. 또한, 레콤비나제 인식 서열의 표적화된 도입은 트랜스제닉 식물, 동물 연구 모델, 생물생산 세포주, 또는 다른 통상의 진핵 세포주의 생성을 포함한 적용을 고도로 가능하게 할 수 있다. 예를 들어, 트랜스제닉 식물에서 다중-플랩 PE-특이적 표적에서의 레콤비나제-매개된 게놈 재배열은 개선된 특성을 갖는 농업 작물을 생성하기 위한 병목 중 하나를 극복할 수 있다[8,9].

[1274] 표 6. 레콤비나제 인식 서열의 다중-플랩 PE-기반 설치를 통해 복구될 수 있는 대규모 게놈 변형과 연관된 유전 질환의 예.

질환	원인
17P 삼염색체증	유전자 중복
샤르코-마리-투스병 유형 1	유전자 중복
스미스-마제니스 증후군	유전자 결실
윌리암스-보이렌 증후군	유전자 결실
드 라 샤펠 증후군	염색체 전위
다운 증후군 (일부 형태)	염색체 전위
A형 혈우병	유전자 역위
헌터 증후군	유전자 역위

[1275]

[1276] 천연 및 조작된 티로신 레콤비나제 (표 7), 대형 세린 인테그라제 (표 8), 세린 레솔바제 (표 9), 및 티로신 인테그라제 (표 10)를 포함한 다수의 SSR 패밀리 구성원이 특징화되었고, 그의 레콤비나제 인식 서열이 기재되었다. 증진된 비율의 게놈 통합을 입증한 변형된 표적 서열이 또한 여러 SSR에 대해 기재되었다[22-30]. 천연 레콤비나제에 더하여, 별개의 특이성을 갖는 프로그램가능한 레콤비나제가 개발되었다[31-40]. PE를 사용하여, 이들 인식 서열 중 1개 이상을 목적하는 적용에 따라 게놈 내 명시된 위치, 예컨대 세이프 하버 유전자좌에 도입할 수 있다[41-43].

[1277] 예를 들어, 프라임 편집에 의한 게놈 내 단일 레콤비나제 인식 서열의 도입은 DNA 공여자 주형과의 통합 재조합을 발생시킬 것이다 (도 64b). 인간 세포에서 강건하게 작동하는 세린 인테그라제는 유전자 통합에 특히 매우 적합할 수 있다[44,45].

[1278] 추가적으로, 2개의 레콤비나제 인식 서열의 도입은 표적의 정체 및 배향에 따라 개재 서열의 결실, 개재 서열의 역위, 염색체 전위, 또는 카세트 교환을 발생시킬 수 있다 (도 64c-f). 레콤비나제 표적과 이미 밀접하게 유사한 내인성 서열을 선택하는 것에 의해, 완전한 레콤비나제 표적을 도입하는데 필요한 편집의 범주가 감소될 것이다.

[1279] 마지막으로, 여러 레콤비나제는 천연 발생 유사부위에서 인간 또는 진핵 게놈 내로 통합되는 것으로 입증되었다[46-64]. 다중-플랩 프라임 편집은 이들 유전자좌를 변형시켜 이들 천연 유사부위에서의 통합 속도를 증진시키거나, 또는 대안적으로 원치 않는 오프-타겟 서열로서 작용할 수 있는 유사부위를 제거하기 위해 사용될 수 있다.

[1280] 본 개시내용은 다중-플랩 PE를 사용하여 진핵 게놈에 레콤비나제 표적 서열을 도입하기 위한 일반적 방법론을 기재하며, 그의 적용은 거의 무한하다. 게놈 편집 반응은 맞춤형 프라임 편집 가이드 RNA (PEgRNA)를 이용하는 CRISPR/Cas9 단백질 및 리버스 트랜스크립타제 도메인의 키메라 융합체인 "프라임 편집제"와 함께 사용하기 위해 의도된다. 더 나아가, 여러 기술을 사용하여 indel의 비율을 낮춤으로써 DNA 주형을 통해 레콤비나제 인식 서열을 도입하기 위해 Cas9 도구 및 상동성-지정 복구 (HDR) 경로가 또한 이용될 수 있다[65-67]. 인간 세포 배양물에서의 개념 증명 실험이 도 65에 제시된다. 일부 실시양태에서, 프라임 편집제는 다중-플랩 프라임 편집제이다.

[1281] 하기 여러 표가 레콤비나제 인식 서열의 PE-지정 설치에 관한 상기 설명에서 인용되고, 사용될 수 있는 예시적인 레콤비나제, 및 다중-플랩 PE에 의해 설치될 수 있는 그의 동족 레콤비나제 인식 서열의 목록을 제공한다.

[1282] 표 7. 티로신 레콤비나제 및 SSR 표적 서열.

레콤비나제	레콤비나제 인식 서열	명칭
Cre	ATAACTTCGTATAGCATACATTATACGAAGTTAT (SEQ ID NO: 517)	*loxP*
Dre	TAACTTTAAATAATGCCAATTATTTAAAGTTA (SEQ ID NO: 518)	*rox*
VCre	TCAATTTCTGAGAACTGTCATTCTCGGAAATTGA (SEQ ID NO: 519)	*loxV*
SCre	CTCGTGTCCGATAACTGTAATTATCGGACATGAT (SEQ ID NO: 520)	*loxS*
Flp	GAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC (SEQ ID NO: 521)	*FRT*
B2	GAGTTTCATTAAGGAATAACTAATTCCCTAATGAA ACTC (SEQ ID NO: 522)	*loxB*
B3	GGTTGCTTAAGAATAAGTAATTCTTAAGCAACC (SEQ ID NO: 523)	*loxB3*
Kw	ACGAAAAATGGTAAGGAATAGACCATTCCTTACC ATTTTTGGT (SEQ ID NO: 524)	
R	TTGATGAAAGAATAACGTATTCTTTCATCAA (SEQ ID NO: 525)	*RS*
TD1-40	GTGCGTCAAATAATAACGTATTATTTGACACTT (SEQ ID NO: 526)	*TDRS*
Vika	AATAGGTCTGAGAACGCCCATTCTCAGACGTATT (SEQ ID NO: 527)	*vox*
Nigri	TGAATGTCCTATAATTACACTTATAGGACATTCA (SEQ ID NO: 528)	*nox*
Panto	GAAACTTTAAATAATAAGTCTTATTTAAAGTTTC (SEQ ID NO: 529)	*pox*
Kd	AAACGATATCAGACATTTGTCTGATAATGCTTCA TTATCAGACAAATGTCTGATATCGTTT (SEQ ID NO: 530)	*loxK*
Fre	ATATATACGTATATAGACATATATACGTATATAT (SEQ ID NO: 531)	*loxH*
CreALSHG	ATAACTCTATATAATGTATGCTATATAGAGTTAT (SEQ ID NO: 532)	*loxM7*
Tre	ACAACATCCTATTACACCCTATATGCCAACATGG (SEQ ID NO: 533)	*loxLTR*
Brec1	AACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT (SEQ ID NO: 534)	*loxBTR*
Cre-R3M3	GATACAACGTATATACCTTTCTATACGTTGTTTA (SEQ ID NO: 535)	*loxK2*

[1283]

[1284] 표 8. 대형 세린 인테그라제 및 SSR 표적 서열.

레콤비나제	레콤비나제 인식 서열 좌측	레콤비나제 인식 서열 우측
Bxb1	GGTTTGTCTGGTCAACCACCGC GGTCTCAGTGGTGTACGGTACA AACC (SEQ ID NO: 536)	GGCTTGTCGACGACGGCGGTCTCC GTCGTCAGGATCAT (SEQ ID NO: 537)
phiC31	GTGCCCCAACTGGGGTAACCTT TGAGTTCTCTCAGTTGGGGG (SEQ ID NO: 538)	TGCGGGTGCCAGGGCGTGCCCTTG GGCTCCCCGGGCGCGTACTCC (SEQ ID NO: 539)
R4	TGTTCCCCAAAGCGATACCACT TGAAGCAGTGGTACTGCTTGTG GGTACA (SEQ ID NO: 540)	GCATGTTCCCCAAAGCGATACCAC TTGAAGCAGTGGTACTGCTTGTGG GTACACTCTGCGGGTG (SEQ ID NO: 541)
phiBT1	GGTGCTGGGTTGTTGTCTCTGG ACAGTGATCCATGGGAAACTA CTCAGCACC (SEQ ID NO: 542)	CAGGTTTTTGACGAAAGTGATCCA GATGATCCAG (SEQ ID NO: 543)
MJ1 (phiFC1)	ATTTTAGGTATATGATTTTGTT TATTAGTGTAAATAACACTATG TACCTAAAAT (SEQ ID NO: 544)	CAAAGGATCACTGAATCAAAAGTA TTGCTCATCCACGCGAAA (SEQ ID NO: 545)
MR11	TTTGTGCGGAACTACGAACAGT TCATTAATACGAAGTGTACAA ACTTCCATACAA (SEQ ID NO: 546)	CGAAAATGTATGGAGGCACTTGTA TCAATATAGGATGTATACCTTCGA AGACACTT (SEQ ID NO: 547)
TP901-1	GAGTTTTTATTTCGTTTATTTCA ATTAAGGTAACTAAAAAACTC CTTTTAAGG (SEQ ID NO: 548)	ATGCCAACACAATTAACATCTCAA TCAAGGTAAATGCTTTTTGCTTTTT TTGC (SEQ ID NO: 549)
A118	TTCCTCGTTTTCTCTCGTTGGA AGAAGAAGAAACGAGAAA (SEQ ID NO: 550)	TTTCGGATCAAGCTATGAAGGACG CAAAGAGGGAACTAAA(SEQ ID NO: 551)
U153	TTCCTCGTTTTCTCTCGTTGGAC GGAAACGAATCGAGAAA (SEQ ID NO: 552)	TTTCGGATCAAGCTATGAAGGACG CAAAGAGGGAACTAAA (SEQ ID NO: 553)
phiRV1	GTAGTGTATCTCACAGGTCCAC GGTTGGCCGTGGACTGCTGAA GAACATTCC (SEQ ID NO: 554)	GAAGGTGTTGGTGCGGGGTTGGCC GTGGTCGAGGTGGGGT (SEQ ID NO: 555)
phi370.1	AAAAAAATACAGCGTTTTTCAT GTACAACTATACTAGTTGTAGT GCCTAAAA (SEQ ID NO: 556)	TTGTAAAGGAGACTGATAATGGCA TGTACAACTATACTCGTCGGTAAA AAGGCA (SEQ ID NO: 557)
TG1	TCCAGCCCAACAGTGTTAGTCT TTGCTCTTACCCAGTTGGGCGG GA (SEQ ID NO: 558)	GATCAGCTCCGCGGGCAAGACCTT TCTCCTTCACGGGGTGGAAGGTC (SEQ ID NO: 559)
WB	CTAGTTTTAAAGTTGGTTATTA	CGGAAGGTAGCGTCAACGATAGGT

[1285]

[1286]

레콤비나제	레콤비나제 인식 서열 좌측	레콤비나제 인식 서열 우측
	GTTACTGTGATATTTATCACGGTACCCAATAACCAATGAAT (SEQ ID NO: 560)	GTAACTGTCGTGTTTGTAACGGTACTTCCAACAGCTGGCGCCGCCAC (SEQ ID NO: 561)
BL3	CAATGAAAAACTAGGCATGTAGAAGTTGTTTGT (SEQ ID NO: 562)	TTTCCACAGACAACTCACGTGGAGGTAGTCAC (SEQ ID NO: 563)
SprA	TGTAGTAAGTATCTTAATATACAGCTTTATCTGTTTTTTAAGATACTTACTACTTT (SEQ ID NO: 564)	CACCCATTGTGTTCACAGGAGATACAGCTTTATCTGTACTGATATTAATGACATGCTG (SEQ ID NO: 565)
phiJoe	AGTTGTGGCCATGTGTCCATCTGGGGGCAGATGGAGACGGGGTCACA (SEQ ID NO: 566)	ATCTGGATGTGGGTGTCCATCTGCGGGCAGACGCCGCAGTCGAAGCACGG (SEQ ID NO: 567)
--	--	ACCTTGATCTCGGTGTCCATCGCCGGGCAGACGCCGCAGTCGAAGCACGG (SEQ ID NO: 568)
phiK38	CCCTAATACGCAAGTCGATAACTCTCCTGGGAGCGTTGACAACTTGCGCACCCTGATCTG (SEQ ID NO: 569)	GAGCGCCGGATCAGGGAGTGGACGGCCTGGGAGCGCTACACGCTGTGGCTGCGGTCGGTGC (SEQ ID NO: 570)
Int2	GCTCATGTATGTGTCTACGCGAGATTCTCGCCCGAGAACTTCTGCAAGGCACTGCTCTTGGCT (SEQ ID NO: 571)	GGACGGCGCAGAAGGGGAGTAGCTCTTCGCCGGACCGTCGACATACTGCTCAGCTCGTC (SEQ ID NO: 572)
Int3	ATGGATAAAAAAATACAGCGTTTTTCATGTACAACTATACTAGTTGTAGTGCCTAAATAATGCTT (SEQ ID NO: 573)	GTTTGTAAAGGAGACTGATAATGGCATGTACAACTATACTCGTCGGTAAAAAGGCATCTTAT (SEQ ID NO: 574)
Int4	AAAAATTACAAAGTTTTCAACCCTTGATTTGAATTAGCGGTCAAATAATTTGTAATTCGTTT (SEQ ID NO: 575)	TTCCAAAGAGCGCCCAACGCGACCTGAAATTTGAATAAGACTGCTGCTTGTGTAAAGGCGATGATT (SEQ ID NO: 576)
Int7	GTGTTATAAACCTGTGTGAGAGTTAAGTTTACATGCCTAACCTTAACTTTTACGCAGGTTCAGCTT (SEQ ID NO: 577)	AGACGAGAAACGTTCCGTCCGTCTGGGTCAGTTGGGCAAAGTTGATGACCGGGTCGTCCGTT (SEQ ID NO: 578)
Int8	TTAATAAACTATGGAAGTATGTACAGTCTTGCAATGTTGAGTGAACAAACTTCCATAATAAAAT (SEQ ID NO: 579)	CAATCATCAGATAACTATGGCGGCACGTGCATTAACCACGGTTGTATCCCGTCTAAAGTACTCGT (SEQ ID NO: 580)
Int9	GTGGTTGTTTTTGTTGGAAGTGTGTATCAGGTATCTGCATAGTT	TTTATATTGCGAAAAATAATTGGCGAACGAGGTAACTGGATACCTCAT

레콤비나제	레콤비나제 인식 서열 좌측	레콤비나제 인식 서열 우측
	ATTCCGAACTTCCAATTA (SEQ ID NO: 581)	CCGCCAATTAAAATTTG (SEQ ID NO: 582)
Int10	GGAAAATATAAATAATTTTAGTAACCTACATCTCAATCAAGGATAGTAAAACTCTCACTCTT (SEQ ID NO: 583)	AGCACGCTGATAATCAGCAAGACCACCAACATTTCCACCAATGTAAAAGCTTTAACCTTAGC (SEQ ID NO: 584)
Int11	GTTTATATGTTTACTAATAAGACGCTCTCAACCCATAAAGTCTTATTAGTAAACATATTTCAACT (SEQ ID NO: 585)	ATGGATTTTGCAGATTCCCAGATGCCCCTACAGAAAGAGGTACAAAACATTTATTGGAATTAATT (SEQ ID NO: 586)
Int12	TTTTTGTATGTTAGTTGTGTCACTGGGTAGACCTAAATAGTGACACAACTGCTATTAAAATTTAA (SEQ ID NO: 587)	GTTCGTGGTAACTATGGGTGGTACAGGTGCCACATTAGTTGTACCATTTATGTTTATGTGGTTAAC (SEQ ID NO: 588)
Int13	CAATAACGGTTGTATTTGTAGAACTTGACCAGTTGTTTTAGTAACATAAATACAACTCCGAATA (SEQ ID NO: 589)	GCATACATTGTTGTTGTTTTTCCAGATCCAGTTGGTCCTGTAAATATAAGCAATCCATGTGAGT (SEQ ID NO: 590)
LI	GTTTAGTATCTCGTTATCTCTCGTTGGAGGGAGAAGAAACGGGATACCAAAA (SEQ ID NO: 591)	TAACTTTTTCGGATCGAGTTATGATGGACGTAAAGAGGGAACAAAGCATCTA (SEQ ID NO: 592)
Peaches	TAGTTTCCAATGTTACAGGAACTGCTGGCAGAATCCAACACATTGGAAGTCG (SEQ ID NO: 593)	CGGTCTCCATCGGGATCTGCTGATCGAGCAGCATGCCGACCA (SEQ ID NO: 594)
Bxz2	TAACCGCAAGTGTACATCCCTCGGCTGGCCGAGACAAGTACAGTTGCGACAG (SEQ ID NO: 595)	CGGTCTCCATCGGGATCTGCTGATCGAGCAGCATGCCGACCA (SEQ ID NO: 596)
SV1	ATGTGGTCCTTTAGATCCACTGACGTGGGTCAGTGTCTCTAAAGGACTCGCG (SEQ ID NO: 597)	CATCAGGGCGGTCAGGCCGTAGATGTGGAAGAAACGGCAGCACGGCGAGGACG (SEQ ID NO: 598)

[1287]

[1288] 표 9. 세린 레솔바제 및 SSR 표적 서열.

레솔바제	레콤비나제 인식 서열 좌측	레콤비나제 인식 서열 우측
Gin	CGTTTCCTGTAAACCGAGGTTTTGGATAAACA (SEQ ID NO: 599)	CGTTTCCTGTAAACCGAGGTTTTGGATAATGG (SEQ ID NO: 600)
Cin	GAGTTCTCTTAAACCAAGGTTTAGGATTGAAA (SEQ ID NO: 601)	GAGTTCTCTTAAACCAAGGTATTGGATAACAG (SEQ ID NO: 602)
Hin	TGGTTCTTGAAAACCAAGGTTTTTGATAAAGC (SEQ ID NO: 603)	AAATTTTCCTTTTTGGAAGGTTTTTGATAACCA (SEQ ID NO: 604)
Min	GCCTTCCCCTAAACCAACGTTTTTATGCCGCC (SEQ ID NO: 605)	GCCTTCCCCCAAACCAAGGTAATCAAGAACGC (SEQ ID NO: 606)
Sin	TTGTGAAATTTGGGTACACCCTAATCATACAA (SEQ ID NO: 607)	CGTATGATTAGGGTGTATATTAATTT (SEQ ID NO: 608)

[1289]

[1290] 표 10. 티로신 인테그라제 및 표적 서열.

인테그라제	*attP*	*attB*
HK022	CAAATGATTTTATTTTGACTAATAA TGACCTACTTACATTAATTTACTGAT AATTAAAGAGATTTTAAATATACAA CTTATTCACCTAAAGGATGACAAAA (SEQ ID NO: 609)	GCACTTTAGGTGAAAAAGGT T (SEQ ID NO: 610)
	TAACATTAATCACTTAAAAATCATC GCATTACACTAATCTGTGGTTAAAT GATAGACTACATAATGCGACAAAA CGCAACATATCCAGTCACTATGAAT CAACTACTTAGATAGTATTAGTGAC CT (SEQ ID NO: 611)	
P22	CTAAGTGGTTTGGGACAAAAATGGG ACATACAAATCTTTGCATCGGTTTG CAAGGCTTTGCATGTCTTTCGAAGA TGGGACGTGTGAGCGCAGGTATGAC GTGGTATGTGTTGACTTAAAAGGTA GTTCTTATAATTCGTAATGCGAAGG TCGTAGGTTCGACTCCTATTATCGG CACCAGTTAAATCAAATACTTACGT ATTATTCGTGCCTTCCTTATTTTTAC TGTGGGACATATTTGGGACAGAAGT ACCAAAAA (SEQ ID NO: 612)	GCAGCGCATTCGTAATGCGA AGGTCGT (SEQ ID NO: 613)
L5	GCGATCCCCATCCGCGACGTGCCAA CTAGGTCTCCTCTCGTCGTGAACAA GGCTACCGGGTTGCAACTCCTGTGC AACTCTCAGGCTTCAACGCGCTTCT ACGACCTGCAATTTCTTTCCACTTA GAGGATGCAGCCGAGAGGGGTAAA AACCTATCTTGACCGGCCCATATGT GGTCGGCAGACACCCATTCTTCCAA ACTAGCTACGCGGGTTCGATTCCCG TCGCCCGCTCCGCTGGTCAGAGGGT GTTTTCGCCCTCTGGCCATTTTTCTT TCCAGGGGTCTGCAACTCTTGTGCG ACTCTTCTGACCTGGGCATACGCGG TTGCAACGCATCCCTGATCTGGCTA CTTTCGATGCTGACAAACGAATAGA GCCCCCCGCCTGCGCGAACAGACG AGGGGCATTCACA (SEQ ID NO: 614)	GAGCGGGCGACGGGAATCG AACCCGCGTAGCTAGTTTGG AAGA (SEQ ID NO: 615)

[1291]

[1292] 다양한 다른 측면에서, 본 개시내용은 1개 이상의 레콤비나제 인식 서열을 설치하기 위해 다중-플랩 PE를 사용하는 방법 및 부위-특이적 재조합에서의 그의 용도에 관한 것이다.

[1293] 일부 실시양태에서, 부위-특이적 재조합은 다양한 재조합 결과, 예컨대 DNA 단편의 절제, 통합, 역위, 또는 교환을 발생시킬 수 있다.

[1294] 일부 실시양태에서, 방법은 2개 이상의 핵산 (예를 들어, DNA) 분자의 2개 이상의 영역의 또는 이들 사이의 재조합을 유도하는데 유용하다. 다른 실시양태에서, 방법은 단일 핵산 분자 (예를 들어, DNA) 내의 2개 이상의 영역의 또는 이들 사이의 재조합을 유도하는데 유용하다.

[1295] 일부 실시양태에서, 개시내용은 (a) 다중-플랩 프라임 편집에 의해 게놈 유전자좌에 레콤비나제 인식 서열을 설치하는 단계; (b) 게놈 유전자좌를 레콤비나제의 존재 하에 레콤비나제 인식 서열을 또한 포함하는 DNA 공여자 주형과 접촉시키는 단계를 포함하는, 부위-특이적 재조합에 의해 공여자 DNA 주형을 통합시키는 방법을 제공한다.

[1296] 다른 실시양태에서, 개시내용은 (a) 다중-플랩 프라임 편집에 의해 게놈 유전자좌에 한 쌍의 레콤비나제 인식 서열을 설치하는 단계; (b) 게놈 유전자좌를 레콤비나제와 접촉시킴으로써, 한 쌍의 레콤비나제 인식 서열 사이의 게놈 영역의 결실을 촉매하는 단계를 포함하는, 부위-특이적 재조합에 의해 게놈 영역을 결실시키는 방법을 제공한다.

[1297] 또 다른 실시양태에서, 개시내용은 (a) 다중-플랩 프라임 편집에 의해 게놈 유전자좌에 한 쌍의 레콤비나제 인

식 서열을 설치하는 단계; (b) 게놈 유전자좌를 레콤비나제와 접촉시킴으로써, 한 쌍의 레콤비나제 인식 서열 사이의 게놈 영역의 역위를 촉매하는 단계를 포함하는, 부위-특이적 재조합에 의해 게놈 영역을 역전시키는 방법을 제공한다.

[1298] 또 다른 실시양태에서, 개시내용은 (a) 프라임 편집에 의해 제1 게놈 유전자좌에 제1 레콤비나제 인식 서열을 설치하는 단계; (b) 다중-플랩 프라임 편집에 의해 제2 게놈 유전자좌에 제2 레콤비나제 인식 서열을 설치하는 단계; (c) 제1 및 제2 게놈 유전자좌를 레콤비나제와 접촉시킴으로써, 제1 및 제2 게놈 유전자좌의 염색체 전위를 촉매하는 단계를 포함하는, 제1 게놈 부위와 제2 게놈 부위 사이의 염색체 전위를 유도하는 방법을 제공한다.

[1299] 다른 실시양태에서, 개시내용은 (a) 다중-플랩 프라임 편집에 의해 제1 게놈 유전자좌에 제1 레콤비나제 인식 서열을 설치하는 단계; (b) 다중-플랩 프라임 편집에 의해 제2 게놈 유전자좌에 제2 레콤비나제 인식 서열을 설치하는 단계; (c) 제1 및 제2 게놈 유전자좌를 제1 및 제2 레콤비나제 인식 서열 및 레콤비나제와 플랭킹된 카세트를 포함하는 공여자 DNA와 접촉시켜, 플랭킹된 게놈 유전자좌 및 DNA 공여자 내 카세트의 교환을 촉매하는 단계를 포함하는, 게놈 유전자좌와 카세트를 포함하는 공여자 DNA 사이의 카세트 교환을 유도하는 방법을 제공한다.

[1300] 게놈 내에 1개 초과의 레콤비나제 인식 서열의 삽입을 수반하는 다양한 실시양태에서, 레콤비나제 인식 서열은 동일하거나 상이할 수 있다. 일부 실시양태에서, 레콤비나제 인식 서열은 동일하다. 다른 실시양태에서, 레콤비나제 인식 서열은 상이하다.

[1301] 다양한 실시양태에서, 레콤비나제는 티로신 레콤비나제, 예컨대 표 7에 제시된 바와 같은 Cre, Dre, Vcre, Scre, Flp, B2, B3, Kw, R, TD1-40, Vika, Nigri, Panto, Kd, Fre, Cre(ALSHG), Tre, Brec1, 또는 Cre-R3M3일 수 있다. 이러한 실시양태에서, 레콤비나제 인식 서열은 사용 중인 레콤비나제에 상응하는 표 7의 RRS일 수 있다.

[1302] 다양한 다른 실시양태에서, 레콤비나제는 대형 세린 레콤비나제, 예컨대 표 8에 제시된 바와 같은 Bxb1, PhiC31, R4, phiBT1, MJ1, MR11, TP901-1, A118, V153, phiRV1, phi370.1, TG1, WB, BL3, SprA, phiJoe, phiK38, Int2, Int3, Int4, Int7, Int8, Int9, Int10, Int11, Int12, Int13, L1, peaches, Bxz2, 또는 SV1일 수 있다. 이러한 실시양태에서, 레콤비나제 인식 서열은 사용 중인 레콤비나제에 상응하는 표 8의 RRS일 수 있다.

[1303] 또 다른 실시양태에서, 레콤비나제는 세린 레콤비나제, 예컨대 표 8에 제시된 바와 같은 Bxb1, PhiC31, R4, phiBT1, MJ1, MR11, TP901-1, A118, V153, phiRV1, phi370.1, TG1, WB, BL3, SprA, phiJoe, phiK38, Int2, Int3, Int4, Int7, Int8, Int9, Int10, Int11, Int12, Int13, L1, peaches, Bxz2, 또는 SV1일 수 있다. 이러한 실시양태에서, 레콤비나제 인식 서열은 사용 중인 레콤비나제에 상응하는 표 8의 RRS일 수 있다.

[1304] 다른 실시양태에서, 레콤비나제는 세린 레솔바제, 예컨대 표 9에 제시된 바와 같은 Gin, Cin, Hin, Min, 또는 Sin일 수 있다. 이러한 실시양태에서, 레콤비나제 인식 서열은 사용 중인 레콤비나제에 상응하는 표 9의 RRS일 수 있다.

[1305] 다양한 다른 실시양태에서, 레콤비나제는 티로신 인테그라제, 예컨대 표 10에 제시된 바와 같은 HK022, P22, 또는 L5일 수 있다. 이러한 실시양태에서, 레콤비나제 인식 서열은 사용 중인 레콤비나제에 상응하는 표 10의 RRS일 수 있다.

[1306] 일부 실시양태에서, 다중-플랩 PE에 의한 부위-특이적 재조합을 위한 임의의 방법이 생체내 또는 시험관내에서 수행될 수 있다. 일부 실시양태에서, 부위-특이적 재조합을 위한 임의의 방법은 세포에서 수행된다 (예를 들어, 세포에서 게놈 DNA를 재조합함). 세포는 원핵 또는 진핵 세포일 수 있다. 세포, 예컨대 진핵 세포는 본원에 기재된 바와 같은 개체, 예컨대 대상체 (예를 들어, 인간 대상체) 내의 것일 수 있다. 본원에 기재된 방법은 시험관내 및 생체내에서, 예를 들어 트랜스제닉 세포, 세포주, 또는 동물의 생성과 관련하여 세포의 유전자 변형, 또는 대상체의 세포에서 게놈 서열의 변경, 예를 들어 유전적 결함의 교정에 유용하다.

[1307] [8] 본원에 기재된 다중-플랩 프라임 편집제의 실시양태

[1308] 하기 비제한적 실시양태는 본 출원에 기재된 바와 같은 이중 프라임 편집제 및 그의 용도의 측면을 기재한다.

[1309] 1. 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템으로서, 여기서 시스템은 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체를 포함하고, 여기서 각각의 제1 및 제2 프라임 편집

제 복합체는 (1) (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 (ii) RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 프라임 편집제; 및 (2) 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함하는 pegRNA를 포함하고, 여기서 제1 프라임 편집제 복합체의 pegRNA의 DNA 합성 주형은 제1 단일 가닥 DNA 서열을 코딩하고, 제2 프라임 편집제 복합체의 pegRNA의 DNA 합성 주형은 제2 단일 가닥 DNA 서열을 코딩하고, 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고, 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 편집될 표적 부위에서의 DNA 서열과 비교하여 편집된 부분을 포함하는 듀플렉스를 형성하고, 이는 편집될 표적 부위 내로 통합되는 것인 시스템.

[1310] 2. 실시양태 1에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 시스템.

[1311] 3. 실시양태 1 또는 2에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 시스템.

[1312] 4. 실시양태 1 또는 2에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 상이한 프라임 편집제를 포함하는 것인 시스템.

[1313] 5. 실시양태 1-4 중 어느 하나에 있어서, napDNAbp가 Cas9 도메인 또는 그의 변이체인 시스템.

[1314] 6. 실시양태 1-5 중 어느 하나에 있어서, napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, Cas9 닉카제 도메인 또는 그의 변이체인 시스템.

[1315] 7. 실시양태 1-4 중 어느 하나에 있어서, napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, CasX, CasY 및 아르고노트로 이루어진 군으로부터 선택된 것인 시스템.

[1316] 8. 실시양태 1-7 중 어느 하나에 있어서, napDNAbp가 닉카제 활성을 갖는 것인 시스템.

[1317] 9. 실시양태 1-4 중 어느 하나에 있어서, napDNAbp가 서열식별번호: 18-88 중 어느 하나의 아미노산 서열 또는 서열식별번호: 18-88 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1318] 10. 실시양태 1-9 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드가 리버스 트랜스크립타제인 시스템.

[1319] 11. 실시양태 1-10 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드가 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159 및 700-766 중 어느 하나의 아미노산 서열 또는 서열식별번호: 89-100, 105-122, 128-129 및 132 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1320] 12. 실시양태 1-11 중 어느 하나에 있어서, 프라임 편집제가 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 시스템.

[1321] 13. 실시양태 12에 있어서, 링커가 서열식별번호: 166-177 중 어느 하나의 아미노산 서열 또는 서열식별번호: 166-177 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1322] 14. 실시양태 12 또는 13에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 시스템.

[1323] 15. 실시양태 1-14 중 어느 하나에 있어서, pegRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 시스템.

[1324] 16. 실시양태 1-15 중 어느 하나에 있어서, 제1 프라임 편집제 복합체의 pegRNA의 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 시스템.

[1325] 17. 실시양태 1-16 중 어느 하나에 있어서, 제2 프라임 편집제 복합체의 pegRNA의 스페이서 서열이 편집될 표

적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 시스템.

[1326] 18. 실시양태 17에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 pegRNA의 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

[1327] 19. 실시양태 18에 있어서, 제1 및 제2 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 시스템.

[1328] 20. 실시양태 19에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1329] 21. 실시양태 19에 있어서, 인접 영역이 편집된 부분을 포함하는 듀플렉스에 의해 대체되는 것인 시스템.

[1330] 22. 실시양태 1-21 중 어느 하나에 있어서, DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적

어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1331] 23. 실시양태 1-22 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 시스템.

[1332] 24. 실시양태 1-22 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 시스템.

[1333] 25. 실시양태 1-24 중 어느 하나에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 시스템.

[1334] 26. 실시양태 25에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이이고, 목적하는 DNA 서열 및 그의 상보체의 길이 이하인 시스템.

[1335] 27. 실시양태 25에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역이 편집된 부분을 포함하는 듀플렉스를 형성하는 것인 시스템.

[1336] 28. 실시양태 1-27 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 시스템.

[1337] 29. 실시양태 21에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 시스템.

[1338] 30. 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템으로서, 상기 시스템은 하기를 포함하고:

[1339] (a) 하기를 포함하는 제1 프라임 편집제 복합체:

[1340] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1341] ii. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[1342] (b) 하기를 포함하는 제2 프라임 편집제 복합체:

[1343] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1344] ii. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[1345] 여기서 제1 PEgRNA는 제1 단일 가닥 DNA 서열을 코딩하는 제1 DNA 합성 주형을 포함하고, 제2 PEgRNA는 제2 단일 가닥 DNA 서열을 코딩하는 제2 DNA 합성 주형을 포함하고, 여기서 이중 가닥 DNA 서열의 제1 가닥 및 제2 가닥은 서로 상보적인 것인 시스템.

[1346] 31. 실시양태 30에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 시스템.

[1347] 32. 실시양태 30에 있어서, 제1 및 제2 단일 가닥 DNA 서열이 각각 다른 것에 대한 상보성 영역을 포함하는 것인 시스템.

[1348] 33. 실시양태 30에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열이 편집된 부분을 포함하는 듀플렉스를 형성하고, 이는 편집될 표적 부위 내로 통합되는 는 것인 시스템.

[1349] 34. 실시양태 30에 있어서, 제1 단일 가닥 DNA 서열 및/또는 제2 단일 가닥 서열이 표적 부위에서의 DNA 서열과 비교하여 서열 상동성을 포함하지 않는 것인 시스템.

[1350] 35. 실시양태 30에 있어서, 제1 PEgRNA가 제1 스페이서 서열, 제1 gRNA 코어 및 제1 프라이머 결합 부위를 추가로 포함하고, 제2 PEgRNA가 제2 스페이서 서열, 제2 gRNA 코어 및 제2 프라이머 결합 부위를 추가로 포함하는 것인 시스템.

[1351] 36. 실시양태 30-35 중 어느 하나에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 시스템.

[1352] 37. 실시양태 30-35 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드가 동일한 것인 시스템.

[1353] 38. 실시양태 30-37 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9 도메인 또는 그의 변이체인 시스템.

[1354] 39. 실시양태 30-38 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, Cas9 닉카제 도메인 또는 그의 변이체인 시스템.

[1355] 40. 실시양태 30-39 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a,

Cas12b1, Cas13a, Cas12c, CasX, CasY 및 아르고노트로 이루어진 군으로부터 선택된 것인 시스템.

[1356] 41. 실시양태 30-40 중 어느 하나에 있어서, napDNAbp가 닉카제 활성을 갖는 것인 시스템.

[1357] 42. 실시양태 30-41 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 서열식별번호: 18-88 중 어느 하나의 아미노산 서열 또는 서열식별번호: 18-88 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1358] 43. 실시양태 30-42 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 리버스 트랜스크립타제인 시스템.

[1359] 44. 실시양태 30-43 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159 및 700-766 중 어느 하나의 아미노산 서열 또는 서열식별번호: 89-100, 105-122, 128-129 및 132 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1360] 45. 실시양태 30-44 중 어느 하나에 있어서, 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나; 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 시스템.

[1361] 46. 실시양태 45에 있어서, 링커가 서열식별번호: 166-177 중 어느 하나의 아미노산 서열 또는 서열식별번호: 166-177 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1362] 47. 실시양태 45 또는 46에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 시스템.

[1363] 48. 실시양태 30-47 중 어느 하나에 있어서, 제1 및/또는 제2 pegRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 시스템.

[1364] 49. 실시양태 30-48 중 어느 하나에 있어서, 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 시스템.

[1365] 50. 실시양태 49에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 시스템.

[1366] 51. 실시양태 50에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 제1 및/또는 제2 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

[1367] 52. 실시양태 51에 있어서, 제1 및 제2 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 시스템.

[1368] 53. 실시양태 52에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티

드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1369] 54. 실시양태 53에 있어서, 인접 영역이 편집된 부분을 포함하는 듀플렉스에 의해 대체되는 것인 시스템.

[1370] 55. 실시양태 30-54 중 어느 하나에 있어서, 제1 및/또는 제2 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1371] 56. 실시양태 30-55 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 시스템.

[1372] 57. 실시양태 30-55 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 시스템.

[1373] 58. 실시양태 30-57 중 어느 하나에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA

서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 시스템.

[1374] 59. 실시양태 30-58 중 어느 하나에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1375] 60. 실시양태 30-59 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

[1376] 61. 실시양태 30-60 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 시스템.

[1377] 62. 실시양태 54에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 시스템.

[1378] 63. 실시양태 1-62 중 어느 하나의 시스템을 코딩하는 폴리뉴클레오티드.

[1379] 64. 실시양태 1-62 중 어느 하나의 제1 및 제2 프라임 편집제를 코딩하는 폴리뉴클레오티드.

[1380] 65. 실시양태 1-29 중 어느 하나의 제1 및/또는 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

[1381] 66. 실시양태 30-62 중 어느 하나의 제1 및/또는 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

[1382] 67. 실시양태 1-29 중 어느 하나의 제1 및 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

[1383] 68. 실시양태 30-62 중 어느 하나의 제1 및 제2 pegRNA를 코딩하는 폴리뉴클레오티드.

[1384] 69. 실시양태 63-68 중 어느 하나의 폴리뉴클레오티드를 포함하는 벡터.

[1385] 70. 실시양태 63-69 중 어느 하나의 폴리뉴클레오티드 또는 실시양태 64의 벡터를 포함하는 세포.

[1386] 71. 실시양태 63-68 중 어느 하나의 폴리뉴클레오티드, 실시양태 69의 벡터 또는 실시양태 70의 세포 및 제약 부형제를 포함하는 제약 조성물.

[1387] 72. 실시양태 63-69 중 어느 하나의 폴리뉴클레오티드 또는 실시양태 70의 벡터 및 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하는 것에 대한 지침서를 포함하는 키트.

[1388] 73. 표적 부위에서 이중 가닥 DNA 서열의 제1 및 제2 상보적 가닥을 동시에 편집하는 방법으로서, 상기 방법은 이중 가닥 DNA 서열을 프라임 편집제 복합체의 쌍과 접촉시키는 것을 포함하고, 상기 쌍은 하기를 포함하고:

[1389] a. 하기를 포함하는 제1 프라임 편집제 복합체:

[1390] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1391] ii. 표적 부위의 상류의 게놈 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[1392] b. 하기를 포함하는 제2 프라임 편집제 복합체:

[1393] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1394] ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[1395] 여기서 제1 프라임 편집제 복합체는 제1 결합 부위에 상보적인 서열에서 제1 닉을 유발하고, 후속하여 제1 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제1 단일 가닥 DNA 서열의 중합을 유발하고;

[1396] 여기서 제2 프라임 편집제 복합체는 제2 결합 부위에 상보적인 서열에서 제2 닉을 유발하고, 후속하여 제2 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제2 단일 가닥 DNA 서열의 중합을 유발하고;

[1397] 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이고, 편집을 포함하는 듀플렉스를 형성하고;

[1398] 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하는 것인 방법.

[1399] 74. 실시양태 73에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 방법.

[1400] 75. 실시양태 73에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 방법.

[1401] 76. 실시양태 73에 있어서, 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체가 상이한 프라임 편집제를 포함하는 것인 방법.

[1402] 77. 실시양태 73-76 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9 도메인 또는 그의 변이체인 방법.

[1403] 78. 실시양태 73-77 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, Cas9 닉카제 도메인 또는 그의 변이체인 방법.

[1404] 79. 실시양태 73-76 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, CasX, CasY 및 아르고노트로 이루어진 군으로부터 선택된 것인 방법.

[1405] 80. 실시양태 73-79 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 닉카제 활성을 갖는 것인 방법.

[1406] 81. 실시양태 73-80 중 어느 하나에 있어서, 제1 및/또는 제2 napDNAbp가 서열식별번호: 18-88 중 어느 하나의 아미노산 서열 또는 서열식별번호: 18-88 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

[1407] 82. 실시양태 73-81 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 리버스 트랜스크립타제인 방법.

[1408] 83. 실시양태 73-82 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 및/또는 제2 폴리펩티드가 서열식별번호: 89-100, 105-122, 128-129, 132, 139, 143, 149, 154, 159 및 700-766 중 어느 하나의 아미노산 서열 또는 서열식별번호: 89-100, 105-122, 128-129 및 132 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

[1409] 84. 실시양태 73-83 중 어느 하나에 있어서, 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나, 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 방법.

[1410] 85. 실시양태 84에 있어서, 링커가 서열식별번호: 166-177 중 어느 하나의 아미노산 서열 또는 서열식별번호: 166-177 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

[1411] 86. 실시양태 84 또는 85에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 방법.

[1412] 87. 실시양태 73-86 중 어느 하나에 있어서, 제1 및/또는 제2 pegRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 방법.

[1413] 88. 실시양태 73-87 중 어느 하나에 있어서, 제1 및/또는 제2 PEgRNA가 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 325-330, 499-505, 101-104, 181-183 및 223-244 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 방법.

[1414] 89. 실시양태 73-88 중 어느 하나에 있어서, 제1 프라임 편집제 복합체의 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 방법.

[1415] 90. 실시양태 73-89 중 어느 하나에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 방법.

[1416] 91. 실시양태 90에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 제1 및 제2 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 방법.

[1417] 92. 실시양태 91에 있어서, 제1 및 제2 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 방법.

[1418] 93. 실시양태 92에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티

드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

[1419] 94. 실시양태 92 또는 93에 있어서, 인접 영역이 편집된 부분을 포함하는 듀플렉스에 의해 대체되는 것인 방법.

[1420] 95. 실시양태 73-94 중 어느 하나에 있어서, 제1 및/또는 제2 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

[1421] 96. 실시양태 73-95 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 방법.

[1422] 97. 실시양태 73-95 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 방법.

[1423] 98. 실시양태 73-97 중 어느 하나에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도

25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 방법.

[1424] 99. 실시양태 73-98 중 어느 하나에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이이고, 목적하는 DNA 서열 및 그의 상보체의 길이 이하인 방법.

[1425] 100. 실시양태 73-99 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 방법.

[1426] 101. 실시양태 73-100 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 방법.

[1427] 102. 실시양태 94에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 방법.

[1428] 103. 실시양태 73-102 중 어느 하나에 있어서, 편집을 포함하는 듀플렉스가 1개 이상의 레콤비나제 부위를 포함하는 것인 방법.

[1429] 104. 실시양태 103에 있어서, 편집될 표적 부위에서 DNA 서열의 레콤비나제-매개된 편집을 촉진하기 위해 레콤비나제를 제공하는 것을 추가로 포함하는 방법.

[1430] 105. 이중 프라임 편집에 사용하기 위한 pegRNA의 쌍으로서, 여기서 쌍은

[1431] a. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA)이며, 여기서 제1 PEgRNA는 제1 스페이서 서열, 제1 gRNA 코어, 제1 DNA 합성 주형 및 제1 프라이머 결합 부위를 포함하고, 여기서 제1 DNA 합성 주형은 편집된 부분을 포함하는 제1 단일 가닥 DNA 서열을 코딩하는 것인 제1 PEgRNA;

[1432] b. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA)이며, 여기서 제2 PEgRNA는 제2 스페이서 서열, 제2 gRNA 코어, 제2 DNA 합성 주형 및 제2 프라이머 결합 부위를 포함하고, 여기서 제2 DNA 합성 주형은 편집된 부분을 포함하는 제2 단일 가닥 DNA

서열을 코딩하고, 여기서 제2 단일 가닥 DNA 서열은 제1 단일 가닥 DNA 서열에 대한 상보성 영역을 포함하는 것인 제2 PEgRNA

[1433] 를 포함하는 것인 pegRNA의 쌍.

[1434] 106. 실시양태 105에 있어서, 제1 PEgRNA가 하기 구조:

[1435] 5'-[제1 gRNA 스페이서 서열]-[제1 gRNA 코어]-[제1 DNA 합성 주형]-[제1 프라이머 결합 부위]-3'를 포함하고,

[1436] 제2 PEgRNA가 하기 구조:

[1437] 5'-[제2 gRNA 스페이서 서열]-[제2 gRNA 코어]-[제2 DNA 합성 주형]-[제2 프라이머 결합 부위]-3'를 포함하는 것인 pegRNA의 쌍.

[1438] 107. 실시양태 105 또는 106에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열이 듀플렉스를 형성하고, 이는 편집될 표적 부위에서 이중 가닥 DNA 서열 내로 통합되는 것인 pegRNA의 쌍.

[1439] 108. 실시양태 107에 있어서, 듀플렉스의 통합이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 pegRNA의 쌍.

[1440] 109. 실시양태 108에 있어서, 삽입이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 pegRNA의 쌍.

[1441] 110. 실시양태 109에 있어서, 결실이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개

뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 pegRNA의 쌍.

[1442] 111. 실시양태 110에 있어서, 대체가 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적

어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 pegRNA의 쌍.

[1443] 112. 실시양태 105-111 중 어느 하나의 pegRNA의 쌍을 코딩하는 폴리뉴클레오티드.

[1444] 113. 세포에서 pegRNA의 쌍을 발현시키기 위한 실시양태 112의 폴리뉴클레오티드를 포함하는 벡터.

[1445] 114. 실시양태 113의 벡터를 포함하는 세포.

[1446] 115. 실시양태 105-111 중 어느 하나의 pegRNA의 쌍, 실시양태 113의 벡터 또는 실시양태 114의 세포 및 제약 부형제를 포함하는 제약 조성물.

[1447] 116. 프라임 편집제 복합체의 쌍으로서, 상기 쌍은

[1448] a. 하기를 포함하는 제1 프라임 편집제 복합체:

[1449] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1450] ii. 표적 부위의 상류의 게놈 DNA 서열의 가닥 상의 제1 표적 서열에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[1451] b. 하기를 포함하는 제2 프라임 편집제 복합체:

[1452] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1453] ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제2 표적 서열에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA)

[1454] 를 포함하는 것인 프라임 편집제 복합체의 쌍.

[1455] 117. 실시양태 116에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체 또는 제1 프라임 편집제 복합체 및 제2 프라임 편집제 복합체 둘 다의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 프라임 편집제 복합체의 쌍.

[1456] 118. 실시양태 116 또는 117의 프라임 편집제 복합체의 쌍을 코딩하는 폴리뉴클레오티드.

[1457] 119. 세포에서 프라임 편집제 복합체의 쌍을 발현시키기 위한 실시양태 118의 폴리뉴클레오티드를 포함하는 벡터.

[1458] 120. 실시양태 119의 벡터를 포함하는 세포.

[1459] 121. 실시양태 116 또는 117의 프라임 편집제 복합체의 쌍, 실시양태 119의 벡터 또는 실시양태 120의 세포 및 제약 부형제를 포함하는 제약 조성물.

[1460] 122. 실시양태 1-62 중 어느 하나 또는 실시양태 68-98 중 어느 하나에 있어서, 편집된 부분을 포함하는 듀플렉스의 통합이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 시스템 또는 방법.

[1461] 123. 실시양태 122에 있어서, 삽입이 펩티드 태그인 시스템 또는 방법.

[1462] 124. 실시양태 123에 있어서, 펩티드 태그가 폴리-히스티딘 (예를 들어, HHHHHH) (서열식별번호: 252-262), FLAG (예를 들어, DYKDDDDK) (서열식별번호: 2), V5 (예를 들어, GKPIPNPLLGLDST) (서열식별번호: 3), GCN4, HA (예를 들어, YPYDVPDYA) (서열식별번호: 5), Myc (예를 들어, EQKLISEED) (서열식별번호: 6) 또는 GST인 시스템 또는 방법.

[1463] 125. 실시양태 123에 있어서, 펩티드 태그가 서열식별번호: 1-6, 245-249, 252-262, 264-273, 275-276, 281, 278-288 및 622로 이루어진 군으로부터 선택된 아미노산 서열을 갖는 것인 시스템 또는 방법.

[1464] 126. 실시양태 122에 있어서, 삽입이 이뮤노에피토프 태그인 시스템 또는 방법.

[1465] 127. 실시양태 126에 있어서, 이뮤노에피토프 태그가 파상풍 톡소이드 (서열식별번호: 396); 디프테리아 독소 돌연변이체 CRM197 (서열식별번호: 630); 볼거리 이뮤노에피토프 1 (서열식별번호: 400); 볼거리 이뮤노에피토프 2 (서열식별번호: 402); 볼거리 이뮤노에피토프 3 (서열식별번호: 404); 풍진 바이러스 (서열식별번호:

406); 헤마글루티닌 (서열식별번호: 408); 뉴라미니다제 (서열식별번호: 410); TAP1 (서열식별번호: 412); TAP2 (서열식별번호: 414); 부류 I HLA에 대한 헤마글루티닌 에피토프 (서열식별번호: 416); 부류 I HLA에 대한 뉴라미니다제 에피토프 (서열식별번호: 418); 부류 II HLA에 대한 헤마글루티닌 에피토프 (서열식별번호: 420); 부류 II HLA에 대한 뉴라미니다제 에피토프 (서열식별번호: 422); 헤마글루티닌 에피토프 H5N1-결합된 부류 I 및 부류 II HLA (서열식별번호: 424); 및 뉴라미니다제 에피토프 H5N1-결합된 부류 I 및 부류 II HLA (서열식별번호: 426)로 이루어진 군으로부터 선택된 것인 시스템 또는 방법.

[1466] 128. 실시양태 122에 있어서, 삽입이 이량체화 도메인인 시스템 또는 방법.

[1467] 129. 실시양태 128에 있어서, 이량체화 도메인이 서열식별번호: 488의 FKBP12, 서열식별번호: 489의 FKBP12-F36V 또는 서열식별번호: 490 또는 493-494의 시클로필린의 소분자 결합 도메인인 시스템 또는 방법.

[1468] 130. 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하기 위한 시스템으로서, 시스템은 하기를 포함하고:

[1469] a) 하기를 포함하는 제1 프라임 편집제 복합체:

[1470] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1471] ii. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[1472] b) 하기를 포함하는 제2 프라임 편집제 복합체:

[1473] iii. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1474] iv. 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[1475] c) 하기를 포함하는 제3 프라임 편집제 복합체:

[1476] iii. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

[1477] iv. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA);

[1478] d) 하기를 포함하는 제4 프라임 편집제 복합체:

[1479] iii. 제4 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

[1480] iv. 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

[1481] 여기서 제1 PEgRNA는 제1 단일 가닥 DNA 서열을 코딩하는 제1 DNA 합성 주형을 포함하고, 제2 PEgRNA는 제2 단일 가닥 DNA 서열을 코딩하는 제2 DNA 합성 주형을 포함하고, 제3 PEgRNA는 제3 단일 가닥 DNA 서열을 코딩하는 제3 DNA 합성 주형을 포함하고, 제4 PEgRNA는 제4 단일 가닥 DNA 서열을 코딩하는 제4 DNA 합성 주형을 포함하고;

[1482] 여기서 제1 및 제3 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고; 여기서 제2 및 제4 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하는 것인 시스템.

[1483] 131. 1개 이상의 이중 가닥 DNA 서열을 편집하기 위한 시스템으로서, 시스템은 하기를 포함하고:

[1484] e) 하기를 포함하는 제1 프라임 편집제 복합체:

[1485] iii. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1486] iv. 편집될 제1 표적 부위에서 제1 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임

편집 가이드 RNA (제1 PEgRNA);

[1487] f) 하기를 포함하는 제2 프라임 편집제 복합체:

[1488] v. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1489] vi. 편집될 제1 표적 부위에서 제1 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[1490] g) 하기를 포함하는 제3 프라임 편집제 복합체:

[1491] v. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

[1492] vi. 편집될 제2 표적 부위에서 제2 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA);

[1493] h) 하기를 포함하는 제4 프라임 편집제 복합체:

[1494] v. 제4 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

[1495] vi. 편집될 제2 표적 부위에서 제2 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

[1496] 여기서 제1 PEgRNA는 제1 단일 가닥 DNA 서열을 코딩하는 제1 DNA 합성 주형을 포함하고, 제2 PEgRNA는 제2 단일 가닥 DNA 서열을 코딩하는 제2 DNA 합성 주형을 포함하고, 제3 PEgRNA는 제3 단일 가닥 DNA 서열을 코딩하는 제3 DNA 합성 주형을 포함하고, 제4 PEgRNA는 제4 단일 가닥 DNA 서열을 코딩하는 제4 DNA 합성 주형을 포함하고;

[1497] 여기서 제1 및 제3 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하고; 여기서 제2 및 제4 단일 가닥 DNA 서열은 각각 다른 것에 대한 상보성 영역을 포함하는 것인 시스템.

[1498] 132. 실시양태 130 또는 131에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및/또는 제4 프라임 편집제 복합체의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 시스템.

[1499] 133. 실시양태 129-132 중 어느 하나에 있어서,

[1500] (i) 제1 PEgRNA가 제1 스페이서 서열, 제1 gRNA 코어 및 제1 프라이머 결합 부위를 추가로 포함하고;

[1501] (ii) 제2 PEgRNA가 제2 스페이서 서열, 제2 gRNA 코어 및 제2 프라이머 결합 부위를 추가로 포함하고;

[1502] (iii) 제3 PEgRNA가 제3 스페이서 서열, 제3 gRNA 코어 및 제3 프라이머 결합 부위를 추가로 포함하고;

[1503] (iv) 제4 PEgRNA가 제4 스페이서 서열, 제4 gRNA 코어, 제4 DNA 합성 주형 및 제4 프라이머 결합 부위를 포함하는 것인 시스템.

[1504] 134. 실시양태 131에 있어서, 제1과 제3 단일 가닥 DNA 서열의 상보성 영역 및/또는 제2와 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

[1505] 135. 실시양태 134에 있어서, 제1 스페이서 서열이 편집될 제1 표적 부위의 상류의 제1 이중 가닥 DNA 서열의 제1 결합 부위에 결합하고, 제2 스페이서 서열이 편집될 제1 표적 부위의 하류의 제1 이중 가닥 DNA 서열 상의 제2 결합 부위에 결합하는 것인 시스템.

[1506] 136. 실시양태 134 또는 135에 있어서, 제3 스페이서 서열이 편집될 제2 표적 부위의 상류의 제2 이중 가닥 DNA 서열의 제1 결합 부위에 결합하고, 제4 스페이서 서열이 편집될 제2 표적 부위의 하류의 제2 이중 가닥 DNA 서열 상의 제2 결합 부위에 결합하는 것인 시스템.

[1507] 137. 실시양태 134-136 중 어느 하나에 있어서, 제1 및 제2 결합 부위가 제1 이중 가닥 DNA의 제1 인접 영역의 2개의 단부를 규정하는 것인 시스템.

[1508] 138. 실시양태 134-137 중 어느 하나에 있어서, 제3 및 제4 결합 부위가 제2 이중 가닥 DNA의 제2 인접 영역의

2개의 단부를 규정하는 것인 시스템.

[1509] 139. 실시양태 138에 있어서, 제1 및 제2 결합 부위 사이 또는 제3 및 제4 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1510] 140. 실시양태 134-140 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열의 상보성 영역 및/또는 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

[1511] 141. 실시양태 140에 있어서, 듀플렉스가 제1 이중 가닥 DNA 및/또는 제2 이중 가닥 DNA에 혼입되는 것인 시스템.

[1512] 142. 실시양태 134-141 중 어느 하나에 있어서, 제1 이중 가닥 DNA 내로 제1 인접 영역의 제2 인접 영역에 의한 대체를 발생시키는 시스템.

[1513] 143. 실시양태 134-141 중 어느 하나에 있어서, 제1 이중 가닥 DNA와 제2 이중 가닥 DNA 사이의 제1 인접 영역 및 제2 인접 영역의 교환을 발생시키는 시스템.

[1514] 144. 실시양태 130-143 중 어느 하나에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및 제4 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 시스템.

[1515] 145. 실시양태 130-144 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1, 제2, 제3 및 제4 폴리펩티드가 동일한 것인 시스템.

[1516] 146. 실시양태 130-145 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9 도메인 또는 그의 변이체인 시스템.

[1517] 147. 실시양태 130-146 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인 또는 Cas9 닉카제 도메인 또는 그의 변이체인 시스템.

[1518] 148. 실시양태 130-147 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c 및 아르고노트로 이루어진 군으로부터 선택된 것인 시스템.

[1519] 149. 실시양태 130-148 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 닉카제 활성을 갖는 것인 시스템.

[1520] 150. 실시양태 130-149 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 서열식별번호: 2-65 중 어느 하나의 아미노산 서열 또는 서열식별번호: 2-65 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1521] 151. 실시양태 130-150 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 리버스 트랜스크립타제인 시스템.

[1522] 152. 실시양태 130-151 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나의 아미노산 서열 또는 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1523] 153. 실시양태 130-152 중 어느 하나에 있어서,

[1524] (i) 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

[1525] (ii) 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

[1526] (iii) 제3 프라임 편집제가 제3 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제3 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

[1527] (iv) 제4 프라임 편집제가 제4 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제4 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 시스템.

[1528] 154. 실시양태 153에 있어서, 링커가 서열식별번호: 119-128 중 어느 하나의 아미노산 서열 또는 서열식별번호: 119-128 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 시스템.

[1529] 155. 실시양태 153 또는 154에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 시스템.

[1530] 156. 실시양태 130-155 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 PEgRNA가 서열식별번호: 192-203 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 192-203 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 시스템.

[1531] 157. 실시양태 130-156 중 어느 하나에 있어서, 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 시스템.

[1532] 158. 실시양태 130-157 중 어느 하나에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 시스템.

[1533] 159. 실시양태 130-158 중 어느 하나에 있어서, 제3 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 것인 시스템.

[1534] 160. 실시양태 130-159 중 어느 하나에 있어서, 제4 프라임 편집제 복합체의 제4 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 것인 시스템.

[1535] 161. 실시양태 160에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 제1 및/또는 제2 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

[1536] 162. 실시양태 161에 있어서, PAM 서열의 존재 하에서의 제3 및 제4 프라임 편집제 복합체의 스페이서 서열의 결합이 각각의 제3 및 제4 프라임 편집제 복합체의 제3 및/또는 제4 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 시스템.

[1537] 163. 실시양태 160 또는 162에 있어서, 제1 및 제3 결합 부위 또는 제2 및 제4 결합 부위가 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 시스템.

[1538] 164. 실시양태 163에 있어서, 제1 및 제3 결합 부위 사이 또는 제2 및 제4 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1539] 165. 실시양태 130-164 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열의 상보성 영역 및/또는 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 시스템.

[1540] 166. 실시양태 164 또는 165에 있어서, 인접 영역이 듀플렉스에 의해 대체되는 것인 시스템.

[1541] 167. 실시양태 164 또는 165에 있어서, 듀플렉스가 표적 부위에서 이중 가닥 DNA에 혼입되는 것인 시스템.

[1542] 168. 실시양태 130-167 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개

뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 시스템.

[1543] 169. 실시양태 130-168 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 시스템.

[1544] 170. 실시양태 130-168 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 시스템.

[1545] 171. 실시양태 130-170 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 단일 가닥 DNA 서열의 상보성 영역이 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열 또는 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 시스템.

[1546] 172. 실시양태 130-171 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 단일 가닥 DNA 서열의 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레

오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이이고, 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열, 및 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열의 길이 이하인 시스템.

[1547] 173. 실시양태 130-172 중 어느 하나에 있어서, 제1, 제2, 제3 및 제4 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 시스템.

[1548] 174. 실시양태 167에 있어서, 편집된 부분을 포함하는 듀플렉스의 혼입이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 시스템.

[1549] 175. 실시양태 130-174 중 어느 하나의 제1, 제2, 제3 및 제4 프라임 편집제를 코딩하는 폴리뉴클레오티드.

[1550] 176. 실시양태 130-174 중 어느 하나의 제1, 제2, 제3 및 제4 PEgRNA를 코딩하는 폴리뉴클레오티드.

[1551] 177. 실시양태 175 또는 176의 폴리뉴클레오티드를 포함하는 벡터.

[1552] 178. 실시양태 175 또는 176의 폴리뉴클레오티드 또는 실시양태 177의 벡터를 포함하는 세포.

[1553] 179. 실시양태 175 또는 176의 폴리뉴클레오티드, 실시양태 177의 벡터 또는 실시양태 178의 세포 및 제약 부형제를 포함하는 제약 조성물.

[1554] 180. 실시양태 175 또는 176의 폴리뉴클레오티드 또는 실시양태 177의 벡터 및 편집될 표적 부위에서 이중 가닥 DNA 서열의 두 가닥을 동시에 편집하는 것에 대한 지침서를 포함하는 키트.

[1555] 181. 표적 부위에서 이중 가닥 DNA 서열의 제1 및 제2 상보적 가닥을 동시에 편집하는 방법으로서, 방법은 이중 가닥 DNA 서열을 프라임 편집제 복합체의 쌍과 접촉시키는 것을 포함하고, 쌍은 하기를 포함하고:

[1556] (a) 하기를 포함하는 제1 프라임 편집제 복합체:

[1557] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1558] ii. 표적 부위의 상류의 게놈 DNA 서열의 제1 가닥 상의 제1 표적 서열에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[1559] (b) 하기를 포함하는 제2 프라임 편집제 복합체:

[1560] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1561] ii. 표적 부위의 상류의 게놈 DNA 서열의 제2 가닥 상의 제2 표적 서열에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[1562] (c) 하기를 포함하는 제3 프라임 편집제 복합체:

[1563] i. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

[1564] ii. 표적 부위의 하류의 게놈 DNA 서열의 제1 가닥 상의 제3 표적 서열에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA);

[1565] (d) 하기를 포함하는 제4 프라임 편집제 복합체:

[1566] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

[1567] ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제4 표적 서열에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

[1568] 여기서 제1 프라임 편집제 복합체는 제1 표적 서열에서 제1 닉을 유발하고, 후속하여 제1 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제1 단일 가닥 DNA 서열의 중합을 유발하고;

[1569] 여기서 제2 프라임 편집제 복합체는 제2 표적 서열에서 제2 닉을 유발하고, 후속하여 제2 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제2 단일 가닥 DNA 서열의 중합을 유발하고;

[1570] 여기서 제3 프라임 편집제 복합체는 제3 표적 서열에서 제3 닉을 유발하고, 후속하여 제3 닉에 의해 형성된 이

용가능한 5'-단부로부터 3'-단부를 갖는 제3 단일 가닥 DNA 서열의 중합을 유발하고;

[1571] 여기서 제4 프라임 편집제 복합체는 제4 표적 서열에서 제4 닉을 유발하고, 후속하여 제4 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제4 단일 가닥 DNA 서열의 중합을 유발하고;

[1572] 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이며 듀플렉스를 형성하고, 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하고;

[1573] 여기서 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이며 듀플렉스를 형성하고, 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하는 것인 방법.

[1574] 182. 실시양태 181에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및/또는 제4 프라임 편집제 복합체의 프라임 편집제가 napDNAbp 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드를 포함하는 융합 단백질인 방법.

[1575] 183. 실시양태 181 또는 182에 있어서, 제1 프라임 편집제 복합체, 제2 프라임 편집제 복합체, 제3 프라임 편집제 복합체 및 제4 프라임 편집제 복합체가 동일한 프라임 편집제를 포함하는 것인 방법.

[1576] 184. 실시양태 181 또는 182에 있어서, 제1, 제2, 제3 및/또는 제4 프라임 편집제 복합체가 상이한 프라임 편집제를 포함하는 것인 방법.

[1577] 185. 실시양태 181-184 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9 도메인 또는 그의 변이체인 방법.

[1578] 186. 실시양태 181-185 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인 또는 Cas9 닉카제 도메인 또는 그의 변이체인 방법.

[1579] 187. 실시양태 181-184 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c로 이루어진 군으로부터 선택된 것인 방법.

[1580] 188. 실시양태 181-187 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 닉카제 활성을 갖는 것인 방법.

[1581] 189. 실시양태 181-188 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 napDNAbp가 서열식별번호: 2-65 중 어느 하나의 아미노산 서열 또는 서열식별번호: 2-65 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

[1582] 190. 실시양태 181-189 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 리버스 트랜스크립타제인 방법.

[1583] 191. 실시양태 181-190 중 어느 하나에 있어서, RNA-의존성 DNA 폴리머라제 활성을 갖는 제1, 제2, 제3 및/또는 제4 폴리펩티드가 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나의 아미노산 서열 또는 서열식별번호: 37, 68-79, 82-98, 81, 98 및 110 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 방법.

[1584] 192. 실시양태 181-191 중 어느 하나에 있어서,

[1585] (i) 제1 프라임 편집제가 제1 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제1 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

[1586] (ii) 제2 프라임 편집제가 제2 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제2 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

[1587] (iii) 제3 프라임 편집제가 제3 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제3 폴리펩티드를 연결하는 링커를 추가로 포함하고/거나;

[1588] (iv) 제4 프라임 편집제가 제4 napDNAbp와 RNA-의존성 DNA 폴리머라제 활성을 갖는 제4 폴리펩티드를 연결하는 링커를 추가로 포함하는 것인 방법.

[1589] 193. 실시양태 192에 있어서, 링커가 서열식별번호: 119-128 중 어느 하나의 아미노산 서열 또는 서열식별번호: 119-128 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 아미노산 서열

을 포함하는 것인 방법.

[1590] 194. 실시양태 192 또는 193에 있어서, 링커가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 아미노산 길이인 방법.

[1591] 195. 실시양태 181-191 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 PEgRNA가 서열식별번호: 192-203 중 어느 하나의 뉴클레오티드 서열 또는 서열식별번호: 192-203 중 어느 하나와 적어도 80%, 85%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 방법.

[1592] 196. 실시양태 181-195 중 어느 하나에 있어서, 제1 프라임 편집제 복합체의 제1 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 것인 방법.

[1593] 197. 실시양태 181-196 중 어느 하나에 있어서, 제2 프라임 편집제 복합체의 제2 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 것인 방법.

[1594] 198. 실시양태 181-197 중 어느 하나에 있어서, 제3 프라임 편집제 복합체의 제3 스페이서 서열이 편집될 표적 부위로부터 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 것인 방법.

[1595] 199. 실시양태 181-198 중 어느 하나에 있어서, 제4 프라임 편집제 복합체의 제4 스페이서 서열이 편집될 표적 부위로부터 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 것인 방법.

[1596] 200. 실시양태 197에 있어서, PAM 서열의 존재 하에서의 제1 및 제2 프라임 편집제 복합체의 제1 및 제2 스페이서 서열의 결합이 각각의 제1 및 제2 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 방법.

[1597] 201. 실시양태 199에 있어서, PAM 서열의 존재 하에서의 제3 및 제4 프라임 편집제 복합체의 제3 및 제4 스페이서 서열의 결합이 각각의 제3 및 제4 프라임 편집제 복합체의 napDNAbp에 의해 각각의 가닥 상의 PAM 서열에 근접한 닉 부위에서 각각 제1 및 제2 가닥의 닉킹을 발생시키는 것인 방법.

[1598] 202. 실시양태 200 또는 201에 있어서, 제1 및 제2 결합 부위 및/또는 제3 및 제4 결합 부위가 편집될 표적 부위를 포함하는 이중 가닥 DNA의 인접 영역의 2개의 단부를 규정하는 것인 방법.

[1599] 203. 실시양태 202에 있어서, 제1 및 제2 결합 부위 사이의 인접 영역 및/또는 제3 및 제4 결합 부위 사이의 인접 영역이 적어도 10개 핵염기 쌍, 적어도 11개 핵염기 쌍, 적어도 12개 핵염기 쌍, 적어도 13개 핵염기 쌍, 적어도 14개 핵염기 쌍, 적어도 15개 핵염기 쌍, 적어도 16개 핵염기 쌍, 적어도 17개 핵염기 쌍, 적어도 18개 핵염기 쌍, 적어도 19개 핵염기 쌍, 적어도 20개 핵염기 쌍, 적어도 21개 핵염기 쌍, 적어도 22개 핵염기 쌍, 적어도 23개 핵염기 쌍, 적어도 24개 핵염기 쌍, 적어도 25개 핵염기 쌍, 적어도 26개 핵염기 쌍, 적어도 27개 핵염기 쌍, 적어도 28개 핵염기 쌍, 적어도 29개 핵염기 쌍, 적어도 30개 핵염기 쌍, 적어도 31개 핵염기 쌍, 적어도 32개 핵염기 쌍, 적어도 33개 핵염기 쌍, 적어도 34개 핵염기 쌍, 적어도 35개 핵염기 쌍, 적어도 36개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 37개 핵염기 쌍, 적어도 38개 핵염기 쌍, 적어도 39개 핵염기 쌍 또는 적어도 40개 핵염기 쌍, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티

드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

[1600] 204. 실시양태 202 또는 203에 있어서, 인접 영역이 듀플렉스에 의해 대체되는 것인 방법.

[1601] 205. 실시양태 181-204 중 어느 하나에 있어서, 제1, 제2, 제3 및/또는 제4 DNA 합성 주형이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

[1602] 206. 실시양태 181-205 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 동일한 길이를 갖는 것인 방법.

[1603] 207. 실시양태 181-206 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 상이한 길이를 갖는 것인 방법.

[1604] 208. 실시양태 181-207 중 어느 하나에 있어서, 상보성 영역이 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열의 총 길이의 적어도 99%, 적어도 95%, 적어도 90%, 적어도 85%, 적어도 80%, 적어도 75%, 적어도 70%, 적어도 65%, 적어도 60%, 적어도 55%, 적어도 50%, 적어도 45%, 적어도 40%, 적어도 35%, 적어도 30%, 적어도 25%, 적어도 20%, 적어도 15% 또는 적어도 10%인 방법.

[1605] 209. 실시양태 181-208 중 어느 하나에 있어서, 상보성 영역이 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티

드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 방법.

[1606] 210. 실시양태 181-209 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열의 상보성 영역이 듀플렉스를 형성하는 것인 방법.

[1607] 211. 실시양태 181-210 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 각각 3' 단부를 갖는 단일 가닥 DNA 가닥인 방법.

[1608] 212. 실시양태 204에 있어서, 편집된 부분을 포함하는 듀플렉스에 의한 인접 영역의 대체가 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 방법.

[1609] 213. 실시양태 181-212 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열이 표적 DNA 서열의 단부 상에 있고, 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 동일한 표적 DNA 서열의 대향 단부 상에 있는 것인 방법.

[1610] 214. 실시양태 213에 있어서, 표적 DNA 서열이 역전된 것인 방법.

[1611] 215. 실시양태 181-214 중 어느 하나에 있어서, 원형 DNA 공여자를 제공하는 것을 추가로 포함하는 방법.

[1612] 216. 실시양태 215에 있어서, 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열이 표적 DNA 서열의 대향 단부 상에 있고, 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 원형 DNA 공여자 상에 있는 것인 방법.

[1613] 217. 실시양태 216에 있어서, 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 원형 DNA 공여자의 부분이 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 표적 서열을 대체하는 것인 방법.

[1614] 218. 실시양태 181-214 중 어느 하나에 있어서, 제1 단일 가닥 DNA 서열 및 제3 단일 가닥 DNA 서열이 제1 핵산 분자 상에 있고, 제2 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열이 제2 핵산 분자 상에 있는 것인 방법.

[1615] 219. 실시양태 218에 있어서, 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 제1 핵산 분자의 부분이 제2 핵산 분자 내로 혼입되고, 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 제2 핵산 분자의 부분이 제1 핵산 분자 내로 혼입되는 것인 방법.

[1616] 220. 실시양태 218 또는 219에 있어서, 제1 핵산 분자가 제1 염색체이고, 제2 핵산 분자가 제2 염색체인 방법.

[1617] 221. 다중-플랩 프라임 편집에 사용하기 위한 PEgRNA의 세트로서, 여기서 PEgRNA는

[1618] (a) 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA)이며, 여기서 제1 PEgRNA는 제1 스페이서 서열, 제1 gRNA 코어, 제1 DNA 합성 주형

및 제1 프라이머 결합 부위를 포함하고, 여기서 제1 DNA 합성 주형은 제1 단일 가닥 DNA 서열을 코딩하는 것인 제1 PEgRNA;

[1619] (b) 편집될 표적 부위의 상류의 이중 가닥 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA)이며, 여기서 제2 PEgRNA는 제2 스페이서 서열, 제2 gRNA 코어, 제2 DNA 합성 주형 및 제2 프라이머 결합 부위를 포함하고, 여기서 제2 DNA 합성 주형은 편집된 부분을 포함하는 제2 단일 가닥 DNA 서열을 코딩하고, 여기서 제2 단일 가닥 DNA 서열은 제1 단일 가닥 DNA 서열에 상보적인 것인 제2 PEgRNA;

[1620] (c) 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제1 가닥 상의 제3 결합 부위에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA)이며, 여기서 제3 PEgRNA는 제3 스페이서 서열, 제3 gRNA 코어, 제3 DNA 합성 주형 및 제3 프라이머 결합 부위를 포함하고, 여기서 제3 DNA 합성 주형은 제3 단일 가닥 DNA 서열을 코딩하는 것인 제3 PEgRNA;

[1621] (d) 편집될 표적 부위의 하류의 이중 가닥 DNA 서열의 제2 가닥 상의 제4 결합 부위에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA)이며, 여기서 제4 PEgRNA는 제4 스페이서 서열, 제4 gRNA 코어, 제4 DNA 합성 주형 및 제4 프라이머 결합 부위를 포함하고, 여기서 제4 DNA 합성 주형은 제4 단일 가닥 DNA 서열을 코딩하고, 여기서 제4 단일 가닥 DNA 서열은 제3 단일 가닥 DNA 서열에 대한 상보성 영역을 포함하는 것인 제4 PEgRNA

[1622] 를 포함하는 것인 PEgRNA의 세트.

[1623] 222. 실시양태 221에 있어서,

[1624] 제1 PEgRNA가 하기 구조:

[1625] 5'-[제1 gRNA 스페이서 서열]-[제1 gRNA 코어]-[제1 DNA 합성 주형]-[제1 프라이머 결합 부위]-3'를 포함하고;

[1626] 제2 PEgRNA가 하기 구조:

[1627] 5'-[제2 gRNA 스페이서 서열]-[제2 gRNA 코어]-[제2 DNA 합성 주형]-[제2 프라이머 결합 부위]-3'를 포함하고;

[1628] 제3 PEgRNA가 하기 구조:

[1629] 5'-[제3 gRNA 스페이서 서열]-[제3 gRNA 코어]-[제3 DNA 합성 주형]-[제3 프라이머 결합 부위]-3'를 포함하고;

[1630] 제4 PEgRNA가 하기 구조:

[1631] 5'-[제4 gRNA 스페이서 서열]-[제4 gRNA 코어]-[제4 DNA 합성 주형]-[제4 프라이머 결합 부위]-3'를 포함하는 것인 PEgRNA의 세트.

[1632] 223. 실시양태 221 또는 222에 있어서, 제1 단일 가닥 DNA 서열과 제2 단일 가닥 DNA 서열 및/또는 제3 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열이 듀플렉스를 형성하고, 이는 편집될 표적 부위에서 이중 가닥 DNA 서열 내로 통합되는 것인 PEgRNA의 세트.

[1633] 224. 실시양태 221-223 중 어느 한 실시양태에 있어서, 듀플렉스의 통합이 표적 부위에서 DNA의 삽입, 결실 또는 대체를 발생시키는 것인 PEgRNA의 세트.

[1634] 225. 실시양태 224에 있어서, 삽입이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개

뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 길이인 PEgRNA의 세트.

[1635] 226. 실시양태 224에 있어서, 결실이 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개 뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 PEgRNA의 세트.

[1636] 227. 실시양태 224에 있어서, 대체가 적어도 1개 뉴클레오티드, 적어도 2개 뉴클레오티드, 적어도 3개 뉴클레오티드, 적어도 4개 뉴클레오티드, 적어도 5개 뉴클레오티드, 적어도 6개 뉴클레오티드, 적어도 7개 뉴클레오티드, 적어도 8개 뉴클레오티드, 적어도 9개 뉴클레오티드, 적어도 10개 뉴클레오티드, 적어도 11개 뉴클레오티드, 적어도 12개 뉴클레오티드, 적어도 13개 뉴클레오티드, 적어도 14개 뉴클레오티드, 적어도 15개 뉴클레오티드, 적어도 16개 뉴클레오티드, 적어도 17개 뉴클레오티드, 적어도 18개 뉴클레오티드, 적어도 19개 뉴클레오티드, 적어도 20개 뉴클레오티드, 적어도 21개 뉴클레오티드, 적어도 22개 뉴클레오티드, 적어도 23개

뉴클레오티드, 적어도 24개 뉴클레오티드, 적어도 25개 뉴클레오티드, 적어도 26개 뉴클레오티드, 적어도 27개 뉴클레오티드, 적어도 28개 뉴클레오티드, 적어도 29개 뉴클레오티드, 적어도 30개 뉴클레오티드, 적어도 31개 뉴클레오티드, 적어도 32개 뉴클레오티드, 적어도 33개 뉴클레오티드, 적어도 34개 뉴클레오티드, 적어도 35개 뉴클레오티드, 적어도 36개 뉴클레오티드, 적어도 37개 뉴클레오티드, 적어도 38개 뉴클레오티드, 적어도 39개 뉴클레오티드, 적어도 40개 뉴클레오티드, 적어도 41개 뉴클레오티드, 적어도 42개 뉴클레오티드, 적어도 43개 뉴클레오티드, 적어도 44개 뉴클레오티드, 적어도 45개 뉴클레오티드, 적어도 46개 뉴클레오티드, 적어도 47개 뉴클레오티드, 적어도 48개 뉴클레오티드, 적어도 49개 뉴클레오티드, 적어도 50개 뉴클레오티드, 적어도 51개 뉴클레오티드, 적어도 52개 뉴클레오티드, 적어도 53개 뉴클레오티드, 적어도 54개 뉴클레오티드, 적어도 55개 뉴클레오티드, 적어도 56개 뉴클레오티드, 적어도 57개 뉴클레오티드, 적어도 58개 뉴클레오티드, 적어도 59개 뉴클레오티드, 적어도 60개 뉴클레오티드, 적어도 61개 뉴클레오티드, 적어도 62개 뉴클레오티드, 적어도 63개 뉴클레오티드, 적어도 64개 뉴클레오티드, 적어도 65개 뉴클레오티드, 적어도 66개 뉴클레오티드, 적어도 67개 뉴클레오티드, 적어도 68개 뉴클레오티드, 적어도 69개 뉴클레오티드, 적어도 70개 뉴클레오티드, 적어도 71개 뉴클레오티드, 적어도 72개 뉴클레오티드, 적어도 73개 뉴클레오티드, 적어도 74개 뉴클레오티드, 적어도 75개 뉴클레오티드, 적어도 76개 뉴클레오티드, 적어도 77개 뉴클레오티드, 적어도 78개 뉴클레오티드, 적어도 79개 뉴클레오티드, 적어도 80개 뉴클레오티드, 적어도 81개 뉴클레오티드, 적어도 82개 뉴클레오티드, 적어도 83개 뉴클레오티드, 적어도 84개 뉴클레오티드, 적어도 85개 뉴클레오티드, 적어도 86개 뉴클레오티드, 적어도 87개 뉴클레오티드, 적어도 88개 뉴클레오티드, 적어도 89개 뉴클레오티드, 적어도 90개 뉴클레오티드, 적어도 91개 뉴클레오티드, 적어도 92개 뉴클레오티드, 적어도 93개 뉴클레오티드, 적어도 94개 뉴클레오티드, 적어도 95개 뉴클레오티드, 적어도 96개 뉴클레오티드, 적어도 97개 뉴클레오티드, 적어도 98개 뉴클레오티드, 적어도 99개 뉴클레오티드, 적어도 100개 뉴클레오티드, 적어도 101개 뉴클레오티드, 적어도 102개 뉴클레오티드, 적어도 103개 뉴클레오티드, 적어도 104개 뉴클레오티드, 적어도 105개 뉴클레오티드, 적어도 106개 뉴클레오티드, 적어도 107개 뉴클레오티드, 적어도 108개 뉴클레오티드, 적어도 109개 뉴클레오티드 또는 적어도 110개 뉴클레오티드 또는 그 초과의 길이인 PEgRNA의 세트.

[1637] 228. 실시양태 221-227 중 어느 한 실시양태의 PEgRNA의 세트를 코딩하는 폴리뉴클레오티드.

[1638] 229. 세포에서 복수의 PEgRNA를 발현시키기 위한 실시양태 228의 폴리뉴클레오티드를 포함하는 벡터.

[1639] 230. 실시양태 229의 벡터를 포함하는 세포.

[1640] 231. 실시양태 221-227 중 어느 하나의 복수의 PEgRNA, 실시양태 229의 벡터 또는 실시양태 230의 세포 및 제약 부형제를 포함하는 제약 조성물.

[1641] 232. 세포 내 게놈을 변형시키는 방법으로서,

[1642] (i) 게놈을 이중-플랩 프라임 편집 시스템과 접촉시키는 단계이며, 상기 시스템은 표적 부위의 제1 가닥 상의 제1 닉 부위에 제1 3' 핵산 플랩을 설치할 수 있는 제1 프라임 편집제/pegRNA 복합체 및 표적 부위의 제2 가닥 상의 제2 닉 부위에 제2 3' 핵산 플랩을 설치할 수 있는 제2 프라임 편집제/pegRNA 복합체를 포함하고,

[1643] 여기서 제1 3' 핵산 플랩 및 제2 3' 핵산 플랩은 듀플렉스를 형성하는 역 상보체 서열이고,

[1644] 여기서 듀플렉스는 제1 레콤비나제 부위를 포함하고,

[1645] 여기서 세포는 제1 및 제2 닉 부위 사이에 위치한 내인성 듀플렉스 대신에 표적 부위에 제1 레콤비나제 부위를 포함하는 듀플렉스를 설치하는 것인 단계;

[1646] (ii) 제1 레콤비나제 부위를 제2 레콤비나제 부위를 포함하는 공여자 핵산과 접촉시키는 단계; 및

[1647] (iii) 게놈을 제1 및 제2 레콤비나제 부위 사이의 재조합이 가능한 레콤비나제와 접촉시킴으로써, 게놈 내 표적 부위에서 공여자 핵산을 통합시키는 단계

[1648] 를 포함하는 방법.

[1649] [9] 방법

[1650] 다중-플랩 프라임 편집 방법

[1651] 한 측면에서, 본 개시내용은, 표적 부위에서 이중 가닥 DNA 서열의 제1 및 제2 상보적 가닥을 동시에 편집하는 방법으로서, 상기 방법은 이중 가닥 DNA 서열을 프라임 편집제 복합체의 쌍과 접촉시키는 것을 포함하고, 상기

쌍은 하기를 포함하고:

[1652] a. 하기를 포함하는 제1 프라임 편집제 복합체:

[1653] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1654] ii. 표적 부위의 상류의 게놈 DNA 서열의 제1 가닥 상의 제1 결합 부위에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[1655] b. 하기를 포함하는 제2 프라임 편집제 복합체:

[1656] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1657] ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제2 결합 부위에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[1658] 여기서 제1 프라임 편집제 복합체는 제1 결합 부위에 상보적인 서열에서 제1 닉을 유발하고, 후속하여 제1 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제1 단일 가닥 DNA 서열의 중합을 유발하고;

[1659] 여기서 제2 프라임 편집제 복합체는 제2 결합 부위에 상보적인 서열에서 제2 닉을 유발하고, 후속하여 제2 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제2 단일 가닥 DNA 서열의 중합을 유발하고;

[1660] 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이고, 편집을 포함하는 듀플렉스를 형성하고;

[1661] 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하는 것인 방법을 제공한다.

[1662] 또 다른 측면에서, 본 개시내용은, 표적 부위에서 이중 가닥 DNA 서열의 제1 및 제2 상보적 가닥을 동시에 편집하는 방법으로서, 방법은 이중 가닥 DNA 서열을 프라임 편집제 복합체의 쌍과 접촉시키는 것을 포함하고, 쌍은 하기를 포함하고:

[1663] (a) 하기를 포함하는 제1 프라임 편집제 복합체:

[1664] i. 제1 핵산 프로그램가능한 DNA 결합 단백질 (제1 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제1 폴리펩티드를 포함하는 제1 프라임 편집제; 및

[1665] ii. 표적 부위의 상류의 게놈 DNA 서열의 제1 가닥 상의 제1 표적 서열에 결합하는 제1 프라임 편집 가이드 RNA (제1 PEgRNA);

[1666] (b) 하기를 포함하는 제2 프라임 편집제 복합체:

[1667] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제2 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제2 폴리펩티드를 포함하는 제2 프라임 편집제; 및

[1668] ii. 표적 부위의 상류의 게놈 DNA 서열의 제2 가닥 상의 제2 표적 서열에 결합하는 제2 프라임 편집 가이드 RNA (제2 PEgRNA);

[1669] (c) 하기를 포함하는 제3 프라임 편집제 복합체:

[1670] i. 제3 핵산 프로그램가능한 DNA 결합 단백질 (제3 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제3 폴리펩티드를 포함하는 제3 프라임 편집제; 및

[1671] ii. 표적 부위의 하류의 게놈 DNA 서열의 제1 가닥 상의 제3 표적 서열에 결합하는 제3 프라임 편집 가이드 RNA (제3 PEgRNA);

[1672] (d) 하기를 포함하는 제4 프라임 편집제 복합체:

[1673] i. 제2 핵산 프로그램가능한 DNA 결합 단백질 (제4 napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 포함하는 제4 폴리펩티드를 포함하는 제4 프라임 편집제; 및

[1674] ii. 표적 부위의 하류의 게놈 DNA 서열의 제2 가닥 상의 제4 표적 서열에 결합하는 제4 프라임 편집 가이드 RNA (제4 PEgRNA);

[1675] 여기서 제1 프라임 편집제 복합체는 제1 표적 서열에서 제1 닉을 유발하고, 후속하여 제1 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제1 단일 가닥 DNA 서열의 중합을 유발하고;

[1676] 여기서 제2 프라임 편집제 복합체는 제2 표적 서열에서 제2 닉을 유발하고, 후속하여 제2 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제2 단일 가닥 DNA 서열의 중합을 유발하고;

[1677] 여기서 제3 프라임 편집제 복합체는 제3 표적 서열에서 제3 닉을 유발하고, 후속하여 제3 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제3 단일 가닥 DNA 서열의 중합을 유발하고;

[1678] 여기서 제4 프라임 편집제 복합체는 제4 표적 서열에서 제4 닉을 유발하고, 후속하여 제4 닉에 의해 형성된 이용가능한 5'-단부로부터 3'-단부를 갖는 제4 단일 가닥 DNA 서열의 중합을 유발하고;

[1679] 여기서 제1 단일 가닥 DNA 서열 및 제2 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이며 듀플렉스를 형성하고, 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하고;

[1680] 여기서 제3 단일 가닥 DNA 서열 및 제4 단일 가닥 DNA 서열은 적어도 상보성 영역에 걸쳐 역 상보체이며 듀플렉스를 형성하고, 여기서 듀플렉스는 이중 가닥 DNA 서열의 닉킹된 제1 및 제2 상보적 가닥을 대체하는 것인 방법을 제공한다.

[1681] 본 명세서는 두 DNA 가닥을 동시에 편집함으로써 플랩 평형 및 비-편집된 상보적 게놈 DNA 가닥 내로의 편집의 후속 혼입과 연관된 도전과제를 다루는 다중-플랩 프라임 편집 시스템 (예를 들어, 이중-플랩 프라임 편집 시스템 및 사중-플랩 프라임 편집 시스템 포함)을 기재한다. 이중-플랩 프라임 편집 시스템에서, 2개의 PEgRNA는 게놈 부위의 대향 가닥을 표적화하고 편집된 DNA 서열을 함유하는 2개의 상보적 3' 플랩의 합성을 지시하는 데 사용된다 (도 90). 사중-플랩 프라임 편집 시스템에서, 4개의 PEgRNA가 사용되고, 4개의 3' 플랩 (이들 중 2개는 서로에 대해 상보적이고, 다른 2개는 서로에 대해 상보적임)의 합성을 지시한다 (도 95, 도 96a 및 도 98a). 전형적 프라임 편집과는 달리, 상보적인 편집된 가닥이 대신 혼성화에 이용가능하기 때문에, 편집된 DNA 가닥의 쌍 (3' 플랩)이 내인성 게놈 DNA에서 5' 플랩과 직접 경쟁하는 것이 요구되지 않는다. 듀플렉스의 두 가닥이 편집된 DNA로서 합성되기 때문에, 다중-플랩 프라임 편집 시스템은 전형적 프라임 편집에 의해 요구되는 비-편집된 상보적 DNA 가닥의 대체에 대한 필요를 제거한다. 대신, 세포 DNA 복구 기구는 단지 쌍형성된 5' 플랩 (원래 게놈 DNA)만을 절제하고, 쌍형성된 3' 플랩 (편집된 DNA)을 유전자좌 내로 라이게이션시킬 필요가 있다. 따라서, 또한 새로 합성된 DNA 가닥에 게놈 DNA에 상동인 서열을 포함할 필요가 없으므로, 새로운 가닥의 선택적 혼성화를 허용하고 최소 게놈 상동성을 함유하는 편집을 용이하게 한다. DNA의 두 가닥을 절단하는 다중-플랩 프라임 편집제의 뉴클레아제-활성 버전은 또한 원래 DNA 서열의 제거를 가속화하는 데 사용될 수 있다.

[1682] 전형적 프라임 편집과 같이, 다중-플랩 프라임 편집은 핵산 프로그램가능한 DNA 결합 단백질 ("napDNAbp") (이는 폴리머라제와 회합하여 (즉, 융합 단백질의 형태로 또는 달리 napDNAbp와 트랜스로 제공됨) 작동함)을 사용하여 특정된 DNA 부위 내로 새로운 유전자 정보를 직접 기록하는 다재다능한 정확한 게놈 편집 방법이며, 여기서 프라임 편집 시스템은 표적 부위를 특정하고 또한 가이드 RNA 상에 (예를 들어, 5' 또는 3' 단부에 또는 가이드 RNA의 내부 부분에) 조작된 연장부 (DNA 또는 RNA)에 의해 대체 DNA 가닥의 형태로 목적하는 편집의 합성의 주형이 되는 프라임 편집 (PE) 가이드 RNA ("PEgRNA")로 프로그램화된다. 목적하는 편집 (예를 들어, 단일 핵염기 치환)을 함유하는 대체 가닥은 (목적하는 편집을 포함하는 것을 제외하고는) 편집될 표적 부위의 내인성 가닥과 동일한 서열을 공유한다. DNA 복구 및/또는 복제 기구를 통해, 표적 부위의 내인성 가닥은 목적하는 편집을 함유하는 새로 합성된 대체 가닥에 의해 대체된다. 일부 경우에, 프라임 편집은 "검색-및-대체" 게놈 편집 기술로서 생각될 수 있으며, 이는 이중 프라임 편집제가, 본원에 기재된 바와 같이, 편집될 목적하는 표적 부위를 검색하고 위치를 찾아낼 뿐만 아니라, 동시에, 상응하는 표적 부위 내인성 DNA 가닥 대신 설치된 목적하는 편집을 함유하는 대체 가닥을 코딩하기 때문이다.

[1683] 이중-플랩 프라임 편집에서, 이중 가닥 DNA 서열은 표적 부위에서 제1 및 제2 프라임 편집제 복합체와 접촉된다. 각각의 복합체는 융합 단백질 및 PEgRNA를 포함한다. 일부 실시양태에서, 각각의 융합 단백질은 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드 (예를 들어, 리버스 트랜스크립타제)를 포함하고, 각각의 PEgRNA는 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 각각의 DNA 합성 주형은 1개 이상의 뉴클레오티드의 편집된 부분을 포함할 수 있는 단일 가닥 DNA 서열을 코딩할 수 있다. 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대해 상보적일 수 있고, 편집될 표적 부위 내로 통합될 수 있는 듀플렉스를 형성할 수 있다. 프라임 편집제 복합체의 다양한 요소 (예를 들어, 융합 단백질, napDNAbp, 폴리머라제, PEgRNA 등)는 본원에 개시된 시스템의 임의의 실시양태를 포함할

수 있다.

[1684] 사중-플랩 프라임 편집에서, 이중 가닥 DNA 서열은 표적 부위에서 제1, 제2, 제3 및 제4 프라임 편집제 복합체와 접촉된다. 각각의 복합체는 융합 단백질 및 PEgRNA를 포함한다. 일부 실시양태에서, 각각의 융합 단백질은 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 및 RNA-의존성 DNA 폴리머라제 활성을 갖는 폴리펩티드 (예를 들어, 리버스 트랜스크립타제)를 포함하고, 각각의 PEgRNA는 스페이서 서열, gRNA 코어, DNA 합성 주형 및 프라이머 결합 부위를 포함한다. 각각의 DNA 합성 주형은 단일 가닥 DNA 서열을 코딩한다. 코딩된 2개의 단일 가닥 DNA 서열은 서로에 대해 상보적일 수 있고, 편집될 표적 부위 내로 통합될 수 있는 듀플렉스를 형성할 수 있다. 프라임 편집제 복합체의 다양한 요소 (예를 들어, 융합 단백질, napDNAbp, 폴리머라제, PEgRNA 등)는 본원에 개시된 시스템의 임의의 실시양태를 포함할 수 있다.

[1685] 본원에 제공된 다중-플랩 프라임 편집 방법은 수많은 적용에 사용될 수 있다. 예를 들어, 이들은 표적 DNA 서열의 역위를 용이하게 하는 데 사용될 수 있다. 본 출원에서, 제1 PEgRNA의 DNA 합성 주형에 의해 코딩된 제1 단일 가닥 DNA 서열 및 제2 PEgRNA의 DNA 합성 주형에 의해 코딩된 제2 단일 가닥 DNA 서열은 표적 DNA 서열의 대향 단부 상에 있고, 제3 PEgRNA의 DNA 합성 주형에 의해 코딩된 제3 단일 가닥 DNA 서열 및 제4 PEgRNA의 DNA 합성 주형에 의해 코딩된 제4 단일 가닥 DNA 서열은 동일한 표적 DNA 서열의 대향 단부 상에 있다.

[1686] 일부 실시양태에서, 본원에 제공된 다중-플랩 프라임 편집 방법은 원형 DNA 공여자를 제공하는 것을 추가로 포함하며, 그의 일부는 표적 부위에서 이중 가닥 핵산 내로 통합될 수 있다. 본 출원에서, 제1 PEgRNA의 DNA 합성 주형에 의해 코딩된 제1 단일 가닥 DNA 서열 및 제3 PEgRNA의 DNA 합성 주형에 의해 코딩된 제3 단일 가닥 DNA 서열은 표적 DNA 서열의 대향 단부 상에 있고, 제2 PEgRNA의 DNA 합성 주형에 의해 코딩된 제2 단일 가닥 DNA 서열 및 제4 PEgRNA의 DNA 합성 주형에 의해 코딩된 제4 단일 가닥 DNA 서열은 원형 DNA 공여자 상에 있다. 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 원형 DNA 공여자의 부분은 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 표적 DNA 서열을 대체하는 듀플렉스를 형성할 수 있다.

[1687] 또 다른 적용에서, 본원에 제공된 다중-플랩 프라임 편집 방법은 제1 핵산 분자 (예를 들어, 제1 염색체)로부터 제2 핵산 분자 (예를 들어, 제2 염색체)로의 표적 DNA 서열의 전위를 허용한다. 본 출원에서, 제1 PEgRNA의 DNA 합성 주형에 의해 코딩된 제1 단일 가닥 DNA 서열 및 제3 PEgRNA의 DNA 합성 주형에 의해 코딩된 제3 단일 가닥 DNA 서열은 제1 핵산 분자 상에 있고, 제2 PEgRNA의 DNA 합성 주형에 의해 코딩된 제2 단일 가닥 DNA 서열 및 제4 PEgRNA의 DNA 합성 주형에 의해 코딩된 제4 단일 가닥 DNA 서열은 제2 핵산 분자 상에 있다. 제1 단일 가닥 DNA 서열과 제3 단일 가닥 DNA 서열 사이의 제1 핵산 분자의 부분은 제2 핵산 분자 내로 혼입될 수 있고, 제2 단일 가닥 DNA 서열과 제4 단일 가닥 DNA 서열 사이의 제2 핵산 분자의 부분은 제1 핵산 분자 내로 혼입된다.

[1688] 다른 방법

[1689] 본 개시내용은 프리온 질환 (예를 들어, 본원의 실시예 5), 트리뉴클레오티드 반복부 확장 질환 (예를 들어, 본원의 실시예 3), 또는 CDKL5 결핍 장애 (CDD) (예를 들어, 본원의 실시예 23)에 예시된 바와 같으나 이에 제한되지는 않는, 본원에 제공된 다중-플랩 프라임 편집 시스템에 의해 교정될 수 있는 점 돌연변이 또는 다른 돌연변이 (예를 들어, 결실, 삽입, 역위, 중복 등)와 연관되거나 그에 의해 유발된 질환으로 진단된 대상체의 치료 방법을 제공한다.

[1690] 실질적으로 임의의 질환-유발 유전적 결함은 다중-플랩 프라임 편집을 사용함으로써 복구될 수 있으며, 이는 적절한 프라임 편집제 융합 단백질 (napDNAbp 및 폴리머라제 (예를 들어, 리버스 트랜스크립타제)를 포함함)을 선택하는 것, 및 (a) 편집 부위를 함유하는 적절한 표적 DNA를 표적화하고, (b) 닉 부위의 바로 하류의 내인성 가닥을 치환 및 대체하는 목적하는 편집을 포함하는, 닉 부위의 3' 단부로부터의 DNA의 단일 가닥의 합성을 위한 주형을 제공하도록 설계된 적절한 PEgRNA를 설계하는 것을 포함한다. 다중-플랩 프라임 편집은 비제한적으로, (a) 뉴클레오티드 서열에 돌연변이-교정 변화를 설치하고, (b) 단백질 및 RNA 태그를 설치하고, (c) 관심 단백질 상에 이뮤노에피토프를 설치하고, (d) 단백질에 유도성 이량체화 도메인을 설치하고, (e) 생체분자의 활성을 변경하기 위해 서열을 설치하거나 제거하고, (f) 특이적 유전자 변화를 지시하기 위해 레콤비나제 표적 부위를 설치하고, (g) 오류-유발 RT를 사용함으로써 표적 서열을 돌연변이유발하는 데 사용될 수 있다.

[1691] 장애를 치료하는 방법은 초기 단계로서 본원에 기재된 방법에 따른 적절한 PEgRNA 및 프라임 편집제 융합 단백질의 설계를 수반할 수 있으며, 이는 고려될 수 있는 다수의 고려사항, 예컨대 하기를 포함한다:

[1692] (a) 표적 서열, 즉 1개 이상의 핵염기 변형을 프라임 편집제로 설치하고자 목적하는 뉴클레오티드 서열;

[1693] (b) 표적 서열 내의 절단 부위의 위치, 즉 프라임 편집제가 단일 가닥 닉을 유도하여 닉의 한 측면 상에 3' 단부 RT 프라이머 서열을 생성하고 닉의 다른 측면 상에 5' 단부 내인성 플랩을 생성할 (이는 궁극적으로 FEN1 또는 그의 등가물에 의해 제거되고 3' ssDNA 플랩에 의해 대체됨) 특이적 핵염기 위치. 절단 부위는 RNA-의존성 DNA 중합 동안 PE 융합 단백질의 폴리머라제 (예를 들어, RT 효소)에 의해 연장되게 되는 3' 단부 프라이머 서열을 생성하여 목적하는 편집을 함유하는 3' ssDNA 플랩을 생성하고, 이는 이어서 표적 서열에서 5' 내인성 DNA 플랩을 대체한다.

[1694] (c) 이용가능한 PAM 서열 (정규 SpCas9 PAM 부위, 뿐만 아니라 확장된 또는 상이한 PAM 특이성을 갖는 Cas9 변이체 및 등가물에 의해 인식되는 비-정규 PAM 부위 포함);

[1695] (d) 이용가능한 PAM 서열과 PAM 가닥 내의 절단 부위의 위치 사이의 간격;

[1696] (e) 사용되는데 이용가능한 프라임 편집제의 특정한 Cas9, Cas9 변이체, 또는 Cas9 등가물 (이는 부분적으로 이용가능한 PAM에 의해 좌우됨);

[1697] (f) 프라이머 결합 부위의 서열 및 길이;

[1698] (g) 편집 주형의 서열 및 길이;

[1699] (h) 상동성 아암의 서열 및 길이;

[1700] (i) 스페이서 서열 및 길이; 및

[1701] (j) gRNA 코어 서열.

[1702] 적합한 PEgRNA, 및 임의로 제2 부위 닉킹을 위한 닉킹-sgRNA 설계 가이드가 상기 고려사항 중 1가지 이상을 고려한 하기 예시적인 단계별 지침 세트에 의해 설계될 수 있다. 단계는 도 70a-70i에 제시된 예를 참조한다.

[1703] 1. 표적 서열 및 편집을 규정한다. 목적하는 편집 (점 돌연변이, 삽입, 결실, 또는 그의 조합)의 위치 주위에 집중된 표적 DNA 영역 (~200bp)의 서열을 검색한다. 도 70a를 참조한다.

[1704] 2. 표적 PAM의 위치를 찾아낸다. 목적하는 편집 위치에 근접한 PAM을 확인한다. PAM은 목적하는 편집 위치에 근접한 DNA의 어느 한 가닥 상에서 확인될 수 있다. 편집 위치에 가까운 PAM이 바람직하지만 (즉, 여기서 닉 부위는 편집 위치로부터 30 nt 미만이거나, 또는 편집 위치로부터 닉 부위까지 29 nt, 28 nt, 27 nt, 26 nt, 25 nt, 24 nt, 23 nt, 22 nt, 21 nt, 20 nt, 19 nt, 18 nt, 17 nt, 16 nt, 15 nt, 14 nt, 13 nt, 12 nt, 11 nt, 10 nt, 9 nt, 8 nt, 7 nt, 6 nt, 5 nt, 4 nt, 3 nt, 또는 2 nt 미만임), 닉을 편집 위치로부터 ≥ 30 nt에 배치하는 프로토스페이서 및 PAM을 사용하여 편집을 설치하는 것이 가능하다. 도 70b를 참조한다.

[1705] 3. 닉 부위를 위치시킨다. 고려되는 각각의 PAM에 대해, 상응하는 닉 부위 및 어떤 가닥 상의 것인지 확인한다. Sp Cas9 H840A 닉카제의 경우, 절단은 PAM-함유 가닥에서 NGG PAM에 대해 5'의 세번째 및 네번째 염기 사이에서 일어난다. 모든 편집된 뉴클레오티드는 닉 부위의 3'에 존재해야 하고, 따라서 적절한 PAM은 PAM-함유 가닥 상의 표적 편집에 대해 5'에 닉을 배치해야 한다. 하기 제시된 예에서, 2종의 가능한 PAM이 존재한다. 단순성을 위해, 나머지 단계는 PAM 1만을 사용한 PEgRNA의 설계를 입증할 것이다. 도 70c를 참조한다.

[1706] 4. 스페이서 서열을 설계한다. SpCas9의 프로토스페이서는 PAM-함유 가닥 상의 NGG PAM에 대해 5'의 20개 뉴클레오티드에 상응한다. 효율적인 Pol III 전사 개시는 G가 제1 전사된 뉴클레오티드일 것을 요구한다. 프로토스페이서의 제1 뉴클레오티드가 G인 경우에, PEgRNA에 대한 스페이서 서열은 단순히 프로토스페이서 서열이다. 프로토스페이서의 제1 뉴클레오티드가 G가 아닌 경우에, PEgRNA의 스페이서 서열은 G에 이어서 프로토스페이서 서열이 후속한다. 도 70d를 참조한다.

[1707] 5. 프라이머 결합 부위 (PBS)를 설계한다. 출발 대립유전자 서열을 사용하여, PAM-함유 가닥 상의 DNA 프라이머를 확인한다. DNA 프라이머의 3' 단부는 닉 부위의 바로 상류에 있는 뉴클레오티드이다 (즉, Sp Cas9의 경우 NGG PAM에 대해 5'의 네번째 염기). PE2 및 PE3과 함께 사용하기 위한 일반적 설계 원리로서, DNA 프라이머에 대해 12 내지 13개 뉴클레오티드의 상보성을 함유하는 PEgRNA 프라이머 결합 부위 (PBS)가 ~40-60% GC 함량을 함유하는 서열에 대해 사용될 수 있다. 낮은 GC 함량을 갖는 서열의 경우에, 보다 긴 (14- 내지 15-nt) PBS가 시험되어야 한다. 보다 높은 GC 함량을 갖는 서열의 경우에, 보다 짧은 (8- 내지 11-nt) PBS가 시험되어야 한다. 최적의 PBS 서열은 GC 함량에 상관없이 실험적으로 결정되어야 한다. 길이-p PBS 서열을 설계하기 위해, 출발 대립유전자 서열을 사용하여 PAM-함유 가닥에서 닉 부위의 5'의 첫번째 p 뉴클레오티드의 역 상보체를 취

한다. 도 70e를 참조한다.

[1708] 6. RT 주형 (또는 DNA 합성 주형)을 설계한다. RT 주형 (또는 폴리머라제가 리버스 트랜스크립타제가 아닌 경우 DNA 합성 주형)은 설계된 편집 및 편집에 인접한 서열에 대한 상동성을 코딩한다. 한 실시양태에서, 이들 영역은 도 3d 및 도 3e의 DNA 합성 주형에 상응하며, 여기서 DNA 합성 주형은 "편집 주형" 및 "상동성 아암"을 포함한다. 최적의 RT 주형 길이는 표적 부위에 기초하여 달라진다. 짧은-범위 편집 (위치 +1 내지 +6)의 경우, 짧은 (9 내지 12 nt), 중간 (13 내지 16 nt), 및 긴 (17 내지 20 nt) RT 주형을 시험하는 것이 권장된다. 긴-범위 편집 (위치 +7 및 그 초과)의 경우, 충분한 3' DNA 플랩 상동성을 허용하기 위해 편집 위치를 지나 적어도 5개의 nt (바람직하게는 10개 이상의 nt) 연장된 RT 주형을 사용하는 것이 권장된다. 긴-범위 편집의 경우, 여러 RT 주형이 기능적 설계를 확인하기 위해 스크리닝되어야 한다. 보다 큰 삽입 및 결실 (≥5 nt)을 위해, RT 주형 내로의 보다 큰 3' 상동성 (~20 nt 이상)의 혼입이 권장된다. 편집 효율은 전형적으로 RT 주형이 역전사된 DNA 생성물에서 마지막 뉴클레오티드로서 G의 합성 (PEgRNA의 RT 주형에서의 C에 상응함)을 코딩하는 경우에 손상된다. 많은 RT 주형이 효율적인 프라임 편집을 지지하기 때문에, RT 주형을 설계할 때 최종 합성된 뉴클레오티드로서 G를 회피하는 것이 권장된다. 길이-r RT 주형 서열을 설계하기 위해, 목적하는 대립유전자 서열을 사용하고, 원래 PAM을 함유하는 가닥에서 닉 부위의 3'의 첫번째 r 뉴클레오티드의 역 상보체를 취한다. SNP 편집과 비교하여, 동일한 길이의 RT 주형을 사용한 삽입 또는 결실 편집은 동일한 상동성을 함유하지 않을 것임을 주목한다. 도 70f를 참조한다.

[1709] 7. 전장 PEgRNA 서열을 조립한다. PEgRNA 성분을 하기 순서 (5'에서 3')로: 스페이서, 스캐폴드, RT 주형 및 PBS로 연결한다. 도 70g를 참조한다.

[1710] 8. PE3에 대한 닉킹-sgRNA를 설계한다. 편집의 상류 및 하류의 비-편집된 가닥 상에서 PAM을 확인한다. 최적의 닉킹 위치는 고도로 유전자좌-의존적이고, 실험적으로 결정되어야 한다. 일반적으로, PEgRNA-유도된 닉에 걸친 위치에 대해 5'에 40 내지 90개의 뉴클레오티드가 배치된 닉은 보다 높은 편집 수율 및 보다 적은 indel로 이어진다. 닉킹 sgRNA는 출발 대립유전자에서 20-nt 프로토스페이서와 매칭되는 스페이서 서열을 가지며, 여기서 프로토스페이서가 G로 시작되지 않는 경우 5'-G가 부가된다. 도 70h를 참조한다.

[1711] 9. PE3b 닉킹-sgRNA를 설계한다. PAM이 상보적 가닥에 존재하고 그의 상응하는 프로토스페이서가 편집을 위해 표적화된 서열과 중첩되는 경우에, 이러한 편집은 PE3b 시스템에 대한 후보일 수 있다. PE3b 시스템에서, 닉킹-sgRNA의 스페이서 서열은 출발 대립유전자가 아닌 목적하는 편집된 대립유전자의 서열과 매칭된다. PE3b 시스템은 편집된 뉴클레오티드(들)가 닉킹-sgRNA 프로토스페이서의 시드 영역 (PAM에 인접한 ~10개의 nt) 내에 속할 때 효율적으로 작동한다. 이는 편집된 가닥의 설치 후까지 상보적 가닥의 닉킹을 방지하여, 표적 DNA에의 결합에 대한 PEgRNA와 sgRNA 사이의 경쟁을 방지한다. PE3b는 또한 두 가닥 상의 동시 닉의 생성을 피하고, 따라서 높은 편집 효율을 유지하면서 indel 형성을 유의하게 감소시킨다. PE3b sgRNA는 필요한 경우에 5' G를 부가하여 목적하는 대립유전자에서 20-nt 프로토스페이서와 매칭되는 스페이서 서열을 가져야 한다. 도 70i를 참조한다.

[1712] 적합한 PEgRNA 및 제2-부위 닉킹 sgRNA를 설계하기 위한 상기 단계별 과정은 어떠한 방식으로도 제한하는 것으로 의도되지 않는다. 본 개시내용은 관련 기술분야의 통상의 기술자에 의해 그로부터 유도될 수 있는 상기 기재된 단계별 과정의 변경을 고려한다.

[1713] 적합한 PEgRNA 및 PE 융합 단백질이 선택/설계되면, 이들은 적합한 방법론에 의해, 예컨대 벡터-기반 형질감염 (PEgRNA 및 PE 융합 단백질을 코딩하는 DNA를 포함하고 벡터로의 형질감염 시 세포 내에서 발현되는 1개 이상의 벡터), 전달 포맷 (예를 들어, 지질 입자, 나노입자)의 PEgRNA와 복합체화된 PE 융합 단백질의 직접 전달 (예를 들어, RNP 전달), 또는 mRNA-기반 전달 시스템에 의해 투여될 수 있다. 이러한 방법은 본 개시내용에서 본원에 기재되고, 임의의 공지된 방법이 이용될 수 있다.

[1714] PEgRNA 및 PE 융합 단백질 (또는 함께, PE 복합체로 지칭됨)은 관심 표적 DNA와 접촉 시 목적하는 편집이 그 안에 설치되도록 하는 치료 유효량으로 세포에 전달될 수 있다.

[1715] 적절한 세포로의 전달이 실현가능한 한 이러한 방법에 의해 임의의 질환이 치료가능한 것으로 생각된다. 관련 기술분야의 통상의 기술자는 의도된 목적 및 의도된 표적 세포에 적합하도록 PE 전달 방법론을 선별 및/또는 선택할 수 있을 것이다.

[1716] 예를 들어, 일부 실시양태에서, 이러한 질환, 예를 들어 상기 기재된 바와 같은 점 돌연변이와 연관된 암을 갖는 대상체에게, 목적하는 유전자 변화를 포함하는 공여자 DNA 분자의 존재 하에 상동성-지정 복구에 의해 매개

되는 바와 같은, 점 돌연변이를 교정하거나 또는 탈활성화 돌연변이를 질환-연관 유전자 내로 도입하는 본원에 기재된 다중-플랩 프라임 편집 시스템의 유효량을 투여하는 것을 포함하는 방법이 제공된다. 일부 실시양태에서, 이러한 질환, 예를 들어 상기 기재된 바와 같은 점 돌연변이와 연관된 암을 갖는 대상체에게, 점 돌연변이를 교정하거나 또는 탈활성화 돌연변이를 질환-연관 유전자 내로 도입하는 본원에 기재된 다중-플랩 프라임 편집 시스템의 유효량을 투여하는 것을 포함하는 방법이 제공된다. 일부 실시양태에서, 질환은 증식성 질환이다. 일부 실시양태에서, 질환은 유전 질환이다. 일부 실시양태에서, 질환은 신생물성 질환이다. 일부 실시양태에서, 질환은 대사 질환이다. 일부 실시양태에서, 질환은 리소솜 축적 질환이다. 점 돌연변이를 교정하거나 또는 탈활성화 돌연변이를 질환-연관 유전자 내로 도입함으로써 치료될 수 있는 다른 질환은 관련 기술분야의 통상의 기술자에게 공지되어 있을 것이며, 개시내용은 이와 관련하여 제한되지 않는다.

[1717] 본 개시내용은 추가의 질환 또는 장애, 예를 들어 TPRT-매개된 유전자 편집에 의해 교정될 수 있는 점 돌연변이와 연관되거나 그에 의해 유발되는 질환 또는 장애의 치료 방법을 제공한다. 일부 이러한 질환은 본원에 기재되어 있고, 본원에 제공된 전략 및 융합 단백질에 의해 치료될 수 있는 추가의 적합한 질환은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다. 예시적인 적합한 질환 및 장애가 하기 열거된다. 각각의 서열 내의 특정 위치 또는 잔기의 넘버링은 사용되는 특정한 단백질 및 넘버링 스킴에 좌우됨을 이해할 것이다. 넘버링은, 예를 들어 성숙 단백질의 전구체 및 성숙 단백질 자체에서 상이할 수 있고, 종마다의 서열의 차이는 넘버링에 영향을 미칠 수 있다. 관련 기술분야의 통상의 기술자는 관련 기술분야에 널리 공지된 방법에 의해, 예를 들어 서열 정렬 및 상동 잔기의 결정에 의해 임의의 상동 단백질 및 각각의 코딩 핵산에서 각각의 잔기를 확인할 수 있을 것이다. 예시적인 적합한 질환 및 장애는 비제한적으로 다음을 포함한다: 2-메틸-3-히드록시부티르산뇨; 3 베타-히드록시스테로이드 데히드로게나제 결핍; 3-메틸클루타콘산뇨; 3-옥소-5 알파-스테로이드 델타 4-데히드로게나제 결핍; 46,XY 성별 반전, 유형 1, 3, 및 5; 5-옥소프롤리나제 결핍; 6-피루보일-테트라히드로프테린 신타제 결핍; 아르스코그 증후군; 아즈 증후군; 연골무발생증 유형 2; 완전색맹 2 및 7; 후천성 긴 QT 증후군; 뇌량무형성 증후군, 쉰젤 유형; 아크로카피토페모랄 이형성증; 선단이골증 2, 호르몬 저항성 동반 또는 비동반; 말단홍색각피증; 말단왜소 이형성증; ACTH-비의존성 거대결정성 부신 증식증 2; 활성화된 PI3K-델타 증후군; 급성 간헐성 포르피린증; 아실-CoA 데히드로게나제 패밀리, 구성원 9의 결핍; 아담스-올리버 증후군 5 및 6; 아데닌 포스포리보실트랜스퍼라제 결핍; 아데닐레이트 키나제 결핍; 아데닐로숙시네이트 리아제 결핍으로 인한 용혈성 빈혈; 청소년 신결석증; 신장-간-췌장 이형성증; 멕켈 증후군 유형 7; 부신백질이영양증; 성인 접합부 수포성 표피박리증; 수포성 표피박리증, 접합부, 국소 변이형; 성인 신경 세로이드 리포푸신증; 성인 신경 세로이드 리포푸신증; 안구운동 실행증을 동반한 성인 발병 운동실조; 성인 증후군; 무섬유소혈증 및 선천성 무섬유소혈증; 상염색체 열성 무감마글로불린혈증 2; 연령-관련 황반 변성 3, 6, 11, 및 12; 에카르디 구티에레스 증후군 1, 4, 및 5; 동창 루푸스 1; 알라질 증후군 1 및 2; 알렉산더병; 알캅톤뇨증; 알란-헌던-두들리 증후군; 범발성 탈모증 선천성; 알퍼스 뇌병증; 알파-1-항트립신 결핍; 상염색체 우성, 상염색체 열성, 및 X-연관 열성 알포트 증후군; 알츠하이머병, 가족성, 3, 경직성 하반신부전마비 및 실행증 동반; 알츠하이머병, 유형, 1, 3, 및 4; 저석회화 유형 및 저성숙 유형, IIA1 불완전 사기질형성증; 아미노아실라제 1 결핍; 아미시 영아 간질 증후군; 아밀로이드생성 트랜스티레틴 아밀로이드증; 아밀로이드 심근병증, 트랜스티레틴-관련; 심근병증; 근위축성 측삭 경화증 유형 1, 6, 15 (전두측두엽 치매 동반 또는 비동반), 22 (전두측두엽 치매 동반 또는 비동반), 및 10; TDP43 함유물을 동반한 전두측두엽 치매, TARDBP-관련; 앤더만 증후군; 안데르센 타윌 증후군; 선천성 긴 QT 증후군; 빈혈, 비구상 용혈성, G6PD 결핍으로 인함; 안젤만 증후근; 소두증을 동반한 중증 신생아-발병 뇌병증; 자폐증에 감수성, X-연관 3; 혈관병증, 유전성, 신병증, 동맥류 및 근육 경련 동반; 안지오텐신 i-전환 효소, 양성 혈청 증가; 무홍채증, 소뇌 운동실조, 및 정신 지체; 손발톱결여증; 항트롬빈 III 결핍; 생식기 기형 및 장애성 스테로이드생성을 동반한 앤틀리-빅슬러 증후군; 대동맥류, 가족성 흉부 4, 6, 및 9; 흉부 대동맥류 및 대동맥 박리; 전신성 평활근 기능장애 증후군; 모야모야병 5; 재생불량성 빈혈; 겉보기 미네랄로코르티코이드 과잉; 아르기나제 결핍; 아르기니노숙시네이트 리아제 결핍; 아로마타제 결핍; 부정맥유발성 우심실 심근병증 유형 5, 8, 및 10; 원발성 가족성 비대성 심근병증; 선천성 다발성 관절만곡증, 원위, X-연관; 관절만곡증 신장 기능장애 담즙정체 증후군; 관절만곡증, 신장 기능장애, 및 담즙정체 2; 아스파라긴 신테타제 결핍; 뉴런 이동의 이상; 비타민 E 결핍을 동반한 운동실조; 운동실조, 감각, 상염색체 우성; 운동실조-모세혈관확장증 증후군; 유전성 암-소인 증후군; 무트랜스페린혈증; 심방 세동, 가족성, 11, 12, 13, 및 16; 심방 중격 결손 2, 4, 및 7 (방실 전도 결함 동반 또는 비동반); 심방 정지 2; 방실 중격 결손 4; 유전성 안구 위축증; ATR-X 증후군; 귓바퀴관절융기 증후군 2; 자가면역 질환, 다기관, 영아-발병; 자가면역 림프증식성 증후군, 유형 1a; 상염색체 우성 발한저하성 외배엽 이형성증; 미토콘드리아 DNA 결실을 동반한 상염색체 우성 진행성 외안근마비 1 및 3; 상염색체 우성 염전 이상긴장증 4; 상염색체 열성 중심핵성 근병증; 상염색체 열

성 선천성 어린선 1, 2, 3, 4A, 및 4B; 상염색체 열성 이완성 피부증 유형 IA 및 1B; 상염색체 열성 발한저하성 외배엽 이형성증 증후군; 외배엽 이형성증 11b; 발한저하성/모발/치아 유형, 상염색체 열성; 상염색체 열성 저인산혈증 골 질환; 악센펠트-리거 증후군 유형 3; 베인브리지-로퍼스 증후군; 바나얀-릴리-루발카바 증후군; PTEN 과오종 종양 증후군; 바레이처-윈터 증후군 1 및 2; 바라카트 증후군; 바르데-비들 증후군 1, 11, 16, 및 19; 무표지 림프구 증후군 유형 2, 상보군 E; 바터 증후군 산전 유형 2; 바터 증후군 유형 3, 저칼슘뇨 동반 3, 및 4; 기저 신경절 석회화, 특발성, 4; 염주털; 양성 가족성 혈뇨; 양성 가족성 신생아 발작 1 및 2; 발작, 양성 가족성 신생아, 1 및/또는 근파동; 발작, 조기 영아 간질성 뇌병증 7; 양성 가족성 신생아-영아 발작; 양성 유전성 무도병; 심근병증을 동반한 양성 견갑비골 근육 이영양증; 버나드-술리에 증후군, 유형 A1 및 A2 (상염색체 우성); 베스트로핀병증, 상염색체 열성; 베타 지중해빈혈; 베스렘 근병증 및 베스렘 근병증 2; 비에티 결정질 각막망막 이영양증; 담즙산 합성 결함, 선천성, 2; 비오티니다제 결핍; 버크 바렐 정신 지체 이형 증후군; 실눈증, 안검하수증, 및 역내안각주름; 블룸 증후군; 보제슨-포스만-레만 증후군; 보우셔 노이하우저 증후군; 단지증 유형 A1 및 A2; 고혈압을 동반한 단지증; 출혈을 동반한 뇌 소혈관 질환; 분지형-쇄 케토산 데히드로게나제 키나제 결핍; 아가미귀 증후군 2 및 3; 유방암, 조기 발병; 유방-난소암, 가족성 1, 2, 및 4; 취약 각막 증후군 2; 브로디 근병증; 땀 염화물 상승 동반 또는 비동반 기관지확장증 3; 브라운-비알레토-반 라에레 증후군 및 브라운-비알레토-반 라에레 증후군 2; 브루가다 증후군; 브루가다 증후군 1; 심실 세동; 발작성 가족성 심실 세동; 브루가다 증후군 및 브루가다 증후군 4; 긴 QT 증후군; 심장 돌연사; 과녁 황반 이영양증; 스타르가르트병 4; 추체-간체 이영양증 12; 수포성 어린선형 홍피증; 번-맥코운 증후군; 칸디다증, 가족성, 2, 5, 6, 및 8; 탄수화물-결핍 당단백질 증후군 유형 I 및 II; 탄산 안히드라제 VA 결핍, 이에 기인한 고암모니아혈증; 결장의 암종; 심장 부정맥; 긴 QT 증후군, LQT1 하위유형; 심뇌근병증, 치명성 영아, 시토크롬 c 옥시다제 결핍으로 인함; 심장얼굴피부 증후군; 심근병증; 다논병; 비대성 심근병증; 좌심실 비치밀 심근병증; 카네발 증후군; 카니 복합증, 유형 1; 카르니틴 아실카르니틴 트랜스로카제 결핍; 카르니틴 팔미토일트랜스퍼라제 I, II, II (후기 발병), 및 II (영아) 결핍; 백내장 1, 4, 상염색체 우성, 상염색체 우성, 다중 유형, 소각막 동반, 코폭-유사, 소아, 소각막 및 글루코스뇨 동반, 및 핵 미만성 비진행성; 카테콜아민성 다형성 심실성 빈맥; 미측 퇴행 증후군; Cd8 결핍, 가족성; 중심핵병; 염색체 1,9 및 16의 동원체 불안정성 및 면역결핍; 진행성 외안근마비를 동반한 영아 소뇌 운동실조 및 소뇌 운동실조, 정신 지체, 및 평형이상 증후군 2; 뇌 아밀로이드 혈관병증, APP-관련; 피질하 경색 및 백질뇌병증 동반 뇌 상염색체 우성 및 열성 동맥병증; 뇌 해면 기형 2; 뇌안구안면골격 증후군 2; 뇌-안구-안면-골격 증후군; 석회화 및 낭을 동반한 대뇌망막 미세혈관병증; 세로이드 리포푸신증 뉴런 2, 6, 7, 및 10; 체디악-히가시 증후군, 체디악-히가시 증후군, 성인 유형; 샤르코-마리-투스병 유형 1B, 2B2, 2C, 2F, 2I, 2U (축삭), 1C (탈수초성), 우성 중간체 C, 열성 중간체 A, 2A2, 4C, 4D, 4H, IF, IVF, 및 X; 견갑비골 척수성 근육 위축; 원위 척수성 근육 위축, 선천성 비진행성; 척수성 근육 위축, 원위, 상염색체 열성, 5; CHARGE 연합증; 소아기 저포스파타제증; 성인 저포스파타제증; 담낭염; 진행성 가족성 간내 담즙정체 3; 담즙정체, 간내, 임신중 3; 콜레스타놀 축적 질환; 콜레스테롤 모노옥시게나제 (측쇄 절단) 결핍; 연골이형성증 블롬스트랜드 유형; 점상 연골이형성증 1, X-연관 열성 및 2 X-연관 우성; CHOPS 증후군; 만성 육아종성 질환, 상염색체 열성 시토크롬 b-양성, 유형 1 및 2; 추들리-맥쿨로우 증후군; 섬모 이상운동증, 원발성, 7, 11, 15, 20 및 22; 시트룰린혈증 유형 I; 시트룰린혈증 유형 I 및 II; 쇄골두개 이골증; C-유사 증후군; 코케인 증후군 유형 A; 조효소 Q10 결핍, 원발성 1, 4, 및 7; 코핀 시리스/지적 장애; 코핀-로우리 증후군; 코헨 증후군; 감기-유발 발한 증후군 1; 콜-카펜터 증후군 2; 육아종을 동반한 복합 세포성 및 체액성 면역 결함; 복합 d-2- 및 l-2-히드록시글루타르산뇨; 복합 말론산 및 메틸말론산뇨; 복합 산화성 인산화 결핍 1, 3, 4, 12, 15, 및 25; 복합 부분 및 완전 17-알파-히드록실라제/17,20-리아제 결핍; 공통 가변성 면역결핍 9; 보체 성분 4, 이의 부분 결핍, 기능장애성 c1 억제제로 인함; 보체 인자 B 결핍; 추체 단색형색각; 추체-간체 이영양증 2 및 6; 추체-간체 이영양증 불완전 사기질형성증; 선천성 부신 증식증 및 선천성 부신 저형성증, X-연관; 선천성 무거핵구성 혈소판감소증; 선천성 무홍채증; 선천성 중앙 저환기; 히르쉬스프룽 질환 3; 선천성 구축성 거미가락증; 선천성 사지 및 안면 구축, 저긴장증, 및 발달 지체; 선천성 글리코실화 장애 유형 1B, 1D, 1G, 1H, 1J, 1K, 1N, 1P, 2C, 2J, 2K, IIm; 선천성 적혈구생성이상 빈혈, 유형 I 및 II; 선천성 안면 외배엽 이형성증; 선천성 적혈구생성 포르피린증; 선천성 전신 지방이영양증 유형 2; 선천적 심장 질환, 다중 유형, 2; 선천적 심장 질환; 단속 대동맥궁; 선천성 지방종성 과도성장, 혈관 기형, 및 표피 모반; 비소세포 폐암; 난소의 신생물; 심장 전도 결손, 비특이적; 선천성 미세융모 위축; 선천성 근육 이영양증; 부분 LAMA2 결핍으로 인한 선천성 근육 이영양증; 뇌 및 안구 기형을 동반한 선천성 근육 이영양증-디스트로글리칸병증, 유형 A2, A7, A8, A11, 및 A14; 정신 지체를 동반한 선천성 근육 이영양증-디스트로글리칸병증, 유형 B2, B3, B5, 및 B15; 정신 지체를 비동반한 선천성 근육 이영양증-디스트로글리칸병증, 유형 B5; 선천성 근육 비대-뇌 증후군; 선천성 근무력 증후군, 아세타졸아미드-반응성; 섬유 유형 불균등을 동반한 선천성 근병증; 선천성 안구 결손증; 선천성 비진행성

야맹증, 유형 1A, 1B, 1C, 1E, 1F, 및 2A; 코프로포르피린증; 납작각막 2; 각막 이영양증, 푸크스 내피, 4; 각막 내피 이영양증 유형 2; 각막 취약성 구형각막, 청색 공막 및 관절 과운동성; 코넬리아 드 랑쥐 증후군 1 및 5; 관상 동맥 질환, 상염색체 우성 2; 관상동맥 심장 질환; 고알파지단백혈증 2; 피질 이형성증, 복합체, 다른 뇌 기형 동반 5 및 6; 피질 기형, 후두; 코르티코스테로이드-결합 글로불린 결핍; 코르티코스테론 메틸옥시다제 유형 2 결핍; 코스텔로 증후군; 코우덴 증후군 1; 편평고; 두개골간 이형성증, 상염색체 우성; 두개골유합증 1 및 4; 두개골유합증 및 치아 기형; 크레아틴 결핍, X-연관; 크루존 증후군; 잠복안구 증후군; 잠복고환, 일측 또는 양측; 쿠싱 지절유합증; 피부 악성 흑색종 1; 골이영양증 및 중증 폐, 위장, 및 뇨 이상을 동반한 이완성 피부증; 청색증, 일과성 신생아 및 비정형 신병증성; 낭성 섬유증; 시스틴뇨; 시토크롬 c 옥시다제 i 결핍; 시토크롬-c 옥시다제 결핍; D-2-히드록시글루타르산뇨 2; 다리에병, 분절성; 미로 무형성 소이증 및 소치증(LAMM)을 동반한 난청; 난청, 상염색체 우성 3a, 4, 12, 13, 15, 상염색체 우성 비증후군성 감각신경성 17, 20, 및 65; 난청, 상염색체 열성 1A, 2, 3, 6, 8, 9, 12, 15, 16, 18b, 22, 28, 31, 44, 49, 63, 77, 86, 및 89; 난청, 와우, 근시 및 지적 손상 동반, 전정 침범 비동반, 상염색체 우성, X-연관 2; 2-메틸부티릴-CoA 데히드로게나제의 결핍; 3-히드록시아실-CoA 데히드로게나제의 결핍; 알파-만노시다제의 결핍; 방향족-L-아미노산 데카르복실라제의 결핍; 비스포스포글리세레이트 뮤타제의 결핍; 부티릴-CoA 데히드로게나제의 결핍; 페록시다제의 결핍; 갈락토키나제의 결핍; 구아니디노아세테이트 메틸트랜스퍼라제의 결핍; 히알루로노글루코사미니다제의 결핍; 리보스-5-포스페이트 이소머라제의 결핍; 스테로이드 11-베타-모노옥시게나제의 결핍; UDP글루코스-헥소스-1-포스페이트 우리딜트랜스퍼라제의 결핍; 크산틴 옥시다제의 결핍; 데제린-소타스병; 샤르코-마리-투스병, 유형 ID 및 IVF; 데제린-소타스 증후군, 상염색체 우성; 수지상 세포, 단핵구, B 림프구, 및 자연 킬러 림프구 결핍; 드뷔쿠아 이형성증 2; 드뷔쿠아 증후군; DFNA 2 비증후군성 청각 상실; 시신경 위축 및 난청을 동반한 당뇨병 및 요붕증; 당뇨병, 유형 2, 및 인슐린-의존성, 20; 다이아몬드-블랙판 빈혈 1, 5, 8, 및 10; 설사 3 (분비 나트륨, 선천성, 증후군성) 및 5 (터프팅 장병증 동반, 선천성); 디카르복실산 아미노산뇨증; 미만성 수장족저각피증, 보트니안 유형; 수족지신장뇌 증후군; 디히드로프테리딘 리덕타제 결핍; 확장성 심근병증 1A, 1AA, 1C, 1G, 1BB, 1DD, 1FF, 1HH, 1I, 1KK, 1N, 1S, 1Y, 및 3B; 좌심실 비치밀 3; 시토크롬 p450 옥시도리덕타제 결핍으로 인한 장애성 스테로이드생성; 원위 관절만곡증 유형 2B; 원위 유전성 모터 뉴런병증 유형 2B; 원위 근병증 마르케스베리-그릭스 유형; 원위 척수성 근육 위축, X-연관 3; 첩모중생증-림프부종 증후군; 피부의 부재를 동반한 우성 이영양성 수포성 표피박리증; 우성 유전성 시신경 위축; 돈나이 바로우 증후군; 도파민 베타 히드록실라제 결핍; 도파민 수용체 d2, 감소된 뇌 밀도; 다울링-데고스병 4; 도인 벌집형 망막 이영양증; 말라티아 레벤티네스; 듀안 증후군 유형 2; 두빈-존슨 증후군; 뒤시엔느 근육 이영양증; 베커 근육 이영양증; 이상피브리노겐혈증; 선천성 이상각화증 상염색체 우성 및 상염색체 우성, 3; 선천성 이상각화증, 상염색체 열성, 1, 3, 4, 및 5; 선천성 이상각화증 X-연관; 이상운동증, 가족성, 안면 근파동 동반; 이상플라스미노겐혈증; 이상긴장증 2 (염전, 상염색체 열성), 3 (염전, X-연관), 5 (도파-반응성 유형), 10, 12, 16, 25, 26 (근간대성); 발작, 양성 가족성 영아, 2; 조기 영아 간질성 뇌병증 2, 4, 7, 9, 10, 11, 13, 및 14; 비정형 레트 증후군; 초기 T 세포 전구세포 급성 림프모구성 백혈병; 외배엽 이형성증 피부 취약성 증후군; 외배엽 이형성증-합지증 증후군 1; 수정체 편위, 고립성 상염색체 열성 및 우성; 가락결손증, 외배엽 이형성증, 및 구순열/구개 증후군 3; 엘러스-단로스 증후군 유형 7 (상염색체 열성), 전형적 유형, 유형 2 (조로), 히드록시리신-결핍, 유형 4, 유형 4 변이형, 및 테나신-X 결핍으로 인함; 아이흐스펠트 유형 선천성 근육 이영양증; 내분비-뇌골이형성증; 증진된 s-추체 증후군; 비대 전정 수도관 증후군; 엔테로키나제 결핍; 사마귀양 표피이형성증; 단순 수포성 표피박리증 및 지대 근육 이영양증, 얼룩덜룩한 색소침착을 동반한 단순 수포성 표피박리증, 유문 폐쇄증을 동반한 단순 수포성 표피박리증, 상염색체 열성 및 유문 폐쇄증을 동반한 단순 수포성 표피박리증; 표피박리성 수장족저 각피증; 가족성 열성 발작 8; 간질, 소아기 부재 2, 12 (특발성 전신, 이에 감수성) 5 (야간 전두엽), 야간 전두엽 유형 1, 부분, 가변성 초점 동반, 진행성 근간대성 3, 및 X-연관, 가변성 학습 불능 및 행동 장애 동반; 간질성 뇌병증, 소아기-발병, 초기 영아, 1, 19, 23, 25, 30, 및 32; 골단 이형성증, 다중, 근시 및 전도성 난청 동반; 간헐성 운동실조 유형 2; 간헐성 통증 증후군, 가족성, 3; 엡스타인 증후군; 퀵트너 증후군; 적혈구생성 프로토포르피린증; 에스트로겐 저항성; 삼출성 유리체망막병증 6; 파브리병 및 심장 변이형 파브리병; 인자 H, VII, X, v 및 인자 viii, 2의 복합 결핍, xiii, 서브유닛, 결핍; 가족성 선종성 폴립증 1 및 3; 두드러기 및 난청을 동반한 가족성 아밀로이드 신병증; 가족성 한랭 두드러기; 충부의 가족성 무형성증; 가족성 양성 천포창; 가족성 유방암; 유방암, 이에 감수성; 골육종; 췌장암 3; 가족성 심근병증; 가족성 한랭 자가염증성 증후군 2; 가족성 결장직장암; 가족성 삼출성 유리체망막병증, X-연관; 가족성 편마비성 편두통 유형 1 및 2; 가족성 고콜레스테롤혈증; 가족성 비후성 심근병증 1, 2, 3, 4, 7, 10, 23 및 24; 가족성 저칼륨혈증-저마그네슘혈증; 가족성 형성부전성, 사구체낭성 신장; 가족성 영아 근무력증; 가족성 소아 통풍; 가족성 지중해열 및 가족성 지중해열, 상염색체 우성; 가족성 공뇌증; 가족성 만발성 피부 포르피린증; 가족성 폐 모세혈관종증; 가족성 신장 글루코

스뇨; 가족성 신장 저요산혈증; 가족성 제한성 심근병증 1; 가족성 유형 1 및 3 고지단백혈증; 판코니 빈혈, 상보군 E, I, N, 및 O; 판코니-빅켈 증후군; 잠두중독증, 이에 감수성; 열성 발작, 가족성, 11; 페인골드 증후군 1; 태아성 헤모글로빈 정량적 형질 유전자좌 1; FG 증후군 및 FG 증후군 4; 외안근의 섬유증, 선천성, 1, 2, 3a (안외 침범 동반 또는 비동반), 3b; 피쉬-아이 질환; 반점 각막 이영양증; 플로팅-하버 증후군; 정신 지체 동반 또는 비동반 언어 장애를 동반한 초점성 간질; 초점성 분절성 사구체경화증 5; 전뇌 결함; 프랭크 테르 하르 증후군; 보론 디 로코 크로바토 증후군; 프레이저 증후군; 윌름스 종양 1; 프리만-쉘돈 증후군; 전두골간단 이형성증 1 및 3; 전두측두엽 치매; 전두측두엽 치매 및/또는 근위축성 측삭 경화증 3 및 4; 전두측두엽 치매 염색체 3-연관 및 전두측두엽 치매 유비퀴틴-양성; 프룩토스-비포스파타제 결핍; 푸르만 증후군; 감마-아미노부티르산 트랜스아미나제 결핍; 감스토르프-볼파르트 증후군; 고셔병 유형 1 및 아급성 뉴런병증성; 주시 마비, 가족성 수평, 진행성 척추측만증 동반; 전신 우성 이영양성 수포성 표피박리증; 열성 발작을 동반한 전신 간질 플러스 3, 유형 1, 유형 2; 간질성 뇌병증 레녹스-가스토 유형; 거대 축삭 신경병증; 글라즈만 혈소판무력증; 녹내장 1, 개방각, e, F, 및 G; 녹내장 3, 원발성 선천성, d; 녹내장, 선천성 및 녹내장, 선천성, 결손증; 녹내장, 원발성 개방각, 소아-발병; 신경교종 감수성 1; 글루코스 수송체 유형 1 결핍 증후군; 글루코스-6-포스페이트 수송 결함; GLUT1 결핍 증후군 2; 간질, 특발성 전신, 이에 감수성, 12; 글루타메이트 포름이미노트랜스퍼라제 결핍; 글루타르산혈증 IIA 및 IIB; 글루타르산뇨, 유형 1; 글루타티온 신테타제 결핍; 글리코겐 축적 질환 0 (근육), II (성인 형태), IXa2, IXc, 유형 1A; 유형 II, 유형 IV, IV (복합 간 및 근병증성), 유형 V, 및 유형 VI; 골드만-파브르 증후군; 고돈 증후군; 골린 증후군; 완전전뇌증 서열; 완전전뇌증 7; 육아종성 질환, 만성, X-연관, 변이형; 난소의 과립막 세포 종양; 회색 혈소판 증후군; 그리셀리 증후군 유형 3; 그뢰노우브 각막 이영양증 유형 I; 성장 및 정신 지체, 하악안면 이골증, 소두증, 및 구개열; 뇌하수체 기형을 동반한 성장 호르몬 결핍; 면역결핍을 동반한 성장 호르몬 무감응; GTP 시클로히드롤라제 I 결핍; 하쥬-체니 증후군; 수족 자궁 증후군; 청각 장애; 혈관종, 모세관 영아; 혈액 신생물; 혈색소증 유형 1, 2B, 및 3; 당뇨병의 미세혈관 합병증 7; 트랜스페린 혈청 수준 정량적 형질 유전자좌 2; 헤모글로빈 H 질환, 비결실; 용혈성 빈혈, 비구상, 글루코스 포스페이트 이소머라제 결핍으로 인함; 혈구포식성 림프조직구증식증, 가족성, 2; 혈구포식성 림프조직구증식증, 가족성, 3; 헤파린 보조인자 II 결핍; 유전성 장병성 말단피부염; 유전성 유방 및 난소암 증후군; 운동실조-모세혈관확장증-유사 장애; 유전성 미만성 위암; 타원체를 동반한 유전성 미만성 백질뇌병증; 유전성 인자 II, IX, VIII 결핍 질환; 유전성 출혈성 모세혈관확장증 유형 2; 무한증을 동반한 유전성 무통각증; 유전성 림프부종 유형 I; 시신경 위축을 동반한 유전성 운동 및 감각 신경병증; 초기 호흡 부전을 동반한 유전성 근병증; 유전성 신경통 근위축; 유전성 비폴립증 결장직장 신생물; 린치 증후군 I 및 II; 유전성 췌장염; 췌장염, 만성, 이에 감수성; 유전성 감각 및 자율 신경병증 유형 IIB 및 IIA; 유전성 철적혈모구성 빈혈; 헤르만스키-푸들라크 증후군 1, 3, 4, 및 6; 착위, 내장, 2, 4, 및 6, 상염색체; 착위, 내장, X-연관; 이소증; 조직구성 수질성 망상증; 조직구증-림프절병증 플러스 증후군; 홀로카르복실라제 신테타제 결핍; 완전전뇌증 2, 3,7, 및 9; 홀트-오람 증후군; MTHFR 결핍으로 인한 호모시스테인혈증, CBS 결핍으로 인한 호모시스틴뇨증, 및 호모시스틴뇨증, 피리독신-반응성; 코발라민 대사에서의 결함으로 인한 호모시스틴뇨증-거대적모구성 빈혈, cblE 보완 유형; 호웰-에반스 증후군; 후를러 증후군; 허친슨-길포드 증후군; 수두증; 고암모니아혈증, 유형 III; 고콜레스테롤혈증 및 고콜레스테롤혈증, 상염색체 열성; 과도놀람증 2 및 과도놀람증 유전성; 고페리틴혈증 백내장 증후군; 고글리신뇨증; 주기열을 동반한 고이뮤노글로불린 D; 메발론산뇨증; 고이뮤노글로불린 E 증후군; 고인슐린혈증성 저혈당증 가족성 3, 4, 및 5; 고인슐린증-고암모니아혈증 증후군; 고리신혈증; 이상긴장증, 다혈구혈증 및 간경변증을 동반한 고망가니즈혈증; 고오르니틴혈증-고암모니아혈증-호모시트룰린뇨증 증후군; 부갑상선기능항진증 1 및 2; 부갑상선기능항진증, 신생아 중증; 고페닐알라닌혈증, bh4-결핍, a, 부분 pts 결핍으로 인함, BH4-결핍, D, 및 비-pku; 정신 지체 증후군을 동반한 고포스파타제증 2, 3, 및 4; 털감소증성 골연골이형성증; 저베타지단백혈증, 가족성, apob32 연관; 저칼슘혈증, 상염색체 우성 1; 저칼슘뇨 고칼슘혈증, 가족성, 유형 1 및 3; 저연골형성; 철 과부하를 동반한 저색소성 소구성 빈혈; 간 내 글리코겐 신테타제의 결핍을 동반한 저혈당증; 후각결여 동반 또는 비동반 저생식선자극 생식선기능저하증 11; 면역 결핍을 동반한 발한저하성 외배엽 이형성증; 발한저하성 X-연관 외배엽 이형성증; 저칼륨혈성 주기성 마비 1 및 2; 저마그네슘혈증 1, 장; 저마그네슘혈증, 발작, 및 정신 지체; 저수초형성 백질이영양증 7; 형성부전성 좌심 증후군; 방실 중격 결손 및 공통 방실 접합부; 요도하열 1 및 2, X-연관; 갑상선기능저하증, 선천성, 비갑상선종성, 1; 털감소증 8 및 12; 털감소증-림프부종-모세혈관확장증 증후군; I 혈액형 시스템; 지멘스 수포성 어린선; 박탈성 어린선; 어린선 조숙 증후군; 특발성 기저 신경절 석회화 5; 특발성 섬유화 폐포염, 만성 형태; 선천성 이상각화증, 상염색체 우성, 2 및 5; 영아기의 특발성 고칼슘혈증; 칼슘 유입 결함으로 인한 T-세포 불활성화를 동반한 면역 기능장애 2; 면역결핍 15, 16, 19, 30, 31C, 38, 40, 8, cd3-제타에서의 결함으로 인함, 과다 IgM 동반 유형 1 및 2, 및 X-연관, 마그네슘 결함, 엡스타인-바르 바이러스 감염 및 신생물 동반; 면역결핍-동원체 불안

정성-안면 기형 증후군 2; 봉입체 근병증 2 및 3; 노나카 근병증; 영아 경련 및 발작성 무도무정위운동증, 가족성; 영아 피질 과골증; 영아 GM1 강글리오시드증; 영아 저포스파타제증; 영아 신결석증; 영아 안진, X-연관; 영아 파킨슨증-이상긴장증; 다중-꼬리 정자 및 과도한 DNA와 연관된 불임; 인슐린 저항성; 인슐린-저항성 당뇨병 및 흑색 극세포증; 인슐린-의존성 당뇨병 분비성 설사 증후군; 간질성 신염, 거대세포핵; 자궁내 성장 지연, 골간단 이형성증, 선천성 부신 저형성증, 및 생식기 기형; 아이오도티로실 커플링 결함; IRAK4 결핍; 이리도전방각이발생증 우성 유형 및 유형 1; 뇌 내 철 축적; 좌골슬개골 이형성증; 도세포 증식증; 고립성 17,20-리아제 결핍; 고립성 루트로핀 결핍; 이소발레릴-CoA 데히드로게나제 결핍; 잔코빅 리베라 증후군; 제벨 랑쥐-닐슨 증후군 2; 주버트 증후군 1, 6, 7, 9/15 (이유전자), 14, 16, 및 17, 및 구안지 증후군 xiv; 헬리츠 중증 접합부 수포성 표피박리증; 소아 GM>1< 강글리오시드증; 소아 폴립증 증후군; 소아 폴립증/유전성 출혈성 모세혈관확장증 증후군; 소아 망막층간분리; 가부키 메이크업 증후군; 칼만 증후군 1, 2, 및 6; 지연된 사춘기; 칸자키병; 카라크 증후군; 카르타게너 증후군; 케니-카페이 증후군 유형 2; 케펜-루빈스키 증후군; 원추각막 1; 모낭 각화증; 선상 수장족저 각화증 1; 킨틀러 증후군; L-2-히드록시글루타르산뇨; 라슨 증후군, 우성 유형; 격자 각막이영양증 유형 III; 레베르 흑암시; 젤베거 증후군; 퍼옥시솜 생물발생 장애; 젤베거 증후군 스펙트럼; 레베르 선천성 흑암시 11, 12, 13, 16, 4, 7, 및 9; 레베르 시신경 위축; 아미노글리코시드-유발 난청; 난청, 비증후군성 감각신경성, 미토콘드리아; 좌심실 비치밀 5; 좌우축 기형; 리병; 미토콘드리아 단쇄 에노일-CoA 히드라타제 1 결핍; 미토콘드리아 복합체 I 결핍으로 인한 리 증후군; 라이너병; 레리 웨일 이상연골골형성증; 치사 선천성 구축 증후군 6; 백혈구 부착 결핍 유형 I 및 III; 백질이영양증, 저수초형성, 11 및 6; 백질뇌병증, 운동실조 동반, 뇌간 및 척수 침범 및 락테이트 상승 동반, 소멸 백질 동반, 및 진행성, 난소 부전 동반; 전체 백색손발톱; 루이 소체 치매; 리히텐스타인-노르 증후군; 리-프라우메니 증후군 1; Lig4 증후군; 지대 근육 이영양증, 유형 1B, 2A, 2B, 2D, C1, C5, C9, C14; 뇌 및 안구 기형을 동반한 선천성 근육 이영양증-디스트로글리칸병증, 유형 A14 및 B14; 리파제 결핍 복합; 지질 단백증; 지방이영양증, 가족성 부분, 유형 2 및 3; 활택뇌증 1, 2 (X-연관), 3, 6 (소두증 동반), X-연관; 피질하 층형 이소증, X-연관; 간부전 급성 영아; 로이스-디에츠 증후군 1, 2, 3; 긴 QT 증후군 1, 2, 2/9, 2/5, (이유전자), 3, 5 및 5, 후천성, 이에 감수성; 폐암; 림프부종, 유전성, id; 림프부종, 원발성, 골수이형성증 동반; 림프증식성 증후군 1, 1 (X-연관), 및 2; 리소솜 산 리파제 결핍; 대두증, 거구증, 안면 이형 증후군; 황반 이영양증, 난황상, 성인-발병; 악성 고온요법 감수성 유형 1; 악성 림프종, 비-호지킨; 악성 흑색종; 전립선의 악성 종양; 하악말단 이골증; 유형 A 또는 B 지방이영양증을 동반한 하악말단 이형성증, 비정형; 하악안면 이골증, 트리처 콜린스 유형, 상염색체 열성; 만노스-결합 단백질 결핍; 메이플 시럽뇨 질환 유형 1A 및 유형 3; 마르덴 워커 유사 증후군; 마르팡 증후군; 마린스코-쇼그렌 증후군; 마트솔프 증후군; 청소년의 성숙기-발병 당뇨병, 유형 1, 유형 2, 유형 11, 유형 3, 및 유형 9; 메이-헤글린 기형; MYH9 관련 장애; 세바스티안 증후군; 맥쿤-올브라이트 증후군; 소마토트로프 선종; 성별 코드-기질 종양; 쿠싱 증후군; 맥쿠식 카우프만 증후군; 맥레오드 신경유극적혈구증가증 증후군; 멕켈-그루버 증후군; 중간쇄 아실-조효소 A 데히드로게나제 결핍; 수모세포종; 피질하 낭을 동반한 거대뇌성 백질뇌병증 1 및 2a; 거대뇌증 선천성 모세혈관확장 대리석 피부증; PIK3CA 관련 과도성장 스펙트럼; 거대뇌증-다소뇌회증-다지증-수두증 증후군 2; 거대적모구성 빈혈, 티아민-반응성, 당뇨병 및 감각신경성 난청 동반; 마이어-골린 증후군 1 및 4; 멜닉-니들스 증후군; 수막종; 정신 지체, X-연관, 3, 21, 30, 및 72; 정신 지체 및 뇌교 및 소뇌 저형성증을 동반한 소두증; 정신 지체 X-연관 증후군 5; 정신 지체, 전방 상악 돌출, 및 사시; 정신 지체, 상염색체 우성 12, 13, 15, 24, 3, 30, 4, 5, 6,및 9; 정신 지체, 상염색체 열성 15, 44, 46, 및 5; 정신 지체, 상동증적 운동, 간질 및/또는 뇌 기형; 정신 지체, 증후군성, 클라에스-젠슨 유형, X-연관; 정신 지체, X-연관, 비특이적, 증후군성, 헤데라 유형, 및 증후군성, wu 유형; 메로신 결핍 선천성 근육 이영양증; 이염성 백질이영양증 소아, 후기 영아, 및 성인 유형; 이염성 백질이영양증; 영양위축성 이형성증; 메트헤모글로빈혈증 유형 I 및 2; 메티오닌 아데노실트랜스퍼라제 결핍, 상염색체 우성; 호모시스틴뇨증을 동반한 메틸말론산혈증; 메틸말론산뇨 cblB 유형; 메틸말로닐-CoA 뮤타제 결핍으로 인한 메틸말론산뇨; 메틸말론산뇨, mut(0) 유형; 소두 골이형성 원발 왜소증 유형 2; 맥락망막병증, 림프부종 또는 정신 지체 동반 또는 비동반 소두증; 소두증, 열공 헤르니아 및 신증후군; 소두증; 뇌량의 저형성증; 경직성 하반신마비 50, 상염색체 열성; 전반적 발달 지연; CNS 저수포형성증; 뇌 위축; 소두증, 정상 지능 및 면역결핍; 소두증-모세관 기형 증후군; 소구성 빈혈; 소안구증 증후군성 5, 7, 및 9; 소안구증, 고립성 3, 5, 6, 8, 및 결손증 동반 6; 구형소수정체증; 편두통, 가족성 뇌기저; 밀러 증후군; 외안근마비를 동반한 미니코어 근병증; 근병증, 코어 동반 선천성; 미첼-릴리 증후군; 미토콘드리아 3-히드록시-3-메틸글루타릴-CoA 신타제 결핍; 미토콘드리아 복합체 I, II, III, III (핵 유형 2, 4, 또는 8) 결핍; 미토콘드리아 DNA 고갈 증후군 11, 12 (심근병증 유형), 2, 4B (MNGIE 유형), 8B (MNGIE 유형); 미토콘드리아 DNA-고갈 증후군 3 및 7, 간뇌 유형, 및 13 (뇌근병증 유형); 미토콘드리아 포스페이트 담체 및 피루베이트 담체 결핍; 미토콘드리아 삼중기능성 단백질 결핍; 긴-쇄 3-히드록시아실-CoA 데히드로게나제 결핍;

미요시 근육 이영양증 1; 근병증, 원위, 전경골 발병 동반; 모어-트라네뱌예르크 증후군; 몰리브데넘 보조인자 결핍, 상보군 A; 모와트-윌슨 증후군; 점액지질증 III 감마; 뮤코폴리사카라이드증 유형 VI, 유형 VI (중증), 및 유형 VII; 뮤코폴리사카라이드증, MPS-I-H/S, MPS-II, MPS-III-A, MPS-III-B, MPS-III-C, MPS-IV-A, MPS-IV-B; 색소성 망막염 73; 강글리오시드증 GM1 유형1 (심장 침범 동반) 3; 다중심성 골용해 신병증; 다중심성 골용해, 소결절증 및 관절병증; 다발성 선천성 기형; 심방 중격 결손 2; 다발성 선천성 기형-저긴장증-발작 증후군 3; 다중 피부 및 점막 정맥 기형; 다발성 내분비 신생물, 유형 1 및 4; 다중 골단 이형성증 5 또는 우성; 다중 위장 폐쇄증; 다중 익상편 증후군 에스코바르 유형; 다중 술파타제 결핍; 다중 익상편 증후군 3; 근육 AMP 구아닌 옥시다제 결핍; 근육 안구 뇌 질환; 근육 이영양증, 선천성, 메가코니알 유형; 근무력증, 가족성 영아, 1; 근무력 증후군, 선천성, 11, 아세틸콜린 수용체 결핍 연관; 근무력 증후군, 선천성, 17, 2A (느린-채널), 4B (빠른-채널), 및 관상 응집체 부재; 미엘로퍼옥시다제 결핍; MYH-연관 폴립증; 자궁내막 암종; 심근경색 1; 근간대성 이상긴장증; 근간대성-무긴장성 간질; 불균일 적색 근섬유를 동반한 간질을 동반한 근간대성경련; 근원섬유성 근병증 1 및 ZASP-관련; 미오글로빈뇨, 급성 재발성, 상염색체 열성; 근신경 위장 뇌병증 증후군; 진행성 외안근마비를 동반한 영아 소뇌 운동실조; 미토콘드리아 DNA 고갈 증후군 4B, MNGIE 유형; 근병증, 중심핵성, 1, 선천성, 파량의 근육 스핀들 동반, 원위, 1, 락트산 산증, 및 철적혈모구성 빈혈 1, 미토콘드리아 진행성, 선천성 백내장, 청각 상실 및 발달 지체 동반, 및 관상 응집체, 2; 근시 6; 근경화증, 상염색체 열성; 선천성 근긴장증; 선천성 근긴장증, 상염색체 우성 및 열성 형태; 손발톱-슬개골 증후군; 난스-호란 증후군; 소안구증 2; 나바호 신경간병증; 네말린 근병증 3 및 9; 신생아 저긴장증; 지적 장애; 발작; 지연된 말 및 언어 발달; 정신 지체, 상염색체 우성 31; 시트린 결핍에 의해 유발된 신생아 간내 담즙정체; 신원성 요붕증, 신원성 요붕증, X-연관; 신석증/골다공증, 저인산혈증, 2; 신결석증 13, 15 및 4; 불임; 뇌-안구-신장 증후군 (신결석증, 안구운동 실행증 및 소뇌 이상); 신증후군, 유형 3, 유형 5, 안구 이상 동반 또는 비동반, 유형 7, 및 유형 9; 네스토르-기예르모 조로증 증후군; 노이-락소바 증후군 1; 뇌 철 축적을 동반한 신경변성 4 및 6; 신경계페리틴병증; 신경섬유종증, 유형 1 및 유형 2; 신경섬유육종; 신경뇌하수체 요붕증; 신경병증, 유전성 감각, 유형 IC; 중성 1 아미노산 수송 결함; 근병증을 동반한 중성 지질 축적 질환; 호중구 면역결핍 증후군; 니콜라이데스-바라이처 증후군; 니만-픽병 유형 C1, C2, 유형 A, 및 유형 C1, 성인 형태; 비-케톤성 고글리신혈증; 누난 증후군 1 및 4, 레오파드 증후군 1; 소아 골수단핵구성 백혈병 동반 또는 비동반 누난 증후군-유사 장애; 정상칼륨혈성 주기성 마비, 칼륨-감수성; 노룸병; 간질, 청각 상실, 및 정신 지체 증후군; 정신 지체, X-연관 102 및 증후군성 13; 비만; 안구 백색증, 유형 I; 눈피부 백색증 유형 1B, 유형 3, 및 유형 4; 눈코치아골격 이형성증; 치아저포스파타제증; 오돈토트리코멜릭 증후군; 오구치병; 치아부족증-결장직장암 증후군; 오피츠 G/BBB 증후군; 시신경 위축 9; 구강-안면-수족지 증후군; 오르니틴 아미노트랜스퍼라제 결핍; 구강안면열 11 및 7, 구순열/구개-외배엽 이형성증 증후군; 오르스타빅 린데만 솔베르크 증후군; 경도 연골이형성증을 동반한 골관절염; 박리성 골연골염; 골형성 부전증 유형 12, 유형 5, 유형 7, 유형 8, 유형 I, 유형 III, 정상 공막 동반, 우성 형태, 열성 주산기 치사; 두개 경화증을 동반한 선상 골병증; 골화석증 상염색체 우성 유형 1 및 2, 열성 4, 열성 1, 열성 6; 가성신경교종을 동반한 골다공증; 귀-구개-수족지 증후군, 유형 I 및 II; 난소 이발생증 1; 난소백질이영양증; 선천성 손발톱비대증 4 및 유형 2; 골의 파제트병, 가족성; 팔리스터-홀 증후군; 수장족저 각피증, 비표피박리성, 초점성 또는 미만성; 췌장 무발생증 및 선천성 심장 질환; 파피용-르페브르 증후군; 부신경절종 3; 폰 오일렌부르크의 선천성 이상근긴장증; 부갑상선 암종; 파킨슨병 14, 15, 19 (소아-발병), 2, 20 (조기 발병), 6, (상염색체 열성 조기 발병, 및 9; 부분 백색증; 부분 하이포크산틴-구아닌 포스포리보실트랜스퍼라제 결핍; 망막 색소 상피의 패턴화된 이영양증; PC-K6a; 펠리제우스-메르츠바허병; 펜드레드 증후군; 말초 탈수초성 신경병증, 중앙 이상수초화; 히르쉬스프룽 질환; 영속성 신생아 당뇨병; 당뇨병, 영속성 신생아, 신경계 양상 동반; 신생아 인슐린-의존성 당뇨병; 청소년의 성숙기-발병 당뇨병, 유형 2; 퍼옥시솜 생물발생 장애 14B, 2A, 4A, 5B, 6A, 7A, 및 7B; 페라울트 증후군 4; 페리 증후군; 영아기의 지속성 고인슐린혈증성 저혈당증; 가족성 고인슐린증; 표현형; 페닐케톤뇨; 크롬친화세포종; 유전성 부신경절종-크롬친화세포종 증후군; 부신경절종 1; 장의 카르시노이드 종양; 코우덴 증후군 3; 포스포글리세레이트 데히드로게나제 결핍; 포스포글리세레이트 키나제 1 결핍; 감광성 모발유황이영양증; 피탄산 축적 질환; 픽병; 피어슨 증후군; 색소성 망막 이영양증; 색소침착 결절성 부신피질 질환, 원발성, 1; 모기질세포종; 피트-홉킨스 증후군; 뇌하수체 의존성 고코르티솔증; 뇌하수체 호르몬 결핍, 복합 1, 2, 3, 및 4; 플라스미노겐 활성인자 억제제 유형 1 결핍; 플라스미노겐 결핍, 유형 I; 혈소판-유형 출혈 장애 15 및 8; 다형피부증, 유전성 섬유화성, 건 구축, 근병증 및 폐 섬유증 동반; 다낭성 신장 질환 2, 성인 유형, 및 영아 유형; 경화성 백질뇌병증을 동반한 다낭성 지질막성 골이형성증; 면역결핍 동반 또는 비동반 폴리글루코산체 근병증 1; 다소뇌회증, 비대칭, 양측 전두두정; 다발신경병증, 청각 상실, 운동실조, 색소성 망막염, 및 백내장; 교뇌소뇌 저형성증 유형 4; 슬와 익상편 증후군; 공뇌증 2; 한공각화증 8, 파종성 표재성 광선 유형; 포르포빌리노겐 신타제 결핍; 만발성 피부 포르피린증; 색소

성 망막염을 동반한 척수후 기둥 운동실조; 후방 극성 백내장 유형 2; 프라더-윌리-유사 증후군; 조기 난소 부전 4, 5, 7, 및 9; 원발성 상염색체 열성 소두증 10, 2, 3, 및 5; 원발성 섬모 이상운동증 24; 원발성 확장성 심근병증; 좌심실 비치밀 6; 4, 좌심실 비치밀 10; 발작성 심방 세동; 원발성 고옥살산뇨, 유형 I, 유형, 및 유형 III; 원발성 비대성 골관절병증, 상염색체 열성 2; 원발성 저마그네슘혈증; 원발성 개방각 녹내장 소아 발병 1; 원발성 폐고혈압; 프림로즈 증후군; 진행성 가족성 심장 차단 유형 1B; 진행성 가족성 간내 담즙정체 2 및 3; 진행성 간내 담즙정체; 운동실조를 동반한 진행성 근간대성경련 간질; 진행성 가성류마티스 이형성증; 진행성 경화성 회백질위축증; 프롤리다제 결핍; 프롤린 데히드로게나제 결핍; 정신분열증 4; 프로페르딘 결핍, X-연관; 프로피온산혈증; 전구단백질 컨버타제 1/3 결핍; 전립선암, 유전성, 2; 적색각이상 결함; 단백뇨; 핀란드 선천성 신증후군; 프로테우스 증후군; 유방 선암종; 가연골무형성 척추골단 이형성증 증후군; 제1형 가성저알도스테론증 상염색체 우성 및 열성 및 유형 2; 가성부갑상선기능저하증 유형 1A, 가가성부갑상선기능저하증; 가성 신생아 부신백질이영양증; 가성원발성 고알도스테론증; 탄성섬유 가성황색종; 영아기의 전신 동맥 석회화 2; 다중 응고 인자 결핍을 동반한 탄성섬유 가성황색종-유사 장애; 건선 감수성 2; PTEN 과오종 종양 증후군; 유전성 출혈성 모세혈관확장증 관련 폐동맥 고혈압; 폐 섬유증 및/또는 골수 부전, 텔로미어-관련, 1 및 3; 폐고혈압, 원발성, 1, 유전성 출혈성 모세혈관확장증 동반; 퓨린-뉴클레오시드 포스포릴라제 결핍; 피루베이트 카르복실라제 결핍; 피루베이트 데히드로게나제 E1-알파 결핍; 적혈구의 피루베이트 키나제 결핍; 레인 증후군; RAS병증; 열성 이영양성 수포성 표피박리증; 손발톱 장애, 비증후군성 선천성, 8; 라이펜스타인 증후군; 신장 이형성증; 신장 카르니틴 수송 결함; 신장 결손증 증후군; 신장 이형성증; 신장 이형성증, 망막 색소성 이영양증, 소뇌 운동실조 및 골격 이형성증; 신세관성 산증, 원위, 상염색체 열성, 후기-발병 감각신경성 청각 상실 동반, 또는 용혈성 빈혈 동반; 신세관성 산증, 근위, 안구 이상 및 정신 지체 동반; 망막 원추 이영양증 3B; 색소성 망막염; 색소성 망막염 10, 11, 12, 14, 15, 17, 및 19; 색소성 망막염 2, 20, 25, 35, 36, 38, 39, 4, 40, 43, 45, 48, 66, 7, 70, 72; 망막모세포종; 레트 장애; 횡문근양 종양 소인 증후군 2; 열공성 망막 박리, 상염색체 우성; 근형부 점상 연골이형성증 유형 2 및 유형 3; 로버츠-SC 해표지증 증후군; 로비노 소라우프 증후군; 로비노 증후군, 상염색체 열성, 상염색체 열성, 단지-신-다지증 동반; 로트문드-톰슨 증후군; 라파딜리노 증후군; RRM2B-관련 미토콘드리아 질환; 루빈스타인-테이비 증후군; 살라병; 샌드호프병, 성인 및 영아 유형; 사르코이드증, 조기 발병; 블라우 증후군; 쉰들러병, 유형 1; 분열뇌증; 정신분열증 15; 슈네켄베켄 이형성증; 슈반세포종증 2; 슈바르츠 얌펠 증후군 유형 1; 공막화각막, 상염색체 열성; 경화협착; 속발성 갑상선기능저하증; 세가와 증후군, 상염색체 열성; 시니어-로켄 증후군 4 및 5; 감각 실조 신경병증, 구음장애, 및 안근부전마비; 세피아프테린 리덕타제 결핍; 참깨 증후군; ADA 결핍으로 인한 중증 복합 면역결핍, 소두증, 성장 지연 및 이온화 방사선에 대한 감수성 동반, 비정형, 상염색체 열성, T 세포-음성, B 세포-양성, NK 세포-음성 또는 NK-양성; 중증 선천성 호중구감소증; 중증 선천성 호중구감소증 3, 상염색체 열성 또는 우성; 중증 선천성 호중구감소증 및 6, 상염색체 열성; 영아기에서의 중증 근간대성 간질; 열성 발작을 동반한 전신 간질 플러스, 유형 1 및 2; 중증 X-연관 근세관성 근병증; 짧은 QT 증후군 3; 비특이적 골격 이상을 동반한 저신장; 저신장, 이도 폐쇄증, 하악 저형성증, 골격 이상; 저신장, 손발톱이형성증, 안면 이형, 및 털감소증; 원발 왜소증; 다지증 동반 또는 비동반 짧은-늑골 흉부 이형성증 11 또는 3; 시알산증 유형 I 및 II; 은 경직성 하반신마비 증후군; 느려진 신경 전도 속도, 상염색체 우성; 스미스-렘리-오피츠 증후군; 스나이더 로빈슨 증후군; 소마토트로프 선종; 프로락틴선종; 가족성, 뇌하수체 선종 소인; 소토스 증후군 1 또는 2; 경직성 운동실조 5, 상염색체 열성, 샤를부아-사기네이 유형, 1, 10, 또는 11, 상염색체 열성; 근위축성 측삭 경화증 유형 5; 경직성 하반신마비 15, 2, 3, 35, 39, 4, 상염색체 우성, 55, 상염색체 열성, 및 5A; 담즙산 합성 결함, 선천성, 3; 정자발생 실패 11, 3, 및 8; 구상적혈구증 유형 4 및 5; 타원체 근병증; 척수성 근육 위축, 하지 우세형 2, 상염색체 우성; 척수성 근육 위축, 유형 II; 척수소뇌성 운동실조 14, 21, 35, 40, 및 6; 척수소뇌성 운동실조 상염색체 열성 1 및 16; 비장 저형성증; 척추손목발목 유합증 증후군; 스폰딜로케이로디스플라시아, 엘러스-단로스 증후군-유사, 면역 조절이상 동반, 아그레칸 유형, 선천성 관절 탈구 동반, 짧은 사지-손 유형, 세다가티안 유형, 추체-간체 이영양증 동반, 및 코즐로우스키 유형; 파라스트렘마틱 왜소증; 스타르가르트병 1; 추체-간체 이영양증 3; 스티클러 증후군 유형 1; 니스트 이형성증; 스티클러 증후군, 유형 1 (비증후군성 안구) 및 4; 자상-연관 혈관병증, 영아-발병; 스토르모르켄 증후군; 스터지-웨버 증후군, 모세관 기형, 선천성, 1; 숙시닐-CoA 아세토아세테이트 트랜스퍼라제 결핍; 수크라제-이소말타제 결핍; 영아 돌연사 증후군; 술파이트 옥시다제 결핍, 고립성; 판막위 대동맥 협착; 계면활성제 대사 기능장애, 폐, 2 및 3; 지절유합증, 근위, 1b; 합지증 세나니 렌즈 유형; 합지증 유형 3; 증후군성 X-연관 정신 지체 16; 내반첨족; 탄지에르병; TARP 증후군; 테이-삭스병, B1 변이형, Gm2-강글리오시드증 (성인), Gm2-강글리오시드증 (성인-발병); 템타미 증후군; 테노리오 증후군; 말단 골이형성증; 테스토스테론 17-베타-데히드로게나제 결핍; 사지절단증, 상염색체 열성; 팔로 사징증; 형성부전성 좌심 증후군 2; 동맥간증; 심장 및 대혈관의 기형; 심실 중격 결손 1; 티엘-벤케 각막 이영양증; 흉부 대동맥류

및 대동맥 박리; 마르파노이드 체형; 3 M 증후군 2; 혈소판감소증, 혈소판 기능장애, 용혈, 및 불균형 글로빈 합성; 혈소판감소증, X-연관; 혈전성향증, 유전성, 단백질 C 결핍으로 인함, 상염색체 우성 및 열성; 갑상선 무발생증; 갑상선암, 여포성; 갑상선 호르몬 대사, 비정상적; 갑상선 호르몬 저항성, 전신, 상염색체 우성; 갑상선중독성 주기성 마비 및 갑상선중독성 주기성 마비 2; 티로트로핀-방출 호르몬 저항성, 전신; 티모시 증후군; TNF 수용체-연관 주기성 발열 증후군 (TRAPS); 치아 무발생증, 선택적, 3 및 4; 토르사드 드 포인트; 타운즈-브룩스-아가미귀신장-유사 증후군; 신생아의 일과성 수포성 피부박리증; 트리처 콜린스 증후군 1; 정신 지체, 왜소증 및 색소성 망막 변성을 동반한 긴속눈썹증; 모발비지절 이형성증 유형 I; 모발비지절 증후군 유형 3; 트리메틸아민뇨; 결절성 경화증 증후군; 림프관근종증; 결절성 경화증 1 및 2; 티로시나제-음성 눈피부 백색증; 티로시나제-양성 눈피부 백색증; 티로신혈증 유형 I; UDP글루코스-4-에피머라제 결핍; 울리히 선천성 근육 이영양증; 중증 사지 결핍을 동반한 척골 및 비골 부재; 업쇼-슐만 증후군; 우로카네이트 히드라타제 결핍; 어셔 증후군, 유형 1, 1B, 1D, 1G, 2A, 2C, 및 2D; 색소성 망막염 39; UV-감수성 증후군; 반더우드 증후군; 반 말데르겜 증후군 2; 헤네캄 림프관확장증-림프부종 증후군 2; 혼합 포르피린증; 낭성 신장 질환을 동반한 뇌실확장증; 베르헤이즈 증후군; 초장쇄 아실-CoA 데히드로게나제 결핍; 방광요관 역류 8; 내장 착위 5, 상염색체; 내장 근병증; 비타민 D-의존성 구루병, 유형 1 및 2; 난황상 이영양증; 폰 빌레브란트 질환 유형 2M 및 유형 3; 바르덴부르크 증후군 유형 1, 4C, 및 2E (신경계 침범 동반); 클라인-바르덴베르크 증후군; 바커-바르부르크 선천성 근육 이영양증; 바르부르크 마이크로 증후군 2 및 4; 사마귀, 저감마글로불린혈증, 감염, 및 골수배출장애; 위버 증후군; 바일-마르케사니 증후군 1 및 3; 바일-마르케사니-유사 증후군; 바이센바허-즈베이뮬러 증후군; 베르드니히-호프만병; 샤르코-마리-투스병; 베르너 증후군; WFS1-관련 장애; 비데만-스타이너 증후군; 윌슨병; 볼프람-유사 증후군, 상염색체 우성; 워스병; 반 부헴병 유형 2; 색소성 건피증, 상보군 b, D군, E군, 및 G군; X-연관 무감마글로불린혈증; X-연관 유전성 운동 및 감각 신경병증; 스테릴-술파타제 결핍을 동반한 X-연관 어린선; X-연관 뇌실주위 이소증; 귀-구개-수족지 증후군, 유형 I; X-연관 중증 복합 면역결핍; 짐메르만-라반트 증후군 및 짐메르만-라반트 증후군 2; 및 층판 분말상 백내장 3.

[1718] 표적 뉴클레오티드 서열은 질환, 장애, 또는 상태와 연관된 표적 서열 (예를 들어, 점 돌연변이)을 포함할 수 있다. 표적 서열은 질환, 장애, 또는 상태와 연관된 T에서 C로의 (또는 A에서 G로의) 점 돌연변이를 포함할 수 있고, 여기서 돌연변이체 C 염기의 탈아미노화는 질환, 장애, 또는 상태와 연관되지 않은 서열에 대한 미스매치 복구-매개된 교정을 발생시킨다. 표적 서열은 질환, 장애, 또는 상태와 연관된 G에서 A로의 (또는 C에서 T로의) 점 돌연변이를 포함할 수 있고, 여기서 돌연변이체 A 염기의 탈아미노화는 질환, 장애, 또는 상태와 연관되지 않은 서열에 대한 미스매치 복구-매개된 교정을 발생시킨다. 표적 서열은 단백질을 코딩할 수 있고, 여기서 점 돌연변이는 코돈 내에 있고, 야생형 코돈과 비교하여 돌연변이체 코돈에 의해 코딩된 아미노산에서 변화를 발생시킨다. 표적 서열은 또한 스플라이스 부위에 있을 수 있고, 점 돌연변이는 야생형 전사체와 비교하여 mRNA 전사체의 스플라이싱에서 변화를 발생시킨다. 또한, 표적은 유전자의 비-코딩 서열, 예컨대 프로모터에 있을 수 있고, 점 돌연변이는 유전자의 증가된 또는 감소된 발현을 발생시킨다.

[1719] 따라서, 일부 측면에서, 돌연변이체 C의 탈아미노화는 돌연변이체 코돈에 의해 코딩된 아미노산의 변화를 발생시키며, 이는 일부 경우에 야생형 아미노산의 발현을 발생시킬 수 있다. 다른 측면에서, 돌연변이체 A의 탈아미노화는 돌연변이체 코돈에 의해 코딩된 아미노산의 변화를 발생시키며, 이는 일부 경우에 야생형 아미노산의 발현을 발생시킬 수 있다.

[1720] 세포를 조성물 또는 rAAV 입자와 접촉시키는 것을 수반하는 본원에 기재된 방법은 시험관내, 생체외, 또는 생체내에서 이루어질 수 있다. 특정 실시양태에서, 접촉 단계는 대상체에서 이루어진다. 특정 실시양태에서, 대상체는 질환, 장애, 또는 상태로 진단되었다.

[1721] 일부 실시양태에서, 본원에 개시된 방법은 포유동물 세포를 조성물 또는 rAAV 입자와 접촉시키는 것을 수반한다. 특정한 실시양태에서, 방법은 망막 세포, 피질 세포 또는 소뇌 세포를 접촉시키는 것을 수반한다.

[1722] 본원에 기재된 방법을 사용하여 전달된 분할 Cas9 단백질 또는 분할 프라임 편집제는 바람직하게는 원래의 Cas9 단백질 또는 프라임 편집제 (즉, 세포에 전달되거나 또는 전체로서 세포에서 발현된 비분할 단백질)와 비교하여 대등한 활성을 갖는다. 예를 들어, 분할 Cas9 단백질 또는 분할 프라임 편집제는 원래의 Cas9 단백질 또는 프라임 편집제의 활성의 적어도 50% (예를 들어, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100%)를 보유한다. 일부 실시양태에서, 분할 Cas9 단백질 또는 분할 프라임 편집제는 원래의 Cas9 단백질 또는 프라임 편집제의 것보다 더 활성이다 (예를 들어, 2배, 5배, 10배, 100배, 1000배, 또는 그 초과).

[1723] 본원에 기재된 조성물은 대상체가 앓고 있는 질환 또는 장애를 치료 및/또는 예방하기 위한 치료 유효량으로 그를 필요로 하는 대상체에게 투여될 수 있다. CRISPR/Cas9-기반 게놈-편집 기술을 사용하여 치료 및/또는 예방될 수 있는 임의의 질환 또는 장애는 본원에 기재된 분할 Cas9 단백질 또는 분할 다중-플랩 프라임 편집제에 의해 치료될 수 있다. 분할 Cas9 단백질 또는 다중-플랩 프라임 편집제를 코딩하는 뉴클레오티드 서열이 gRNA를 추가로 코딩하지 않는 경우에, gRNA를 코딩하는 개별 핵산 벡터가 본원에 기재된 조성물과 함께 투여될 수 있는 것으로 이해되어야 한다.

[1724] 예시적인 적합한 질환, 장애 또는 상태는, 비제한적으로, 낭성 섬유증, 페닐케톤뇨, 표피박리성 과다각화증 (EHK), 만성 폐쇄성 폐 질환 (COPD), 샤르코-마리-투스병 유형 4J, 신경모세포종 (NB), 폰 빌레브란트병 (vWD), 선천성 근긴장증, 유전성 신장 아밀로이드증, 확장성 심근병증, 유전성 림프부종, 가족성 알츠하이머병, 프리온 질환, 만성 영아 신경계 피부 관절 증후군 (CINCA), 선천성 난청, 니만-픽병 유형 C (NPC) 질환, 및 데스민-관련 근병증 (DRM)으로 이루어진 군으로부터 선택된 질환 또는 장애를 포함한다. 특정한 실시양태에서, 질환 또는 상태는 니만-픽병 유형 C (NPC) 질환이다.

[1725] 일부 실시양태에서, 질환, 장애 또는 상태는 NPC 유전자, DNMT1 유전자, PCSK9 유전자, 또는 TMC1 유전자에서의 점 돌연변이와 연관된다. 특정 실시양태에서, 점 돌연변이는 NPC에서의 T3182C 돌연변이이며, 이는 I1061T 아미노산 치환을 발생시킨다.

[1726] 특정 실시양태에서, 점 돌연변이는 TMC1에서의 A545G 돌연변이이며, 이는 Y182C 아미노산 치환을 발생시킨다. TMC1은 내이의 감각 유모 세포에서 기계감수성 이온 채널을 형성하고 정상적인 청각 기능에 요구되는 단백질을 코딩한다. Y182C 아미노산 치환은 선천성 난청과 연관된다.

[1727] 일부 실시양태에서, 질환, 장애 또는 상태는 유전자의 코딩 영역 내에 정지 코돈, 예를 들어 조기 정지 코돈을 생성하는 점 돌연변이와 연관된다.

[1728] 추가의 예시적인 질환, 장애 및 상태는 낭성 섬유증 (예를 들어, 문헌 [Schwank et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell stem cell. 2013; 13: 653-658; 및 Wu et al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. Cell stem cell. 2013; 13: 659-662] 참조, 이들 중 어느 것도 유전적 결함을 교정하기 위해 데아미나제 융합 단백질을 사용하지 않음); 페닐케톤뇨 - 예를 들어, 페닐알라닌 히드록실라제 유전자에서의 위치 835 (마우스) 또는 240 (인간) 또는 상동 잔기에서 페닐알라닌에서 세린으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [McDonald et al., Genomics. 1997; 39:402-405] 참조; 버나드-술리에 증후군 (BSS) - 예를 들어, 혈소판 막당단백질 IX에서의 위치 55 또는 상동 잔기에서 페닐알라닌에서 세린으로의 돌연변이, 또는 잔기 24 또는 상동 잔기에서 시스테인에서 아르기닌으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Noris et al., British Journal of Haematology. 1997; 97: 312-320, 및 Ali et al., Hematol. 2014; 93: 381-384] 참조; 표피박리성 과다각화증 (EHK) - 예를 들어, 케라틴 1에서의 위치 160 또는 161 (개시인자 메티오닌을 카운팅하는 경우) 또는 상동 잔기에서 류신에서 프롤린으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Chipev et al., Cell. 1992; 70: 821-828] 참조, 또한 www[dot]uniprot[dot]org에서 유니프롯 데이터베이스에서 수탁 번호 P04264 참조; 만성 폐쇄성 폐 질환 (COPD) - 예를 들어, α_1-항트립신의 프로세싱된 형태에서의 위치 54 또는 55 (개시인자 메티오닌을 카운팅하는 경우) 또는 상동 잔기에서, 또는 비프로세싱된 형태에서의 잔기 78 또는 상동 잔기에서 류신에서 프롤린으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Poller et al., Genomics. 1993; 17: 740-743] 참조, 또한 유니프롯 데이터베이스에서 수탁 번호 P01011 참조; 샤르코-마리-투스병 유형 4J - 예를 들어, FIG4에서의 위치 41 또는 상동 잔기에서 이소류신에서 트레오닌으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Lenk et al., PLoS Genetics. 2011; 7: e1002104] 참조; 신경모세포종 (NB) - 예를 들어, 카스파제-9에서의 위치 197 또는 상동 잔기에서 류신에서 프롤린으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Kundu et al., 3 Biotech. 2013, 3:225-234] 참조; 폰 빌레브란트병 (vWD) - 예를 들어, 폰 빌레브란트 인자의 프로세싱된 형태에서의 위치 509 또는 상동 잔기에서, 또는 폰 빌레브란트 인자의 비프로세싱된 형태에서의 위치 1272 또는 상동 잔기에서 시스테인에서 아르기닌으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Lavergne et al., Br. J. Haematol. 1992] 참조, 또한 유니프롯 데이터베이스에서 수탁 번호 P04275 참조; 82: 66-72; 선천성 근긴장증 - 예를 들어, 근육 클로라이드 채널 유전자 CLCN1에서의 위치 277 또는 상동 잔기에서 시스테인에서 아르기닌으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Weinberger et al., The J. of Physiology. 2012; 590: 3449-3464] 참조; 유전성 신장 아밀로이드증 - 예를 들어, 아포지단백질 AII의 프로세싱된 형태에서의 위치 78 또는 상동 잔기에서, 또는 비프로세싱된 형태에서의 위치 101 또는 상동 잔기에서

정지 코돈에서 아르기닌으로의 돌연변이 (T>C 돌연변이) - 예를 들어, 문헌 [Yazaki et al., Kidney Int. 2003; 64: 11-16] 참조; 확장성 심근병증 (DCM) - 예를 들어, FOXD4 유전자에서의 위치 148 또는 상동 잔기에서 트립토판에서 아르기닌으로의 돌연변이 (T>C 돌연변이), 예를 들어, 문헌 [Minoretti et al., Int. J. of Mol. Med. 2007; 19: 369-372] 참조; 유전성 림프부종 - 예를 들어, VEGFR3 티로신 키나제에서의 위치 1035 또는 상동 잔기에서 히스티딘에서 아르기닌으로의 돌연변이 (A>G 돌연변이), 예를 들어, 문헌 [Irrthum et al., Am. J. Hum. Genet. 2000; 67: 295-301] 참조; 가족성 알츠하이머병 - 예를 들어, 프레세닐린1에서의 위치 143 또는 상동 잔기에서 이소류신에서 발린으로의 돌연변이 (A>G 돌연변이), 예를 들어, 문헌 [Gallo et al., J. Alzheimer's disease. 2011; 25: 425-431] 참조; 프리온 질환 - 예를 들어, 프리온 단백질에서의 위치 129 또는 상동 잔기에서 메티오닌에서 발린으로의 돌연변이 (A>G 돌연변이) - 예를 들어, 문헌 [Lewis et al., J. of General Virology. 2006; 87: 2443-2449] 참조; 만성 영아 신경학적 피부 관절 증후군 (CINCA) - 예를 들어, 크리오피린에서의 위치 570 또는 상동 잔기에서 티로신에서 시스테인으로의 돌연변이 (A>G 돌연변이) - 예를 들어, 문헌 [Fujisawa et al., Blood. 2007; 109: 2903-2911] 참조; 및 데스민-관련 근병증 (DRM) - 예를 들어, αβ 크리스탈린에서의 위치 120 또는 상동 잔기에서 아르기닌에서 글리신으로의 돌연변이 (A>G 돌연변이) - 예를 들어, 문헌 [Kumar et al., J. Biol. Chem. 1999; 274: 24137-24141] 참조를 포함한다. 모든 참고문헌 및 데이터베이스 엔트리의 전체 내용은 본원에 참조로 포함된다.

[1729] 트리뉴클레오티드 반복부 확장 질환

[1730] 트리뉴클레오티드 반복부 확장은 헌팅톤병, 유약 X 증후군, 및 프리드라이히 운동실조를 포함한 다수의 인간 질환과 연관된다. 가장 흔한 트리뉴클레오티드 반복부는 CAG 트리플렛을 함유하지만, GAA 트리플렛 (프리드라이히 운동실조) 및 CGG 트리플렛 (유약 X 증후군)이 또한 발생한다. 확장에 대한 소인을 유전하거나, 또는 이미 확장된 모 대립유전자를 획득하는 것은 질환을 획득할 가능성을 증가시킨다. 트리뉴클레오티드 반복부의 병원성 확장은 가설상 다중-플랩 프라임 편집을 사용하여 교정될 수 있다.

[1731] 반복 영역의 상류 영역은 RNA-가이드된 뉴클레아제에 의해 닉킹될 수 있고, 이는 이어서 도 1g 또는 도 22에 약술된 일반적 메카니즘에 따라 건강한 수의 반복부 (이는 특정한 유전자 및 질환에 의존함)를 함유하는 새로운 DNA 가닥의 합성을 프라이밍하는 데 사용될 수 있다. 반복 서열 후에, 반복부의 다른 단부 (적색 가닥)에 인접한 서열의 정체와 매칭되는 짧은 상동성 스트레치가 부가된다. TPRT 시스템에 의한 새로 합성된 가닥의 침입, 및 내인성 DNA의 새로 합성된 플랩으로의 후속 대체는 축소된 반복 대립유전자로 이어진다. 용어 "축소된"은 뉴클레오티드 반복 영역의 길이를 단축시켜 트리뉴클레오티드 반복 영역을 복구하는 것을 지칭한다.

[1732] 본원에 기재된 다중-플랩 프라임 편집 시스템 (예를 들어, 이중-플랩 및 사중-플랩 프라임 편집 시스템)은 헌팅톤병 및 다른 트리뉴클레오티드 반복 장애와 같은 상태를 치료하기 위해 트리뉴클레오티드 반복 돌연변이 (또는 "트리플렛 확장 질환")를 축소시키는데 사용될 수 있다. 트리뉴클레오티드 반복부 확장 장애는 발달 신경생물학을 수반하고 종종 인지 뿐만 아니라 감각-운동 기능에 영향을 미치는 복잡한 진행성 장애이다. 장애는 유전학적 표현촉진을 나타낸다 (즉, 각 세대에 따라 중증도가 증가됨). DNA 확장 또는 축소는 통상적으로 감수분열적으로 (즉, 배우자발생 시간 동안, 또는 배아 발생 초기에) 일어나고, 종종 성별-편향을 가지며, 이는 일부 유전자가 여성을 통해 유전될 때에만 확장되고, 다른 것은 남성을 통해서만 확장된다는 것을 의미한다. 인간에서, 트리뉴클레오티드 반복부 확장 장애는 전사 또는 번역 수준에서 유전자 침묵을 유발할 수 있고, 이는 본질적으로 유전자 기능을 녹아웃시킨다. 대안적으로, 트리뉴클레오티드 반복부 확장 장애는 종종 우성-음성 방식으로 (예를 들어 폴리-글루타민 질환) 단백질 기능을 제거하거나 변화시키는 대형의 반복 아미노산 서열로 생성된 변경된 단백질을 유발할 수 있다.

[1733] 이론에 얽매이는 것을 원하지는 않지만, 트리플렛 확장은 DNA 복제 동안 또는 DNA 복구 합성 동안 슬리피지에 의해 유발된다. 탠덤 반복부는 서로 동일한 서열을 갖기 때문에, 2개의 DNA 가닥 사이의 염기 쌍형성은 서열을 따라 다중 지점에서 일어날 수 있다. 이는 DNA 복제 또는 DNA 복구 합성 동안 "루프 아웃" 구조의 형성으로 이어질 수 있다. 이는 반복 서열의 반복 카피를 유도하여, 반복부의 수를 확장시킬 수 있다. 하이브리드 RNA:DNA 중간체를 수반하는 추가의 메카니즘이 제안되었다. 다중-플랩 프라임 편집은 1개 이상의 또는 문제의 반복 코돈 트리플렛을 결실시킴으로써 이들 트리플렛 확장 영역을 감소 또는 제거하는 데 사용될 수 있다. 이러한 용도의 한 실시양태에서, 도 23은 프라임 편집으로 트리뉴클레오티드 반복 서열을 축소 또는 감소시키기 위한 PEgRNA 설계의 개략도를 제공한다.

[1734] 다중-플랩 프라임 편집은 절단 부위에 적절하게 표적화된 PEgRNA를 포함하는 프라임 편집제로 트리플렛 반복 영역의 상류 영역을 닉킹함으로써 트리플렛 확장 영역을 축소시키도록 구현될 수 있다. 이어서, 프라임 편집제는

건강한 수의 트리플렛 반복부 (이는 특정한 유전자 및 질환에 의존함)를 코딩하는 주형 (즉, 그의 편집 주형)으로서 PEgRNA에 기초하여 새로운 DNA 가닥 (ssDNA 플랩)을 합성한다. 건강한 트리플렛 반복 서열을 포함하는 새로 합성된 ssDNA 가닥은 또한 반복부의 다른 단부 (적색 가닥)에 인접한 서열에 매칭되는 짧은 상동성 스트레치 (즉, 상동성 아암)를 포함하도록 합성된다. 새로 합성된 가닥의 침입, 및 내인성 DNA의 새로 합성된 ssDNA 플랩으로의 후속 대체는 축소된 반복 대립유전자로 이어진다.

[1735] 특정한 트리뉴클레오티드 확장 장애에 따라, "트리뉴클레오티드 반복부 확장 단백질"에서 결함-유도 트리플렛 확장이 발생할 수 있다. 트리뉴클레오티드 반복부 확장 단백질은 트리뉴클레오티드 반복부 확장 장애의 발생에 대한 감수성, 트리뉴클레오티드 반복부 확장 장애의 존재, 트리뉴클레오티드 반복부 확장 장애의 중증도 또는 그의 임의의 조합과 연관된 단백질의 다양한 세트이다. 트리뉴클레오티드 반복부 확장 장애는 반복부의 유형에 의해 결정되는 2개의 카테고리로 나뉜다. 가장 흔한 반복부는 트리플렛 CAG이며, 이는 유전자의 코딩 영역에 존재하는 경우에 아미노산 글루타민 (Q)을 코딩한다. 따라서, 이들 장애는 폴리글루타민 (폴리Q) 장애로 지칭되고, 하기 질환을 포함한다: 헌팅톤병 (HD); 척수연수성 근육 위축 (SBMA); 척수소뇌성 운동실조 (SCA 유형 1, 2, 3, 6, 7, 및 17); 및 치상핵적핵-담창구시상하핵 위축 (DRPLA). 나머지 트리뉴클레오티드 반복부 확장 장애는 CAG 트리플렛을 수반하지 않거나 또는 CAG 트리플렛이 유전자의 코딩 영역에 있지 않고, 따라서 비-폴리글루타민 장애로 지칭된다. 비-폴리글루타민 장애는 유약 X 증후군 (FRAXA); 유약 XE 정신 지체 (FRAXE); 프리드라이히 운동실조 (FRDA); 근긴장성 이영양증 (DM); 및 척수소뇌 운동실조 (SCA 유형 8, 및 12)를 포함한다.

[1736] 트리뉴클레오티드 반복부 확장 장애와 연관된 단백질은 트리뉴클레오티드 반복부 확장 장애에 대한 트리뉴클레오티드 반복부 확장 장애와 연관된 단백질의 실험적 연관에 기초하여 선택될 수 있다. 예를 들어, 트리뉴클레오티드 반복부 확장 장애와 연관된 단백질의 생산율 또는 순환 농도는 트리뉴클레오티드 반복부 확장 장애를 갖는 집단에서 트리뉴클레오티드 반복부 확장 장애가 결여된 집단에 비해 상승 또는 저하될 수 있다. 단백질 수준의 차이는 웨스턴 블롯, 면역조직화학적 염색, 효소 연결 면역흡착 검정 (ELISA) 및 질량 분광측정법을 포함하나 이에 제한되지는 않는 단백질체학 기술을 사용하여 평가될 수 있다. 대안적으로, 트리뉴클레오티드 반복부 확장 장애와 연관된 단백질은 DNA 마이크로어레이 분석, 유전자 발현의 연속 분석 (SAGE), 및 정량적 실시간 폴리머라제 연쇄 반응 (Q-PCR)을 포함하나 이에 제한되지는 않는 게놈 기술을 사용하여 단백질을 코딩하는 유전자의 유전자 발현 프로파일을 수득함으로써 확인될 수 있다.

[1737] 다중-플랩 프라임 편집에 의해 교정될 수 있는 트리뉴클레오티드 반복부 확장 장애와 연관된 단백질의 비제한적 예는 AR (안드로겐 수용체), FMR1 (유약 X 정신 지체 1), HTT (헌팅틴), DMPK (근긴장성 이영양증-단백질 키나제), FXN (프라탁신), ATXN2 (아탁신 2), ATN1 (아트로핀 1), FEN1 (플랩 구조-특이적 엔도뉴클레아제 1), TNRC6A (트리뉴클레오티드 반복부 함유 6A), PABPN1 (폴리(A) 결합 단백질, 핵 1), JPH3 (준크토필린 3), MED15 (매개체 착물 서브유닛 15), ATXN1 (아탁신 1), ATXN3 (아탁신 3), TBP (TATA 박스 결합 단백질), CACNA1A (칼슘 채널, 전압-의존성, P/Q 유형, 알파 1A 서브유닛), ATXN80S (ATXN8 대향 가닥 (비-단백질 코딩)), PPP2R2B (단백질 포스파타제 2, 조절 서브유닛 B, 베타), ATXN7 (아탁신 7), TNRC6B (트리뉴클레오티드 반복부 함유 6B), TNRC6C (트리뉴클레오티드 반복부 함유 6C), CELF3 (CUGBP, Elav-유사 패밀리 구성원 3), MAB21L1 (mab-21-유사 1 (씨. 엘레간스(C. elegans))), MSH2 (mutS 상동체 2, 결장암, 비폴립증 유형 1 (이. 콜라이)), TMEM185A (막횡단 단백질 185A), SIX5 (6 호메오박스 5), CNPY3 (캐노피 3 상동체 (제브라피쉬)), FRAXE (유약 부위, 폴산 유형, 희귀, fra(X)(q28) E), GNB2 (구아닌 뉴클레오티드 결합 단백질 (G 단백질), 베타 폴리펩티드 2), RPL14 (리보솜 단백질 L14), ATXN8 (아탁신 8), INSR (인슐린 수용체), TTR (트랜스티레틴), EP400 (E1A 결합 단백질 p400), GIGYF2 (GRB10 상호작용 GYF 단백질 2), OGG1 (8-옥소구아닌 DNA 글리코실라제), STC1 (스탄니오칼신 1), CNDP1 (카르노신 디펩티다제 1 (메탈로펩티다제 M20 패밀리)), C10orf2 (염색체 10 오픈 리딩 프레임 2), MAML3 마스터마인드-유사 3 (드로소필라), DKC1 (선천성 이상각화증 1, 디스케린), PAXIP1 (PAX 상호작용 (전사-활성화 도메인과) 단백질 1), CASK (칼슘/칼모듈린-의존성 세린 단백질 키나제 (MAGUK 패밀리)), MAPT (미세관-연관 단백질 타우), SP1 (Sp1 전사 인자), POLG (폴리머라제 (DNA 지시됨), 감마), AFF2 (AF4/FMR2 패밀리, 구성원 2), THBS1 (트롬보스폰딘 1), TP53 (종양 단백질 p53), ESR1 (에스트로겐 수용체 1), CGGBP1 (CGG 트리플렛 반복부 결합 단백질 1), ABT1 (기저 전사 활성인자 1), KLK3 (칼리크레인-관련 펩티다제 3), PRNP (프리온 단백질), JUN (jun 종양유전자), KCNN3 (칼륨 중간/작은 전도도 칼슘-활성화된 채널, 서브패밀리 N, 구성원 3), BAX (BCL2-연관 X 단백질), FRAXA (유약 부위, 폴산 유형, 희귀, fra(X)(q27.3) A (대고환증, 정신 지체)), KBTBD10 (켈히 반복부 및 BTB (POZ) 도메인 함유 10), MBNL1 (머슬블라인드-유사 (드로소필라)), RAD51 (RAD51 상동체 (RecA 상동체, 이. 콜라이) (에스. 세레비지아에)), NCOA3 (핵 수용체 보조활성인자 3), ERDA1 (확장된 반복부 도메인, CAG/CTG 1), TSC1 (결절성 경화증 1), COMP (연골 올리고머 매트릭스 단백질), GCLC (글루타메이트-시스테인 리가제, 촉매 서브유닛), RRAD (Ras-관련 당뇨병 연

관), MSH3 (mutS 상동체 3 (이. 콜라이)), DRD2 (도파민 수용체 D2), CD44 (CD44 분자 (인디안 혈액형)), CTCF (CCCTC-결합 인자 (아연 핑거 단백질)), CCND1 (시클린 D1), CLSPN (클라스핀 상동체 (크세노푸스 라에비스)), MEF2A (근세포 인핸서 인자 2A), PTPRU (단백질 티로신 포스파타제, 수용체 유형, U), GAPDH (글리세르알데히드-3-포스페이트 데히드로게나제), TRIM22 (3부 모티프-함유 22), WT1 (윌름스 종양 1), AHR (아릴 탄화수소 수용체), GPX1 (글루타티온 퍼옥시다제 1), TPMT (티오퓨린 S-메틸트랜스퍼라제), NDP (노리병 (가성신경교종)), ARX (아리스탈레스 관련 호메오박스), MUS81 (MUS81 엔도뉴클레아제 상동체 (에스. 세레비지아에)), TYR (티로시나제 (눈피부 백색증 IA)), EGR1 (조기 성장 반응 1), UNG (우라실-DNA 글리코실라제), NUMBL (numb 상동체 (드로소필라)-유사), FABP2 (지방산 결합 단백질 2, 장), EN2 (엔그라일드 호메오박스 2), CRYGC (크리스탈린, 감마 C), SRP14 (신호 인식 입자 14 kDa (상동 Alu RNA 결합 단백질)), CRYGB (크리스탈린, 감마 B), PDCD1 (프로그램화된 세포 사멸 1), HOXA1 (호메오박스 A1), ATXN2L (아탁신 2-유사), PMS2 (PMS2 감수분열 분리후 증가 2 (에스. 세레비지아에)), GLA (갈락토시다제, 알파), CBL (Cas-Br-M (뮤린) 동종향성 레트로바이러스 형질전환 서열), FTH1 (페리틴, 중쇄 폴리펩티드 1), IL12RB2 (인터류킨 12 수용체, 베타 2), OTX2 (오르토덴티클 호메오박스 2), HOXA5 (호메오박스 A5), POLG2 (폴리머라제 (DNA 지시됨), 감마 2, 보조 서브유닛), DLX2 (디스탈-레스 호메오박스 2), SIRPA (신호-조절 단백질 알파), OTX1 (오르토덴티클 호메오박스 1), AHRR (아릴-탄화수소 수용체 리프레서), MANF (중간뇌 성상세포-유도된 신경영양 인자), TMEM158 (막횡단 단백질 158 (유전자/위유전자)), 및 ENSG00000078687을 포함한다.

[1738] 특정한 측면에서, 본 개시내용은 확장 반복 장애 (또한 반복부 확장 장애 또는 트리뉴클레오티드 반복 장애로도 공지됨)로 진단된 대상체의 치료를 위한 TPRT-기반 방법을 제공한다. 확장 반복 장애는 미소위성체 반복부가 한계 길이를 넘어 확장되는 경우에 발생한다. 현재, 적어도 30종의 유전 질환이 반복부 확장에 의해 유발되는 것으로 여겨진다. 이러한 다양한 장애 군의 과학적 이해는 1990년대 초반에 트리뉴클레오티드 반복부가 유약 X, 척수성 및 연수성 근육 위축, 근긴장성 이영양증, 및 헌팅톤병 (문헌 [Nelson et al, "The unstable repeats - three evolving faces of neurological disease," Neuron, March 6, 2013, Vol.77; 825-843], 이는 본원에 참조로 포함됨), 뿐만 아니라 하우 리버 증후군, 야콥센 증후군, 치상핵적핵-담창구시상하핵 위축 (DRPLA), 마차도-요셉병, 합다지증 (SPD II), 수족 생식기 증후군 (HFGS), 쇄골두개 이형성증 (CCD), 완전전뇌증 장애 (HPE), 선천성 중추성 과다호흡 증후군 (CCHS), ARX-비증후군성 X-연관 정신 지체 (XLMR), 및 안인두 근육 이영양증 (OPMD)을 포함한 여러 주요 유전성 상태의 기저를 이룬다는 발견으로 인해 이루어졌다 (참조). 미소위성체 반복부 불안정성은 이들 상태의 특징인 것으로 밝혀졌으며, 이는 예상된 바와 같이 - 반복부 확장은 각각의 연속 세대에서 발생할 수 있는 현상이며, 이는 자손에서 보다 중증의 표현형 및 보다 이른 발병 연령으로 이어진다. 반복부 확장은 여러 상이한 메카니즘을 통해 질환을 유발하는 것으로 여겨진다. 즉, 확장은 유전자, mRNA 전사체, 및/또는 코딩된 단백질의 수준에서 세포 기능을 방해할 수 있다. 일부 조건에서, 돌연변이는 반복부-함유 유전자를 침묵시킴으로써 기능 상실 메카니즘을 통해 작용한다. 다른 실시양태에서, 질환은 기능 획득 메카니즘으로 인해 발생하며, 그에 의해 mRNA 전사체 또는 단백질은 새로운 이상 기능을 갖는다.

[1739] 한 실시양태에서, 트리뉴클레오티드 반복 장애를 치료하는 방법이 도 23에 도시된다. 일반적으로, 접근법은 프라임 편집 과정의 메카니즘을 통해 내인성 이환 트리뉴클레오티드 반복부 서열을 대체하도록 의도된 목적하는 건강한 대체 트리뉴클레오티드 반복부 서열을 코딩하는 영역을 포함하는 연장된 gRNA와 조합하여 TPRT 게놈 편집 (즉, 다중-플랩 프라임 편집)을 사용하는 것을 수반한다. TPRT 게놈 편집 (즉, 다중-플랩 프라임 편집)에 의한 트리뉴클레오티드 반복 서열 및 트리뉴클레오티드 반복부 축소를 위한 예시적인 gRNA 설계의 개략도가 도 23에 제시된다.

[1740] 프리온 질환

[1741] 다중-플랩 프라임 편집은 또한 프리온 질환 과정 동안 미스폴딩되는 프리온 단백질 (PRNP) 내로의 1개 이상의 보호 돌연변이의 설치를 통해 프리온 질환의 진행을 예방하거나 중지시키는데 사용될 수 있다. 프리온 질환 또는 전염성 해면상 뇌병증 (TSE)은 인간 및 동물 둘 다에 영향을 미치는 한 패밀리의 희귀 진행성 신경변성 장애이다. 이들은 긴 인큐베이션 기간, 뉴런 손실과 연관된 특징적인 해면상 변화, 및 염증 반응 유도 실패에 의해 구별된다.

[1742] 인간에서, 프리온 질환은 크로이츠펠트-야콥병 (CJD), 변종 크로이츠펠트-야콥병 (vCJD), 게르스트만-스트라우슬러-샤잉커 증후군, 치명적 가족성 불면증, 및 쿠루병을 포함한다. 동물에서, 프리온 질환은 소 해면상 뇌병증 (BSE 또는 "광우병"), 만성 소모성 질환 (CWD), 스크래피, 전염성 밍크 뇌병증, 고양이 해면상 뇌병증, 및 유제류 해면상 뇌병증을 포함한다. 다중-플랩 프라임 편집은 이들 프리온 질환 중 어느 하나의 진행을 예방하

거나 중지시키기 위해 프리온 단백질 내에 보호 점 돌연변이를 설치하는 데 사용될 수 있다.

[1743] 전형적 CJD는 인간 프리온 질환이다. 이는 특징적인 임상 및 진단 특색을 갖는 신경변성 장애이다. 이 질환은 급속 진행성이고, 항상 치명적이다. 이 질환에 의한 감염은 통상적으로 질병의 발병 1년 내에 사망으로 이어진다. CJD는 프리온 단백질로 공지되어 있는 세포 당단백질의 비정상적 이소형에 의해 유발되는 것으로 여겨지는 급속 진행성, 불변의 치명적 신경변성 장애이다. CJD는 전세계적으로 발생하며, 미국을 포함한 많은 국가에서 추정되는 연간 발생률은 인구 백만명당 약 1건의 사례인 것으로 보고되어 있다. 대다수의 CJD 환자는 통상적으로 질병의 발병 1년 내에 사망한다. CJD는 인간 및 동물에서 발생하는 다른 프리온 질환과 함께 전염성 해면상 뇌병증 (TSE)으로 분류된다. 환자의 약 85%에서, CJD는 인지가능한 전달 패턴 없이 산발성 질환으로서 발생한다. 프리온 단백질 유전자의 유전성 돌연변이로 인해, 보다 적은 비율의 환자 (5 내지 15%)에서 CJD가 발생한다. 이들 유전성 형태는 게르스트만-스트라우슬러-샤잉커 증후군 및 치명적 가족성 불면증을 포함한다. CJD에 대한 치료는 현재 공지되어 있지 않다.

[1744] 변종 크로이츠펠트-야콥병 (vCJD)은 1996년에 영국에서 최초로 기재된 프리온 질환이다. 소에서 프리온 질환의 발생을 담당하는 작용제인 소 해면상 뇌병증 (BSE 또는 '광우'병)이 인간에서 vCJD의 발생을 담당하는 동일한 작용제라는 강력한 과학적 증거가 현재 존재한다. 변종 CJD (vCJD)는 전형적 CJD (종종 간단히 CJD로 불림)와 동일한 질환이 아니다. 이는 전형적 CJD와 상이한 임상적 및 병리학적 특징을 갖는다. 각각의 질환은 또한 프리온 단백질 유전자의 특정한 유전자 프로파일을 갖는다. 둘 다의 장애는 수년간 측정되는 비통상적으로 긴 인큐베이션 기간을 갖는 불변의 치명적인 뇌 질환이며, 프리온으로 불리는 비통상적인 전염성 작용제에 의해 유발된다. vCJD에 대한 치료는 현재 공지되어 있지 않다.

[1745] BSE (소 해면상 뇌병증 또는 "광우병")는 프리온으로 불리는 비통상적 전염성 작용제에 의한 감염으로부터 발생하는 소의 진행성 신경계 장애이다. 전염성 작용제의 성질은 잘 이해되어 있지 않다. 현재, 가장 허용되는 이론은 작용제가 프리온 단백질로 공지된 정상 단백질의 변형된 형태라는 것이다. 아직 이해되지 않는 이유로, 정상 프리온 단백질은 이후 소의 중추 신경계를 손상시키는 병원성 (유해) 형태로 변화된다. 상이한 BSE 계통: 영국에서의 발생의 원인이 된 정형의 또는 전형적 BSE 계통 및 2종의 비정형 계통 (H 및 L 계통)이 존재한다는 증거가 증가하고 있다. BSE에 대한 치료는 현재 공지되어 있지 않다.

[1746] 만성 소모성 질환 (CWD)은 사슴, 엘크, 순록, 시카 사슴 및 무스에 영향을 미치는 프리온 질환이다. 이는 캐나다 및 미국을 비롯한 북아메리카, 노르웨이 및 대한민국의 일부 지역에서 발견되었다. 감염된 동물에서 현저한 체중 감소 (소모), 비틀거림, 무기력 및 다른 신경계 증상을 포함할 수 있는 증상이 발생하기 전까지 1년 넘게 걸릴 수 있다. CWD는 모든 연령의 동물에 영향을 미칠 수 있고, 일부 감염된 동물에서는 질환이 전혀 발생하지 않고 사망할 수 있다. CWD는 동물에게 치명적이고, 치료 또는 백신은 없다.

[1747] TSE의 병원체는 프리온인 것으로 여겨진다. 용어 "프리온"은 전염성이고 뇌에서 가장 풍부하게 발견되는 프리온 단백질로 불리는 특정의 정상 세포성 단백질의 비정상적인 폴딩을 유도할 수 있는 비정상적, 병원체를 지칭한다. 이들 정상 프리온 단백질의 기능은 여전히 완전히 이해되지 않고 있다. 프리온 단백질의 비정상적인 폴딩은 뇌 손상 및 질환의 특징적인 징후 및 증상으로 이어진다. 프리온 질환은 통상적으로 급속 진행성이고, 항상 치명적이다.

[1748] 본원에 사용된 용어 "프리온"은 인간 및 동물에서 질환 (해면상 뇌병증)을 유발하는 것으로 공지된 감염성 입자를 의미할 것이다. 용어 "프리온"은 단어 "단백질" 및 "감염"의 축약형이고, 입자는 주로, 독점적이지는 않지만, 입체형태를 변화시켜 $PRNP^{Sc}$가 되는 $PRNP^{C}$를 발현하는 PRNP 유전자에 의해 코딩된 $PRNP^{Sc}$ 분자로 구성된다. 프리온은 박테리아, 바이러스 및 바이로이드와 구별된다. 공지된 프리온은 동물을 감염시켜 양 및 염소의 신경계의 전염성 변성 질환인 스크래피 뿐만 아니라 소 해면상 뇌병증 (BSE) 또는 광우병 및 고양이의 고양이 해면상 뇌병증을 유발하는 것을 포함한다. 인간에게 영향을 미치는 것으로 공지된 상기 논의된 바와 같은 4종의 프리온 질환은 (1) 쿠루병, (2) 크로이츠펠트-야콥병 (CJD), (3) 게르스트만-스트라우슬러-샤잉커병 (GSS), 및 (4) 치명적 가족성 불면증 (FFI)이다. 본원에 사용된 바와 같은 프리온은 사용되는 임의의 동물--및 특히 인간 및 가축에서 이들 질환 또는 그 밖의 것을 모두 또는 그 중 임의의 것을 유발하는 프리온의 모든 형태를 포함한다.

[1749] 일반적으로, 그리고 이론에 얽매이는 것을 원하지는 않지만, 프리온 질환은 프리온 단백질의 미스폴딩에 의해 유발된다. 이러한 질환-종종 침착 질환으로 불림-프리온 단백질의 미스폴딩은 하기와 같이 설명될 수 있다. A가 단량체 또는 올리고머 상태에서 의도된 생리학적 역할을 수행하는 정상적으로 합성된 유전자 생성물이라면, A*는 현저한 입체형태적 변화를 겪기에 적격인 A의 입체형태적으로 활성화된 형태이고, B는 다량체 조립을 선호

하는 입체형태적으로 변경된 상태 (즉, 침착을 형성하는 미스폴딩된 형태)이고, B_n은 병원성이고 비교적 재순환시키기 어려운 다량체 물질이다. 프리온 질환의 경우, $PRNP^C$ 및 $PRNP^{Sc}$는 상태 A 및 B_n에 상응하고, 여기서 A는 주로 나선형 및 단량체이고, B_n은 β-풍부 및 다량체이다.

[1750] 프리온 단백질에서의 특정 돌연변이는 프리온 질환의 증가된 위험과 연관될 수 있는 것으로 공지되어 있다. 반대로, 프리온 단백질에서의 특정 돌연변이는 사실상 보호성일 수 있다. 문헌 [Bagynszky et al., "Characterization of mutations in PRNP (prion) gene and their possible roles in neurodegenerative diseases," Neuropsychiatr Dis Treat., 2018; 14: 2067-2085]을 참조하며, 그의 내용은 본원에 참조로 포함된다.

[1751] PRNP (NCBI RefSeq No. NP_000302.1 (서열식별번호: 291))-인간 프리온 단백질--은 염색체 20 (4686151-4701588)에 위치하는 16 kb 길이의 유전자에 의해 코딩된다. 이는 2개의 엑손을 함유하고, 엑손 2는 253개 아미노산 (AA) 길이의 PrP 단백질을 코딩하는 오픈 리딩 프레임을 보유한다. 엑손 1은 전사 개시 부위로서 작용할 수 있는 비코딩 엑손이다. 번역후 변형은 처음 22개의 AA N-말단 단편 (NTF) 및 마지막 23개의 AA C-말단 단편 (CTF)의 제거를 발생시킨다. NTF는 내형질 세망 (ER)으로의 PrP 수송 후에 절단되는 한편, CTF (글리코실포스파티딜이노시톨 [GPI] 신호 펩티드 [GPI-SP])는 GPI 앵커에 의해 절단된다. GPI 앵커는 PrP 단백질 수송에 수반될 수 있다. 이는 또한 프리온 단백질을 세포 막 외부 표면으로 부착시키는 역할을 할 수 있다. 정상 PrP는 긴 N-말단 루프 (옥타펩티드 반복 영역을 함유함), 2개의 짧은 β 시트, 3개의 α 나선, 및 C-말단 영역 (GPI 앵커를 함유함)으로 구성된다. PrP의 절단은 세포 막에 고정된 208 AA 길이의 당단백질을 생성한다.

[1752] PRNP (NP_000302.1)의 253개 아미노산 서열은 하기와 같다:

[1753] MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQ
PHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAGA
AAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNF
VHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSS
PPVILLISFLIFLIVG (SEQ ID NO: 291).

[1754] PRNP (NP_000302.1)의 253개 아미노산 서열은 하기 뉴클레오티드 서열 (NCBI Ref. Seq No. NM_000311.5, "호모 사피엔스 프리온 단백질 (PRNP), 전사체 변이체 1, mRNA)에 의해 코딩되며, 하기와 같다:

GCGAACCTTGGCTGCTGGATGCTGGTTCTCTTTGTGGCCACATGGAGTGACCTGGGCCTCT
GCAAGAAGCGCCCGAAGCCTGGAGGATGGAACACTGGGGGCAGCCGATACCCGGGGCAG
GGCAGCCCTGGAGGCAACCGCTACCCACCTCAGGGCGGTGGTGGCTGGGGGCAGCCTCA
TGGTGGTGGCTGGGGGCAGCCTCATGGTGGTGGCTGGGGGCAGCCCCATGGTGGTGGCTG
GGGACAGCCTCATGGTGGTGGCTGGGGTCAAGGAGGTGGCACCCACAGTCAGTGGAACA
AGCCGAGTAAGCCAAAAACCAACATGAAGCACATGGCTGGTGCTGCAGCAGCTGGGGCA
GTGGTGGGGGGCCTTGGCGGCTACATGCTGGGAAGTGCCATGAGCAGGCCCATCATACAT
TTCGGCAGTGACTATGAGGACCGTTACTATCGTGAAAACATGCACCGTTACCCCAACCAAG
TGTACTACAGGCCCATGGATGAGTACAGCAACCAGAACAACTTTGTGCACGACTGCGTCA
ATATCACAATCAAGCAGCACACGGTCACCACAACCACCAAGGGGGAGAACTTCACCGAG
ACCGACGTTAAGATGATGGAGCGCGTGGTTGAGCAGATGTGTATCACCCAGTACGAGAGG
GAATCTCAGGCCTATTACCAGAGAGGATCGAGCATGGTCCTCTTCTCCTCTCCACCTGTGA
TCCTCCTGATCTCTTTCCTCATCTTCCTGATAGTGGGATGAGGAAGGTCTTCCTGTTTTCAC
CATCTTTCTAATCTTTTTCCAGCTTGAGGGAGGCGGTATCCACCTGCAGCCCTTTTAGTGGT
GGTGTCTCACTCTTTCTTCTCTCTTTGTCCCGGATAGGCTAATCAATACCCTTGGCACTGAT
GGGCACTGGAAAACATAGAGTAGACCTGAGATGCTGGTCAAGCCCCCTTTGATTGAGTTC
ATCATGAGCCGTTGCTAATGCCAGGCCAGTAAAAGTATAACAGCAAATAACCATTGGTTAA
TCTGGACTTATTTTTGGACTTAGTGCAACAGGTTGAGGCTAAAACAAATCTCAGAACAGTC
TGAAATACCTTTGCCTGGATACCTCTGGCTCCTTCAGCAGCTAGAGCTCAGTATACTAATGC
CCTATCTTAGTAGAGATTTCATAGCTATTTAGAGATATTTTCCATTTTAAGAAAACCCGACAA
CATTTCTGCCAGGTTTGTTAGGAGGCCACATGATACTTATTCAAAAAAATCCTAGAGATTCT
TAGCTCTTGGGATGCAGGCTCAGCCCGCTGGAGCATGAGCTCTGTGTGTACCGAGAACTG
GGGTGATGTTTTACTTTTCACAGTATGGGCTACACAGCAGCTGTTCAACAAGAGTAAATAT
TGTCACAACACTGAACCTCTGGCTAGAGGACATATTCACAGTGAACATAACTGTAACATAT
ATGAAAGGCTTCTGGGACTTGAAATCAAATGTTTGGGAATGGTGCCCTTGGAGGCAACCT
CCCATTTTAGATGTTTAAAGGACCCTATATGTGGCATTCCTTTCTTTAAACTATAGGTAATTA
AGGCAGCTGAAAAGTAAATTGCCTTCTAGACACTGAAGGCAAATCTCCTTTGTCCATTTAC
CTGGAAACCAGAATGATTTTGACATACAGGAGAGCTGCAGTTGTGAAAGCACCATCATCA
TAGAGGATGATGTAATTAAAAAATGGTCAGTGTGCAAAGAAAAGAACTGCTTGCATTTCTT
TATTTCTGTCTCATAATTGTCAAAAACCAGAATTAGGTCAAGTTCATAGTTTCTGTAATTGG
CTTTTGAATCAAAGAATAGGGAGACAATCTAAAAAATATCTTAGGTTGGAGATGACAGAA
ATATGATTGATTTGAAGTGGAAAAAGAAATTCTGTTAATGTTAATTAAAGTAAAATTATTCC
CTGAATTGTTTGATATTGTCACCTAGCAGATATGTATTACTTTTCTGCAATGTTATTATTGGCT
TGCACTTTGTGAGTATTCTATGTAAAAATATATATGTATATAAAATATATATTGCATAGGACAG
ACTTAGGAGTTTTGTTTAGAGCAGTTAACATCTGAAGTGTCTAATGCATTAACTTTTGTAAG
GTACTGAATACTTAATATGTGGGAAACCCTTTTGCGTGGTCCTTAGGCTTACAATGTGCACT
GAATCGTTTCATGTAAGAATCCAAAGTGGACACCATTAACAGGTCTTTGAAATATGCATGT
ACTTTATATTTTCTATATTTGTAACTTTGCATGTTCTTGTTTTGTTATATAAAAAAATTGTAAA
TGTTTAATATCTGACTGAAATTAAACGAGCGAAGATGAGCACCA (SEQ ID NO: 292)

[1755]

[1756] CJD 및 FFI와 연관된 PRNP (NP_000302.1)에 대한 돌연변이 부위는 하기와 같이 보고되어 있다. 이들 돌연변이는 본원에 개시된 다중-플랩 프라임 편집제를 사용하여 제거 또는 설치될 수 있다.

돌연변이	CJD 프리온 질환과 연관된 돌연변이체 PRNP 의 아미노산 서열 (문헌 [BAGYNSZKY ET AL., 2018]의 표 1 참조) (PRNP NP_000302.1 의 SEQ ID NO: 291 대비)
D178N	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHNCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3940)
T188K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHKVTTTTKGENFTETDVKMMER VVEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3941)
E196K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGKNFTETDVKMMER VVEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3942)
E196A	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGANFTETDVKMMER VVEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 296)
E200K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTKTDVKMMER VVEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3943)
E200G	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTGTDVKMMER VVEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 298)
V203I	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDIKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3944)
R208H	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMEH VVEQMCITQYERESQAYYQRGSSMVLFSSPPV(SEQ ID NO: 3945)
V210I	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV IEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3946)
E211Q	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH

[1757]

	RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VQQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3947)
M232R	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSRVLFSSPPV (SEQ ID NO: 3948)

[1758]

[1759] GSS와 연관된 PRNP (NP_000302.1) (서열식별번호: 291)에 대한 돌연변이 부위는 하기와 같이 보고되어 있다:

돌연변이	GSS 프리온 질환과 연관된 돌연변이체 PRNP 의 아미노산 서열 (문헌 [BAGYNSZKY ET AL., 2018]의 표 2 참조) (PRNP NP_000302.1 의 SEQ ID NO: 291 대비)
P102L	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKLS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3949)
P105L	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KLKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3950)
A117V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAVAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHR YPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3951)
G131V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLVSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3952)
V176G	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFGHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3953)
H187R	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQRTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3954)
	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 487)
F198S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENSTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3955)
D202N	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETNVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3956)
Q212P	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH

[1760]

	RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEPMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3957)
Q217R	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITRYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3958)
M232T	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSTVLFSSPPV (SEQ ID NO: 3959)

[1761]

[1762] 프리온 질환에 대한 가능한 보호 성질과 연관된 PRNP (NP_000302.1) (서열식별번호: 291)에 대한 돌연변이 부위는 하기와 같다:

돌연변이	프리온 질환에 대한 보호 성질과 연관된 돌연변이체 PRNP 의 아미노산 서열 (문헌 [BAGYNSZKY ET AL., 2018]의 표 4 참조) (PRNP NP_000302.1 의 SEQ ID NO: 291 대비)
G127S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGSYMLGSAMSRPIIHFGSDYEDRYYRENMHR YPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3960)
G127V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGVYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3961)
M129V	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYVLGSAMSRPIIHFGSDYEDRYYRENMHR YPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3962)
D167G	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMGEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3963)
D167N	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMNEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3964)
N171S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSSQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3965)
E219K	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYKRESQAYYQRGSSMVLFSSPPV (SEQ ID NO: 3966)
P238S	MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQG GGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPS KPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMH RYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERV VEQMCITQYERESQAYYQRGSSMVLFSSSPV (SEQ ID NO: 3967)

[1763]

[1764] 따라서, 다양한 실시양태에서, 다중-플랩 프라임 편집은 PRNP에서 프리온 질환과 연관된 돌연변이를 제거하거나 또는 PRNP 내에 프리온 질환에 대해 보호성인 것으로 간주되는 돌연변이를 설치하기 위해 사용될 수 있다. 예를 들어, 다중-플랩 프라임 편집은 PRNP 단백질 (NP_000302.1의 PRNP 대비) (서열식별번호: 291)에서 D178N, V180I, T188K, E196K, E196A, E200K, E200G, V203I, R208H, V210I, E211Q, I215V, 또는 M232R 돌연변이를 제거 또는 회복시키기 위해 사용될 수 있다. 다른 실시양태에서, 다중-플랩 프라임 편집은 PRNP 단백질 (NP_000302.1의 PRNP 대비) (서열식별번호: 291)에서 P102L, P105L, A117V, G131V, V176G, H187R, F198S, D202N, Q212P, Q217R, 또는 M232T 돌연변이를 제거 또는 회복시키기 위해 사용될 수 있다. 프라임 편집을 사용하여 PRNP에서 이러한 돌연변이의 존재를 제거 또는 교정함으로써, 프리온 질환의 위험을 감소 또는 제거할 수 있다.

[1765] 다른 실시양태에서, 다중-플랩 프라임 편집은 PRNP 내에 1종 이상의 프리온 질환에 대한 보호 효과와 연관된 보호 돌연변이를 설치하기 위해 사용될 수 있다. 예를 들어, 다중-플랩 프라임 편집은 PRNP (NP_000302.1의 PRNP 대비) (서열식별번호: 291) 내에 G127S, G127V, M129V, D167G, D167N, N171S, E219K, 또는 P238S 보호 돌연변이를 설치하기 위해 사용될 수 있다. 또 다른 실시양태에서, 보호 돌연변이는 PRNP (NP_000302.1의 PRNP 대비) (서열식별번호: 291) 내의 G127, G127, M129, D167, D167, N171, E219, 또는 P238에 설치된 임의의 대안적 아미노산일 수 있다.

[1766] 특정한 실시양태에서, 다중-플랩 프라임 편집은 도 27에 예시되고 실시예 5에 논의된 바와 같이, PRNP 내에 G127V 보호 돌연변이를 설치하기 위해 사용될 수 있다.

[1767] 또 다른 실시양태에서, 프라임 편집은 PRNP 내에 E219K 보호 돌연변이를 설치하기 위해 사용될 수 있다.

[1768] PRNP 단백질 및 보호 돌연변이 부위는 포유동물에서 보존되므로, 인간 질환을 치료하는 것에 더하여, 이는 또한

프리온 질환에 면역성인 소 및 양을 생성하는 데 사용될 수 있거나, 또는 심지어 프리온 질환을 앓고 있는 동물의 야생 집단을 치유하는 것을 돕는데 사용될 수 있다. 프라임 편집은 인간 세포에서 자연 발생 보호 대립유전자의 ~25% 설치를 달성하는데 이미 사용되어 왔고, 이전 마우스 실험은 이러한 설치 수준이 대부분의 프리온 질환으로부터 면역을 유발하기에 충분하다는 것을 나타낸다. 이러한 방법은 대부분의 세포 유형에서 이러한 대립유전자를 이러한 높은 효율로 설치하기 위한 최초의, 잠재적으로 유일한 현재의 방식이다. 또 다른 가능한 치료 전략은 프라임 편집 또는 다중-플랩 프라임 편집을 사용하여 유전자에 초기 정지 코돈을 설치함으로써 PRNP의 발현을 감소 또는 제거하는 것이다.

[1769] PEgRNA 설계를 위해 본원에 기재된 원리를 사용하여, 목적하는 보호 돌연변이를 설치하기 위해, 또는 PRNP로부터 프리온 질환-연관 돌연변이를 제거하기 위해 적절한 PEgRNA를 설계할 수 있다. 예를 들어, 하기 목록의 PEgRNA를 사용하여 인간 PRNP 내에 G127V 보호 대립유전자 및 E219K 보호 대립유전자, 뿐만 아니라 다양한 동물의 PRNP 내에 G127V 보호 대립유전자를 설치할 수 있다.

[1770] [10] 제약 조성물

[1771] 본 개시내용의 다른 측면은 본원에 기재된 다중-플랩 프라임 편집 시스템의 다양한 성분 중 임의의 것 (예를 들어, napDNAbp, 리버스 트랜스크립타제, 융합 단백질 (예를 들어, napDNAbp 및 리버스 트랜스크립타제를 포함함), 연장된 가이드 RNA, 및 융합 단백질 및 연장된 가이드 RNA를 포함하는 복합체, 뿐만 아니라 다중-플랩 프라임 편집 과정을 편집된 생성물 형성쪽으로 구동하는 것을 돕기 위한 보조 요소, 예컨대 제2 가닥 닉킹 성분 및 5' 내인성 DNA 플랩 제거 엔도뉴클레아제를 포함하나 이에 제한되지는 않음)을 포함하는 제약 조성물에 관한 것이다.

[1772] 본원에 사용된 용어 "제약 조성물"은 제약 용도를 위해 제제화된 조성물을 지칭한다. 일부 실시양태에서, 제약 조성물은 제약상 허용되는 담체를 추가로 포함한다. 일부 실시양태에서, 제약 조성물은 추가의 작용제 (예를 들어, 특이적 전달, 반감기 증가, 또는 다른 치료 화합물을 위함)를 포함한다.

[1773] 본원에 사용된 용어 "제약상 허용되는 담체"는 신체의 한 부위 (예를 들어, 전달 부위)로부터 또 다른 부위 (예를 들어, 신체의 기관, 조직 또는 부분)로 화합물을 운반 또는 수송하는데 수반되는 제약상 허용되는 물질, 조성물 또는 비히클, 예컨대 액체 또는 고체 충전제, 희석제, 부형제, 제조 보조제 (예를 들어, 윤활제, 활석, 스테아르산마그네슘, 스테아르산칼슘 또는 스테아르산아연, 또는 스테아르산), 또는 용매 캡슐화 물질을 의미한다. 제약상 허용되는 담체는 제제의 다른 성분과 상용성이고 대상체의 조직에 유해하지 않다는 의미에서 "허용되는" 것이다 (예를 들어, 생리학상 상용성, 멸균, 생리학적 pH 등). 제약상 허용되는 담체로서의 역할을 할 수 있는 물질의 일부 예는 다음을 포함한다: (1) 당, 예컨대 락토스, 글루코스 및 수크로스; (2) 전분, 예컨대 옥수수 전분 및 감자 전분; (3) 셀룰로스 및 그의 유도체, 예컨대 소듐 카르복시메틸 셀룰로스, 메틸셀룰로스, 에틸 셀룰로스, 미세결정질 셀룰로스 및 셀룰로스 아세테이트; (4) 분말화 트라가칸트; (5) 맥아; (6) 젤라틴; (7) 윤활제, 예컨대 스테아르산마그네슘, 소듐 라우릴 술페이트 및 활석; (8) 부형제, 예컨대 코코아 버터 및 좌제 왁스; (9) 오일, 예컨대 땅콩 오일, 목화씨 오일, 홍화 오일, 참깨 오일, 올리브 오일, 옥수수 오일 및 대두 오일; (10) 글리콜, 예컨대 프로필렌 글리콜; (11) 폴리올, 예컨대 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜 (PEG); (12) 에스테르, 예컨대 에틸 올레에이트 및 에틸 라우레이트; (13) 한천; (14) 완충제, 예컨대 수산화마그네슘 및 수산화알루미늄; (15) 알긴산; (16) 발열원-무함유 물; (17) 등장성 염수; (18) 링거액; (19) 에틸 알콜; (20) pH 완충 용액; (21) 폴리에스테르, 폴리카르보네이트 및/또는 폴리무수물; (22) 벌킹제, 예컨대 폴리펩티드 및 아미노산 (23) 혈청 성분, 예컨대 혈청 알부민, HDL 및 LDL; (22) C2-C12 알콜, 예컨대 에탄올; 및 (23) 제약 제제에 사용되는 다른 비-독성 상용성 물질. 습윤제, 착색제, 이형제, 코팅제, 감미제, 향미제, 퍼퓸제, 보존제 및 항산화제가 또한 제제에 존재할 수 있다. "부형제", "담체", "제약상 허용되는 담체" 등과 같은 용어는 본원에서 상호교환가능하게 사용된다.

[1774] 일부 실시양태에서, 제약 조성물은 대상체에게 전달하기 위해, 예를 들어 유전자 편집을 위해 제제화된다. 본원에 기재된 제약 조성물의 적합한 투여 경로는 비제한적으로 국소, 피하, 경피, 피내, 병변내, 관절내, 복강내, 방광내, 경점막, 치은, 치아내, 와우내, 경고실, 기관내, 경막외, 척수강내, 근육내, 정맥내, 혈관내, 골내, 안구주위, 종양내, 뇌내, 및 뇌실내 투여를 포함한다.

[1775] 일부 실시양태에서, 본원에 기재된 제약 조성물은 이환 부위 (예를 들어, 종양 부위)에 국부로 투여된다. 일부 실시양태에서, 본원에 기재된 제약 조성물은 주사에 의해, 카테터에 의해, 좌제에 의해, 또는 이식물에 의해 대상체에게 투여되며, 이식물은 막, 예컨대 시알라스틱 막, 또는 섬유를 포함한 다공성, 비-다공성, 또는 젤라틴성 물질이다.

[1776] 다른 실시양태에서, 본원에 기재된 제약 조성물은 제어 방출 시스템으로 전달된다. 한 실시양태에서, 펌프가 사용될 수 있다 (예를 들어, 문헌 [Langer, 1990, Science 249:1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al., 1989, N. Engl. J. Med. 321:574] 참조). 또 다른 실시양태에서, 중합체 물질이 사용될 수 있다. (예를 들어, 문헌 [Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61] 참조. 또한 문헌 [Levy et al., 1985, Science 228:190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71:105] 참조). 다른 제어 방출 시스템은, 예를 들어 상기 문헌 [Langer]에 논의되어 있다.

[1777] 일부 실시양태에서, 제약 조성물은 대상체, 예를 들어 인간에게 정맥내 또는 피하 투여하기에 적합화된 조성물로서 상용 절차에 따라 제제화된다. 일부 실시양태에서, 주사에 의한 투여를 위한 제약 조성물은 멸균 등장성 수성 완충제 중의 용액이다. 필요한 경우에, 제약은 또한 가용화제 및 주사 부위에서의 통증을 완화시키기 위한 국부 마취제, 예컨대 리그노카인을 포함할 수 있다. 일반적으로, 성분은 단위 투여 형태로, 예를 들어 활성제의 양을 나타내는 기밀 용기, 예컨대 앰플 또는 사쉐 내의 건조 동결건조 분말 또는 무수 농축물로서 개별적으로 또는 함께 혼합되어 공급된다. 제약이 주입에 의해 투여되는 경우에, 이는 멸균 제약 등급 물 또는 염수를 함유하는 주입 병으로 분배될 수 있다. 제약 조성물이 주사에 의해 투여되는 경우에, 성분이 투여 전에 혼합될 수 있도록 주사용 멸균수 또는 염수의 앰플이 제공될 수 있다.

[1778] 전신 투여를 위한 제약 조성물은 액체, 예를 들어 멸균 염수, 락테이트화 링거 또는 행크 용액일 수 있다. 또한, 제약 조성물은 고체 형태일 수 있고, 사용 직전에 재용해 또는 현탁될 수 있다. 동결건조 형태가 또한 고려된다.

[1779] 제약 조성물은 지질 입자 또는 소포, 예컨대 리포솜 또는 미세결정 내에 함유될 수 있으며, 이는 또한 비경구 투여에 적합하다. 입자는 조성물이 그 안에 함유되는 한, 임의의 적합한 구조, 예컨대 단층 또는 다층일 수 있다. 화합물은 융합생성 지질 디올레오일포스파티딜에탄올아민 (DOPE), 낮은 수준 (5-10 mol%)의 양이온성 지질을 함유하는 "안정화된 플라스미드-지질 입자" (SPLP)에 포획될 수 있고, 폴리에틸렌글리콜 (PEG) 코팅에 의해 안정화될 수 있다 (Zhang Y. P. et al., Gene Ther. 1999, 6:1438-47). 양으로 하전된 지질, 예컨대 N-[1-(2,3-디올레오일옥시)프로필]-N,N,N-트리메틸-암모늄메틸술페이트, 또는 "DOTAP"가 이러한 입자 및 소포에 특히 바람직하다. 이러한 지질 입자의 제조는 널리 공지되어 있다. 예를 들어, 미국 특허 번호 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; 및 4,921,757을 참조하고; 이들 각각은 본원에 참조로 포함된다.

[1780] 본원에 기재된 제약 조성물은, 예를 들어 단위 용량으로서 투여 또는 패키징될 수 있다. 용어 "단위 용량"은 본 개시내용의 제약 조성물과 관련하여 사용되는 경우에 대상체에 대한 단일 투여량으로서 적합한 물리적 이산 단위를 지칭하며, 각각의 단위는 요구되는 희석제; 즉, 담체 또는 비히클과 함께 목적하는 치료 효과를 생성하도록 계산된 미리 결정된 양의 활성 물질을 함유한다.

[1781] 추가로, 제약 조성물은 (a) 동결건조 형태의 본 발명의 화합물을 함유하는 용기 및 (b) 주사용 제약상 허용되는 희석제 (예를 들어, 멸균수)를 함유하는 제2 용기를 포함하는 제약 키트로서 제공될 수 있다. 제약상 허용되는 희석제는 본 발명의 동결건조 화합물의 재구성 또는 희석에 사용될 수 있다. 인간 투여를 위한 제조, 사용 또는 판매의 정부 기관에 의한 승인을 반영하는, 제약 또는 생물학적 제품의 제조, 사용 또는 판매를 규제하는 정부 기관에 의해 규정된 형태의 통지서가 이러한 용기(들)와 임의로 회합될 수 있다.

[1782] 또 다른 측면에서, 상기 기재된 질환의 치료에 유용한 물질을 함유하는 제조 물품이 포함된다. 일부 실시양태에서, 제조 물품은 용기 및 라벨을 포함한다. 적합한 용기는, 예를 들어 병, 바이알, 시린지 및 시험 튜브를 포함한다. 용기는 다양한 물질, 예컨대 유리 또는 플라스틱으로 형성될 수 있다. 일부 실시양태에서, 용기는 본원에 기재된 질환을 치료하는데 효과적인 조성물을 보유하고, 멸균 접근 포트를 가질 수 있다. 예를 들어, 용기는 피하 주사 바늘로 뚫을 수 있는 마개를 갖는 정맥내 용액 백 또는 바이알일 수 있다. 조성물 중 활성제는 본 발명의 화합물이다. 일부 실시양태에서, 용기 상의 또는 용기와 회합된 라벨은 조성물이 선택된 질환을 치료하는 데 사용된다는 것을 표시한다. 제조 물품은 제약상 허용되는 완충제, 예컨대 포스페이트-완충 염수, 링거액 또는 덱스트로스 용액을 포함하는 제2 용기를 추가로 포함할 수 있다. 이는 다른 완충제, 희석제, 필터, 바늘, 시린지, 및 사용에 대한 지침서를 갖는 패키지 삽입물을 포함한, 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[1783] [11] 키트, 세포, 벡터, 및 전달

[1784] 키트

[1785] 본 개시내용의 조성물은 키트로 조립될 수 있다. 일부 실시양태에서, 키트는 본원에 기재된 다중-플랩 프라임 편집제의 발현을 위한 핵산 벡터를 포함한다. 다른 실시양태에서, 키트는 Cas9 단백질 또는 프라임 편집제를 목적하는 표적 서열에 표적화하기 위해, 적절한 가이드 뉴클레오티드 서열 (예를 들어, PEgRNA 및 제2-부위 gRNA) 또는 이러한 가이드 뉴클레오티드 서열의 발현을 위한 핵산 벡터를 추가로 포함한다.

[1786] 본원에 기재된 키트는 본원에 기재된 방법을 수행하기 위한 성분을 수용하는 1개 이상의 용기 및 임의로 사용에 대한 지침서를 포함할 수 있다. 본원에 기재된 키트 중 임의의 것은 검정 방법을 수행하는데 필요한 성분을 추가로 포함할 수 있다. 키트의 각각의 성분은, 적용가능한 경우에, 액체 형태 (예를 들어, 용액) 또는 고체 형태 (예를 들어, 건조 분말)로 제공될 수 있다. 특정 경우에, 성분 중 일부는, 예를 들어 키트와 함께 제공될 수 있거나 또는 제공되지 않을 수 있는 적합한 용매 또는 다른 종 (예를 들어, 물)의 첨가에 의해 (예를 들어, 활성 형태로) 재구성가능하거나 또는 달리 가공가능할 수 있다.

[1787] 일부 실시양태에서, 키트는 제공된 성분의 사용에 대한 지침서 및/또는 홍보물을 임의로 포함할 수 있다. 본원에 사용된 "지침서"는 지침서 및/또는 홍보물의 성분을 규정할 수 있고, 전형적으로 개시내용의 패키징 상의 또는 그와 회합된 서면 지침서를 수반할 수 있다. 지침서는 또한 사용자가 지침서를 키트와 연관된 것으로 명확하게 인식하게 할 임의의 방식, 예를 들어 시청각 (예를 들어, 비디오테이프, DVD 등), 인터넷 및/또는 웹-기반 통신 등으로 제공되는 임의의 구두 또는 전자 지침서를 포함할 수 있다. 서면 지침서는 제약 또는 생물학적 제품의 제조, 사용, 또는 판매를 규제하는 정부 기관에 의해 규정된 형태일 수 있으며, 이는 또한 동물 투여를 위한 제조, 사용 또는 판매에 대한 정부 기관의 승인을 반영할 수 있다. 본원에 사용된 "홍보된"은 개시내용와 연관된, 교육, 병원 및 다른 임상 지침, 과학적 탐구, 약물 발견 또는 개발, 학술 연구, 제약 판매를 포함한 제약 산업 활동, 및 임의의 형태의 서면, 구두 및 전자 커뮤니케이션을 포함한 임의의 광고 또는 다른 홍보 활동의 방법을 포함한 영업을 수행하는 모든 방법을 포함한다. 추가적으로, 키트는 본원에 기재된 바와 같은 구체적 적용에 따라 다른 성분을 포함할 수 있다.

[1788] 키트는 1개 이상의 용기에 본원에 기재된 성분 중 임의의 1종 이상을 함유할 수 있다. 성분은 멸균 제조되고, 시린지 내에 패키징되고, 냉장 운송될 수 있다. 대안적으로, 이는 저장을 위해 바이알 또는 다른 용기에 수용될 수 있다. 제2 용기는 멸균 제조된 다른 성분을 가질 수 있다. 대안적으로, 키트는 바이알, 튜브 또는 다른 용기 내에서 예비혼합되고 운송된 활성제를 포함할 수 있다.

[1789] 키트는 다양한 형태, 예컨대 블리스터 파우치, 수축 랩핑된 파우치, 진공 밀봉가능한 파우치, 밀봉가능한 열성형된 트레이, 또는 파우치 내에 느슨하게 패킹된 부속품을 갖는 유사한 파우치 또는 트레이 형태, 1개 이상의 튜브, 용기, 박스 또는 백을 가질 수 있다. 키트는 부속품이 첨가된 후에 멸균될 수 있고, 이에 의해 용기 내의 개별 부속품을 달리 꺼내는 것이 가능하다. 키트는 임의의 적절한 멸균 기술, 예컨대 방사선 멸균, 가열 멸균, 또는 관련 기술분야에 공지된 다른 멸균 방법을 사용하여 멸균될 수 있다. 키트는 또한 특정한 적용에 따라 다른 성분, 예를 들어, 용기, 세포 배지, 염, 완충제, 시약, 시린지, 바늘, 소독제를 적용 또는 제거하기 위한 직물, 예컨대 거즈, 일회용 장갑, 투여 전의 작용제를 위한 지지체 등을 포함할 수 있다. 본 개시내용의 일부 측면은 본원에 기재된 다중-플랩 프라임 편집 시스템 (예를 들어, 이중 프라임 편집 및 사중 프라임 편집 시스템)의 다양한 성분 (예를 들어, napDNAbp, 리버스 트랜스크립타제, 폴리머라제, 융합 단백질 (예를 들어, napDNAbp 및 리버스 트랜스크립타제 (또는 보다 넓게는, 폴리머라제)를 포함함), 연장된 가이드 RNA, 및 융합 단백질 및 연장된 가이드 RNA를 포함하는 복합체, 뿐만 아니라 다중-플랩 프라임 편집 과정을 편집된 생성물 형성쪽으로 구동하는 것을 돕기 위한 보조 요소, 예컨대 제2 가닥 닉킹 성분 (예를 들어, 제2 가닥 닉킹 gRNA) 및 5' 내인성 DNA 플랩 제거 엔도뉴클레아제를 포함하나 이에 제한되지는 않음)을 코딩하는 뉴클레오티드 서열을 포함하는 핵산 구축물을 포함하는 키트를 제공한다. 일부 실시양태에서, 뉴클레오티드 서열(들)은 다중-플랩 프라임 편집 시스템 성분의 발현을 구동하는 이종 프로모터 (또는 단일 초과의 프로모터)를 포함한다.

[1790] 본 개시내용의 다른 측면은 본원에 기재된 다중-플랩 프라임 편집 시스템의 다양한 성분을 코딩하는 1개 이상의 핵산 구축물을 포함하는, 예를 들어 표적 DNA 서열을 변형시킬 수 있는 다중-플랩 프라임 편집 시스템의 성분을 코딩하는 뉴클레오티드 서열을 포함하는 키트를 제공한다. 일부 실시양태에서, 뉴클레오티드 서열은 다중-플랩 프라임 편집 시스템 성분의 발현을 구동하는 이종 프로모터를 포함한다.

[1791] 본 개시내용의 일부 측면은 (a) 리버스 트랜스크립타제에 융합된 napDNAbp (예를 들어, Cas9 도메인)를 코딩하

는 뉴클레오티드 서열 및 (b) (a)의 서열의 발현을 구동하는 이종 프로모터를 포함하는 핵산 구축물을 포함하는 키트를 제공한다.

[1792] 세포

[1793] 본원에 기재된 조성물 중 임의의 것을 함유할 수 있는 세포는 원핵 세포 및 진핵 세포를 포함한다. 본원에 기재된 방법은 Cas9 단백질 또는 다중-플랩 프라임 편집제를 진핵 세포 (예를 들어, 포유동물 세포, 예컨대 인간 세포) 내로 전달하는 데 사용된다. 일부 실시양태에서, 세포는 시험관내의 것 (예를 들어, 배양된 세포)이다. 일부 실시양태에서, 세포는 생체내 (예를 들어, 대상체, 예컨대 인간 대상체 내)의 것이다. 일부 실시양태에서, 세포는 생체외의 것 (예를 들어, 대상체로부터 단리되고, 동일한 또는 상이한 대상체에게 다시 투여될 수 있는 것)이다.

[1794] 본 개시내용의 포유동물 세포는 인간 세포, 영장류 세포 (예를 들어, 베로 세포), 래트 세포 (예를 들어, GH3 세포, OC23 세포) 또는 마우스 세포 (예를 들어, MC3T3 세포)를 포함한다. 비제한적으로, 인간 배아 신장 (HEK) 세포, HeLa 세포, 국립 암 연구소의 60 암 세포주 (NCI60)로부터의 암 세포, DU145 (전립선암) 세포, Lncap (전립선암) 세포, MCF-7 (유방암) 세포, MDA-MB-438 (유방암) 세포, PC3 (전립선암) 세포, T47D (유방암) 세포, THP-1 (급성 골수성 백혈병) 세포, U87 (교모세포종) 세포, SHSY5Y 인간 신경모세포종 세포 (골수종으로부터 클로닝됨) 및 Saos-2 (골암) 세포를 포함한 다양한 인간 세포주가 존재한다. 일부 실시양태에서, rAAV 벡터는 인간 배아 신장 (HEK) 세포 (예를 들어, HEK 293 또는 HEK 293T 세포) 내로 전달된다. 일부 실시양태에서, rAAV 벡터는 줄기 세포 (예를 들어, 인간 줄기 세포), 예컨대, 예를 들어, 만능 줄기 세포 (예를 들어, 인간 유도된 만능 줄기 세포 (hiPSC)를 포함한 인간 만능 줄기 세포) 내로 전달된다. 줄기 세포는 배양 시 무한 기간 동안 분열하고 특화 세포를 생성하는 능력을 갖는 세포를 지칭한다. 만능 줄기 세포는 유기체의 모든 조직으로 분화할 수 있지만, 단독으로는 완전한 유기체 발생을 지속할 수 없는 줄기 세포의 유형을 지칭한다. 인간 유도된 만능 줄기 세포는 배아 줄기 세포의 정의 특성을 유지하는데 중요한 유전자 및 인자를 발현하도록 강제됨으로써 배아 줄기 세포-유사 상태로 재프로그램화된 체세포 (예를 들어, 성숙 또는 성체)를 지칭한다 (예를 들어, 본원에 참조로 포함된 문헌 [Takahashi and Yamanaka, Cell 126 (4): 663-76, 2006] 참조). 인간 유도된 만능 줄기 세포 세포는 줄기 세포 마커를 발현하고, 모든 3가지 배엽층 (외배엽, 내배엽, 중배엽)의 세포 특징을 생성할 수 있다.

[1795] 본 개시내용에 따라 사용될 수 있는 세포주의 추가의 비제한적 예는 293-T, 293-T, 3T3, 4T1, 721, 9L, A-549, A172, A20, A253, A2780, A2780ADR, A2780cis, A431, ALC, B16, B35, BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C2C12, C3H-10T1/2, C6, C6/36, Cal-27, CGR8, CHO, CML T1, CMT, COR-L23, COR-L23/5010, COR-L23/CPR, COR-L23/R23, COS-7, COV-434, CT26, D17, DH82, DU145, DuCaP, E14Tg2a, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, Hepa1c1c7, 하이 파이브 세포, HL-60, HMEC, HT-29, HUVEC, J558L 세포, Jurkat, JY 세포, K562 세포, KCL22, KG1, Ku812, KYO1, LNCap, Ma-Mel 1, 2, 3....48, MC-38, MCF-10A, MCF-7, MDA-MB-231, MDA-MB-435, MDA-MB-468, MDCK II, MG63, MONO-MAC 6, MOR/0.2R, MRC5, MTD-1A, MyEnd, NALM-1, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NW-145, OPCN/OPCT Peer, PNT-1A/PNT 2, PTK2, Raji, RBL 세포, RenCa, RIN-5F, RMA/RMAS, S2, Saos-2 세포, Sf21, Sf9, SiHa, SKBR3, SKOV-3, T-47D, T2, T84, THP1, U373, U87, U937, VCaP, WM39, WT-49, X63, YAC-1 및 YAR 세포를 포함한다.

[1796] 본 개시내용의 일부 측면은 본원에 개시된 구축물 중 임의의 것을 포함하는 세포를 제공한다. 일부 실시양태에서, 숙주 세포는 본원에 기재된 1개 이상의 벡터로 일시적으로 또는 비-일시적으로 형질감염된다. 일부 실시양태에서, 세포는 대상체에서 자연 발생하는 바와 같이 형질감염된다. 일부 실시양태에서, 형질감염된 세포는 대상체로부터 채취된다. 일부 실시양태에서, 세포는 대상체로부터 채취한 세포, 예컨대 세포주로부터 유래된다. 조직 배양을 위한 매우 다양한 세포주가 관련 기술분야에 공지되어 있다. 세포주의 예는 C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 원숭이 신장 상피, BALB/3T3 마우스 배아 섬유모세포, 3T3 스위스, 3T3-L1, 132-d5 인간 태아 섬유모세포; 10.1 마우스 섬유모세포, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A 172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 세포, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-

434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY 세포, K562 세포, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK 11, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT 세포주, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 세포, Sf-9, SkBr3, T2, T-47D, T84, THP1 세포주, U373, U87, U937, VCaP, 베로 세포, WM39, WT-49, X63, YAC-1, YAR, 및 그의 트랜스제닉 변종을 포함하나 이에 제한되지는 않는다.

[1797] 세포주는 관련 기술분야의 통상의 기술자에게 공지된 다양한 공급원으로부터 입수가능하다 (예를 들어, 아메리칸 타입 컬쳐 콜렉션 (ATCC) (버지니아주 마나사스) 참조). 일부 실시양태에서, 본원에 기재된 1개 이상의 벡터로 형질감염된 세포는 1개 이상의 벡터-유래 서열을 포함하는 새로운 세포주를 확립하는 데 사용된다. 일부 실시양태에서, 본원에 기재된 바와 같은 CRISPR 시스템의 성분으로 일시적으로 형질감염되고 (예컨대 1개 이상의 벡터의 일시적 형질감염, 또는 RNA로의 형질감염에 의함), CRISPR 복합체의 활성을 통해 변형된 세포는 변형을 함유하지만 임의의 다른 외인성 서열은 결여된 세포를 포함하는 새로운 세포주를 확립하는 데 사용된다. 일부 실시양태에서, 본원에 기재된 1개 이상의 벡터로 일시적으로 또는 비-일시적으로 형질감염된 세포, 또는 이러한 세포로부터 유래된 세포주는 1개 이상의 시험 화합물을 평가하는 데 사용된다.

[1798] 벡터

[1799] 본 개시내용의 일부 측면은 본원에 기재된 다중-플랩 프라임 편집제 또는 그의 성분, 예를 들어 분할 Cas9 단백질 또는 분할 핵염기 다중-플랩 프라임 편집제를 세포 내로 전달하기 위해 재조합 바이러스 벡터 (예를 들어, 아데노-연관 바이러스 벡터, 아데노바이러스 벡터, 또는 단순 포진 바이러스 벡터)를 사용하는 것에 관한 것이다. 분할-PE 접근법의 경우, PE 융합 단백질의 N-말단 부분 및 PE 융합체의 C-말단 부분은 별개의 재조합 바이러스 벡터 (예를 들어, 아데노-연관 바이러스 벡터, 아데노바이러스 벡터, 또는 단순 포진 바이러스 벡터)에 의해 동일한 세포 내로 전달되는데, 이는 전장 Cas9 단백질 또는 다중-플랩 프라임 편집제가 다양한 바이러스 벡터, 예를 들어 rAAV의 패키징 한계 (~4.9 kb)를 초과하기 때문이다.

[1800] 따라서, 한 실시양태에서, 개시내용은 분할 다중-플랩 프라임 편집제 융합 단백질 또는 그의 분할 성분을 전달할 수 있는 벡터를 고려한다. 일부 실시양태에서, 분할 Cas9 단백질 또는 분할 프라임 편집제를 세포 (예를 들어, 포유동물 세포, 인간 세포) 내로 전달하기 위한 조성물이 제공된다. 일부 실시양태에서, 본 개시내용의 조성물은 (i) C-말단에서 인테인-N에 융합된 Cas9 단백질 또는 프라임 편집제의 N-말단 부분을 코딩하는 제1 뉴클레오티드 서열을 포함하는 제1 재조합 아데노-연관 바이러스 (rAAV) 입자; 및 (ii) Cas9 단백질 또는 프라임 편집제의 C-말단 부분의 N-말단에 융합된 인테인-C를 코딩하는 제2 뉴클레오티드 서열을 포함하는 제2 재조합 아데노-연관 바이러스 (rAAV) 입자를 포함한다. 본 개시내용의 rAAV 입자는 바이러스 캡시드 단백질 내에 캡시드화된 rAAV 벡터 (즉, rAAV의 재조합 게놈)를 포함한다.

[1801] 일부 실시양태에서, rAAV 벡터는 (1) 본원에 기재된 바와 같은 임의의 형태의 분할 Cas9 단백질 또는 분할 다중-플랩 프라임 편집제의 N-말단 부분 또는 C-말단 부분을 코딩하는 제1 또는 제2 뉴클레오티드 서열을 포함하는 이종 핵산 영역, (2) 이종 핵산 영역의 발현을 용이하게 하는 서열 (예를 들어, 프로모터)을 포함하는 1개 이상의 뉴클레오티드 서열, 및 (3) 세포의 게놈 내로의 이종 핵산 영역의 통합을 용이하게 하는 서열을 포함하는 1개 이상의 핵산 영역을 (임의로 발현을 용이하게 하는 서열을 포함하는 1개 이상의 핵산 영역과 함께) 포함한다. 일부 실시양태에서, 통합을 용이하게 하는 바이러스 서열은 역전된 말단 반복부 (ITR) 서열을 포함한다. 일부 실시양태에서, 분할 Cas9 단백질 또는 분할 다중-플랩 프라임 편집제의 N-말단 부분 또는 C-말단 부분을 코딩하는 제1 또는 제2 뉴클레오티드 서열은 ITR 서열이 각 측면에 플랭킹된다. 일부 실시양태에서, 핵산 벡터는 ITR이 플랭킹된 영역 내에 또는 영역 외부에 함유된, 본원에 기재된 바와 같은 AAV Rep 단백질을 코딩하는 영역을 추가로 포함한다. ITR 서열은 임의의 AAV 혈청형 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10)으로부터 유래될 수 있거나 또는 1종 초과의 혈청형으로부터 유래될 수 있다. 일부 실시양태에서, ITR 서열은 AAV2 또는 AAV6으로부터 유래된다.

[1802] 따라서, 일부 실시양태에서, 본원에 개시된 rAAV 입자는 적어도 1개의 rAAV2 입자, rAAV6 입자, rAAV8 입자, rPHP.B 입자, rPHP.eB 입자, 또는 rAAV9 입자, 또는 그의 변이체를 포함한다. 특정한 실시양태에서, 개시된 rAAV 입자는 rPHP.B 입자, rPHP.eB 입자, rAAV9 입자이다.

[1803] ITR 서열 및 ITR 서열을 함유하는 플라스미드는 관련 기술분야에 공지되어 있고, 상업적으로 입수가능하다 (예를 들어, 펜실베니아주 필라델피아 소재의 벡터 바이오랩스(Vector Biolabs); 캘리포니아주 샌디에고 소재의 셀

바이오랩스(Cellbiolabs); 캘리포니아주 산타 클라라 소재의 애질런트 테크놀로지스(Agilent Technologies); 및 매사추세츠주 캠브리지 소재의 애드진(Addgene)으로부터 입수가능한 제품 및 서비스; 및 문헌 [Gene delivery to skeletal muscle results in sustained expression and systemic delivery of a therapeutic protein. Kessler PD, Podsakoff GM, Chen X, McQuiston SA, Colosi PC, Matelis LA, Kurtzman GJ, Byrne BJ. Proc Natl Acad Sci USA. 1996 Nov 26;93(24):14082-7; and Curtis A. Machida. Methods in Molecular Medicine™. Viral Vectors for Gene Therapy Methods and Protocols. 10.1385/1-59259-304-6:201 ⓒ Humana Press Inc. 2003. Chapter 10. Targeted Integration by Adeno-Associated Virus. Matthew D. Weitzman, Samuel M. Young Jr., Toni Cathomen and Richard Jude Samulski]; 미국 특허 번호 5,139,941 및 5,962,313 참조, 이들은 모두 본원에 참조로 포함됨).

[1804] 일부 실시양태에서, 본 개시내용의 rAAV 벡터는 이종 핵산 영역의 발현을 제어하기 위한 1개 이상의 조절 요소 (예를 들어, 프로모터, 전사 종결인자, 및/또는 다른 조절 요소)를 포함한다. 일부 실시양태에서, 제1 및/또는 제2 뉴클레오티드 서열은 1개 이상 (예를 들어, 1, 2, 3, 4, 5개 또는 그 초과)의 전사 종결인자에 작동가능하게 연결된다. 본 개시내용에 따라 사용될 수 있는 전사 종결인자의 비제한적 예는 소 성장 호르몬 유전자 (bGH), 인간 성장 호르몬 유전자 (hGH), SV40, CW3, ϕ, 또는 그의 조합의 전사 종결인자를 포함한다. 여러 전사 종결인자의 효율은 분할 Cas9 단백질 또는 분할 다중-플랩 프라임 편집제의 발현 수준에서의 그의 각각의 효과를 결정하기 위해 시험되었다. 일부 실시양태에서, 본 개시내용에 사용된 전사 종결인자는 bGH 전사 종결인자이다. 일부 실시양태에서, rAAV 벡터는 우드척 간염 바이러스 전사후 조절 요소 (WPRE)를 추가로 포함한다. 특정 실시양태에서, WPRE는 말단절단된 WPRE 서열, 예컨대 "W3"이다. 일부 실시양태에서, WPRE는 전사 종결인자의 5'에 삽입된다. 이러한 서열은 전사되는 경우에, 특히 바이러스 벡터로부터의 발현을 증진시키는 3차 구조를 생성한다.

[1805] 일부 실시양태에서, 본원에 사용된 벡터는 PE 융합 단백질, 또는 그의 성분 중 임의의 것 (예를 들어, napDNAbp, 링커, 또는 폴리머라제)을 코딩할 수 있다. 또한, 본원에 사용된 벡터는 PEgRNA, 및/또는 제2 가닥 닉킹을 위한 보조 gRNA를 코딩할 수 있다. 벡터는 세포에서 1개 이상의 코딩 서열의 발현을 구동할 수 있다. 일부 실시양태에서, 세포는 원핵 세포, 예컨대, 예를 들어, 박테리아 세포일 수 있다. 일부 실시양태에서, 세포는 진핵 세포, 예컨대, 예를 들어 효모, 식물, 곤충, 또는 포유동물 세포일 수 있다. 일부 실시양태에서, 진핵 세포는 포유동물 세포일 수 있다. 일부 실시양태에서, 진핵 세포는 설치류 세포일 수 있다. 일부 실시양태에서, 진핵 세포는 인간 세포일 수 있다. 상이한 유형의 세포에서 발현을 구동하기에 적합한 프로모터는 관련 기술분야에 공지되어 있다. 일부 실시양태에서, 프로모터는 야생형일 수 있다. 다른 실시양태에서, 프로모터는 보다 효율적이거나 효과적인 발현을 위해 변형될 수 있다. 또 다른 실시양태에서, 프로모터는 말단절단되어도 그의 기능을 보유할 수 있다. 예를 들어, 프로모터는 벡터를 바이러스 내로 적절히 패키징하는데 적합한 정상 크기 또는 감소된 크기를 가질 수 있다.

[1806] 일부 실시양태에서, 프라임 편집제 벡터에 사용될 수 있는 프로모터는 구성적, 유도성, 또는 조직-특이적일 수 있다. 일부 실시양태에서, 프로모터는 구성적 프로모터일 수 있다. 비제한적인 예시적인 구성적 프로모터는 시토메갈로바이러스 극초기 프로모터 (CMV), 원숭이 바이러스 (SV40) 프로모터, 아데노바이러스 주요 후기 (MLP) 프로모터, 라우스 육종 바이러스(RSV) 프로모터, 마우스 유방 종양 바이러스 (MMTV) 프로모터, 포스포글리세레이트 키나제 (PGK) 프로모터, 신장 인자-알파 (EF1a) 프로모터, 유비퀴틴 프로모터, 액틴 프로모터, 튜불린 프로모터, 이뮤노글로불린 프로모터, 그의 기능적 단편, 또는 상기 중 임의의 것의 조합을 포함한다. 일부 실시양태에서, 프로모터는 CMV 프로모터일 수 있다. 일부 실시양태에서, 프로모터는 말단절단된 CMV 프로모터일 수 있다. 다른 실시양태에서, 프로모터는 EF1a 프로모터일 수 있다. 일부 실시양태에서, 프로모터는 유도성 프로모터일 수 있다. 비제한적인 예시적인 유도성 프로모터는 열 쇼크, 광, 화학물질, 펩티드, 금속, 스테로이드, 항생제, 또는 알콜에 의해 유도가능한 것을 포함한다. 일부 실시양태에서, 유도성 프로모터는 낮은 기저 (비-유도된) 발현 수준을 갖는 것, 예컨대, 예를 들어, 테트-온(Tet-On)® 프로모터 (클론테크)일 수 있다. 일부 실시양태에서, 프로모터는 조직-특이적 프로모터일 수 있다. 일부 실시양태에서, 조직-특이적 프로모터는 간 조직에서 독점적으로 또는 우세하게 발현된다. 비제한적인 예시적인 조직-특이적 프로모터는 B29 프로모터, CD14 프로모터, CD43 프로모터, CD45 프로모터, CD68 프로모터, 데스민 프로모터, 엘라스타제-1 프로모터, 엔도글린 프로모터, 피브로넥틴 프로모터, Flt-1 프로모터, GFAP 프로모터, GPIIb 프로모터, ICAM-2 프로모터, INF-β 프로모터, Mb 프로모터, Nphs1 프로모터, OG-2 프로모터, SP-B 프로모터, SYN1 프로모터, 및 WASP 프로모터를 포함한다.

[1807] 일부 실시양태에서, 프라임 편집제 벡터 (예를 들어, 프라임 편집제 융합 단백질 및/또는 PEgRNA, 및/또는 보조

제2 가닥 닉킹 gRNA를 코딩하는 임의의 벡터를 포함함)는 표적 세포에 전달된 후에만 발현을 시작하는 유도성 프로모터를 포함할 수 있다. 비제한적인 예시적인 유도성 프로모터는 열 쇼크, 광, 화학물질, 펩티드, 금속, 스테로이드, 항생제, 또는 알콜에 의해 유도가능한 것을 포함한다. 일부 실시양태에서, 유도성 프로모터는 낮은 기저 (비-유도된) 발현 수준을 갖는 것, 예컨대, 예를 들어, 테트-온® 프로모터 (클론테크)일 수 있다.

[1808] 추가의 실시양태에서, 프라임 편집제 벡터 (예를 들어, 프라임 편집제 융합 단백질 및/또는 PEgRNA, 및/또는 보조 제2 가닥 닉킹 gRNA를 코딩하는 임의의 벡터를 포함함)는 특정 조직 내로 전달된 후에만 발현을 시작하기 위해 조직-특이적 프로모터를 포함할 수 있다. 비제한적인 예시적인 조직-특이적 프로모터는 B29 프로모터, CD14 프로모터, CD43 프로모터, CD45 프로모터, CD68 프로모터, 데스민 프로모터, 엘라스타제-1 프로모터, 엔도글린 프로모터, 피브로넥틴 프로모터, Flt-1 프로모터, GFAP 프로모터, GPIIb 프로모터, ICAM-2 프로모터, INF-β 프로모터, Mb 프로모터, Nphs1 프로모터, OG-2 프로모터, SP-B 프로모터, SYN1 프로모터, 및 WASP 프로모터를 포함한다.

[1809] 일부 실시양태에서, PEgRNA (또는 다중-플랩 프라임 편집과 관련하여 사용되는 임의의 가이드 RNA)를 코딩하는 뉴클레오티드 서열은 적어도 1개의 전사 또는 번역 제어 서열에 작동가능하게 연결될 수 있다. 일부 실시양태에서, 가이드 RNA를 코딩하는 뉴클레오티드 서열은 적어도 1개의 프로모터에 작동가능하게 연결될 수 있다. 일부 실시양태에서, 프로모터는 RNA 폴리머라제 III (Pol III)에 의해 인식될 수 있다. Pol III 프로모터의 비제한적 예는 U6, HI 및 tRNA 프로모터를 포함한다. 일부 실시양태에서, 가이드 RNA를 코딩하는 뉴클레오티드 서열은 마우스 또는 인간 U6 프로모터에 작동가능하게 연결될 수 있다. 다른 실시양태에서, 가이드 RNA를 코딩하는 뉴클레오티드 서열은 마우스 또는 인간 HI 프로모터에 작동가능하게 연결될 수 있다. 일부 실시양태에서, 가이드 RNA를 코딩하는 뉴클레오티드 서열은 마우스 또는 인간 tRNA 프로모터에 작동가능하게 연결될 수 있다. 1개 초과의 가이드 RNA를 갖는 실시양태에서, 발현을 구동하는 데 사용되는 프로모터는 동일하거나 상이할 수 있다. 일부 실시양태에서, 가이드 RNA의 crRNA를 코딩하는 뉴클레오티드 및 가이드 RNA의 tracr RNA를 코딩하는 뉴클레오티드는 동일한 벡터 상에 제공될 수 있다. 일부 실시양태에서, crRNA를 코딩하는 뉴클레오티드 및 tracr RNA를 코딩하는 뉴클레오티드는 동일한 프로모터에 의해 구동될 수 있다. 일부 실시양태에서, crRNA 및 tracr RNA는 단일 전사체로 전사될 수 있다. 예를 들어, crRNA 및 tracr RNA는 단일 전사체로부터 프로세싱되어 이중-분자 가이드 RNA를 형성할 수 있다. 대안적으로, crRNA 및 tracr RNA는 단일-분자 가이드 RNA로 전사될 수 있다.

[1810] 일부 실시양태에서, 가이드 RNA를 코딩하는 뉴클레오티드 서열은 PE 융합 단백질을 코딩하는 뉴클레오티드 서열을 포함하는 동일한 벡터 상에 위치할 수 있다. 일부 실시양태에서, 가이드 RNA 및 PE 융합 단백질의 발현은 그의 상응하는 프로모터에 의해 구동될 수 있다. 일부 실시양태에서, 가이드 RNA의 발현은 PE 융합 단백질의 발현을 구동하는 동일한 프로모터에 의해 구동될 수 있다. 일부 실시양태에서, 가이드 RNA 및 PE 융합 단백질 전사체는 단일 전사체 내에 함유될 수 있다. 예를 들어, 가이드 RNA는 Cas9 단백질 전사체의 비번역 영역 (UTR) 내에 있을 수 있다. 일부 실시양태에서, 가이드 RNA는 PE 융합 단백질 전사체의 5' UTR 내에 있을 수 있다. 다른 실시양태에서, 가이드 RNA는 PE 융합 단백질 전사체의 3' UTR 내에 있을 수 있다. 일부 실시양태에서, PE 융합 단백질 전사체의 세포내 반감기는 가이드 RNA를 그의 3' UTR 내에 함유시켜 그의 3' UTR의 길이를 단축시킴으로써 감소될 수 있다. 추가의 실시양태에서, 가이드 RNA는 PE 융합 단백질 전사체의 인트론 내에 있을 수 있다. 일부 실시양태에서, 가이드 RNA가 전사체로부터 적절하게 스플라이싱되도록 가이드 RNA가 위치하는 인트론에 적합한 스플라이스 부위가 부가될 수 있다. 일부 실시양태에서, 동일한 벡터 상에서의 아주 근접한 Cas9 단백질 및 가이드 RNA의 발현은 CRISPR 복합체의 보다 효율적인 형성을 용이하게 할 수 있다.

[1811] 다중-플랩 프라임 편집제 벡터 시스템은 1개의 벡터, 또는 2개의 벡터, 또는 3개의 벡터, 또는 4개의 벡터, 또는 5개의 벡터, 또는 그 초과를 포함할 수 있다. 일부 실시양태에서, 벡터 시스템은 PE 융합 단백질 및 PEgRNA 둘 다를 코딩하는 1개의 단일 벡터를 포함할 수 있다. 다른 실시양태에서, 벡터 시스템은 2개의 벡터를 포함할 수 있고, 여기서 하나의 벡터는 PE 융합 단백질을 코딩하고, 다른 벡터는 PEgRNA를 코딩한다. 추가의 실시양태에서, 벡터 시스템은 3개의 벡터를 포함할 수 있고, 여기서 제3 벡터는 본원의 방법에 사용되는 제2 가닥 닉킹 gRNA를 코딩한다.

[1812] 일부 실시양태에서, rAAV 입자 (본원에서 고려되는 임의의 형태)를 포함하는 조성물은 제약상 허용되는 담체를 추가로 포함한다. 일부 실시양태에서, 조성물은 인간 또는 동물 대상체에게 투여하기 위한 적절한 제약 비히클 중에 제제화된다.

[1813] 제약상 허용되는 담체로서의 역할을 할 수 있는 물질의 일부 예는 다음을 포함한다: (1) 당, 예컨대 락토스, 글

루코스 및 수크로스; (2) 전분, 예컨대 옥수수 전분 및 감자 전분; (3) 셀룰로스 및 그의 유도체, 예컨대 소듐 카르복시메틸 셀룰로스, 메틸셀룰로스, 에틸 셀룰로스, 미세결정질 셀룰로스 및 셀룰로스 아세테이트; (4) 분말화 트라가칸트; (5) 맥아; (6) 젤라틴; (7) 윤활제, 예컨대 스테아르산마그네슘, 소듐 라우릴 술페이트 및 활석; (8) 부형제, 예컨대 코코아 버터 및 좌제 왁스; (9) 오일, 예컨대 땅콩 오일, 목화씨 오일, 홍화 오일, 참깨 오일, 올리브 오일, 옥수수 오일 및 대두 오일; (10) 글리콜, 예컨대 프로필렌 글리콜; (11) 폴리올, 예컨대 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜 (PEG); (12) 에스테르, 예컨대 에틸 올레에이트 및 에틸 라우레이트; (13) 한천; (14) 완충제, 예컨대 수산화마그네슘 및 수산화알루미늄; (15) 알긴산; (16) 발열원-무함유 물; (17) 등장성 염수; (18) 링거액; (19) 에틸 알콜; (20) pH 완충 용액; (21) 폴리에스테르, 폴리카르보네이트 및/또는 폴리무수물; (22) 벌킹제, 예컨대 폴리펩티드 및 아미노산 (23) 혈청 성분, 예컨대 혈청 알부민, HDL 및 LDL; (22) C2-C12 알콜, 예컨대 에탄올; 및 (23) 제약 제제에 사용되는 다른 비-독성 상용성 물질. 습윤제, 착색제, 이형제, 코팅제, 감미제, 향미제, 퍼퓸제, 보존제 및 항산화제가 또한 제제에 존재할 수 있다. "부형제", "담체", "제약상 허용되는 담체" 등과 같은 용어는 본원에서 상호교환가능하게 사용된다.

[1814] 전달 방법

[1815] 일부 측면에서, 본 발명은 본원에 기재된 다중-플랩 프라임 편집제의 다양한 성분을 코딩하는 1개 이상의 폴리뉴클레오티드, 예컨대 본원에 기재된 바와 같은 1개 이상의 벡터, 그의 1개 이상의 전사체, 및/또는 그로부터 전사된 1개 이상의 단백질을 숙주 세포에 전달하는 것을 포함하는 방법을 제공한다. 일부 측면에서, 본 발명은 이러한 방법에 의해 생산된 세포, 및 이러한 세포를 포함하거나 또는 그로부터 생산된 유기체 (예컨대 동물, 식물, 또는 진균)를 추가로 제공한다. 일부 실시양태에서, 가이드 서열과 조합된 (및 임의로 그와 복합체화된) 본원에 기재된 바와 같은 염기 편집제가 세포에 전달된다.

[1816] 예시적인 전달 전략은 본원의 다른 곳에 기재되어 있으며, 이는 벡터-기반 전략, PE 리보핵단백질 복합체 전달, 및 mRNA 방법에 의한 PE의 전달을 포함한다.

[1817] 일부 실시양태에서, 제공된 전달 방법은 뉴클레오펙션, 미세주사, 바이오리스틱, 비로솜, 리포솜, 이뮤노리포솜, 다가양이온 또는 지질:핵산 접합체, 네이키드 DNA, 인공 비리온, 및 DNA의 작용제-증진된 흡수를 포함한다.

[1818] 핵산의 예시적인 전달 방법은 리포펙션, 뉴클레오펙션, 전기천공, 안정한 게놈 통합 (예를 들어, 피기백), 미세주사, 바이오리스틱, 비로솜, 리포솜, 이뮤노리포솜, 다가양이온 또는 지질:핵산 접합체, 네이키드 DNA, 인공 비리온, 및 DNA의 작용제-증진된 흡수를 포함한다. 리포펙션은, 예를 들어 미국 특허 번호 5,049,386, 4,946,787; 및 4,897,355에 기재되어 있고, 리포펙션 시약은 상업적으로 판매된다 (예를 들어, 트랜스펙탐(Transfectam)™, 리포펙틴(Lipofectin)™ 및 SF 세포주 4D-뉴클레오펙터 X 키트™ (론자(Lonza))). 폴리뉴클레오티드의 효율적인 수용체-인식 리포펙션에 적합한 양이온성 및 중성 지질은 파이너(Feigner), WO 91/17424; WO 91/16024의 것을 포함한다. 전달은 세포 (예를 들어, 시험관내 또는 생체외 투여) 또는 표적 조직 (예를 들어, 생체내 투여)에 대한 것일 수 있다. 전달은 RNP 복합체의 사용을 통해 달성될 수 있다.

[1819] 표적화된 리포솜, 예컨대 면역지질 복합체를 포함한 지질:핵산 복합체의 제조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 [Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992)]; 미국 특허 번호 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 및 4,946,787 참조).

[1820] 다른 실시양태에서, 본원에 제공된 전달 방법 및 벡터는 RNP 복합체이다. 융합 단백질의 RNP 전달은 염기 편집의 DNA 특이성을 현저하게 증가시킨다. 융합 단백질의 RNP 전달은 온- 및 오프-타겟 DNA 편집의 탈커플링으로 이어진다. RNP 전달은 플라스미드 전달과 대등한 온-타겟 편집을 유지시키면서 비-반복 부위에서의 오프-타겟 편집을 제거하고, 심지어 고도로 반복적인 VEGFA 부위에서도 오프-타겟 DNA 편집을 크게 감소시킨다 2. 문헌 [Rees, H.A. et al., Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery, Nat. Commun. 8, 15790 (2017)], 2016년 12월 27일에 허여된 미국 특허 번호 9,526,784, 및 2017년 8월 22일에 허여된 미국 특허 번호 9,737,604를 참조하며, 이들 각각은 본원에 참조로 포함된다.

[1821] 핵산을 세포에 전달하기 위한 추가의 방법은 관련 기술분야의 통상의 기술자에게 공지되어 있다. 예를 들어,

본원에 참조로 포함된 US 2003/0087817을 참조한다.

[1822] 본 개시내용의 다른 측면은 다중-플랩 프라임 편집제 구축물을 세포 내로 전달하여 세포 내에서 완전하고 기능적인 프라임 편집제를 형성하는 방법을 제공한다. 예를 들어, 일부 실시양태에서, 세포는 본원에 기재된 조성물 (예를 들어, 분할 Cas9 또는 분할 프라임 편집제를 코딩하는 뉴클레오티드 서열 또는 이러한 뉴클레오티드 서열을 포함하는 핵산 벡터를 함유하는 AAV 입자를 포함하는 조성물)과 접촉된다. 일부 실시양태에서, 접촉은 이러한 뉴클레오티드 서열의 세포 내로의 전달을 발생시키며, 여기서 Cas9 단백질 또는 프라임 편집제의 N-말단 부분 및 Cas9 단백질 또는 프라임 편집제의 C-말단 부분은 세포에서 발현되고, 연결되어, 완전한 Cas9 단백질 또는 완전한 프라임 편집제를 형성한다.

[1823] 본원에 제공된 임의의 rAAV 입자, 핵산 분자 또는 조성물은 임의의 적합한 방식으로, 안정적으로 또는 일시적으로 세포 내로 도입될 수 있음을 인지하여야 한다. 일부 실시양태에서, 개시된 단백질은 세포를 형질감염시킬 수 있다. 일부 실시양태에서, 세포는 핵산 분자로 형질도입 또는 형질감염될 수 있다. 예를 들어, 세포는 (예를 들어, 분할 단백질을 코딩하는 바이러스로) 형질도입될 수 있거나, 또는 분할 단백질을 코딩하는 핵산 분자로 (예를 들어, 분할 단백질을 코딩하는 플라스미드로), 또는 1개 이상의 핵산 분자를 코딩하는 바이러스 게놈을 함유하는 rAAV 입자로 형질감염될 수 있다. 이러한 형질도입은 안정한 또는 일시적 형질도입일 수 있다. 일부 실시양태에서, 분할 단백질을 발현하거나 또는 분할 단백질을 함유하는 세포는, 예를 들어 분할 Cas9 (예를 들어, nCas9) 단백질의 전달 시 1개 이상의 가이드 RNA 서열로 형질도입 또는 형질감염될 수 있다. 일부 실시양태에서, 분할 단백질을 발현하는 플라스미드는 전기천공, 일시적 (예를 들어, 리포펙션) 및 안정한 게놈 통합 (예를 들어, 피기백) 및 바이러스 형질도입 또는 관련 기술분야의 통상의 기술자에게 공지된 다른 방법을 통해 세포 내로 도입될 수 있다.

[1824] 특정 실시양태에서, 본원에 제공된 조성물은 지질 및/또는 중합체를 포함한다. 특정 실시양태에서, 지질 및/또는 중합체는 양이온성이다. 이러한 지질 입자의 제조는 널리 공지되어 있다. 예를 들어, 미국 특허 번호 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; 4,921,757; 및 9,737,604를 참조하고; 이들 각각은 본원에 참조로 포함된다.

[1825] 가이드 RNA 서열은 15-100개 뉴클레오티드 길이일 수 있고, 표적 뉴클레오티드 서열에 상보적인 적어도 10, 적어도 15, 또는 적어도 20개의 인접 뉴클레오티드의 서열을 포함한다. 가이드 RNA는 표적 뉴클레오티드 서열에 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40개의 인접 뉴클레오티드의 서열을 포함할 수 있다. 가이드 RNA는 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 또는 50개 뉴클레오티드 길이일 수 있다.

[1826] 일부 실시양태에서, 표적 뉴클레오티드 서열은 게놈, 예를 들어 진핵 게놈 내의 DNA 서열이다. 특정 실시양태에서, 표적 뉴클레오티드 서열은 포유동물 (예를 들어, 인간) 게놈 내에 있다.

[1827] 본 개시내용의 조성물은 예를 들어 단위 용량으로서 투여 또는 패키징될 수 있다. 용어 "단위 용량"은 본 개시내용의 제약 조성물과 관련하여 사용되는 경우에 대상체에 대한 단일 투여량으로서 적합한 물리적 이산 단위를 지칭하며, 각각의 단위는 요구되는 희석제, 즉, 담체 또는 비히클과 함께 목적하는 치료 효과를 생성하도록 계산된 미리 결정된 양의 활성 물질을 함유한다.

[1828] 질환 또는 장애의 치료는 질환의 발생 또는 진행의 지연, 또는 질환 중증도의 감소를 포함한다. 질환을 치료하는 것은 반드시 치유적 결과를 요구하는 것은 아니다.

[1829] 본원에 사용된 질환의 발생의 "지연"은 질환의 진행을 연기, 방해, 둔화, 지체, 안정화, 및/또는 늦추는 것을 의미한다. 이러한 지연은 치료될 질환 및/또는 개체의 병력에 따라 다양한 시간 길이일 수 있다. 질환의 발생을 "지연" 또는 완화시키거나, 또는 질환의 발병을 지연시키는 방법은, 방법을 사용하지 않는 것과 비교하여, 주어진 시간 프레임에서 질환의 1종 이상의 증상이 발생할 확률을 감소시키고/거나 주어진 시간 프레임에서 증상의 정도를 감소시키는 방법이다. 이러한 비교는 전형적으로 통계적으로 유의한 결과를 제공하기에 충분한 다수의 대상체를 사용하는 임상 연구에 기초한다.

[1830] 질환의 "발생" 또는 "진행"은 질환의 초기 징후 및/또는 뒤이은 진행을 의미한다. 질환의 발생은 관련 기술분야에 널리 공지된 바와 같은 표준 임상 기술을 사용하여 검출가능하고 평가될 수 있다. 그러나, 발생은 또한 검출불가능할 수 있는 진행을 지칭한다. 본 개시내용의 목적상, 발생 또는 진행은 증상의 생물학적 과정을 지칭한다. "발생"은 출현, 재발, 및 발병을 포함한다.

[1831] 본원에 사용된 질환의 "발병" 또는 "출현"은 초기 발병 및/또는 재발을 포함한다. 치료될 질환의 유형 또는 질환의 부위에 따라, 의약 분야의 통상의 기술자에게 공지된 통상적인 방법을 사용하여 단리된 폴리펩티드 또는 제약 조성물을 대상체에게 투여할 수 있다.

[1832] 추가의 상술 없이, 관련 기술분야의 통상의 기술자는 상기 설명에 기초하여 본 개시내용을 그의 최대 정도로 이용할 수 있는 것으로 여겨진다. 따라서, 하기 구체적 실시양태는 단지 예시적이며, 어떠한 방식으로도 나머지 개시내용을 제한하지 않는 것으로 해석되어야 한다. 본원에 인용된 모든 공개물은 본원에 언급된 목적 또는 주제를 위해 참조로 포함된다.

[1833] 실시예

[1834] 실시예 1. 게놈 내 정확한 뉴클레오티드 변화의 설치를 위한 프라임 편집 (PE)

[1835] 목적은 포유동물 게놈에서 단일 뉴클레오티드 변화의 정확하고 일반적인 설치를 위해 형질전환 게놈 편집 기술을 개발하는 것이다. 이 기술은 조사자가 사실상 임의의 포유동물 유전자에서 단일 뉴클레오티드 변경의 효과를 연구할 수 있게 하고, 인간 환자에서 병원성 점 돌연변이를 교정하기 위한 치료적 개입을 잠재적으로 가능하게 할 것이다.

[1836] 게놈 편집을 위한 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) 시스템의 채택은 생명 과학을 혁신시켰다[1-3]. CRISPR을 사용한 유전자 파괴가 이제 상용적이지만, 단일 뉴클레오티드 편집의 정확한 설치는 다수의 질환-원인 돌연변이를 연구 또는 교정하는데 필요함에도 불구하고 주요 도전과제로 남아있다. 상동성 지정 복구 (HDR)는 이러한 편집을 달성할 수 있지만, 낮은 효율 (종종 <5%), 공여자 DNA 복구 주형에 대한 요건, 및 이중 가닥 DNA 파괴 (DSB) 형성의 유해 효과를 겪는다. 최근에, 리우(Liu) 실험실은 염기 편집을 개발하였고, 이는 DSB 없이 효율적인 단일 뉴클레오티드 편집을 달성한다. 염기 편집제 (BE)는 CRISPR 시스템을 염기-변형 데아미나제 효소와 조합하여 표적 C·G 또는 A·T 염기 쌍을 각각 A·T 또는 G·C로 전환시킨다[4-6]. 전세계 연구원들에 의해 이미 널리 사용되기는 하였지만 (애드진에 의해 > 5,000개의 리우 실험실 BE 구축물이 배포됨), 현재의 BE는 12개의 가능한 염기 쌍 전환 중 단지 4개만을 가능하게 하고, 작은 삽입 또는 결실은 교정할 수 없다. 더욱이, 염기 편집의 표적화 범주는 표적 염기에 인접한 비-표적 C 또는 A 염기의 편집 ("방관자 편집")에 의해, 그리고 PAM 서열이 표적 염기로부터 15±2 bp에 존재해야 한다는 요건에 의해 제한된다. 따라서, 이들 제한을 극복하는 것은 게놈 편집의 기본 연구 및 치료 용도를 크게 확장시킬 것이다.

[1837] 여기서, 염기 편집의 주요 제한을 극복하면서 그의 많은 이익-즉, 이중 가닥 파괴 및 공여자 DNA 복구 주형의 회피-을 제공하는 새로운 정밀 편집 접근법을 개발하는 것이 제안된다. 이러한 자명한 목적을 달성하기 위해, 표적-프라이밍 역전사 (TPRT)를 사용하여 표적 게놈 부위에 편집된 DNA 가닥을 직접 설치하는 것을 목표로 한다. 본원에 논의된 설계에서, CRISPR 가이드 RNA (gRNA)는 회합된 리버스 트랜스크립타제 (RT) 효소에 의해 실행될, 돌연변이유발 DNA 가닥 합성을 코딩하는 주형을 운반하도록 조작될 것이다. CRISPR 뉴클레아제 (Cas9)-닉킹된 표적 부위 DNA는 역전사를 위한 프라이머로서의 역할을 하여, 임의의 목적하는 뉴클레오티드 편집의 직접 혼입을 가능하게 할 것이다.

[1838] 섹션 1

[1839] 돌연변이유발 DNA 가닥의 가이드 RNA-주형화된 역전사를 확립한다. 이전의 연구는 DNA 절단 후에 그러나 복합체 해리 전에, Cas9가 비-표적 DNA 가닥을 방출하여 유리 3' 말단을 노출시킨다는 것을 보여주었다. 이 DNA 가닥은 폴리머라제 효소에 의해 연장부에 접근가능하고, gRNA는 DNA 합성을 위한 주형으로서의 역할을 하도록 그의 5' 또는 3' 말단의 연장부를 통해 조작될 수 있는 것으로 가설화된다. 예비 시험관내 연구에서, Cas9:gRNA-결합된 복합체 내의 닉킹된 DNA 가닥은 주형으로서 결합된 gRNA를 사용하여 역전사를 실제로 프라이밍할 수 있다는 것이 확립되었다 (RT 효소는 트랜스로). 다음으로, 상이한 gRNA 링커, 프라이머 결합 부위, 및 합성 주형을 연구하여 시험관내에서 최적 설계 규칙을 결정할 것이다. 이어서, 트랜스로 또는 Cas9에 대한 융합체로서 작용하는 상이한 RT 효소가 시험관내에서 평가될 것이다. 마지막으로, 세포에서 효율적인 결합 및 절단 활성을 보유하는 조작된 gRNA 설계가 확인될 것이다. 이러한 목표의 성공적인 입증은 세포에서 돌연변이유발 가닥 합성을 수행하기 위한 토대를 제공할 것이다.

[1840] 섹션 2

[1841] 인간 세포에서 프라임 편집을 확립한다. DNA 프로세싱 및 복구 메카니즘에 기초하여, 돌연변이유발 DNA 가닥 (단일 가닥 플랩)이 표적 뉴클레오티드의 특이적이고 효율적인 편집을 지시하는 데 사용될 수 있는 것으로 가설

화된다. 예비 연구를 장려함에 있어서, 이러한 전략에 대한 실행가능성은 돌연변이유발 플랩을 함유하는 모델 플라스미드 기질에 의한 편집을 입증함으로써 확립되었다. 목표 1과 공동으로, 돌연변이유발 플랩의 길이, 서열 조성, 표적 뉴클레오티드 정체, 및 3' 말단을 체계적으로 변화시킴으로써 복구 결과를 추가로 평가할 것이다. 작은 1 내지 3개의 뉴클레오티드 삽입 및 결실이 또한 시험될 것이다. 병행하여, 그리고 목표 1로부터 구축하여, 융합 단백질 및 비-공유 동원 전략을 포함한 Cas9-RT 아키텍처가 평가될 것이다. Cas9-RT 아키텍처 및 연장된 gRNA는 인간 게놈 내의 다중 표적 부위에서 세포 편집에 대해 검정될 것이고, 이어서 높은 효율을 위해 최적화될 것이다. 성공적인 경우에, 이 목표는 기초 과학 적용을 위한 TPRT 게놈 편집 (즉, 프라임 편집)을 즉시 확립할 것이다.

[1842] 섹션 3

[1843] 배양된 인간 세포에서 병원성 돌연변이의 부위-특이적 편집을 달성한다. 이러한 기술의 잠재적 일반성은 BE에 의해 현재 교정가능하지 않은 전환 돌연변이 및 indel의 편집을 가능하게 할 수 있다. 목표 1 및 목표 2의 결과에 따라, 베타 글로빈에서의 겸상 적혈구 질환 창시자 돌연변이 (교정을 위해 A·T에서 T·A로의 전환이 필요함) 및 ATP7B에서의 가장 보편적인 윌슨병 돌연변이 (교정을 위해 G·C에서 T·A로의 전환이 필요함)를 포함한 병원성 전환 돌연변이가 배양된 인간 세포에서 표적화될 것이다. 낭성 섬유증을 유발하는 CFTR에서의 3-뉴클레오티드 ΔF508 결실을 포함한 작은 삽입 및 결실 돌연변이의 교정을 또한 검사할 것이다. 성공적인 경우에, 이는 이들 중요한 인간 질환을 다루는 강력한 치료 접근법을 개발하는 토대를 마련할 것이다.

[1844] 접근법

[1845] 목적은 표적화된 게놈 부위에 점 돌연변이를 직접 설치하는 게놈 편집 전략을 개발하는 것이다. 기술 개발 단계에서, 노력은 TPRT 기능성을 CRISPR/Cas 시스템 내로 혼입시키기 위한 단백질 및 RNA 조작에 포커스를 맞출 것이다. 시험관내 검정을 사용하여 기초부터 구축되는 TPRT의 각각의 단계의 기능을 주의깊게 프로빙할 것이다 (목표 1). 제2 포커스 영역은 모델 기질 및 조작된 CRISPR/Cas 시스템의 조합을 사용하여 포유동물 세포에서의 편집 결과를 평가할 것이다 (목표 2). 마지막으로, 적용 단계는 다른 방법에 의한 게놈 편집에 난치성인 돌연변이를 교정하기 위해 기술을 사용할 것이다 (목표 3).

[1846] 일반적 편집 설계가 도 1a-1b에 제시된다. Cas9 닉카제는 HNH 뉴클레아제 도메인에 대한 불활성화 돌연변이를 함유하여 (Spy Cas9 H840A 또는 N863A), DNA 절단을 PAM 함유 가닥 (비-표적 가닥)으로 제한한다. 가이드 RNA (gRNA)는 역전사를 위한 주형을 함유하도록 조작된다 (설계는 슬라이드 5에서 상세화됨). gRNA의 5' 연장부가 제시되지만, 3' 연장부가 또한 구현될 수 있다. Cas9 닉카제는 C-말단 또는 N-말단을 통해 리버스 트랜스크립타제 (RT) 효소에 융합된다. gRNA:Cas9-RT 복합체는 관심 DNA 영역을 표적화하고, 비-표적 가닥을 대체한 후에 R 루프를 형성한다. Cas9는 비-표적 DNA 가닥을 닉킹한다. 닉킹된 가닥의 방출은 주형으로서 연장된 gRNA를 사용하여 역전사를 프라이밍하는 데 적격인 유리 3'-OH 말단을 노출시킨다. 이러한 DNA 합성 반응은 융합된 RT 효소에 의해 수행된다. gRNA 주형은 편집을 위해 표적화된 뉴클레오티드를 제외하고는, 원래의 DNA 듀플렉스와 상동인 DNA 서열을 코딩한다. 역전사의 생성물은 목적하는 편집을 코딩하는 단일 가닥 DNA 플랩이다. 유리 3' 말단을 함유하는 이 플랩은 인접한 DNA 가닥과 평형화되어 5' 플랩 종을 생성할 수 있다. 후자의 종은 지체 가닥 DNA 합성 동안 오카자키 단편으로부터 5' 플랩을 자연적으로 절제하고, 긴-패치 염기 절제 복구 동안 발생하는 가닥 대체 합성 후에 5' 플랩을 제거하는 효소인 FEN1 (플랩 엔도뉴클레아제 1)에 대한 효율적인 기질로서의 역할을 하는 것으로 가설화된다. 닉킹된 DNA의 라이게이션은 미스매칭된 염기 쌍을 생성한다. 이 중간체는 원래의 염기 쌍으로의 복귀 또는 미스매치 복구 (MMR) 과정을 통한 목적하는 편집된 염기 쌍으로의 전환을 겪을 수 있다. 대안적으로, 반보존적 DNA 복제는 각각의 복귀 및 편집의 1개의 카피를 생성할 수 있다.

[1847] 실시예 2 - 프라임 편집: 이중 가닥 DNA 파괴가 없는, 인간 세포에서의 고도로 다재다능하고 정확한 검색-및-대체 게놈 편집

[1848] 현재의 게놈 편집 방법은 프로그램가능한 뉴클레아제를 사용하여 이중 가닥 DNA 파괴의 부산물을 동반하면서 표적 유전자를 붕괴시키거나, 결실시키거나, 또는 삽입하고, 염기 편집제를 사용하여 표적 유전자좌에 4개의 전이 점 돌연변이를 설치할 수 있다. 그러나, 작은 삽입, 작은 결실, 및 8개의 전환 점 돌연변이는 집합적으로 대부분의 병원성 유전자 변이체를 나타내지만, 대부분의 세포 유형에서 과잉의 부산물 없이 효율적으로 교정될 수 없다. 표적 부위를 특정하고 또한 목적하는 편집을 코딩하는 조작된 프라임 편집 가이드 RNA (PEgRNA)에 의해 프로그램화된, 조작된 리버스 트랜스크립타제에 융합된 촉매적으로 손상된 Cas9를 사용하여 새로운 유전자 정보를 특정된 DNA 부위 내로 직접 기록하는, 고도로 다재다능하고 정확한 게놈 편집 방법인 프라임 편집이 본원에 기재된다. 프라임 편집이 표적화된 삽입, 결실, 모든 12가지의 가능한 유형의 점 돌연변이, 및 그의 조합을 효

율적으로 (전형적으로, 비분류된 세포에서 20-60%, 최대 77%), 적은 부산물 (전형적으로, 1-10%)로, 이중 가닥 파괴 또는 공여자 DNA 주형을 필요로 하지 않으면서 만들 수 있다는 것을 확립하기 위해 인간 세포에서 175가지 초과의 별개의 편집을 수행하였다. 프라임 편집은 겸상 적혈구 질환 (HBB에서 A·T에서 T·A로의 전환을 필요로 함) 및 테이-삭스병 (HEXA에서 4-염기 결실을 필요로 함)의 1차 유전적 원인을 교정하기 위해 인간 세포에서 적용되었으며, 둘 다의 경우에서 병원성 게놈 대립유전자는 야생형으로, 최소 부산물로 효율적으로 복귀되었다. 또한 프라임 편집을 사용하여 이들 병원성 HBB 전환 및 HEXA 삽입 돌연변이를 갖는 인간 세포주를 생성하고, 프리온 질환에 대한 저항성을 부여하는 (G·C에서 T·A로의 전환을 필요로 함) PRNP 내에 G127V 돌연변이를 설치하고, His6 태그, FLAG 에피토프 태그, 및 연장된 LoxP 부위를 인간 세포의 표적 유전자좌 내로 효율적으로 삽입하였다. 프라임 편집은 HDR에 비해 효율 및 생성물 순도 이점, 및 염기 편집과 비교하여 상보적 강점 및 약점을 제공한다. 3개의 별개의 염기-쌍형성 사건을 필요로 하는 그의 검색-및-대체 메카니즘과 일치하게, 프라임 편집은 Cas9보다 공지된 Cas9 오프-타겟 부위에서 오프-타겟 DNA 변형에 훨씬 덜 취약하다. 프라임 편집은 게놈 편집의 범주 및 능력을 실질적으로 확장시키고, 원칙적으로 공지된 병원성 인간 유전자 변이체의 ~89%를 교정할 수 있다.

[1849] 임의의 살아있는 세포 또는 유기체의 게놈에서 실질적으로 임의의 표적화된 변화를 만드는 능력은 생명 과학의 오랜 열망이다. 게놈 편집 기술의 급속한 진보에도 불구하고, 질환과 연관된 >75,000개의 공지된 인간 유전자 변이체의 대다수는[111] 대부분 치료상 적절한 세포에서 교정되거나 설치될 수 없다 (도 38a). 프로그램가능한 뉴클레아제, 예컨대 CRISPR-Cas9는 표적 부위에서 삽입 및 결실 (indel)의 혼합을 유도함으로써 유전자를 파괴할 수 있는 이중 가닥 DNA 파괴 (DSB)를 만든다[112-114]. 뉴클레아제는 또한 상동성-비의존성 과정을 통해 표적 유전자를 결실시키거나[115,116], 또는 외인성 유전자를 삽입하는데[117-119] 사용될 수 있다. 그러나, 이중 가닥 DNA 파괴는 또한 생성물의 복합 혼합물, 전위[120], 및 p53 활성화[121,122]를 비롯한 바람직하지 않은 결과와 연관된다. 또한, 대다수의 병원성 대립유전자는 교정을 위해 훨씬 더 정확한 편집 기술을 필요로 하는 작은 삽입, 결실, 또는 염기 치환에 의해 그의 비-병원성 대응물과 상이하다 (도 38a). 뉴클레아제-유도된 DSB에 의해 자극된 상동성-지정 복구 (HDR)[123]는 다양한 정확한 DNA 변화를 설치하는데 널리 사용되어 왔다. 그러나, HDR은 외인성 공여자 DNA 복구 주형에 의존하고, 전형적으로 DSB의 단부-연결 복구로부터 과량의 indel 부산물을 생성하고, 대부분의 치료상 적절한 세포 유형에서 비효율적이다 (T 세포 및 일부 줄기 세포는 중요한 예외임)[124,125]. DSB-매개된 게놈 편집의 효율 및 정밀도를 증진시키는 것은 유망한 노력 포커스로 남아있지만[126-130], 이들 도전과제는 대안적 정밀 게놈 편집 전략의 탐구를 필요로 한다.

[1850] 염기 편집은 포유동물을 포함한 매우 다양한 세포 유형 및 유기체에서 DSB를 필요로 하지 않으면서 4가지 유형의 전이 돌연변이 (C에서 T로, G에서 A로, A에서 G로, 및 T에서 C로)를 효율적으로 설치하거나 교정할 수 있지만[128-131], 현재 8가지 전환 돌연변이 (C에서 A로, C에서 G로, G에서 C로, G에서 T로, A에서 C로, A에서 T로, T에서 A로, 및 T에서 G로) 중 임의의 것, 예컨대 겸상 적혈구 질환의 가장 흔한 원인을 직접 교정하는데 필요한 T·A에서 A·T 돌연변이 (HBB E6V)는 달성할 수 없다[132]. 또한, 표적 결실, 예컨대 테이-삭스병을 유발하는 4-염기 중복의 제거 (HEXA 1278+TATC)[133], 또는 표적화된 삽입, 예컨대 낭성 섬유증의 가장 흔한 원인을 직접 교정하는데 요구되는 정확한 3-염기 삽입 (CFTR ΔF508)[134]을 수행하는 DSB-무함유 방법은 보고된 바 없다. 따라서, 표적화된 전환 점 돌연변이, 삽입, 및 결실은 이들이 집합적으로 대부분의 공지된 병원성 대립유전자를 차지하지만, 대부분의 세포 유형에서 과잉의 부산물 없이 효율적으로 설치하거나 교정하기 어렵다 (도 38a).

[1851] 이중 가닥 DNA 파괴 또는 공여자 DNA 주형을 필요로 하지 않으면서 인간 세포 내 표적화된 유전자좌에서 표적화된 삽입, 결실, 및 모든 12개의 가능한 염기-대-염기 전환을 매개하는 새로운 "검색-및-대체" 게놈 편집 기술인 프라임 편집의 개발이 본원에 기재된다. 초기에 PE1에 의해 예시된 프라임 편집제는 프로그램가능한 닉카제에 융합된 리버스 트랜스크립타제 및 프라임 편집 연장된 가이드 RNA (PEgRNA)를 사용하여 유전자 정보를 PEgRNA 상의 연장부로부터 표적 게놈 유전자좌 내로 직접 카피한다. 제2 세대 프라임 편집제 (PE2)는 조작된 리버스 트랜스크립타제를 사용하여 최소 (전형적으로 <2%) indel 형성 하에 편집 효율을 실질적으로 증가시키는 한편, 제3 세대 PE3 시스템은 제2 가이드 RNA를 부가하여 비-편집된 가닥을 닉킹함으로써, 비-편집된 가닥의 대체를 선호하고, 추가로 편집 효율을 약 1-10% indel 형성 하에 인간 세포에서 전형적으로 약 20-50%로 증가시킨다. PE3은 최적화된 Cas9 뉴클레아제-개시된 HDR과 비교하여 훨씬 더 적은 부산물 및 더 높거나 유사한 효율을 제공

하고, 현-세대 염기 편집제와 비교하여 상보적 강점 및 약점을 제공한다.

[1852] PE3은 HBB E6V의 야생형 HBB로의 효율적인 전환, HEXA 1278+TATC를 야생형 HEXA로 복원하기 위한 삽입된 TATC의 결실, 프리온 질환[135]에 대한 저항성을 부여하는 (G·C에서 T·A로의 전환을 필요로 함) G127V 돌연변이의 PRNP 내 설치, 및 His_6 태그 (18 bp), FLAG 에피토프 태그 (24 bp), 및 Cre-매개된 재조합을 위한 연장된 LoxP 부위 (44 bp)의 표적화된 삽입을 달성하기 위해 인간 HEK293T 세포 내의 게놈 유전자좌에 적용되었다. 프라임 편집은 또한 3종의 다른 인간 세포주, 뿐만 아니라 유사분열후 1차 마우스 피질 뉴런에서 다양한 효율로 성공적이었다. 초기 닉과 편집 위치 사이의 거리에서의 높은 유연성 정도로 인해, 프라임 편집은 Cas9의 PAM 요건에 의해 실질적으로 제약받지 않고, 원칙적으로 대다수의 게놈 유전자좌를 표적화할 수 있다. 오프-타겟 프라임 편집은 공지된 Cas9 오프-타겟 유전자좌에서 오프-타겟 Cas9 편집보다 훨씬 드물고, 이는 아마도 생산적 프라임 편집이 일어나기 위한 3개의 별개의 DNA 염기 쌍형성 사건의 요건 때문일 것이다. DSB 또는 공여자 DNA 주형에 대한 필요 없이 매우 다양한 게놈 유전자좌에서 정확한 표적화된 삽입, 결실, 및 모든 12가지의 가능한 부류의 점 돌연변이를 가능하게 함으로써, 프라임 편집은 많은 유전자 변이체의 연구 및 교정을 진행시킬 잠재력을 갖는다.

[1853] 결과

[1854] 연장된 가이드 RNA로부터 표적 DNA 유전자좌로 정보를 전달하기 위한 전략

[1855] Cas9는 표적 DNA 부위에 혼성화하는 스페이서 서열을 함유하는 가이드 RNA를 사용하여 DNA를 표적화한다[112-114,136,137]. 목표는 천연 CRISPR 시스템에서와 같이 DNA 표적을 특정하고[138,139], 또한 표적 유전자좌에서 상응하는 DNA 뉴클레오티드를 대체하는 새로운 유전자 정보를 함유하도록 가이드 RNA를 조작하는 것이었다. 연장된 가이드 RNA로부터 특정된 DNA 부위로의 유전자 정보의 직접 전달에 이은 원래의 비편집된 DNA의 대체는, 원칙적으로 DSB 또는 공여자 DNA 주형에 대한 의존없이 살아있는 세포에 표적화된 DNA 서열 변화를 설치하는 일반적 수단을 제공할 수 있다. 이러한 직접 정보 전달을 달성하기 위해, 목표는 표적 부위에서 닉킹된 게놈 DNA를 사용하여 3'-히드록실 기를 노출시키고, 조작된 가이드 RNA (이하에서 프라임 편집 가이드 RNA 또는 PEgRNA로 지칭됨) 상의 연장부로부터 표적 부위 내로 직접 유전자 정보의 역전사를 프라이밍하는 것이었다 (도 38a).

[1856] 일부 천연 이동형 유전 요소에 의해 사용되는 메카니즘과 유사한[140], 닉킹 및 역전사의 이들 초기 단계는 한 가닥 상에 2개의 중복 단일 가닥 DNA 플랩: 비편집된 DNA 서열을 함유하는 5' 플랩, 및 PEgRNA로부터 카피된 편집된 서열을 함유하는 3' 플랩을 갖는 분지형 중간체를 생성한다 (도 38b). 성공적인 편집을 달성하기 위해, 이 분지형 중간체는 편집된 3' 플랩이 비편집된 5' 플랩을 대체하도록 분해되어야 한다. 편집된 3' 플랩이 비편집된 가닥과 더 적은 염기 쌍을 만들 수 있기 때문에 5' 플랩과 비편집된 가닥의 혼성화가 열역학적으로 선호될 가능성이 있지만, 5' 플랩은 지체-가닥 DNA 합성 및 긴-패치 염기 절제 복구 동안 생성된 5' 플랩을 절제하는 구조-특이적 엔도뉴클레아제, 예컨대 FEN1[141]에 대한 바람직한 기질이다. 우선적 5' 플랩 절제 및 3' 플랩 라이게이션이 편집된 DNA 가닥의 혼입을 구동시켜, 1개의 편집된 가닥 및 1개의 비편집된 가닥을 함유하는 헤테로듀플렉스 DNA를 생성할 수 있는 것으로 추론되었다 (도 38b).

[1857] 편집의 영구적 설치는 편집된 가닥 내의 정보를 상보적 DNA 가닥에 카피하는 방식으로 2개의 DNA 가닥 사이의 미스매치를 분해하는 후속 DNA 복구로부터 발생할 수 있다 (도 38c). DNA 염기 편집의 효율을 최대화하기 위해 개발된 유사한 전략에 기초하여[131-133], 비-편집된 DNA 가닥을, 이중 가닥 파괴 형성을 최소화하기 위해 초기 닉의 부위로부터 충분히 멀리에서 닉킹하는 것은 DNA 복구를 비-편집된 가닥을 우선적으로 대체하는 것으로 편향시킬 수 있는 것으로 고려되었다.

[1858] 시험관내 및 효모 세포에서의 프라임 편집 단계의 검증

[1859] Cas9의 RuvC 뉴클레아제 도메인에 의한 PAM-함유 DNA 가닥의 절단 후에, 이 가닥의 PAM-원위 단편은 달리 안정한 Cas9:sgRNA:DNA 복합체로부터 해리될 수 있다[143]. 이러한 유리된 가닥의 3' 단부는 DNA 중합을 프라이밍하기 위해 충분히 접근가능할 수 있는 것으로 가설화되었다. 가이드 RNA 조작 노력[144-146] 및 Cas9:sgRNA:DNA 복합체의 결정 구조[147-149]는 sgRNA의 5' 및 3' 말단이 Cas9:sgRNA 활성을 없애지 않고 연장될 수 있음을 시사한다. PEgRNA는 2개의 중요한 성분: 닉킹된 DNA 가닥의 3' 단부가 PEgRNA에 혼성화하도록 하는 프라이머 결합 부위 (PBS), 및 닉킹된 DNA 가닥의 3' 단부가 폴리머라제에 의해 RNA 주형에 걸쳐 연장됨에 따라 게놈 DNA 부위 내로

직접 카피될 목적하는 편집을 함유하는 리버스 트랜스크립타제 (RT) 주형을 포함하도록 sgRNA를 연장시킴으로써 설계되었다 (도 38c).

[1860] 이들 가설을 정제된 에스. 피오게네스 Cas9 단백질을 사용하여 시험관내에서 시험하였다. PBS 서열 (5 내지 6개의 뉴클레오티드, nt) 및 RT 주형 (7 내지 22개의 nt)을 갖는 어느 하나의 말단 상의 sgRNA를 연장시킴으로써 일련의 PEgRNA 후보를 구축하였다. 5'-연장된 PEgRNA는 표적 DNA에 대한 Cas9 결합을 지시하고, 5'-연장된 PEgRNA 및 3'-연장된 PEgRNA 둘 다는 시험관내에서 Cas9-매개된 표적 닉킹 및 포유동물 세포에서 DNA 절단 활성을 지지한다는 것이 확인되었다 (도 44a-44c). 이들 후보 PEgRNA 설계를 사전-닉킹된 5'-Cy5-표지된 dsDNA 기질, 촉매 기능상실 Cas9 (dCas9), 및 몰로니 뮤린 백혈병 바이러스 (M-MLV) 리버스 트랜스크립타제의 상업적 변이체를 사용하여 시험하였다 (도 44d). 모든 성분이 존재할 때, RT 주형을 따른 역전사와 일치하는, 형광 표지된 DNA 가닥의 겔 이동성을 갖는 보다 긴 DNA 생성물로의 효율적인 전환이 관찰되었다 (도 38d, 도 44d-44e). 5'-연장된 또는 3'-연장된 PEgRNA에 의해 목적하는 길이의 생성물이 형성되었다 (도 38d-38e). dCas9의 생략은 DNA 주형 상에서의 리버스 트랜스크립타제-매개된 DNA 중합으로부터 유래된 닉 번역 생성물을 발생시켰고, PEgRNA 정보 전달은 없었다 (도 38d). PEgRNA를 통상적인 sgRNA로 대체한 경우 어떠한 DNA 중합 생성물도 관찰되지 않았고, 이는 PEgRNA의 PBS 및 RT 주형 성분의 필요성을 확인시켜준다 (도 38d). 이들 결과는 Cas9-매개된 DNA 용융이, 닉킹되는 경우에 5'-연장된 또는 3'-연장된 PEgRNA로부터의 역전사를 프라이밍하는 데 적격인 단일 가닥 R-루프를 노출시킨다는 것을 입증한다.

[1861] 다음으로, 비-닉킹된 dsDNA 기질을 PAM-함유 가닥만을 독점적으로 닉킹하는 Cas9 닉카제 (H840A 돌연변이체)를 사용하여 시험하였다[112]. 이들 반응에서, 5'-연장된 PEgRNA는 아마도 손상된 Cas9 닉카제 활성으로 인해 역전사 생성물을 비효율적으로 생성하였다 (도 44f). 그러나, 3'-연장된 PEgRNA는 강건한 Cas9 닉킹 및 효율적인 역전사를 가능하게 하였다 (도 38e). 3'-연장된 PEgRNA의 사용은, 원칙적으로 역전사가 PEgRNA의 나머지 내의 어디에서든 종결되는 잠재력에도 불구하고, 단지 단일의 명백한 생성물만을 생성하였다. Cas9 닉카제, RT, 및 3'-연장된 PEgRNA와의 반응 생성물의 DNA 서열분석은 완전한 RT 주형 서열이 DNA 기질로 역전사되었음을 밝혀내었다 (도 44g). 이들 실험은 3'-연장된 PEgRNA가 Cas9 닉카제 활성을 지시하는 능력을 보유하면서 새로운 DNA 가닥의 역전사의 주형이 될 수 있다는 것을 확립하였다.

[1862] 시험관내 PEgRNA-프로그램화된 역전사에 의해 생산된 3' 플랩의 진핵 세포 DNA 복구 결과를 평가하기 위해, 시험관내 리포터 플라스미드 기질 상의 PEgRNA, Cas9 닉카제, 및 RT를 사용하여 DNA 닉킹 및 역전사를 수행한 후, 반응 생성물로 효모 (에스. 세레비지아에) 세포를 형질전환시켰다 (도 45a). 고무적으로, 조기 정지 코돈을 교정하는 T·A에서 A·T 전환을 코딩하는 3'-연장된 PEgRNA로 플라스미드를 시험관내 편집한 경우에, 효모 형질전환체의 37%가 GFP 및 mCherry 단백질 둘 다를 발현하였다 (도 38f, 도 45c). 도 38e 및 도 44f의 결과와 일치하게, 5'-연장된 PEgRNA를 사용하여 시험관내 수행된 편집 반응은 3'-연장된 PEgRNA를 사용한 것보다 더 적은 GFP 및 mCherry 이중-양성 콜로니 (9%)를 생성하였다 (도 38f 및 도 45d). 또한 프레임시프트 돌연변이를 교정하기 위해 단일 뉴클레오티드를 삽입하거나 (15% 이중-양성 형질전환체) 또는 단일 뉴클레오티드를 결실시키는 (29% 이중-양성 형질전환체) 3'-연장된 PEgRNA를 사용하여 생산적 편집을 관찰하였다 (도 38f 및 도 45e-45f). 이중-양성 효모 콜로니로부터 회수한 편집된 플라스미드의 DNA 서열분석으로 코딩된 전환 편집이 목적하는 서열 위치에서 일어났음을 확인하였다 (도 45g). 이들 결과는 진핵 세포에서의 DNA 복구가 전환, 삽입, 및 결실을 포함한 정확한 DNA 편집을 혼입하기 위해 프라임 편집으로부터 발생하는 3' DNA 플랩을 분해할 수 있다는 것을 입증한다.

[1863] 프라임 편집제 1 (PE1)의 설계

[1864] 시험관내 및 효모에서의 결과에 의해 고무되어, 포유동물 세포에서 게놈 DNA를 편집할 수 있는 최소 개수의 성분을 갖는 프라임 편집 시스템의 개발이 추구되었다. 3'-연장된 PEgRNA (이하 간단히 PEgRNA로 지칭됨, 도 39a) 및 가요성 링커를 통한 리버스 트랜스크립타제에의 Cas9 H840A의 직접 융합체가 기능적 2-성분 프라임 편집 시스템을 구성할 수 있는 것으로 가설화되었다. HEK293T (불멸화 인간 배아 신장) 세포를 Cas9 H840A 닉카제의 어느 하나의 말단에 대한 야생형 M-MLV 리버스 트랜스크립타제의 융합체를 코딩하는 하나의 플라스미드, 뿐만 아니라 PEgRNA를 코딩하는 제2 플라스미드로 형질감염시켰다. 초기 시도는 HEK3 표적 유전자좌에서 검출가능한 T·A에서 A·T로의 전환으로 이어지지 않았다.

[1865] 그러나, PEgRNA에서의 PBS의 8-15개 염기로의 연장 (도 39a)은 HEK3 표적 부위에서의 검출가능한 T·A에서 A·T로의 편집으로 이어졌고 (도 39b), N-말단 RT-Cas9 닉카제 융합체 (1.3% 최대 T·A에서 A·T로의 전환)와 비교하여 RT가 Cas9 닉카제의 C-말단에 융합된 프라임 편집제 구축물의 경우에 효율이 더 높았다 (3.7% 최대 T·A에

서 A·T로의 전환, PBS 길이는 8-15 nt 범위임) (도 39b; 본원의 모든 포유동물 세포 데이터는 달리 명시되지 않는 한 선택 또는 분류 없이 전체 처리된 세포 집단에 대한 값을 보고함). 이들 결과는 Cas9에 융합된 야생형 M-MLV RT가 트랜스로 공급된 M-MLV RT의 상업적 변이체를 사용한 경우 시험관내 요구되는 것과 비교하여 인간 세포에서의 게놈 편집을 위해 보다 긴 PBS 서열을 필요로 한다는 것을 시사한다. Cas9 H840A 닉카제의 C-말단에 융합된 이러한 제1 세대 야생형 M-MLV 리버스 트랜스크립타제를 PE1로 지정하였다.

[1866] PEgRNA에 의해 특정된 4개의 추가의 게놈 표적 부위에 전환 점 돌연변이를 정확하게 도입하는 PE1의 능력 (도 39c)을 시험하였다. HEK3 유전자좌에서의 편집과 유사하게, 이들 게놈 부위에서의 효율은 PBS 길이에 좌우되었고, 최대 편집 효율은 0.7-5.5% 범위였다 (도 39c). PE1로부터의 indel은 낮았고, 각각의 부위의 편집 효율을 최대화하는 조건 하에 5개의 부위에 대해 평균 0.2±0.1%였다 (도 46a). PE1은 또한 HEK3 유전자좌에서 단일-뉴클레오티드 결실 (4.0% 효율), 단일-뉴클레오티드 삽입 (9.7%), 및 3-뉴클레오티드 삽입 (17%)에 의해 예시되는, 표적화된 삽입 및 결실을 설치할 수 있었다 (도 39c). 이들 결과는 이중 가닥 DNA 파괴 또는 DNA 주형을 필요로 하지 않으면서 표적화된 전환, 삽입, 및 결실을 직접 설치하는 PE1의 능력을 확립한다.

[1867] 프라임 편집제 2 (PE2)의 설계

[1868] PE1이 HEK293T 세포 내의 여러 유전자좌에서 다양한 편집을 설치할 수 있지만, 편집 효율은 일반적으로 낮았다 (전형적으로 ≤5%) (도 39c). PE1에서 리버스 트랜스크립타제를 조작하는 것이 프라임 편집 복합체의 고유한 입체형태적 제약 내에서 DNA 합성의 효율을 개선시켜, 보다 높은 게놈 편집 수율을 생성할 수 있는 것으로 가설화되었다. 효소 열안정성[150,151], 프로세싱성[150], 및 DNA:RNA 헤테로듀플렉스 기질 친화도[152]를 증가시키고, RNaseH 활성을 불활성화시키는[153] M-MLV RT 돌연변이가 이전에 보고되었다. 인간 세포에서의 그의 프라임 편집 효율을 평가하기 위해 다양한 리버스 트랜스크립타제 돌연변이를 함유하는 19종의 PE1 변이체를 구축하였다.

[1869] 먼저, 실험실 진화로부터 이전에 발생된 일련의 M-MLV RT 변이체를 숭온에서 역전사를 지지하는 그의 능력[150]에 대해 조사하였다. 이들 아미노산 치환 중 3개 (D200N, L603W, 및 T330P)의 M-MLV RT 내로의 연속 도입 (이하에서 M3으로 지칭됨)은 PE1의 것과 비교하여 HEK293T 세포에서 5개의 게놈 유전자좌에 걸쳐 전환 및 삽입 편집 효율의 6.8배 평균 증가를 발생시켰다 (도 47a-47s).

[1870] 다음으로, M3과 조합하여, 주형:PBS 복합체에 대한 결합, 효소 프로세싱성, 및 열안정성을 증진시키는 것으로 이전에 밝혀진 추가의 리버스 트랜스크립타제 돌연변이[152]를 시험하였다. 분석된 14개의 추가의 돌연변이체 중에서, M3 돌연변이에 더하여 T306K 및 W313F 치환을 갖는 변이체가 인간 세포에서 5개의 게놈 부위에 걸쳐 6개의 전환 또는 삽입 편집에 대해 M3과 비교하여 편집 효율을 추가로 1.3배 내지 3.0배 개선시켰다 (도 47a-47s). PE1 아키텍처 내로 혼입된 M-MLV 리버스 트랜스크립타제의 이러한 오중돌연변이체 (Cas9 H840A-M-MLV RT (D200N L603W T330P T306K W313F))는 이후 PE2로 지칭된다.

[1871] PE2는 PE1보다 실질적으로 더 높은 효율로 단일-뉴클레오티드 전환, 삽입, 및 결실 돌연변이를 설치하고 (도 39c), 보다 짧은 PBS PEgRNA 서열과 상용성이고 (도 39c), 이는 일시적 게놈 DNA:PBS 복합체에 생산적으로 관여하는 증진된 능력과 일치한다. 평균적으로, PE2는 PE1에 비해 프라임 편집 점 돌연변이 효율의 1.6- 내지 5.1배 개선을 발생시켰고 (도 39c), 일부 경우에는 편집 수율을 최대 46배까지 현저하게 개선시켰다 (도 47f 및 도 47i). PE2는 또한 PE1보다 더 효율적으로 표적화된 삽입 및 결실을 발생시켜, HEK3 유전자좌에서의 24-bp FLAG 에피토프 태그의 표적화된 삽입을 4.5% 효율로 달성하였고, 이는 PE1에 의한 이러한 삽입의 설치 효율에 비해 15배 개선된 것이며 (도 47d), HEK3에서 1-bp 결실을 8.6% 효율로 매개하였고, 이는 PE1의 것보다 2.1배 더 높았다 (도 39c). 이들 결과는 PE2를 PE1보다 더 효율적인 프라임 편집제로서 확립한다.

[1872] PEgRNA 특색의 최적화

[1873] PEgRNA 아키텍처와 프라임 편집 효율 사이의 관계를 HEK293T 세포의 5개의 게놈 유전자좌에서 PE2를 사용하여 체계적으로 프로빙하였다 (도 39c). 일반적으로, 보다 낮은 GC 함량을 갖는 프라이밍 부위는 보다 긴 PBS 서열 (EMX1 및 RNF2, 닉의 처음 10 nt 상류에서 각각 40% 및 30% GC 함량을 함유함)을 필요로 하는 반면, 보다 큰 GC 함량을 갖는 것은 보다 짧은 PBS 서열 (HEK4 및 FANCF, 닉의 처음 10 nt 상류에서 각각 80% 및 60% GC 함량을 함유함)을 사용한 프라임 편집을 지지하였고 (도 39c), 이는 닉킹된 DNA 가닥의 PEgRNA PBS에 대한 혼성화를 위한 에너지 요건과 일치한다. PBS 길이 또는 GC 함량 수준은 프라임 편집 효율을 엄격하게 예측하지 못했고, 다른 인자, 예컨대 DNA 프라이머에서의 2차 구조 또는 PEgRNA 연장부가 또한 편집 활성에 영향을 미칠 수 있다. 전형적인 표적 서열에 대해 ~13 nt의 PBS 길이로 시작하고, 서열이 ~40-60% GC 함량에서 벗어나는 경우에 상이

한 PBS 길이를 조사하는 것이 권장된다. 필요한 경우에, 최적 PBS 서열은 실험적으로 결정되어야 한다.

[1874] 다음으로, PEgRNA의 RT 주형 부분의 성능 결정인자를 연구하였다. 10-20 nt 길이 범위의 RT 주형을 갖는 PEgRNA를 PE2를 사용하여 5개의 게놈 표적 부위에서 체계적으로 평가하였고 (도 39d), 3개의 게놈 부위에서는 31 nt만큼 긴 RT 주형을 사용하여 평가하였다 (도 48a-48c). PBS 길이와 같이, RT 주형 길이도 또한 프라임 편집 효율을 최대화하기 위해 달라질 수 있지만, 일반적으로 ≥10 nt 길이의 많은 RT 주형 길이가 보다 효율적인 프라임 편집을 지지한다 (도 39d). 일부 표적 부위는 보다 높은 편집 효율을 달성하기 위해 보다 긴 RT 주형 (>15 nt) (FANCF, EMX1)을 선호하는 반면, 다른 유전자좌는 짧은 RT 주형 (HEK3, HEK4)을 선호하기 때문에 (도 39d), PEgRNA를 최적화할 때 ~10-16 nt로 시작하여 짧은 RT 주형 및 긴 RT 주형 둘 다를 시험하는 것이 권장된다.

[1875] 중요한 것으로, sgRNA 스캐폴드의 말단 헤어핀에 인접한 뉴클레오티드로서 C를 배치하는 RT 주형은 일반적으로 유사한 길이의 RT 주형을 갖는 다른 PEgRNA와 비교하여 보다 낮은 편집 효율을 발생시켰다 (도 48a-48c). Cas9에 결합된 sgRNA의 구조에 기초하여[148,149], 정규 sgRNA의 3' 연장부의 제1 뉴클레오티드로서 C의 존재는 천연적으로 Cas9에서 Tyr 1356과 파이 스택을 형성하는 뉴클레오티드인 G81과 쌍형성하고 sgRNA A68과의 비-정규 염기쌍에 의해 sgRNA 스캐폴드 폴드를 파괴할 수 있는 것으로 추측되었다. 많은 RT 주형 길이가 프라임 편집을 지지하기 때문에, 3' 연장부의 제1 염기 (RT 주형의 마지막 역전사된 염기)가 C가 아닌 PEgRNA를 선택하는 것이 권장된다.

[1876] 프라임 편집제 3 시스템 (PE3 및 PE3b)의 설계

[1877] PE2가 PEgRNA로부터의 유전자 정보를 PE1보다 더 효율적으로 표적 유전자좌로 전달할 수 있는 한편, 세포가 1개의 편집된 가닥 및 1개의 비편집된 가닥에 의해 생성된 헤테로듀플렉스 DNA를 분해하는 방식은 편집이 지속적인지 여부를 결정한다. 염기 편집의 이전 개발은 시토신 또는 아데닌 탈아미노화의 초기 생성물이 1개의 편집된 가닥 및 1개의 비-편집된 가닥을 함유하는 헤테로듀플렉스 DNA이기 때문에 유사한 도전과제에 직면하였다. 염기 편집의 효율을 증가시키기 위해, Cas9 D10A 닉카제를 사용하여 닉을 비-편집된 가닥 내로 도입하고, 편집된 가닥을 주형으로서 사용하여 DNA 복구를 그 가닥으로 지시하였다[129,130,142]. 프라임 편집 효율을 증진시키기 위해 이 원리를 이용하기 위해, PE2에 이미 존재하는 Cas9 H840A 닉카제 및 세포에 의한 비-편집된 가닥의 우선적 대체를 유도하기 위한 단순 sgRNA를 사용하여 비-편집된 가닥을 닉킹하는 유사한 전략을 시험하였다 (도 40a). 프라임 편집을 개시하기 위해 편집된 DNA 가닥도 또한 닉킹되기 때문에, indel로 이어지는 이중 가닥 DNA 파괴의 생산을 최소화하기 위해 비-편집된 가닥 상에서 다양한 sgRNA-프로그램화된 닉 위치를 시험하였다.

[1878] 이러한 PE3 전략은 먼저 PAM의 5' 또는 3'의 PEgRNA-유도된 닉의 부위로부터 14 내지 116개 염기에 위치한 닉을 유도하는 sgRNA를 스크리닝함으로써 HEK293T 세포의 5개의 게놈 부위에서 시험하였다. 시험된 5개의 부위 중 4개에서, 비-편집된 가닥을 닉킹하는 것은 PE2 시스템과 비교하여 indel-무함유 프라임 편집 생성물의 양을 1.5-내지 4.2배만큼, 55%만큼 높게 증가시켰다 (도 40b). 최적 닉킹 위치는 게놈 부위에 따라 달라졌지만, PEgRNA-유도된 닉으로부터 대략 40-90 bp의 PAM의 3' (도 40b에서 양의 거리)에 위치한 닉은 일반적으로 과도한 indel 형성 없이 프라임 편집 효율의 유리한 증가 (평균 41%)를 생성하였다 (시험된 5개의 부위 각각에 대해 가장 높은 편집 효율을 생성하는 sgRNA의 경우 6.8% 평균 indel) (도 40b). 예상된 바와 같이, 일부 부위에서, 비-편집된 가닥 닉을 PEgRNA-유도된 닉의 40 bp 내에 배치하면, indel 형성에서 최대 22%의 큰 증가가 초래되었는데 (도 40b), 이는 아마도 두 가닥을 함께 가까이에서 닉킹하는 것으로부터 이중 가닥 파괴가 형성되기 때문일 것이다. 그러나, 다른 부위에서, PEgRNA-유도된 닉으로부터 14 bp만큼 가까운 닉킹은 단지 5% indel을 생산하였고 (도 40b), 이는 유전자좌-의존성 인자가 근위 이중 닉이 이중 가닥 DNA 파괴로 전환되는 것을 제어한다는 것을 시사한다. 하나의 시험된 부위 (HEK4)에서, 상보적 가닥 닉은 이익을 제공하지 않거나, 또는 심지어 PEgRNA-유도된 닉으로부터 >70 bp의 거리에 위치하는 경우에도 편집 효율을 능가하는 indel 수준 (최대 26%)을 발생시켰으며, 이는 그 부위에서 편집된 가닥이 세포에 의해 닉킹되거나 또는 비효율적으로 라이게이션되는 비통상적인 경향과 일치하였다. PEgRNA-매개된 닉으로부터 대략 50 bp의 비-편집된 가닥 닉으로 시작하고, indel 빈도가 허용되는 수준을 초과하는 경우에 대안적 닉 위치를 시험하는 것이 권장된다.

[1879] 상보적 가닥 닉킹이 프라임 편집 효율을 어떻게 개선시키는지에 대한 이러한 모델 (도 40a)은 편집된 가닥 플랩 분해 후에만 비-편집된 가닥을 닉킹하는 것이 공동 닉의 존재를 최소화하여, 계속 진행되면 indel을 형성하는 이중 가닥 파괴의 빈도를 감소시킬 수 있을 것으로 예측하였다. 비-편집된 가닥 닉킹에 대한 시간적 제어를 달성하기 위해, 편집된 가닥과 매칭되지만 원래의 대립유전자와는 매칭되지 않는 스페이서 서열을 갖는 sgRNA를 설계하였다. 이후 PE3b로 지칭되는 이 전략을 사용하면, 스페이서와 비편집된 대립유전자 사이의 미스매치는

PAM 가닥 상에서 편집 사건이 발생한 후까지 sgRNA에 의한 닉킹을 선호하지 않을 것이다. 이러한 PE3b 접근법을 HEK293T 세포 내의 3개의 게놈 부위에서 5개의 상이한 편집으로 시험하였고, 결과를 PE2 및 PE3 시스템에 의해 달성된 것과 비교하였다. 모든 경우에, PE3b는 PE3과 비교하여 전체 편집 효율의 임의의 명백한 감소 없이, PE3과 비교하여 실질적으로 더 낮은 수준의 indel (3.5- 내지 30배, 평균 12배 더 낮은 indel, 또는 0.85%)과 연관되었다 (도 40c). 따라서, 편집이 제2 프로토스페이서 내에 설정된 경우에, PE3b 시스템은 PE2와 비교하여 편집 효율을 여전히 개선시키면서, 종종 PE3의 것과 유사한 수준으로 indel을 감소시킬 수 있었다 (도 40c).

[1880] 종합하면, 이들 발견은 PE3 시스템 (Cas9 닉카제-최적화된 리버스 트랜스크립타제 + PEgRNA + sgRNA)이 PE2와 비교하여 편집 효율을 ~3배 개선시킨다는 것을 확립하였다 (도 40b-40c). PE3은 PE3의 추가의 닉킹 활성을 고려하여 예상된 바와 같이 PE2보다 더 넓은 범위의 indel을 동반하였다. PE3의 사용은 프라임 편집 효율을 우선시할 때 권장된다. indel의 최소화가 중요한 경우에, PE2는 ~10배 더 낮은 indel 빈도를 제공한다. 비-편집된 가닥을 닉킹하기 위해 설치된 편집을 인식하는 sgRNA를 사용하는 것이 가능한 경우에, PE3b 시스템은 indel 형성을 크게 감소시키면서 PE3-유사 편집 수준을 달성할 수 있다.

[1881] PE3을 사용한 프라임 편집의 표적화 범주 및 다재다능성을 입증하기 위해, 10-뉴클레오티드 RT 주형을 갖는 PE3 및 PEgRNA를 사용하여 HEK3 표적 부위의 +1 내지 +8 위치 (PEgRNA-유도된 닉의 3'의 제1 염기를 위치 +1로서 카운팅함)에 걸친 모든 가능한 단일 뉴클레오티드 치환의 설치를 조사하였다 (도 41a). 집합적으로, 이들 24개의 별개의 편집은 모든 4개의 전이 돌연변이 및 모든 8개의 전환 돌연변이를 커버하고, 평균 33±7.9% (14% 내지 48% 범위)의 편집 효율 (indel은 함유하지 않음)로, 평균 7.5±1.8% indel로 진행된다.

[1882] 중요한 것으로, 장거리 RT 주형은 또한 PE3과 효율적인 프라임 편집을 일으킬 수 있다. 예를 들어, PE3과 34-nt RT 주형을 사용하여, HEK3 유전자좌 내의 위치 +12, +14, +17, +20, +23, +24, +26, +30, 및 +33 (PEgRNA-유도된 닉으로부터 12 내지 33개 염기)에 평균 36±8.7% 효율 및 8.6±2.0% indel로 점 돌연변이를 설치하였다 (도 41b). 다른 유전자좌에서는 +10 위치를 넘어서는 편집이 시도되지 않았지만, 3개의 대안적 부위에서 ≥30 nt인 다른 RT 주형도 또한 효율적인 편집을 지지한다 (도 48a-c). 긴 RT 주형의 실행가능성은 초기 닉 부위로부터 수십개의 뉴클레오티드에 대한 효율적인 프라임 편집을 가능하게 하였다. 어느 하나의 DNA 가닥 상의 NGG PAM은 효율적인 프라임 편집을 지지하는 PAM과 편집 사이의 최대 거리보다 훨씬 적은, 평균 ~8 bp마다 발생하기 때문에, 프라임 편집은 다른 정밀 게놈 편집 방법과 대조적으로 근처 PAM 서열의 이용가능성에 의해 실질적으로 제약받지 않는다[125,142,154]. RNA 2차 구조와 프라임 편집 효율 사이의 추정되는 관계를 고려하면, 긴-범위 편집을 위해 PEgRNA를 설계할 때, 편집 효율을 최적화하기 위해 다양한 길이의 RT 주형, 및 필요한 경우에, 서열 조성 (예를 들어, 동의 코돈)을 시험하는 것이 현명하다.

[1883] 전이 및 전환 점 돌연변이를 도입하기 위해 PE3 시스템의 범주 및 한계를 추가로 시험하기 위해, 6개의 추가의 게놈 표적 부위에 걸쳐 모든 12개의 가능한 유형의 점 돌연변이를 커버하는 72개의 추가의 편집 (도 41c-41h)을 시험하였다. 전반적으로, indel-무함유 편집 효율은 평균 25±14%인 한편, indel 형성은 평균 8.3±7.5%였다. PEgRNA RT 주형이 PAM 서열을 포함하였기 때문에, 프라임 편집은 PAM 서열에 대해 변화를 유도할 수 있었다. 이들 경우에, 보다 높은 편집 효율 (평균 39±9.7%) 및 보다 낮은 indel 생성 (평균 5.0±2.9%)이 관찰되었다 (도 41a-41k, 위치 +5 또는 +6에서의 점 돌연변이). PAM 편집에 대한 이러한 효율의 증가 및 indel 형성의 감소는 상보적 가닥의 복구 전에 Cas9 닉카제가 편집된 가닥에 재결합하고 닉킹하지 못하는 것으로부터 발생할 수 있다. 프라임 편집은 편집 효율의 명백한 상실 없이 조합 편집을 지지하기 때문에, 가능한 경우에 다른 목적하는 변화에 더하여 PAM을 편집하는 것이 권장된다.

[1884] 다음으로, PE3을 사용하여 7개의 게놈 부위에서 14개의 표적화된 작은 삽입 및 14개의 표적화된 작은 결실을 수행하였다 (도 41i). 표적화된 1-bp 삽입은 32±9.8%의 평균 효율로 진행되었고, 3-bp 삽입은 39±16%의 평균 효율로 설치되었다. 표적화된 1-bp 및 3-bp 결실이 또한 효율적이었고, 각각 29±14% 및 32±11%의 평균 수율로 진행되었다. (표적화된 삽입 또는 결실을 넘어선) indel 생성은 평균 6.8±5.4%였다. 위치 +1과 +6 사이에 도입된 삽입 및 결실은 PAM의 위치 또는 구조를 변경시키기 때문에, PAM을 편집하는 점 돌연변이와 유사하게, 상보적 가닥의 복구 전에 Cas9 닉카제가 편집된 DNA 가닥에 재결합하고 닉킹하지 못함으로 인해 이 범위에서의 삽입 및 결실 편집이 전형적으로 보다 효율적인 것으로 추측되었다.

[1885] PE3을 또한 HEK3 부위에서 5 bp 내지 80 bp의 보다 큰 정확한 결실을 매개하는 그의 능력에 대해 시험하였다 (도 41j). 13-nt PBS, 및 각각 표적 유전자좌에 대해 29, 24, 또는 19 bp의 상동성을 함유하는 RT 주형을 사용한 경우에 5-, 10-, 및 15-bp 결실에 대해 매우 높은 편집 효율 (52 내지 78%)이 관찰되었다. 26-nt RT 주형을 사용하는 것은 25 bp의 보다 큰 결실을 72±4.2% 효율로 지지한 한편, 20-nt RT 주형은 80-bp 결실을 52±

3.8% 효율로 가능하게 하였다. 이들 표적화된 결실은 평균 11±4.8%의 indel 빈도를 동반하였다 (도 41j).

[1886] 마지막으로, 삽입 및 결실, 삽입 및 점 돌연변이, 결실 및 점 돌연변이, 또는 3개의 게놈 부위에 걸친 2개의 점 돌연변이로 이루어진 동일한 표적 유전자좌에서의 다중 편집의 12개의 조합을 매개하는 PE3의 능력을 시험하였다. 이들 조합 편집은 평균 55%의 표적 편집, 6.4% indel로 매우 효율적이었고 (도 41k), 이는 개별 표적 부위에서 높은 효율 및 낮은 indel 빈도로 정밀 삽입, 결실, 및 점 돌연변이의 조합을 만드는 프라임 편집의 능력을 입증하였다.

[1887] 종합하면, 도 41a-41k의 예는 7개의 인간 게놈 유전자좌에 걸친 156개의 별개의 전이, 전환, 삽입, 결실, 및 조합 편집을 나타낸다. 이들 발견은 프라임 편집의 다재다능성, 정밀도, 및 표적화 유연성을 확립한다.

[1888] 염기 편집과 비교한 프라임 편집

[1889] 현-세대 시티딘 염기 편집제 (CBE) 및 아데닌 염기 편집제 (ABE)는 높은 효율 및 낮은 indel로 C·G에서 T·A 전이 돌연변이 및 A·T에서 G·C 전이 돌연변이를 설치할 수 있다[129,130,142]. 염기 편집의 적용은 원치않는 방관자 편집을 발생시키는 염기 편집 활성 윈도우 (전형적으로 ~5-bp 폭) 내의 다중 시티딘 또는 아데닌 염기의 존재에 의해[129,130,142,155,] 또는 표적 뉴클레오티드로부터 대략 15±2 nt에 위치한 PAM의 부재에 의해[142,156] 제한될 수 있다. 프라임 편집은 방관자 편집 없이 전이 돌연변이의 정확한 설치에 특히 유용할 수 있거나, 또는 적합하게 위치한 PAM의 결여가 CBE 또는 ABE 활성 윈도우 내에 표적 뉴클레오티드를 유리하게 위치시키는 것을 배제하는 경우에 특히 유용할 수 있는 것으로 예상되었다.

[1890] 닉카제 활성이 없는 (BE2max) 또는 닉카제 활성이 있는 (BE4max) 최적화된 CBE[157]를 사용하거나, 또는 유사한 PE2 및 PE3 프라임 편집 시스템을 사용하여 정규 염기 편집 윈도우 (프로토스페이서 위치 4-8, PAM을 위치 21-23으로서 카운팅함)에서 다중 표적 시티딘을 함유하는 3개의 게놈 유전자좌를 편집함으로써 프라임 편집 및 시토신 염기 편집을 비교하였다. 3개의 부위의 염기 편집 윈도우 내의 9개의 총 표적 시토신 중에서, BE4max는 염기 편집 윈도우의 중심 (프로토스페이서 위치 5-7, 도 42a)의 염기에 대해 PE3보다 평균 2.2배 더 높은 총 C·G에서 T·A 전환을 발생시켰다. 마찬가지로, 비-닉킹 BE2max는 이들 잘-위치된 염기에서 PE2를 평균 1.4배 능가하였다 (도 42a). 그러나, 염기 편집 윈도우의 중심을 넘어선 시토신에 대해서는, PE3은 BE4max를 2.7배 능가하였고, PE2는 BE2max를 2.0배 능가하였다 (PE3의 경우 40±17% 대 BE4max의 경우 15±18%, 및 PE2의 경우 22±11% 대 BE2max의 경우 11±13%의 평균 편집). 전체적으로, PE2의 경우 indel 빈도는 매우 낮았고 (평균 0.86±0.47%), PE3의 경우 BE4max의 것과 유사하거나 그보다 약간 더 높았다 (BE4max 범위: 2.5% 내지 14%; PE3 범위: 2.5% 내지 21%) (도 42b).

[1891] 정확한 C·G에서 T·A 편집의 설치에 대해 염기 편집의 효율을 프라임 편집과 비교할 경우 (임의의 방관자 편집 부재), 프라임 편집의 효율은, 대부분의 게놈 DNA 부위와 유사하게, ~5-bp 염기 편집 윈도우 내에 다중 시토신을 함유하는 상기 부위에서 염기 편집의 효율을 크게 초과하였다 (도 42c). 이들 부위, 예컨대 프로토스페이서 위치 C5, C6, 및 C7에 시토신을 함유하는 EMX1에서, BE4max는 방관자 편집 없이 오직 단일 표적 염기 쌍 전환만을 함유하는 생성물은 거의 생성하지 않았다. 대조적으로, 이 부위에서의 프라임 편집은 임의의 위치 또는 위치의 조합 (C5, C6, C7, C5+C6, C6+C7, C5+C7, 또는 C5+C6+C7)에서 C·G에서 T·A 편집을 선택적으로 설치하는 데 사용될 수 있었다 (도 42c). 모든 정확한 1-염기 또는 2-염기 편집 (즉, 임의의 다른 근처 염기를 변형시키지 않는 편집)은 각각 BE4max 또는 BE2보다 PE3 또는 PE2에 의해 훨씬 더 효율적이었던 반면에, 3-염기 C·G에서 T·A 편집은 BE4max에 의해 더 효율적이었으며 (도 42c), 이는 활성 윈도우 내의 모든 표적 염기를 편집하는 염기 편집제의 경향을 반영한다. 종합하면, 이들 결과는 시토신 염기 편집제가 PE2 또는 PE3보다 최적으로 위치된 표적 염기에서 보다 높은 수준의 편집을 발생시킬 수 있지만, 프라임 편집은 비-최적으로 위치된 표적 염기에서 염기 편집을 능가할 수 있고, 다중 편집가능한 염기로 훨씬 더 높은 정밀도로 편집할 수 있음을 입증한다.

[1892] A·T에서 G·C 편집을 최적화된 비-닉킹 ABE (Cas9 닉카제 대신에 dCas9를 갖는 ABEmax[152], 이하 ABEdmax로 지칭됨) 대 PE2에 의해, 및 최적화된 닉킹 아데닌 염기 편집제 ABEmax 대 PE3에 의해 2개의 게놈 유전자좌에서 비교하였다. 염기 편집 윈도우에 2개의 표적 아데닌을 함유하는 부위에서 (HEK3), ABE는 A5의 전환의 경우 PE2 또는 PE3보다 더 효율적이었지만, PE3은 ABEmax 편집 윈도우의 변연에 있는 A8의 전환의 경우에 더 효율적이었다 (도 42d). 단일 아데닌만을 전환시키는 정밀도 편집의 효율을 비교할 경우, PE3은 A5 및 A8 둘 다에서 ABEmax를 능가하였다 (도 42e). 전반적으로, ABE는 프라임 편집제보다 HEK3에서 훨씬 더 적은 indel을 생산하였다

(ABEdmax의 경우 0.19±0.02% 대 PE2의 경우 1.5±0.46%, 및 ABEmax의 경우 0.53±0.16% 대 PE3의 경우 11±2.3%, 도 42f). 단지 단일 A만이 염기 편집 윈도우 내에 존재하는 FANCF에서, ABE2 및 ABEmax는 총 표적 염기쌍 전환에서 그의 프라임 편집 대응물을 1.8- 내지 2.9배만큼 능가하였고, 여기서 염기 편집 및 프라임 편집 둘 다로부터의 사실상 모든 편집된 생성물은 오직 정확한 편집만을 함유하였다 (도 42d-42e). HEK3 부위와 같이, ABE는 FANCF 부위에서 훨씬 더 적은 indel을 생산하였다 (도 42f).

[1893] 집합적으로, 이들 결과는 염기 편집 및 프라임 편집이 표적화된 전이 돌연변이를 제조하기 위한 상보적 강점 및 약점을 제공한다는 것을 나타낸다. 단일 표적 뉴클레오티드가 염기 편집 윈도우 내에 존재하는 경우, 또는 방관자 편집이 허용되는 경우에, 현재 염기 편집제가 전형적으로 보다 효율적이고, 프라임 편집제보다 더 적은 indel을 생성한다. 다중 시토신 또는 아데닌이 존재하고 방관자 편집이 바람직하지 않은 경우, 또는 표적 염기가 이용가능한 PAM에 비해 염기 편집에 불량하게 위치하는 경우에, 프라임 편집제가 실질적인 이점을 제공한다.

[1894] 오프-타겟 프라임 편집

[1895] 생산적 편집을 발생시키기 위해, 프라임 편집은 Cas9 도메인이 결합하는 것에 대해 상보적인 표적 유전자좌:PEgRNA 스페이서, PEgRNA-프라이밍된 역전사를 개시하는 것에 대해 상보적인 표적 유전자좌:PEgRNA PBS, 및 플랩 분해에 대해 상보적인 표적 유전자좌:리버스 트랜스크립타제 생성물을 필요로 한다. 이들 3가지 별개의 DNA 혼성화 요건은 다른 게놈 편집 방법의 것과 비교하여 오프-타겟 프라임 편집을 최소화할 수 있는 것으로 가설화되었다. 이러한 가능성을 시험하기 위해, HEK293T 세포를 4개의 온-타겟 게놈 유전자좌를 표적화하도록 설계된 PE3 또는 PE2 및 16개의 총 PEgRNA로, Cas9 및 동일한 프로토스페이서를 표적화하는 4개의 상응하는 sgRNA로, 또는 Cas9 및 동일한 16개의 PEgRNA로 처리하였다. 이들 4개의 표적 유전자좌는, 이들 각각이 HEK293T 세포에서 Cas9 및 상응하는 온-타겟 sgRNA가 실질적인 오프-타겟 DNA 변형을 유발하는 것으로 공지된 적어도 4개의 잘-특징화된 오프-타겟 부위를 갖기 때문에 선택되었다[118,159]. 처리 후에, 각각의 온-타겟 스페이서에 대한 4개의 온-타겟 유전자좌 및 상위 4개의 공지된 Cas9 오프-타겟 부위를 총 16개의 오프-타겟 부위에 대해 서열분석하였다 (표 1).

[1896] 이전의 연구와 일치하게[118], Cas9 및 4개의 표적 sgRNA는 이전에 보고된 모든 16개의 오프-타겟 유전자좌를 변형시켰다 (도 42g). HEK3 표적 유전자좌에 대한 4개의 오프-타겟 부위 중 Cas9 오프-타겟 변형 효율은 평균 16%였다. Cas9 및 HEK4를 표적화하는 sgRNA는 4개의 시험된 공지된 오프-타겟 부위의 평균 60% 변형을 발생시켰다. 마찬가지로, EMX1 및 FANCF에 대한 오프-타겟 부위는 Cas9:sgRNA에 의해 각각 48% 및 4.3%의 평균 빈도로 변형되었다 (도 42g). Cas9 뉴클레아제를 갖는 PEgRNA는 sgRNA와 비교하여 평균적으로 유사한 (1- 내지 1.5배 더 낮은) 효율로 온-타겟 부위를 변형시킨 반면, Cas9 뉴클레아제를 갖는 PEgRNA는 sgRNA 보다 ~4배 더 낮은 평균 효율로 오프-타겟 부위를 변형시켰음에 주목한다.

[1897] 놀랍게도, 이들 4개의 표적 스페이서를 함유하는 동일한 16개의 시험된 PEgRNA를 갖는 PE3 또는 PE2는 훨씬 더 낮은 오프-타겟 편집을 발생시켰다 (도 42h). Cas9+sgRNA에 의해 오프-타겟 편집을 겪은 것으로 공지된 16개의 부위 중에서, PE3+PEgRNA 또는 PE2+PEgRNA는 16개의 오프-타겟 부위 중 단지 3개에서 검출가능한 오프-타겟 프라임 편집을 발생시켰으며, 16개 중 단지 1개만이 ≥1%의 오프-타겟 편집 효율을 나타내었다 (도 42h). 이들 16개의 공지된 Cas9 오프-타겟 부위에서 HEK3, HEK4, EMX1, 및 FANCF를 표적화하는 PEgRNA의 경우 평균 오프-타겟 프라임 편집은 각각 <0.1%, <2.2±5.2%, <0.1%, 및 <0.13±0.11%였다 (도 42h). 특히, Cas9+PEgRNA1이 97% 효율로 편집하는 HEK4 오프-타겟 3 부위에서, PE2+PEgRNA1은 동일한 스페이서 서열을 공유함에도 불구하고 오직 0.7% 오프-타겟 편집만을 발생시켰고, 이는 Cas9 편집과 비교하여 프라임 편집에 요구되는 2개의 추가의 DNA 혼성화 사건이 오프-타겟 편집을 얼마나 크게 감소시킬 수 있는지 입증한다. 종합하면, 이들 결과는 PE3 및 PEgRNA가 동일한 프로토스페이서를 표적화하는 Cas9 및 sgRNA보다 인간 세포에서 훨씬 더 낮은 오프-타겟 DNA 편집을 유도한다는 것을 시사한다.

[1898] 3'-연장된 PEgRNA의 역전사는 원칙적으로 가이드 RNA 스캐폴드로 진행될 수 있다. 생성된 3' 플랩이 그의 3' 단부에서 비편집된 DNA 가닥과의 상보성의 결여에도 불구하고 표적 유전자좌 내로 혼입되는 경우에, 그 결과는 indel 빈도에 기여하는 PEgRNA 스캐폴드 뉴클레오티드의 삽입이다. 본 발명자들은 HEK293T 세포의 4개의 유전자좌에서 66회의 PE3-매개된 편집 실험으로부터의 서열분석 데이터를 분석하였고, 평균 1.7±1.5%의 임의의 수의 PEgRNA 스캐폴드 뉴클레오티드의 총 삽입으로, 낮은 빈도의 PEgRNA 스캐폴드 삽입을 관찰하였다 (도 56a-56d). Cas9 도메인 결합으로 인한 리버스 트랜스크립타제에 대한 가이드 RNA 스캐폴드의 접근불가능성, 뿐만 아니라 PEgRNA 스캐폴드 역전사로부터 생성된 3' 플랩의 미스매칭된 3' 단부의 플랩 분해 동안의 세포 절제는 PEgRNA 스캐폴드 뉴클레오티드를 혼입시키는 생성물을 최소화하는 것으로 추측된다. 이러한 사건은 드물지만,

PEgRNA 스캐폴드 혼입을 최소화하는 PEgRNA 또는 프라임 편집제 단백질을 조작하기 위한 향후 노력은 indel 빈도를 추가로 감소시킬 수 있다.

[1899] 일부 염기 편집제에서 데아미나제는 Cas9-비의존성 방식으로 작용하여, 제1-세대 CBE (ABE는 아님) 사이의 낮은 수준이지만 광범위한 오프-타겟 DNA 편집[160-162] 및 제1-세대 CBE 및 ABE 사이의 오프-타겟 RNA 편집[163-165]을 발생시킬 수 있지만, 조작된 데아미나제를 갖는 보다 새로운 CBE 및 ABE 변이체는 Cas9-비의존성 오프-타겟 DNA 및 RNA 편집을 크게 감소시킨다[163-165]. 프라임 편집제는 염기-변형 효소, 예컨대 데아미나제가 결여되어 있고, 따라서 Cas9-비의존성 방식으로 DNA 또는 RNA 염기를 변형시키는 고유 능력을 갖지 않는다.

[1900] 프라임 편집제 내의 리버스 트랜스크립타제 도메인은 원칙적으로 세포에서 적절하게 프라이밍된 RNA 또는 DNA 주형을 프로세싱할 수 있지만, 레트로트랜스포손, 예컨대 LINE-1 패밀리[166], 내인성 레트로바이러스[167,168], 및 인간 텔로머라제의 레트로트랜스포손은 모두 활성 내인성 인간 리버스 트랜스크립타제를 제공하였음에 주목하였다. 인간 세포에서의 그의 천연 존재는 리버스 트랜스크립타제 활성 자체가 실질적으로 독성이 아님을 시사한다. 실제로, 리버스 트랜스크립타제를 불활성화시키는 dCas9, Cas9 H840A 닉카제, 또는 R110S+K103L를 갖는 PE2 (PE2-dRT) 돌연변이를 발현하는 대조군의 것과 비교하여 HEK293T 세포 생존율에서 PE3-의존적 차이는 관찰되지 않았다[169] (도 49a-49b).

[1901] 상기 데이터 및 분석에도 불구하고, 비편향된, 게놈-전반 방식으로 오프-타겟 프라임 편집을 평가하고, 뿐만 아니라 프라임 편집제 또는 프라임 편집 중간체에서 리버스 트랜스크립타제 변이체가 세포에 영향을 미칠 수 있는 정도를 특징화하기 위한 추가의 연구가 필요하다.

[1902] 인간 세포에서의 프라임 편집 병원성 전환, 삽입, 및 결실 돌연변이

[1903] 인간 세포에서 유전 질환을 유발하는 전환, 작은 삽입, 및 작은 결실 돌연변이를 직접 설치하거나 또는 이를 교정하는 PE3의 능력을 시험하였다. 겸상 적혈구 질환은 가장 통상적으로 베타-글로빈에서 Glu6→Val의 돌연변이를 발생시키는 HBB에서의 A·T에서 T·A 전환 돌연변이에 의해 유발된다. 생체외 조혈 줄기 세포를 Cas9 뉴클레아제 및 HDR을 위한 공여자 DNA 주형으로 처리한 다음, 편집된 세포를 풍부화하고, 이식하고, 생착시키는 것은 겸상 적혈구 질환의 치료를 위한 유망한 잠재적 전략이다[170]. 그러나, 이러한 접근법은 정확하게 편집된 HBB 대립유전자에 더하여 많은 indel-함유 부산물을 여전히 생성한다[170-171]. 염기 편집제는 일반적으로 훨씬 더 적은 indel을 생산하지만, 이는 현재 HBB의 정상 서열을 직접 복원하는데 필요한 T·A에서 A·T 전환 돌연변이를 만들 수 없다.

[1904] PE3을 사용하여 HEK293T 세포에서 44% 효율 및 4.8% indel로 HBB E6V 돌연변이를 설치하였다 (도 43a). PE3-처리된 세포의 혼합물로부터, 본 발명자들은 HBB E6V 대립유전자에 대해 동형접합 (삼배체)인 6종의 HEK293T 세포주를 단리하였으며 (도 53a-53d), 이는 병원성 돌연변이를 갖는 인간 세포주를 생성하는 프라임 편집의 능력을 입증한다. HBB E6V 대립유전자를 야생형 HBB로 교정하기 위해, 본 발명자들은 동형접합 HBB E6V HEK293T 세포를 PE3, 및 HBB E6V 돌연변이를 야생형 HBB로 직접 복귀시키도록 프로그램화된 PEgRNA로 처리하였다. 총 14개의 PEgRNA 설계를 시험하였다. 3일 후, DNA 서열분석은 모든 14개의 PEgRNA가 PE3과 조합된 경우 HBB E6V의 야생형 HBB로의 효율적인 교정 (≥26% 야생형 HBB, indel 부재), 및 평균 2.8±0.70%의 indel 수준을 제공함을 밝혀내었다 (도 50a). 최상의 PEgRNA는 HBB E6V를 야생형으로, 2.4% indel로 52% 교정시켰다 (도 43a). PEgRNA에 의해 인식되는 PAM을 변형시키는 침묵 돌연변이의 도입은 편집 효율 및 생성물 순도를 1.4% indel로 58% 교정까지 약간 개선시켰다 (도 43a). 이들 결과는 프라임 편집이 인간 세포주에서 병원성 전환 점 돌연변이를 높은 효율 및 최소 부산물로 설치하고 교정할 수 있음을 확립한다.

[1905] 테이-삭스병은 가장 종종 HEXA 유전자 내로의 4-bp 삽입에 의해 유발된다 (HEXA 1278+TATC)[136]. PE3을 사용하여 HEK293T 세포 내에 이러한 4-bp 삽입을 31% 효율 및 0.8% indel로 설치하였고 (도 43b), HEXA 1278+TATC 대립유전자에 대해 동형접합인 2개의 HEK293T 세포주를 단리하였다 (도 53a-53d). 이들 세포를 사용하여, 야생형 대립유전자로의 완벽한 복귀에 의해 또는 PAM을 파괴하고 침묵 돌연변이를 설치하는 이동된 4-bp 결실에 의해 HEXA에서의 병원성 삽입을 교정하기 위한 PE3 또는 PE3b 시스템과 함께 43개의 PEgRNA 및 3개의 닉킹 sgRNA를 시험하였다 (도 50b). 시험된 43개의 PEgRNA 중 19개는 ≥20% 편집을 발생시켰다. PE3 또는 PE3b 및 최상의 PEgRNA를 사용한 야생형 HEXA로의 완벽한 교정은 유사한 평균 효율 (PE3의 경우 30% 대 PE3b의 경우 33%)로 진행되었지만, PE3b 시스템은 5.3배 더 적은 indel 생성물 (PE3의 경우 1.7% 대 PE3b의 경우 0.32%)을 동반하였

다 (도 43b 및 도 50b). 이들 발견은 포유동물 세포에서 병원성 대립유전자를 효율적으로 및 최소의 부산물로 설치하거나 교정하는 정확한 작은 삽입 및 결실을 만드는 프라임 편집의 능력을 입증한다.

[1906] 마지막으로, 인간 프리온 단백질 (PrP)을 코딩하는 유전자인 PRNP에의 보호 SNP의 설치를 시험하였다. PrP 미스폴딩은 PRNP 유전자에서의 유전성 우성 돌연변이를 통해, 또는 미스폴딩된 PrP에 대한 노출을 통해 자발적으로 발생할 수 있는 진행성 및 치명적 신경변성 프리온 질환을 유발한다[172]. 자연 발생 PRNP G127V 돌연변이체 대립유전자는 인간[138] 및 마우스[173]에서 프리온 질환에 대한 저항성을 부여한다. PE3을 사용하여 HEK293T 세포에서 인간 PRNP 대립유전자 내로 G127V를 설치하였으며, 이는 G·C에서 T·A 전환을 필요로 한다. 4개의 PEgRNA 및 3개의 닉킹 sgRNA를 PE3 시스템으로 평가하였다. 가장 효과적인 PE3 및 PEgRNA에 대한 노출 3일 후에, DNA 서열분석은 53±11%의 G127V 돌연변이 설치 효율 및 1.7±0.7%의 indel 수준을 밝혀내었다 (도 43c). 종합하면, 이들 결과는 질환에 대한 저항성을 효율적으로 유발하거나 부여하고 최소의 부산물을 갖는, 인간 세포에서 전환, 삽입, 또는 결실 돌연변이를 설치하거나 교정하는 프라임 편집의 능력을 확립한다.

[1907] 다양한 인간 세포주 및 1차 마우스 뉴런에서의 프라임 편집

[1908] 다음으로, 프라임 편집을 3종의 추가의 인간 세포주에서 내인성 부위를 편집하는 그의 능력에 대해 시험하였다. K562 (백혈병성 골수) 세포에서, PE3을 사용하여 HEK3, EMX1, 및 FANCF 부위에서의 전환 편집, 뿐만 아니라 HEK3에서의 6xHis 태그의 18-bp 삽입을 수행하였다. 이들 4개의 PE3-매개된 편집 각각에 대해 15-30%의 평균 편집 효율이 관찰되었으며, indel은 평균 0.85-2.2%였다 (도 43a). U2OS (골육종) 세포에서, HEK3 및 FANCF에서의 전환 돌연변이, 뿐만 아니라 HEK3 내로의 3-bp 삽입 및 6xHis 태그 삽입이 7.9-22% 편집 효율로 설치되었고, 이는 indel 형성을 10- 내지 76배 초과하였다 (도 43a). 마지막으로, HeLa (자궁경부암) 세포에서, HEK3 내로의 3-bp 삽입은 12% 평균 효율 및 1.3% indel로 수행되었다 (도 43a). 집합적으로, 이들 데이터는, 편집 효율이 세포 유형에 따라 달라지고 일반적으로 HEK293T 세포에서보다 덜 효율적이기는 하지만, HEK293T 세포 이외의 다중 세포주가 프라임 편집을 지지한다는 것을 나타낸다. 편집:indel 비는 모든 시험된 인간 세포주에서 높게 유지되었다.

[1909] 프라임 편집이 유사분열후, 말단 분화된 1차 세포에서 가능한지 결정하기 위해, E18.5 마우스로부터 수거한 1차 피질 뉴런을 이중 분할-PE3 렌티바이러스 전달 시스템으로 형질도입하였고, 여기서 분할-인테인 스플라이싱[203]은 N-말단 및 C-말단 절반으로부터 PE2 단백질을 재구성하며, 각각은 개별 바이러스로 전달되었다. 유사분열후 뉴런에 대한 편집을 제한하기 위해, 성숙 뉴런에 고도로 특이적인 인간 시냅신 프로모터[204]를 사용하여 PE2 단백질 성분 둘 다의 발현을 구동시켰다. GFP를 자기-절단 P2A 펩티드[205]를 통해 PE2의 N-말단 절반에 융합시켰다. 이중 바이러스 형질도입 2주 후에 뉴런으로부터의 핵을 단리하고, 직접 서열분석하거나, 또는 서열분석 전에 GFP 발현에 대해 분류하였다. DNMT1 유전자좌에서 평균 7.1±1.2%로 전환을 설치하는 프라임 편집이, 분류된 핵에서 0.58±0.14% 평균 indel로 (도 43d) 관찰되었다. 동일한 분할-인테인 이중 렌티바이러스 시스템에서의 Cas9 뉴클레아제는 분류된 피질 뉴런 핵 사이에서 31±5.5% indel을 발생시켰다 (도 43d). 이들 데이터는 유사분열후, 말단 분화된 1차 세포가 프라임 편집을 지지할 수 있음을 나타내고, 따라서 프라임 편집이 세포 복제를 필요로 하지 않는다는 것을 확립한다.

[1910] Cas9-개시된 HDR과 비교한 프라임 편집

[1911] PE3의 성능을 HDR[128]을 지지하는 유사분열 세포주에서 최적화된 Cas9-개시된 HDR[128,125]의 성능과 비교하였다. HEK293T, HeLa, K562 및 U2OS 세포를 Cas9 뉴클레아제, sgRNA, 및 다양한 전환 및 삽입 편집을 설치하도록 설계된 ssDNA 공여자 올리고뉴클레오티드 주형으로 처리하였다 (도 43e-43g, 및 도 51a-51f). 모든 경우에 Cas9-개시된 HDR은 성공적으로 목적하는 편집을 설치하였지만, 이중 가닥 파괴를 유발하는 처리로부터 예상된 바와 같이, 훨씬 더 높은 수준의 부산물 (우세하게 indel)을 가졌다. HEK293T 세포에서 PE3을 사용한 것은, HBB E6V 설치 및 교정이 각각 42% 및 58% 평균 편집 효율 및 2.6% 및 1.4% 평균 indel로 진행되었다 (도 43e 및 도 43g). 대조적으로, Cas9 뉴클레아제 및 HDR 주형을 사용한 동일한 편집은 79% 및 51% 평균 indel 빈도로 5.2% 및 6.7% 평균 편집 효율을 발생시켰다 (도 43e 및 도 43g). 유사하게, PE3은 PRNP G127V를 53% 효율 및 1.7% indel로 설치한 반면, Cas9-개시된 HDR은 이러한 돌연변이를 6.9% 효율 및 53% indel로 설치하였다 (도 43e 및 도 43g). 따라서, HBB E6V 설치, HBB E6V 교정, 및 PRNP G127V 설치에 대한 편집:indel의 비는 평균적으로 Cas9-개시된 HDR보다 PE3의 경우 270배 더 높았다.

[1912] HEK293T 이외의 인간 세포주에서의 PE3과 HDR 사이의 비교는 PE3 편집 효율이 더 낮긴 하지만 유사한 결과를 나

타내었다. 예를 들어, K562 세포에서, HEK3 내로의 PE3-매개된 3-bp 삽입은 Cas9-개시된 HDR의 경우의 17% 편집 및 72% indel과 비교하여 25% 효율 및 2.8% indel로 진행되었고, 40배의 편집:indel 비 이익으로 PE3이 선호된다 (도 43f-43g). U2OS 세포에서, PE3은 이러한 3-bp 삽입을 22% 효율 및 2.2% indel로 수행한 반면, Cas9-개시된 HDR은 49배 더 낮은 편집:indel 비인 15% 편집 및 74% indel을 발생시켰다 (도 43f-43g). HeLa 세포에서, PE3은 이러한 삽입을 12% 효율 및 1.3% indel로 행한 반면, Cas9-개시된 HDR의 경우에는 3.0% 편집 및 69% indel이었고, 이는 210배 편집:indel 비 차이였다 (도 43f-43g). 집합적으로, 이들 데이터는 HDR이 전형적으로 시험된 4종의 세포주에서 PE3과 유사하거나 또는 그보다 더 낮은 편집 효율 및 훨씬 더 높은 indel을 발생시킨다는 것을 나타내었다 (도 51a-51f).

[1913] 논의 및 향후 방향

[1914] DNA 서열을 단일-뉴클레오티드 정밀도로 삽입하는 능력은 특히 프라임 편집 능력을 가능하게 한다. 예를 들어, PE3을 사용하여 HEK293T 세포의 HEK3 유전자좌 내로 His_6 태그 (18 bp, 65% 평균 효율), FLAG 에피토프 태그 (24 bp, 18% 평균 효율), 및 Cre 레콤비나제에 대한 천연 기질인 연장된 LoxP 부위 (44 bp, 23% 평균 효율)를 정확하게 삽입하였다. 평균 indel은 이들 실시예에 대해 3.0% 내지 5.9% 범위였다 (도 43h). 많은 생명공학, 합성 생물학, 및 치료 용도는 새로운 DNA 서열을 살아있는 세포의 관심 표적 부위 내로 효율적이고 정확하게 도입하는 능력으로부터 발생하는 것으로 고려된다.

[1915] 집합적으로, 본원에 기재된 프라임 편집 실험은 명백한 이중 가닥 DNA 파괴를 만들지 않으면서 PAM의 출발의 3 bp 상류 내지 29 bp 하류 범위의 위치에서 인간 및 마우스 게놈의 12개의 내인성 유전자좌에 걸쳐, 최대 44 bp의 18개의 삽입, 최대 80 bp의 22개의 결실, 77개의 전환을 포함한 113개의 점 돌연변이, 및 18개의 조합 편집을 설치하였다. 이들 결과는 현저하게 다재다능한 게놈 편집 방법으로서 프라임 편집을 확립한다. 클린바에서의 삽입, 결실, indel, 및 중복의 압도적인 다수 (85-99%)가 ≤30 bp이기 때문에 (도 52a-52d), 원칙적으로 프라임 편집은 클린바에서의 현재 공지된 75,122개의 병원성 인간 유전자 변이체 (도 38a에서의 전이, 전환, 삽입, 결실, indel, 및 중복)의 최대 ~89%를 교정할 수 있고, 이는 카피수 획득 또는 손실에 의해 유발된 질환을 호전시키는 추가의 잠재력을 갖는다.

[1916] 중요한 것으로, 임의의 목적하는 편집에 대해, 본원에서 광범위하게 입증된 바와 같이, 프라임 편집의 유연성은 PEgRNA-유도된 닉 위치, sgRNA-유도된 제2 닉 위치, PBS 길이, RT 주형 길이, 및 먼저 편집할 가닥의 많은 가능한 선택을 제공한다. 다른 정밀 게놈 편집 방법에 대해 전형적으로 이용가능한 보다 제한된 옵션과 대조되는 이러한 유연성은[125,142,154], 다양한 프라임 편집 전략을 포괄한 14 및 43개의 PEgRNA를 시험하여 각각 병원성 HBB 및 HEXA 대립유전자의 교정을 최적화한 도 50a-50b에 제시된 바와 같이, 편집 효율, 생성물 순도, DNA 특이성, 또는 다른 파라미터를 주어진 적용의 필요에 적합하도록 최적화되게 한다.

[1917] 프라임 편집을 추가로 이해하고 개선시키기 위해 많은 추가의 연구가 필요하다. 프라임 편집제 시스템의 추가의 변형은 그의 상용성을 다른 세포 유형, 예컨대 유사분열후 세포를 포함하도록 확장시키는데 요구될 수 있다. 프라임 편집을 바이러스 및 비-바이러스 시험관내 및 생체내 전달 전략과 결부시키는 것은 유전 질환의 연구 및 치료를 포함한 광범위한 적용을 가능하게 하기 위해 프라임 편집의 잠재력을 충분히 조사하는데 필요하다. 그러나, 이중 가닥 파괴 또는 HDR을 필요로 하지 않으면서 포유동물 세포의 게놈에서 고도로 정확한 표적화된 전이, 전환, 작은 삽입, 및 작은 결실을 가능하게 함으로써, 프라임 편집은 게놈 편집의 범주를 실질적으로 확장시키는 새로운 "검색-및-대체" 능력을 제공한다.

[1918] 방법

[1919] 일반적 방법

[1920] 달리 나타내지 않는 한, 퓨전 U 그린 멀티플렉스(Phusion U Green Multiplex) PCR 마스터 믹스 (써모피셔 사이언티픽) 또는 Q5 핫 스타트 하이-피델리티(Q5 Hot Start High-Fidelity) 2x 마스터 믹스 (뉴 잉글랜드 바이오랩스)를 사용한 PCR에 의해 DNA 증폭을 수행하였다. Cy5-표지된 DNA 올리고뉴클레오티드를 포함한 DNA 올리고뉴클레오티드, dCas9 단백질, 및 Cas9 H840A 단백질을 인티그레이티드 DNA 테크놀로지스로(Integrated DNA Technologies)부터 입수하였다. 효모 리포터 플라스미드를 이전에 기재된 플라스미즈64(plasmids64)로부터 유도하고, 깁슨 조립 방법에 의해 클로닝하였다. 본원에 사용된 모든 포유동물 편집제 플라스미드를 이전에 기재된 바와 같은 USER 클로닝 방법을 사용하여 조립하였다[175]. 어닐링된 올리고뉴클레오티드를 BsmBI-소화된 수용자 벡터 내로 라이게이션하여 sgRNA를 발현하는 플라스미드를 구축하였다. PEgRNA를 발현하는 플라스미드는 맞

춤 수용자 플라스미드를 사용하여 깁슨 조립 또는 골든 게이트 조립에 의해 구축하였다 (보충 '골든 게이트 조립' 개요 참조). 본원에 사용된 sgRNA 및 PEgRNA 구축물의 서열은 표 2A-2C 및 표 3A-3R에 열거된다. 포유동물 세포 실험을 위한 모든 벡터는 플라스미드 플러스 미디프렙(Plasmid Plus Midiprep) 키트 (퀴아젠) 또는 퓨어일드(PureYield) 플라스미드 미니프렙 키트 (프로메가)를 사용하여 정제하였고, 이는 내독소 제거 단계를 포함한다. 살아있는 동물을 사용한 모든 실험은 브로드 인스티튜트 동물 실험 윤리 위원회(Broad Institute Institutional and Animal Care and Use Committees)의 승인을 받았다. 야생형 C57BL/6 마우스를 찰스 리버(Charles River) (#027)로부터 입수하였다.

[1921] 시험관내 생화학적 검정

[1922] PEgRNA 및 sgRNA를 T7 프로모터 서열을 함유하는 PCR-증폭된 주형으로부터 하이스크라이브 T7 시험관내 전사 키트 (뉴 잉글랜드 바이오랩스)를 사용하여 시험관내 전사시켰다. RNA를 변성 우레아 PAGE에 의해 정제하고, 사용 전에 분석용 겔에 의해 품질-확인하였다. 5'-Cy5-표지된 DNA 듀플렉스 기질을 비-닉킹된 기질의 경우 2개의 올리고뉴클레오티드 (Cy5-AVA024 및 AVA025; 1:1.1 비) 또는 사전-닉킹된 기질의 경우 3개의 올리고뉴클레오티드 (Cy5-AVA023, AVA025 및 AVA026; 1:1.1:1.1)를 사용하여 95℃로 3분 동안 가열한 다음 실온으로 천천히 냉각시킴으로써 어닐링하였다 (표 2A-2C). Cas9 절단 및 역전사 반응을 dNTP (20 mM HEPES-K, pH 7.5; 100 mM KCl; 5% 글리세롤; 0.2 mM EDTA, pH 8.0; 3 mM MgCl2; 0.5 mM dNTP 믹스; 5 mM DTT)가 보충된 1x 절단 완충제[205] 중에서 수행하였다. dCas9 또는 Cas9 H840A (최종 5 μM) 및 sgRNA 또는 PEgRNA (최종 5 μM)를 실온에서 5 μL 반응 혼합물 중에서 10분 동안 사전-인큐베이션한 후, 듀플렉스 DNA 기질 (최종 400 nM)을 첨가하고, 이어서 적용가능한 경우에 개시되지 않은 M-MLV RT 변이체인 슈퍼스크립트 III 리버스 트랜스크립타제 (써모피셔 사이언티픽)를 첨가하였다. 반응을 37℃에서 1시간 동안 수행한 다음, 물로 10 μL의 부피로 희석하고, 0.2 μL의 프로테이나제 K 용액 (20 mg/mL, 써모피셔 사이언티픽)으로 처리하고, 실온에서 30분 동안 인큐베이션하였다. 95℃에서 10분 동안 열 불활성화시킨 후, 반응 생성물을 2x 포름아미드 겔 로딩 완충제 (90% 포름아미드; 10% 글리세롤; 0.01% 브로모페놀 블루)와 합하고, 95℃에서 5분 동안 변성시키고, 변성 우레아-PAGE 겔 (15% TBE-우레아, 55℃, 200V)에 의해 분리하였다. DNA 생성물을 타이푼 FLA 7000 생체분자 영상화기를 사용하여 Cy5 형광 신호에 의해 가시화하였다.

[1923] 사전-인큐베이션된 dCas9:sgRNA 또는 dCas9:PEgRNA 복합체 (최종 5 nM 내지 1 μM 농도 범위) 및 Cy5-표지된 듀플렉스 DNA (Cy5-AVA024 및 AVA025; 20 nM 최종)를 사용하여 1x 결합 완충제 (1x 절단 완충제 + 10 μg/mL 헤파린) 중에서 전기영동 이동성 변화 검정을 수행하였다. 37℃에서 15분 인큐베이션 후에, 샘플을 천연 PAGE 겔 (10% TBE)에 의해 분석하고, Cy5 형광에 대해 영상화하였다.

[1924] 역전사 생성물의 DNA 서열분석을 위해, 형광 밴드를 절제하고, 우레아-PAGE 겔로부터 정제한 다음, 제조업체의 프로토콜에 따라 dGTP 또는 dATP의 존재 하에 말단 트랜스퍼라제 (TdT; 뉴 잉글랜드 바이오랩스)로 3' 테일링하였다. 테일링된 DNA 생성물을 결합 완충제 (40% 포화 수성 구아니디늄 클로라이드 + 60% 이소프로판올)로 10배 희석하고, 퀴아퀵(QIAquick) 스핀 칼럼 (퀴아젠(Qiagen))에 의해 정제한 다음, 프라이머 AVA134 (A-테일링된 생성물) 또는 AVA135 (G-테일링된 생성물)를 사용하는 클레나우 단편 (뉴 잉글랜드 바이오랩스)에 의한 프라이머 연장을 위한 주형으로서 사용하였다 (표 2A-2C). 프라이머 AVA110 및 AVA122를 사용하여 10 사이클 동안 PCR에 의해 연장부를 증폭시킨 후, 생어 방법을 사용하여 AVA037로 서열분석하였다 (표 2A-2C).

[1925] 효모 형광 리포터 검정

[1926] 인-프레임 정지 코돈, +1 프레임시프트, 또는 -1 프레임시프트를 함유하는 이중 형광 리포터 플라스미드를 상기 기재된 바와 같이 시험관내에서 5'-연장된 PEgRNA 또는 3'-연장된 PEgRNA 프라임 편집 반응에 적용하였다. 37℃에서 1시간 동안 인큐베이션한 후, 반응물을 물로 희석하고, 플라스미드 DNA를 0.3 M 아세트산나트륨 및 70% 에탄올로 침전시켰다. 이전에 기재된 바와 같이 전기천공에 의해 재현탁된 DNA로 에스. 세레비지아에를 형질전환시키고[67], 류신이 없는 합성 완전 배지 (SC(글루코스), L-) 상에 플레이팅하였다. GFP 및 mCherry 형광 신호를 타이푼 FLA 7000 생체분자 영상화기를 사용하여 콜로니로부터 가시화하였다.

[1927] 일반적 포유동물 세포 배양 조건

[1928] HEK293T (ATCC CRL-3216), U2OS (ATTC HTB-96), K562 (CCL-243), 및 HeLa (CCL-2) 세포를 ATCC로부터 구입하고, 각각 10% (v/v) 태아 소 혈청 (깁코(Gibco), 검증됨) 및 1x 페니실린 스트렙토마이신 (코닝(Corning))으로 보충된 둘베코 변형 이글 배지 (DMEM) 플러스 글루타맥스 (써모피셔 사이언티픽), 맥코이 5A 배지 (깁코), RPMI 배지 1640 플러스 글루타맥스 (깁코), 또는 이글 최소 필수 배지 (EMEM, ATCC)에서 각각 배양

및 계대배양하였다. 모든 세포 유형을 인큐베이션하고, 유지시키고, 5% CO_2 하에 37℃에서 배양하였다. 세포주는 그의 각각의 공급업체에 의해 인증되었고, 미코플라스마에 대해 음성으로 시험되었다.

[1929] HEK293T 조직 배양 형질감염 프로토콜 및 게놈 DNA 제조

[1930] 성장시킨 HEK293T 세포를 48-웰 폴리-D-리신 코팅된 플레이트 (코닝) 상에 시딩하였다. 시딩 16 내지 24시간 후에, 세포를 제조업체의 프로토콜에 따라 리포펙타민 2000 (써모피셔 사이언티픽) 1 μL 및 PE 플라스미드 750 ng, PEgRNA 플라스미드 250 ng, 및 sgRNA 플라스미드 83 ng (PE3 및 PE3b의 경우)으로 대략 60% 전면생장률로 형질감염시켰다. 달리 언급되지 않는 한, 세포를 형질감염 후 3일 배양하였고, 그 후에 배지를 제거하고, 세포를 1x PBS 용액 (써모피셔 사이언티픽)으로 세척하고, 150 μL의 새로 제조된 용해 완충제 (10 mM 트리스-HCl, pH 7.5; 0.05% SDS; 25 μg/mL 프로테이나제 K (써모피셔 사이언티픽))를 조직 배양 플레이트의 각각의 웰에 직접 첨가함으로써 게놈 DNA를 추출하였다. 게놈 DNA 혼합물을 37℃에서 1 내지 2시간 동안 인큐베이션한 후, 80℃ 효소 불활성화 단계를 30분 동안 수행하였다. 포유동물 세포 게놈 DNA 증폭에 사용된 프라이머를 표 4에 열거하였다. HEK293T 세포에서의 HDR 실험을 위해, 231 ng의 뉴클레아제-발현 플라스미드, 69 ng의 sgRNA 발현 플라스미드, 50 ng (1.51 pmol)의 100-nt ssDNA 공여자 주형 (PAGE-정제됨; 인티그레이티드 DNA 테크놀로지스)을 웰당 1.4 μL 리포펙타민 2000 (써모피셔)을 사용하여 리포펙션시켰다. 모든 HDR 실험으로부터의 게놈 DNA를 제조업체의 프로토콜에 따라 아젠코트 디엔어드밴스 키트(Agencourt DNAdvance Kit) (베크만 쿨터)를 사용하여 정제하였다.

[1931] 게놈 DNA 샘플의 고처리량 DNA 서열분석

[1932] 관심 게놈 부위를 게놈 DNA 샘플로부터 증폭시키고, 하기 변형과 더불어 이전에 기재된 바와 같이 일루미나 MiSeq 상에서 서열분석하였다[129,130]. 간략하게, 일루미나 정방향 및 역방향 어댑터를 함유하는 증폭 프라이머 (표 4)를 관심 게놈 영역을 증폭시키는 제1 라운드의 PCR (PCR 1)에 사용하였다. 25 μL PCR 1 반응을 0.5 μM의 각각의 정방향 및 역방향 프라이머, 1 μL의 게놈 DNA 추출물 및 12.5 μL의 퓨전 U 그린 멀티플렉스 PCR 마스터 믹스를 사용하여 수행하였다. PCR 반응을 하기와 같이 수행하였다: 98℃에서 2분, 이어서 [98℃에서 10초, 61℃에서 20초, 및 72℃에서 30초]의 30 사이클, 이어서 최종 72℃에서 2분 연장. 고유한 일루미나 바코딩 프라이머 쌍을 2차 PCR 반응 (PCR 2)에서 각각의 샘플에 첨가하였다. 구체적으로, 25 μL의 주어진 PCR 2 반응물은 0.5 μM의 각각의 고유한 정방향 및 역방향 일루미나 바코딩 프라이머 쌍, 1 μL의 비정제 PCR 1 반응 혼합물, 및 12.5 μL의 퓨전 U 그린 멀티플렉스 PCR 2x 마스터 믹스를 함유하였다. 바코딩 PCR 2 반응을 하기와 같이 수행하였다: 98℃에서 2분, 이어서 [98℃에서 10초, 61℃에서 20초, 및 72℃에서 30초]의 12 사이클, 이어서 최종 72℃에서 2분 연장. PCR 생성물을 1.5% 아가로스 겔에서 전기영동에 의해 분석적으로 평가하였다. PCR 2 생성물 (공통 앰플리콘에 의해 풀링됨)을 퀴아퀵 겔 추출 키트 (퀴아젠)를 사용하여 1.5% 아가로스 겔에 의한 전기영동에 의해, 40 μL의 물로 용리시켜 정제하였다. DNA 농도를 형광측정 정량화 (큐비트(Qubit), 써모피셔 사이언티픽) 또는 qPCR (카파 라이브러리 정량화 키트- 일루미나, 카파 바이오시스템즈(KAPA Biosystems))에 의해 측정하고, 일루미나 MiSeq 기기 상에서 제조업체의 프로토콜에 따라 서열분석하였다.

[1933] 서열분석 판독물을 MiSeq 리포터 (일루미나)를 사용하여 탈멀티플렉스화하였다. CRISPResso2[178]를 사용하여 참조 서열에 대한 앰플리콘 서열의 정렬을 수행하였다. 점 돌연변이 편집의 정량화에 대해, CRISPResso2를 표준 모드로 "폐기_indel_판독물"을 온으로 하여 실행하였다. 편집 효율을 (비-폐기된 판독물에서의 명시된 점 돌연변이의 빈도) x (비-폐기된 판독물의 #) ÷ 총 판독물로서 계산하였다. 삽입 또는 결실 편집에 대해서는, CRISPResso2를 예상된 대립유전자로서 목적하는 대립유전자를 사용하여 HDR 모드로 (e flag), "폐기_indel_판독물"을 온으로 하여 실행하였다. 편집 수율은 HDR 정렬된 판독물의 수를 총 판독물로 나눔으로써 계산하였다. 모든 편집에 대해, indel 수율은 폐기된 판독물의 수를 총 판독물로 나눔으로써 계산하였다.

[1934] U2OS, K562, 및 HeLa 세포의 뉴클레오펙션

[1935] K562, HeLa, 및 U2OS 세포를 사용하여 모든 실험에서 뉴클레오펙션을 수행하였다. 이들 세포 유형에서의 PE 조건을 위해, 800ng 프라임 편집제-발현 플라스미드, 200ng PEgRNA-발현 플라스미드, 및 83 ng 닉킹 플라스미드를 16-웰 뉴클레오큐벳 스트립 (론자)에서 20 μL의 최종 부피로 뉴클레오펙션시켰다. 이들 3가지 세포 유형에서 HDR 조건을 위해, 350 ng 뉴클레아제-발현 플라스미드, 150 ng sgRNA-발현 플라스미드 및 200 pmol (6.6 μg) 100-nt ssDNA 공여자 주형 (PAGE-정제됨; 인티그레이티드 DNA 테크놀로지스)을 16-웰 뉴클레오큐벳 스트립 (론자)에서 샘플당 20 μL의 최종 부피로 뉴클레오펙션시켰다. K562 세포를 SF 세포주 4D-뉴클레오펙터 X 키트 (론자)를 사용하여 샘플당 5 x 10^5개 세포 (프로그램 FF-120)로 제조업체 프로토콜에 따라 뉴클레오펙션시켰다.

U2OS 세포를 SE 세포주 4D-뉴클레오펙터 X 키트 (론자)를 사용하여 샘플당 3-4 x 10^5개 세포 (프로그램 DN-100)로 제조업체 프로토콜에 따라 뉴클레오펙션시켰다. HeLa 세포를 SE 세포주 4D-뉴클레오펙터 X 키트 (론자)를 사용하여 샘플당 2 x 10^5개 세포 (프로그램 CN-114)로 제조업체 프로토콜에 따라 뉴클레오펙션시켰다. 세포를 게놈 DNA 추출을 위해 뉴클레오펙션 72시간 후에 수거하였다.

[1936] HDR 실험을 위한 게놈 DNA 추출

[1937] HEK293T, HEK293T HBB E6V, K562, U2OS, 및 HeLa 세포에서의 모든 HDR 비교 실험으로부터의 게놈 DNA를 제조업체의 프로토콜에 따라 아젠코트 디엔어드밴스 키트 (베크만 쿨터)를 사용하여 정제하였다.

[1938] PE2, PE3, BE2, BE4max, ABEdmax, 및 ABEmax 사이의 비교

[1939] HEK293T 세포를 48-웰 폴리-D-리신 코팅된 플레이트 (코닝) 상에 시딩하였다. 16 내지 24시간 후, 세포를 대략 60% 전면생장률로 형질감염시켰다. CBE 또는 ABE 구축물을 사용한 염기 편집을 위해, 세포를 염기 편집제 플라스미드 750 ng, sgRNA 발현 플라스미드 250 ng, 및 리포펙타민 2000 (써모피셔 사이언티픽) 1 μL로 형질감염시켰다. PE 형질감염을 상기 기재된 바와 같이 수행하였다. PE 및 BE에 대한 게놈 DNA 추출을 상기 기재된 바와 같이 수행하였다.

[1940] 공지된 Cas9 오프-타겟 부위에서의 PE3 활성의 결정

[1941] 공지된 Cas9 오프-타겟 부위에서의 PE3 오프-타겟 편집 활성을 평가하기 위해, PE3으로의 형질감염 3일 후에 HEK293T 세포로부터 추출한 게놈 DNA를 16개의 이전에 보고된 Cas9 오프-타겟 게놈 부위의 PCR 증폭을 위한 주형으로서 사용하였다[118,159] (HEK3, EMX1, FANCF, 및 HEK4 스페이서에 대한 각각 상위 4개의 오프-타겟 부위; 프라이머 서열은 표 4에 열거됨). 이들 게놈 DNA 샘플은 도 41a-41k에 제시된 온-타겟 PE3 편집 활성을 정량화하는 데 사용된 것과 동일하였고; PEgRNA 및 닉킹 sgRNA 서열은 표 3A-3R에 열거된다. 오프-타겟 부위의 PCR 증폭 후에, 앰플리콘을 상기 기재된 바와 같이 일루미나 MiSeq 플랫폼 상에서 서열분석하였다 (HTS 분석). Cas9 뉴클레아제, Cas9 H840A 닉카제, dCas9, 및 PE2-dRT 온-타겟 및 오프-타겟 편집 활성을 결정하기 위해, HEK293T 세포를 750 ng의 편집제 플라스미드 (Cas9 뉴클레아제, Cas9 H840A 닉카제, dCas9, 또는 PE2-dRT), 250 ng의 PEgRNA 또는 sgRNA 플라스미드, 및 1 μL의 리포펙타민 2000으로 형질감염시켰다. 게놈 DNA를 상기 기재된 바와 같이 형질감염 3일 후에 세포로부터 단리하였다. 온-타겟 및 오프-타겟 게놈 유전자좌를 표 4의 프라이머 서열을 사용하여 PCR에 의해 증폭시키고, 일루미나 MiSeq 상에서 서열분석하였다.

[1942] CRISPResso2[178]를 사용하여 HTS 데이터 분석을 수행하였다. Cas9 뉴클레아제, Cas9 H840A 닉카제, 및 dCas9의 편집 효율을 indel을 함유하는 총 서열분석 판독물의 퍼센트로서 정량화하였다. PE3 및 PE3-dRT 오프-타겟의 정량화를 위해, 정렬된 서열분석 판독물을 Cas9 닉 부위에서 개시된 PEgRNA 역전사의 예상 생성물과 일치하는 점 돌연변이, 삽입, 또는 결실에 대해 검사하였다. 샘플 내의 총 판독물 중에서 <0.1% 전체 빈도로 발생하는 단일 뉴클레오티드 변경은 분석에서 제외하였다. ≥ 0.1%의 빈도로 발생하였고 또한 PEgRNA-코딩된 편집과 부분적으로 일치하는 단일 뉴클레오티드 변경을 함유하는 판독물의 경우에는, 변이체가 동일한 스페이서를 함유하지만 상이한 편집을 코딩하는 PEgRNA로 처리된 샘플과 비교하여 유의하게 더 높은 수준으로 발생하였는지 t-검정 (독립표본, 단측, α = 0.5)을 사용하여 결정하였다. 서열분석 오류에서의 차이를 피하기 위해, 동일한 MiSeq 실행 내에서 동시에 서열분석된 샘플들 사이에 비교를 수행하였다. p-값 > 0.05의 기준을 충족시키지 않는 변이체는 제외하였다. 이어서, 오프-타겟 PE3 편집 활성을 상기 기준을 충족시키는 총 서열분석 판독물의 백분율로서 계산하였다.

[1943] Cas9-개시된 HDR을 사용한 HBB E6V 돌연변이를 함유하는 HEK293T 세포주의 생성

[1944] HEK293T 세포를 48-웰 플레이트에 시딩하고, 1.5 μL의 리포펙타민 2000, 300 ng의 Cas9 D10A 닉카제 플라스미드, 100 ng의 sgRNA 플라스미드, 및 200 ng의 100-mer ssDNA 공여자 주형으로 대략 60% 전면생장률로 형질감염시켰다 (표 5). 형질감염 3일 후에, 배지를 신선한 배지로 교체하였다. 형질감염 4일 후에, 세포를 30 μL의 트리플(TrypLE) 용액을 사용하여 해리시키고, 1.5 mL의 배지 중에 현탁시켰다. 단일 세포를 형광-활성화 세포 분류 (FACS) (베크만-쿨터 아스트리오스)에 의해 2개의 96-웰 플레이트의 개별 웰 내로 단리하였다. 대표적인 FACS 분류 실시예에 대해서는 도 53a-53b를 참조한다. 세포를 상기 기재된 바와 같은 게놈 DNA 서열분석 전에 14일 동안 확장시켰다. 단리된 클론 집단 중에서 어떠한 것도 HBB E6V 돌연변이에 대해 동형접합인 것으로 발견되지 않았으므로, 부분적으로 편집된 세포주에서 리포펙션, 분류, 및 성장에 의한 제2 라운드의 편집을 반복

하여 E6V 대립유전자에 대해 동형접합인 세포주를 수득하였다.

[1945] PE3을 사용한 HBB E6V 돌연변이를 함유하는 HEK293T 세포주의 생성

[1946] 항생제의 부재 하에 성장시킨 2.5 x 104개의 HEK293T 세포를 48-웰 폴리-D-리신 코팅된 플레이트 (코닝) 상에 시딩하였다. 시딩 16 내지 24시간 후에, 세포를 제조업체의 프로토콜에 따라 리포펙타민 2000 (써모피셔 사이언티픽) 1 μL 및 PE2-P2A-GFP 플라스미드 750 ng, PEgRNA 플라스미드 250 ng, 및 sgRNA 플라스미드 83 ng으로 대략 70% 전면생장률로 형질감염시켰다. 형질감염 3일 후, 세포를 포스페이트-완충 염수 (깁코)로 세척하고, 트리플 익스프레스(TrypLE Express) (깁코)를 사용하여 해리시켰다. 이어서, 세포를 10% (v/v) FBS (깁코)로 보충된 DMEM 플러스 글루타맥스 (써모피셔 사이언티픽)로 희석하고, 분류 전에 35-μm 세포 스트레이너 (코닝)를 통과시켰다. 유동 세포측정법을 LE-MA900 세포 분류기 (소니(Sony)) 상에서 수행하였다. 분류 15분 전에 세포를 3 nM DAPI (바이오레전드(BioLegend))로 처리하였다. 더블렛 배제에 대해 게이팅한 후, GFP-음성 대조군 세포 집단의 것을 초과하는 GFP 형광을 갖는 단일 DAPI-음성 세포를 10% FBS로 보충된 글루타맥스를 함유하는 사전-냉각된 DMEM으로 충전된 96-웰 편평-바닥 세포 배양 플레이트 (코닝) 내로 분류하였다. 대표적인 FACS 분류 실시예에 대해서는 도 53a-53b를 참조한다. 세포를 10일 동안 배양한 후, 상기 기재된 바와 같이 게놈 DNA 추출하고 HTS에 의해 특징화하였다. HBB에서 E6V 돌연변이에 대해 동형접합인 총 6개의 클론 세포주가 확인되었다.

[1947] PE3을 사용한 HEXA 1278+TATC 삽입을 함유하는 HEK293T 세포주의 생성

[1948] HEXA 1278+TATC 대립유전자를 함유하는 HEK293T 세포를 HBB E6V 세포주의 생성에 대해 상기 기재된 프로토콜에 따라 생성하였고; PEgRNA 및 sgRNA 서열을 표 2A-2C에 도 43a-43h 부표제 하에 열거한다. 형질감염 및 분류 후에, 세포를 10일 동안 배양한 후, 상기 기재된 바와 같이 게놈 DNA 추출하고 HTS에 의해 특징화하였다. 50% HEXA 1278+TATC 대립유전자를 함유하는 2개의 이형접합 세포주를 단리하고, 100% HEXA 1278+TATC 대립유전자를 함유하는 2개의 동형접합 세포주를 회수하였다.

[1949] 세포 생존율 검정

[1950] HEK293T 세포를 48-웰 플레이트에 시딩하고, 상기 기재된 바와 같이 750 ng의 편집제 플라스미드 (PE3, PE3 R110S K103L, Cas9 H840A 닉카제, 또는 dCas9), 250 ng의 HEK3-표적화 PEgRNA 플라스미드, 및 1 μL의 리포펙타민 2000으로 대략 70% 전면생장률로 형질감염시켰다. 세포 생존율을 셀타이터-글로 2.0 검정 (프로메가)을 사용하여 제조업체의 프로토콜에 따라 형질감염후 24시간마다 3일 동안 측정하였다. 발광을 96-웰 편평-바닥 폴리스티렌 마이크로플레이트 (코닝)에서 M1000 프로 마이크로플레이트 판독기 (테칸)를 사용하여 1-초 통합 시간으로 측정하였다.

[1951] 렌티바이러스 생산

[1952] 렌티바이러스를 이전에 기재된 바와 같이 생산하였다[206]. 신속하게 분열 중인 HEK293T 세포 (ATCC; 미국 버지니아주 마나사스)의 T-75 플라스크를 제조업체의 지시에 따라 퓨진 에이치디(FuGENE HD) (프로메가, 미국 위스콘신주 매디슨)를 사용하여 인테인-분할 PE2 편집제를 보유하는 변형된 렌티CRISPR_v2 게놈과 조합된 렌티바이러스 생산 헬퍼 플라스미드 pVSV-G 및 psPAX2로 형질감염시켰다. 4개의 분할-인테인 편집제 구축물을 설계하였다: 1) Npu N-인테인, 자기-절단 P2A 펩티드, 및 GFP-KASH에 융합된 Cas9 H840A 닉카제의 N-말단 부분 (1-573) 및 U6-PEgRNA 발현 카세트를 코딩하는 바이러스 게놈; 2) PE2의 C-말단 나머지에 융합된 Npu C-인테인을 코딩하는 바이러스 게놈; 3) Cas9 대조군에 대한 Cas9의 C-말단 나머지에 융합된 Npu C-인테인을 코딩하는 바이러스 게놈; 및 4) DNMT1에 대한 닉킹 sgRNA. 분할-인테인은 트랜스 스플라이싱을 매개하여 PE2 또는 Cas9의 2개의 절반을 연결하는 한편, P2A GFP-KASH는 핵 막-국재화된 GFP의 공동-번역 생산을 가능하게 한다. 48시간 후, 상청액을 수집하고, 500 g에서 5분 동안 원심분리하여 세포 파편을 제거하고, 0.45 μm 필터를 사용하여 여과하였다. 여과된 상청액을 PEG-it 바이러스 침전 용액 (시스템 바이오사이언시스(System Biosciences), 미국 캘리포니아주 팔로 알토)을 제조업체의 지침에 따라 사용하여 농축시켰다. 생성된 펠릿을 Opti-MEM (써모피셔 사이언티픽, 미국 매사추세츠주 월섬) 중에 원래 배지 부피의 1%를 사용하여 재현탁시켰다. 재현탁된 펠릿을 급속-동결시키고, 사용시까지 -80℃에서 저장하였다.

[1953] 마우스 1차 피질 뉴런 절개 및 배양

[1954] E18.5 해리된 피질 배양물을 시한-임신 C57BL/6 마우스 (찰스 리버)로부터 수거하였다. CO2에 의한 안락사에 이은 단두 후, 임신한 마우스로부터 배아를 수거하였다. 피질 캡을 페니실린/스트렙토마이신 (라이프 테크놀로

지스(Life Technologies))으로 보충된 빙냉 하이버네이트-E에서 해부하였다. 빙냉 하이버네이트-E로 헹군 후, 조직을 37℃에서 8분 동안 파파인/DNase (워싱턴/시그마) 중에서 소화시켰다. 조직을 DNase가 보충된 NBActiv4 (브레인비츠(BrainBits))에서 연화처리하였다. 세포를 카운팅하고, 24-웰 플레이트에 웰당 100,000개 세포로 플레이팅하였다. 배지의 절반을 1주에 2회 교체하였다.

[1955] 1차 뉴런에서의 프라임 편집 및 핵 단리

[1956] DIV 1에서, 15 μL의 렌티바이러스를 10:10:1 비의 N-말단:C-말단:닉킹 sgRNA로 첨가하였다. DIV 14에서, 뉴런 핵을 제조업체의 프로토콜에 따라 EZ-PREP 완충제 (시그마 D8938)를 사용하여 단리하였다. 모든 단계는 얼음 상에서 또는 4℃에서 수행하였다. 해리된 배양물로부터 배지를 제거하고, 배양물을 빙냉 PBS로 세척하였다. PBS를 흡인하고, 200 μL EZ-PREP 용액으로 대체하였다. 얼음 상에서 5-분 인큐베이션한 후, EZ-PREP를 웰의 표면을 가로질러 피펫팅하여 나머지 세포를 축출하였다. 샘플을 500 g에서 5분 동안 원심분리하고, 상청액을 제거하였다. 샘플을 200 μL EZ-PREP로 세척하고, 500 g에서 5분 동안 다시 원심분리하였다. 샘플을 1xPBS 중 100 μg/mL BSA 및 3.33 μM 비브란트 다이사이클 루비(Vybrant DyeCycle Ruby) (써모피셔)로 이루어진 200 μL 빙냉 핵 현탁 완충제 (NSB) 중에 부드럽게 피펫팅하여 재현탁시킨 다음, 500 g에서 5분 동안 원심분리하였다. 상청액을 제거하고, 핵을 100 μL NSB 중에 재현탁시키고, 브로드 인스티튜트 유동 세포측정 설비에서 모플로 아스트리오스(MoFlo Astrios) (베크만 쿨터)를 사용하여 100 μL 아젠코트 디엔어드밴스 용해 완충제 내로 분류하였다. 게놈 DNA를 제조업체 아젠코트 디엔어드밴스 지침에 따라 정제하였다.

[1957] RNA-서열분석 및 데이터 분석

[1958] HEK293T 세포를 PRNP-표적화 또는 HEXA-표적화 PEgRNA 및 PE2, PE2-dRT, 또는 Cas9 H840A 닉카제로 공동-형질감염시켰다. 형질감염 72시간 후에, 총 RNA를 세포로부터 트리졸(TRIzol) 시약 (써모피셔)을 사용하여 수거하고, 온-칼럼 DNaseI 처리를 포함하는 RN이지 미니 키트 (퀴아젠)로 정제하였다. 리보솜을 총 RNA로부터 TruSeq 가닥 총 RNA 라이브러리 프렙 키트 (일루미나)의 rRNA 제거 프로토콜을 사용하여 고갈시키고, 후속적으로 RNA클린 XP 비드 (베크만 쿨터)로 세척하였다. 서열분석 라이브러리를 제조업체의 프로토콜에 따라 스마터 프렙엑스 아폴로(SMARTer PrepX Apollo) NGS 라이브러리 프렙 시스템 (타카라(Takara)) 상에서 리보-고갈된 RNA를 사용하여 제조하였다. 생성된 라이브러리를 2200 테이프스테이션(2200 TapeStation) (애질런트 테크놀로지스) 상에서 가시화하고, 큐비트 dsDNA HS 검정 (써모피셔)을 사용하여 정규화하고, NextSeq 550 상에서 고출력 v2 유동 셀 (일루미나)을 사용하여 75-bp 쌍형성된-단부 판독물로서 서열분석하였다. bcl2fastq2 버전 2.20으로 Fastq 파일을 생성하고, TrimGalore 버전 0.6.2 (github.com/FelixKrueger/TrimGalore)를 사용하여 트리밍하여 저품질 염기, 비쌍형성 서열, 및 어댑터 서열을 제거하였다. 트리밍된 판독물을 RSEM 버전 1.3.1[207]을 사용하여 맞춤 Cas9 H840A 유전자 입력을 갖는 호모 사피엔스 게놈 조립 GRCh[148]에 대해 정렬하였다. 림마-붐(limma-voom)[208] 패키지를 사용하여 유전자 발현 수준을 정규화하고, 배치 효과 교정 하에 차등 발현 분석을 수행하였다. FDR-교정된 p-값 <0.05 및 배수-변화 > 2 컷오프를 갖는 차등 발현된 유전자를 호출하고, 결과를 R에서 가시화하였다.

[1959] 클린바 분석

[1960] 클린바 변이체 요약본을 NCBI (2019년 7월 15일 엑세스)로부터 다운로드하고, 그 안에 담긴 정보를 모든 하류 분석에 사용하였다. 모든 보고된 변이체의 목록을 대립유전자 ID에 의해 필터링하여 중복을 제거하고, 임상 유의성에 의해 필터링하여 분석을 병원성 변이체로 제한하였다. 삽입, 결실 등인 병원성 변이체의 분율을 계산하기 위해 병원성 변이체의 목록을 변이체 유형에 의해 순차적으로 필터링하였다. 보고된 참조 대립유전자 및 대안적 대립유전자에 기초하여 단일 뉴클레오티드 변이체 (SNV)를 2개의 카테고리 (전이 및 전환)로 분리하였다. 참조 대립유전자 또는 대안적 대립유전자가 보고되지 않은 SNV는 분석에서 제외하였다.

[1961] 보고된 삽입, 결실, 및 중복의 길이는 참조/대안적 대립유전자, 변이체 시작/정지 위치, 또는 변이체 명칭에서의 적절한 확인 정보를 사용하여 계산하였다. 상기 정보 중 임의의 것이 보고되지 않은 변이체는 분석에서 제외하였다. 보고된 indel (참조 게놈에 비해 삽입 및 결실 둘 다를 포함하는 단일 변이체)의 길이는 참조 대립유전자 및 대안적 대립유전자 사이의 최상의 쌍별 정렬에서 미스매치 또는 갭의 수를 결정함으로써 계산하였다. 변이체 길이의 빈도 분포를 그래프패드 프리즘 8을 사용하여 계산하였다.

[1962] 데이터 이용가능성

[1963] 고처리량 서열분석 데이터는 NCBI 서열 판독 아카이브 데이터베이스에 기탁되어 있다. PE1, PE2/PE3, 및

PEgRNA 발현 벡터를 코딩하는 플라스미드는 애드진으로부터 입수가능할 것이다.

[1964] 코드 이용가능성

[1965] PEgRNA 스캐폴드 삽입을 정량화하는 데 사용된 스크립트를 도 60a-60b에 제공한다.

[1966] 보충 정보: 표 및 서열

[1967] 표 1: HEK3, HEK4, EMX1, 및 FANCF 온-타겟 및 오프-타겟 부위에서의 프라임 편집제, Cas9 뉴클레아제, Cas9 H840A 닉카제, 및 PE2-dRT의 활성. PE2/PE3 편집은 % 프라임 편집으로서 % indel (괄호 안)과 함께 제시된다. 상위 4개의 이전에 특징화된 오프-타겟 부위에서 Cas9, Cas9 H840A 닉카제 (nCas9), 및 PE2-dRT에 대한 % indel이 제시된다[179,180]. sgRNA 및 PEgRNA 서열은 도 42a-42h 표제 하에 표 3A-3R에서 확인할 수 있다. 모든 값은 3회의 독립적인 생물학적 반복의 평균이다.

PE

부위 \ pegRNA	HEK3 (PE3) ---	1	2	3	4	HEK4 (PE2) ---	1	2	3	4	EMX1 (PE3) ---	1	2	3	4	FANCF (PE3) ---	1	2	3	4
온-타겟		44.2 (31.5)	51.2 (8.8)	48.4 (16.5)	48.4 (3.3)		18.2 (0.5)	14.4 (1.5)	9.8 (2.0)	7.9 (2.2)		28.5 (3.5)	14.1 (2.4)	35.7 (3.3)	15.4 (2.5)		56.8 (9.3)	38.4 (16.7)	42.8 (13.6)	47.6 (12.0)
오프-타겟 1		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)		<0.1 (<0.1)	0.4 (<0.1)	<0.1 (<0.1)	0.4 (<0.1)		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)		<0.1 (0.1)	0.6 (0.1)	<0.1 (0.1)	<0.1 (0.1)
오프-타겟 2		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)		<0.1 (<0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (0.1)	<0.1 (<0.1)
오프-타겟 3		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)		0.2 (0.5)	0.6 (1.5)	19.2 (0.5)	7.9 (3.5)		<0.1 (0.3)	<0.1 (0.3)	<0.1 (0.3)	<0.1 (0.3)		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.11)
오프-타겟 4		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.1)	<0.1 (<0.1)		<0.1 (<0.1)	<0.1 (0.2)	<0.1 (0.2)	<0.1 (<0.1)		<0.1 (0.1)	<0.1 (0.1)	<0.1 (0.2)	<0.1 (0.1)		<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.11)

Cas9

부위 \ pegRNA	HEK3 sgRNA	1	2	3	4	HEK4 sgRNA	1	2	3	4	EMX1 sgRNA	1	2	3	4	FANCF sgRNA	1	2	3	4
온-타겟	91.8	87.5	89.2	89.1	86.8	71.8	68.8	72.8	72.8	70.9	85.6	79.7	79.6	76.6	76.0	78.7	55.9	58.3	51.8	52.8
오프-타겟 1	17.2	1.9	5.5	5.2	1.8	54.2	39.5	48.4	49.7	49.2	81.1	63.5	48.1	53.0	59.5	12.6	1.9	1.9	1.7	1.7
오프-타겟 2	38.0	6.5	12.6	11.8	4.7	42.5	19.5	29.4	27.3	30.3	58.3	12.0	6.0	8.2	12.9	1.1	0.2	0.2	0.2	0.1
오프-타겟 3	8.8	0.6	1.7	1.5	0.5	98.1	98.9	97.3	97.6	97.5	14.8	4.2	3.1	3.8	4.8	2.4	0.2	<0.1	0.2	0.2
오프-타겟 4	0.3	<0.1	<0.1	0.1	<0.1	46.3	16.9	28.0	27.5	29.7	39.5	1.3	0.9	0.6	1.3	1.0	0.2	0.2	0.2	0.2

nCas9

부위 \ pegRNA	HEK3 ---	1	2	3	4	HEK4 ---	1	2	3	4	EMX1 ---	1	2	3	4	FANCF ---	1	2	3	4
오프-타겟 1		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1		0.1	0.1	<0.1	0.1		<0.1	<0.1	<0.1	<0.1
오프-타겟 2		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	< 0.1	<0.1		0.3	0.1	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1
오프-타겟 3		<0.1	<0.1	<0.1	<0.1		0.3	0.5	0.7	0.7		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1
오프-타겟 4		<0.1	<0.1	<0.1	<0.1		<0.1	0.1	<0.1	<0.1		1.6	0.1	0.1	0.4		<0.1	<0.1	<0.1	<0.1

PE2-dRT

부위 \ pegRNA	HEK3 ---	1	2	3	4	HEK4 ---	1	2	3	4	EMX1 ---	1	2	3	4	FANCF ---	1	2	3	4
오프-타겟 1		<0.1	<0.1	<0.1	<0.1		0.1	0.2	<0.1	<0.1		0.1	0.1	0.1	0.1		<0.1	<0.1	<0.1	0.1
오프-타겟 2		<0.1	<0.1	<0.1	<0.1		<0.1	0.1	<0.1	<0.1		<0.1	0.3	<0.1	<0.1		<0.1	<0.1	<0.1	<0.1
오프-타겟 3		<0.1	<0.1	<0.1	<0.1		<0.1	0.1	1.4	0.9		<0.1	<0.1	<0.1	<0.1		<0.1	<0.1	<0.1	0 .1
오프-타겟 4		<0.1	<0.1	<0.1	<0.1		0.1	0.1	0.1	0.2		0.9	0.1	0.1	0.2		<0.1	<0.1	<0.1	0 .1

[1968]

[1969] 표 2A-2C: 시험관내 실험에 사용된 DNA 올리고뉴클레오티드, PEgRNA, 및 sgRNA의 서열.

[1970] 표 2A: DNA 올리고뉴클레오티드

올리고뉴클레오티드	서열
AVA023	5CY5-CCTGGGTCAATCCTTGGGGCCCAGACTGAGCACG (SEQ ID NO: 374)
AVA024	5CY5-CCTGGGTCAATCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGG (SEQ ID NO: 375)
AVA025	5PHOS-CCTTTCCTCTGCCATCACGTGCTCAGTCTGGGCCCCAAGGATTGACCCAGG (SEQ ID NO: 376)
AVA026	5PHOS-TGATGGCAGAGGAAAGG (SEQ ID NO: 377)
AVA037	GCAGGCTTTAAAGGAACCAATTC (SEQ ID NO: 378)
AVA110	GCAGGCTTTAAAGGAACCAATTCCCTGGGTCAATCCTTGGGGC (SEQ ID NO: 379)
AVA122	CTCTGGAGGATCTAGCGGAG (SEQ ID NO: 380)
AVA134	CTCTGGAGGATCTAGCGGAGTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 381)
AVA135	CTCTGGAGGATCTAGCGGAGCCCCCCCCCCCCCC (SEQ ID NO: 382)

[1971]

[1972] 표 2B: 5'-연장된 PEgRNA

PEGRNA	스페이서 서열	5′ 연장부 서열	링커 길이 (NT)	PBS 길이 (NT)	RT 주형 길이 (NT)
PEGRNA 1	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 383)	GGCTAACCGTGCCATT TGATCAGGTCA (SEQ ID NO: 429)	15	5	7
PEGRNA 2	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 384)	GGCTAACCGTGCAAA TTAACAAACTAA(SEQ ID NO: 430)	15	5	7
PEGRNA 3	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 385)	GGCCATCTCGTGCAA ATTAACAAACTAA(SE Q ID NO: 431)	15	5	8
PEGRNA 4	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 386)	GGTCCTCTGCCATCTC GTGCAAATTAACAAA CTAA(SEQ ID NO: 432)	15	5	15
PEGRNA 5	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 387)	GGCTTCCTTTCCTCTG CCATCTCGTGCAAATT AACAAACTAA(SEQ ID NO: 433)	15	5	22
5′-PEGRNA_RT_7_A	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 388)	GGCTAACCGTGCCATT TGATCAGGTCA(SEQ ID NO: 434)	15	5	7
5′-PEGRNA_RT_7_B	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 389)	GGCTAACCGTGCAAA TTAACAAACTAA(SEQ ID NO: 435)	15	5	7
5′-PEGRNA_RT_8	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 390)	GGCCATCTCGTGCAA ATTAACAAACTAA(SE Q ID NO: 436)	15	5	8
5′-PEGRNA_RT_15	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 391)	GGTCCTCTGCCATCTC GTGCAAATTAACAAA CTAA(SEQ ID NO: 437)	15	5	15
5′-PEGRNA_RT_22	GGCCCAGACTG AGCACGTGA (SEQ ID NO: 392)	GGCTTCCTTTCCTCTG CCATCTCGTGCAAATT AACAAACTAA(SEQ ID NO: 438)	15	5	22

[1973]

[1974] 표 2C: 3'-연장된 PEgRNA

PEGRNA	스페이서 서열	3′ 연장부 서열	PBS 길이 (NT)	RT 주형 길이 (NT)
3′-PEGRNA_10	GGCCCAGACTGAG CACGTGA(SEQ ID NO: 506)	TCTGCCATCTCG TGCTC (SEQ ID NO: 439)	7	10
3′-PEGRNA_효모_TTOA	GGCCCAGACTGAG CACGTGA(SEQ ID NO: 507)	TCTGCCATCTCG TGCTC(SEQ ID NO: 440)	7	10
3′-PEGRNA_효모_+1AINS	GGCCCAGACTGAG CACGTGA(SEQ ID NO: 508)	TCTGCCATCATC GTGCTC(SEQ ID NO: 441)	7	11
3′-PEGRNA_효모_+1TDEL	GGCCCAGACTGAG CACGTGA(SEQ ID NO: 509)	TCTGCCATCCGT GCTC(SEQ ID NO: 442)	7	9

[1975]

[1976] 표 3A-3R: 포유동물 세포 실험에 사용된 PEgRNA 및 sgRNA의 서열. 모든 서열은 5'에서 3' 배향으로 제시된다. PEgRNA를 구축하기 위해, 하기 열거된 스페이서 서열을 sgRNA 스캐폴드의 5' 단부에 부가하고, 프라이머 결합 부위 및 RT 주형을 함유하는 하기 열거된 3' 연장부를 sgRNA 스캐폴드의 3' 단부에 부가하였다. sgRNA 스캐폴드 서열은

[1977] GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (서열식별번호: 131)이다.

[1978] 표 3A: 도 39a-39d PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 2890-2996)	3′ 연장부 (SEQ ID NO:2997-3103)	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_2B-C_8	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCA	8	10
HEK3_2B-C_9	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAG	9	10
HEK3_2B-C_10	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGT	10	10
HEK3_2B-C_11	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTC	11	10
HEK3_2B-C_12	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCT	12	10
HEK3_2B-C_13	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_2B-C_14	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGG	14	10
HEK3_2B-C_15	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGGG	15	10
HEK3_2C_16	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGGGC	16	10
HEK3_2C_17	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTGGGCC	17	10
EMX1_2C_9	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGC	9	13
EMX1_2C_10	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCT	10	13
EMX1_2C_11	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTC	11	13
EMX1_2C_12	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCG	12	13
EMX1_2C_13	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGG	13	13
EMX1_2C_14	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGA	14	13
EMX1_2C_15	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGAC	15	13
EMX1_2C_16	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGACT	16	13
EMX1_2C_17	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGACTC	17	13
FANCF_2C_8	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGA	8	17

[1979]

FANCF_2C_9	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAA	9	17
FANCF_2C_10	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAG	10	17
FANCF_2C_11	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGG	11	17
FANCF_2C_12	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGG	12	17
FANCF_2C_13	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	17
FANCF_2C_14	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGAT	14	17
FANCF_2C_15	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATT	15	17
FANCF_2C_16	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATTC	16	17
FANCF_2C_17	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATTCC	17	17
RNF2_2C_9	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACT	9	11
RNF2_2C_10	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTA	10	11
RNF2_2C_11	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAA	11	11
RNF2_2C_12	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAG	12	11
RNF2_2C_13	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGA	13	11
RNF2_2C_14	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGAT	14	11
RNF2_2C_15	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATG	15	11
RNF2_2C_16	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATGA	16	11
RNF2_2C_17	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATGAC	17	11
HEK4_2C_7	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAG	7	13
HEK4_2C_8	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGC	8	13
HEK4_2C_9	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCC	9	13
HEK4_2C_10	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCG	10	13
HEK4_2C_11	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGC	11	13
HEK4_2C_12	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCA	12	13

[1980]

[1981]

HEK4_2C_13	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCAG	13	13
HEK4_2C_14	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCAGT	14	13
HEK4_2C_15	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGCAGTG	15	13
HEK3_2C_1TDEL	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	10
HEK3_2C_1AINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCATCGTGCTCAGTCTG	13	10
HEK3_2C_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	10
HEK3_2D_10	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_2D_11	GGCCCAGACTGAGCACGTGA	CTCTGCCATCTCGTGCTCAGTCTG	13	11
HEK3_2D_12	GGCCCAGACTGAGCACGTGA	CCTCTGCCATCTCGTGCTCAGTCTG	13	12
HEK3_2D_13	GGCCCAGACTGAGCACGTGA	TCCTCTGCCATCTCGTGCTCAGTCTG	13	13
HEK3_2D_14	GGCCCAGACTGAGCACGTGA	TTCCTCTGCCATCTCGTGCTCAGTCTG	13	14
HEK3_2D_15	GGCCCAGACTGAGCACGTGA	TTTCCTCTGCCATCTCGTGCTCAGTCTG	13	15
HEK3_2D_16	GGCCCAGACTGAGCACGTGA	CTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	16
HEK3_2D_17	GGCCCAGACTGAGCACGTGA	CCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	17
HEK3_2D_18	GGCCCAGACTGAGCACGTGA	TCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	18
HEK3_2D_19	GGCCCAGACTGAGCACGTGA	TTCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	19
HEK3_2D_20	GGCCCAGACTGAGCACGTGA	CTTCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	20
EMX1_2D_10	GAGTCCGAGCAGAAGAAGAA	GGAGCCCTTGTTCTTCTGCTCGG	13	10
EMX1_2D_11	GAGTCCGAGCAGAAGAAGAA	GGGAGCCCTTGTTCTTCTGCTCGG	13	11
EMX1_2D_12	GAGTCCGAGCAGAAGAAGAA	TGGGAGCCCTTGTTCTTCTGCTCGG	13	12
EMX1_2D_13	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCCCTTGTTCTTCTGCTCGG	13	13
EMX1_2D_14	GAGTCCGAGCAGAAGAAGAA	GATGGGAGCCCTTGTTCTTCTGCTCGG	13	14
EMX1_2D_15	GAGTCCGAGCAGAAGAAGAA	TGATGGGAGCCCTTGTTCTTCTGCTCGG	13	15
EMX1_2D_16	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	16

EMX1_2D_17	GAGTCCGAGCAGAAGAAGAA	TGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	17
EMX1_2D_18	GAGTCCGAGCAGAAGAAGAA	ATGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	18
EMX1_2D_19	GAGTCCGAGCAGAAGAAGAA	GATGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	19
EMX1_2D_20	GAGTCCGAGCAGAAGAAGAA	TGATGTGATGGGAGCCCTTGTTCTTCTGCTCGG	13	20
FANCF_2D_10	GGAATCCCTTCTGCAGCACC	CGATCAAGGTGCTGCAGAAGGGA	13	10
FANCF_2D_11	GGAATCCCTTCTGCAGCACC	GCGATCAAGGTGCTGCAGAAGGGA	13	11
FANCF_2D_12	GGAATCCCTTCTGCAGCACC	AGCGATCAAGGTGCTGCAGAAGGGA	13	12
FANCF_2D_13	GGAATCCCTTCTGCAGCACC	AAGCGATCAAGGTGCTGCAGAAGGGA	13	13
FANCF_2D_14	GGAATCCCTTCTGCAGCACC	AAAGCGATCAAGGTGCTGCAGAAGGGA	13	14
FANCF_2D_15	GGAATCCCTTCTGCAGCACC	AAAAGCGATCAAGGTGCTGCAGAAGGGA	13	15
FANCF_2D_16	GGAATCCCTTCTGCAGCACC	GAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	16
FANCF_2D_17	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	17
FANCF_2D_18	GGAATCCCTTCTGCAGCACC	CGGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	18
FANCF_2D_19	GGAATCCCTTCTGCAGCACC	TCGGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	19
FANCF_2D_20	GGAATCCCTTCTGCAGCACC	CTCGGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	20
RNF2_2D_10	GTCATCTTAGTCATTACCTG	AACACCTCATGTAATGACTAAGATG	15	10
RNF2_2D_11	GTCATCTTAGTCATTACCTG	GAACACCTCATGTAATGACTAAGATG	15	11
RNF2_2D_12	GTCATCTTAGTCATTACCTG	CGAACACCTCATGTAATGACTAAGATG	15	12
RNF2_2D_13	GTCATCTTAGTCATTACCTG	ACGAACACCTCATGTAATGACTAAGATG	15	13
RNF2_2D_14	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
RNF2_2D_15	GTCATCTTAGTCATTACCTG	CAACGAACACCTCATGTAATGACTAAGATG	15	15
RNF2_2D_16	GTCATCTTAGTCATTACCTG	ACAACGAACACCTCATGTAATGACTAAGATG	15	16
RNF2_2D_17	GTCATCTTAGTCATTACCTG	TACAACGAACACCTCATGTAATGACTAAGATG	15	17
RNF2_2D_18	GTCATCTTAGTCATTACCTG	TTACAACGAACACCTCATGTAATGACTAAGATG	15	18

[1982]

RNF2_2D_19	GTCATCTTAGTCATTACCTG	GTTACAACGAACACCTCATGTAATGACTAAGATG	15	19
RNF2_2D_20	GTCATCTTAGTCATTACCTG	AGTTACAACGAACACCTCATGTAATGACTAAGATG	15	20
HEK4_2D_7	GGCACTGCGGCTGGAGGTGG	ACCCCAACCTCCAGCCGC	11	7
HEK4_2D_8	GGCACTGCGGCTGGAGGTGG	AACCCCAACCTCCAGCCGC	11	8
HEK4_2D_9	GGCACTGCGGCTGGAGGTGG	TAACCCCAACCTCCAGCCGC	11	9
HEK4_2D_10	GGCACTGCGGCTGGAGGTGG	TTAACCCCAACCTCCAGCCGC	11	10
HEK4_2D_11	GGCACTGCGGCTGGAGGTGG	TTTAACCCCAACCTCCAGCCGC	11	11
HEK4_2D_12	GGCACTGCGGCTGGAGGTGG	CTTTAACCCCAACCTCCAGCCGC	11	12
HEK4_2D_13	GGCACTGCGGCTGGAGGTGG	GCTTTAACCCCAACCTCCAGCCGC	11	13
HEK4_2D_14	GGCACTGCGGCTGGAGGTGG	CGCTTTAACCCCAACCTCCAGCCGC	11	14
HEK4_2D_15	GGCACTGCGGCTGGAGGTGG	CCGCTTTAACCCCAACCTCCAGCCGC	11	15
HEK4_2D_16	GGCACTGCGGCTGGAGGTGG	TCCGCTTTAACCCCAACCTCCAGCCGC	11	16
HEK4_2D_17	GGCACTGCGGCTGGAGGTGG	CTCCGCTTTAACCCCAACCTCCAGCCGC	11	17
HEK4_2D_18	GGCACTGCGGCTGGAGGTGG	CTCCGCTTTAACCCCAACCTCCAGCCGC	11	18
HEK4_2D_19	GGCACTGCGGCTGGAGGTGG	CTCCGCTTTAACCCCAACCTCCAGCCGC	11	19

[1983]

[1984] 표 3B: 도 40a-40c PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 3104-3112)	3′ 연장부 (SEQ ID NO: 3113-3121)	PBS 길이 (NT)	RT 주형 길이 (NT)
RNF2_3B	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
EMX1_3B	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCACTTCTTCTTCTGCTCGGAC	15	13
FANCF_3B	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGATT	15	17
HE3_3B	GGCCCAGACTGAGCACGTGA	TCTGCCATGACGTGCTCAGTCTG	13	10
HEK4_3B	GGCACTGCGGCTGGAGGTGG	TTAACCCCAACCTCCAGCC	9	10
RNF2_3C_4ATOC	GTCATCTTAGTCATTACCTG	AACGAACACCGCAGGTAATGACTAAGATG	15	14
RNF2_3C_4ATOG	GTCATCTTAGTCATTACCTG	AACGAACACCCCAGGTAATGACTAAGATG	15	14
FANCF_3C_5GTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGGTGCTGCAGAAGGGA	13	17
FANCF_3C_7ATOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGAGCCAGGTGCTGCAGAAGGGAT	14	17

[1985]

[1986] 표 3C: 도 40a-40c 닉킹 sgRNA 서열

닉킹 SGRNA	스페이서 서열	SEQ ID NO:
RNF2_2B_+41	GTCAACCATTAAGCAAAACAT	**3122**
RNF2_2B_+67	GTCTCAGGCTGTGCAGACAAA	**3123**
EMX1_2B_-116	GGGGCACAGATGAGAAACTC	**3124**
EMX1_2B_-57	GCCGTTTGTACTTTGTCCTC	**3125**
EMX1_2B_+14	GCGCCACCGGTTGATGTGAT	**3126**
EMX1_2B_+27	GCTTCGTGGCAATGCGCCAC	**3127**
EMX1_2B_+53	GACATCGATGTCCTCCCCAT	**3128**
EMX1_2B_+80	GTGGTTGCCCACCCTAGTCAT	**3129**
FANCF_2B_-78	GCGACTCTCTGCGTACTGAT	**3130**
FANCF_2B_-50	GCCCTACTTCCGCTTTCACCT	**3131**
FANCF_2B_-27	GGATTCCATGAGGTGCGCGA	**3132**
FANCF_2B_-17	GCTGCAGAAGGGATTCCATG	**3133**
FANCF_2B_+21	GCTTGAGACCGCCAGAAGCT	**3134**
FANCF_2B_+48	GGGGTCCCAGGTGCTGACGT	**3135**
HEK3_2B_-108	GCAGAAATAGACTAATTGCA	**3136**
HEK3_2B_-38	GGATTGACCCAGGCCAGGGC	**3137**
HEK3_2B_+26	GACGCCCTCTGGAGGAAGCA	**3138**
HEK3_2B_+37	GCTGTCCTGCGACGCCCTC	**3139**
HEK3_2B_+63	GCACATACTAGCCCCTGTCT	**3140**
HEK3_2B_+90	GTCAACCAGTATCCCGGTGC	**3141**
HEK4_2B_-95	TCCCTTCCTTCCACCCAGCC	**3142**
HEK4_2B_-51	CCCTGCCTGTCATCCTGCTT	**3143**
HEK4_2B_-26	GCAGTGCCACCGGGGCGCCG	**3144**
HEK4_2B_+52	GCGGGGGCTCAGAGAGGGCA	**3145**
HEK4_2B_+74	GAGACACACACACAGGCCTGG	**3146**
RNF2_2C_+41	GTCAACCATTAAGCAAAACAT	**3147**
RNF2_2C_4ATOC_+5	GTGAGTTACAACGAACACCGC	**3148**
RNF2_2C_4ATOG_+5	GTGAGTTACAACGAACACCCC	**3149**
FANCF_2C_+48	GGGGTCCCAGGTGCTGACGT	**3150**
FANCF_2C_5GTOT_+7	GAAGCTCGGAAAAGCGATCA	**3151**
FANCF_2C_7ATOC_+7	GAAGCTCGGAAAAGCGAGCC	**3152**
HEK3_2C_+90	GTCAACCAGTATCCCGGTGC	**3153**

[1987]

[1988] 표 3D: 도 41a-41k PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 3154-3304)	3′ 연장부 (SEQ ID NO: 3305-3455)	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_4A_1TTOA	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_4A_1TTOC	GGCCCAGACTGAGCACGTGA	TCTGCCATCGCGTGCTCAGTCTG	13	10
HEK3_4A_1TTOG	GGCCCAGACTGAGCACGTGA	TCTGCCATCCCGTGCTCAGTCTG	13	10
HEK3_4A_2GTOA	GGCCCAGACTGAGCACGTGA	TCTGCCATTACGTGCTCAGTCTG	13	10
HEK3_4A_2GTOC	GGCCCAGACTGAGCACGTGA	TCTGCCATGACGTGCTCAGTCTG	13	10
HEK3_4A_2GTOT	GGCCCAGACTGAGCACGTGA	TCTGCCATAACGTGCTCAGTCTG	13	10
HEK3_4A_3ATOC	GGCCCAGACTGAGCACGTGA	TCTGCCAGCACGTGCTCAGTCTG	13	10
HEK3_4A_3ATOG	GGCCCAGACTGAGCACGTGA	TCTGCCACCACGTGCTCAGTCTG	13	10
HEK3_4A_3ATOT	GGCCCAGACTGAGCACGTGA	TCTGCCAACACGTGCTCAGTCTG	13	10
HEK3_4A_4TTOA	GGCCCAGACTGAGCACGTGA	TCTGCCTTCACGTGCTCAGTCTG	13	10
HEK3_4A_4TTOC	GGCCCAGACTGAGCACGTGA	TCTGCCGTCACGTGCTCAGTCTG	13	10
HEK3_4A_4TTOG	GGCCCAGACTGAGCACGTGA	TCTGCCCTCACGTGCTCAGTCTG	13	10
HEK3_4A_5GTOA	GGCCCAGACTGAGCACGTGA	TCTGCTATCACGTGCTCAGTCTG	13	10
HEK3_4A_5GTOC	GGCCCAGACTGAGCACGTGA	TCTGCGATCACGTGCTCAGTCTG	13	10
HEK3_4A_5GTOT	GGCCCAGACTGAGCACGTGA	TCTGCAATCACGTGCTCAGTCTG	13	10
HEK3_4A_6GTOA	GGCCCAGACTGAGCACGTGA	TCTGTCATCACGTGCTCAGTCTG	13	10
HEK3_4A_6GTOC	GGCCCAGACTGAGCACGTGA	TCTGGCATCACGTGCTCAGTCTG	13	10
HEK3_4A_6GTOT	GGCCCAGACTGAGCACGTGA	TCTGACATCACGTGCTCAGTCTG	13	10
HEK3_4A_7CTOA	GGCCCAGACTGAGCACGTGA	TCTTCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_7CTOG	GGCCCAGACTGAGCACGTGA	TCTCCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_7CTOT	GGCCCAGACTGAGCACGTGA	TCTACCATCACGTGCTCAGTCTG	13	10

[1989]

HEK3_4A_8ATOC	GGCCCAGACTGAGCACGTGA	TCGGCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_8ATOG	GGCCCAGACTGAGCACGTGA	TCCGCCATCACGTGCTCAGTCTG	13	10
HEK3_4A_8ATOT	GGCCCAGACTGAGCACGTGA	TCAGCCATCACGTGCTCAGTCTG	13	10
HEK3_4B_1TTOA	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCTCGTGCTCAGTCTG	13	34
HEK3_4B_12GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTGCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_14ATOT	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTATCCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_17GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTGCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_20GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGGTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_23CTOG	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCACGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_24TTOA	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCTGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_26CTOG	GGCCCAGACTGAGCACGTGA	TGGAGGAACCAGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_30CTOG	GGCCCAGACTGAGCACGTGA	TGGACGAAGCAGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
HEK3_4B_33CTOG	GGCCCAGACTGAGCACGTGA	TCGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTG	13	34
RNF2_4C_1CTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
RNF2_4C_1CTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCACGTAATGACTAAGATG	15	14
RNF2_4C_1CTOT	GTCATCTTAGTCATTACCTG	AACGAACACCTCAAGTAATGACTAAGATG	15	14
RNF2_4C_2TTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCTGGTAATGACTAAGATG	15	14
RNF2_4C_2TTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCCGGTAATGACTAAGATG	15	14
RNF2_4C_3GTOC	GTCATCTTAGTCATTACCTG	AACGAACACCTGAGGTAATGACTAAGATG	15	14

[1990]

RNF2_4C_4ATOC	GTCATCTTAGTCATTACCTG	AACGAACACCGCAGGTAATGACTAAGATG	15	14
RNF2_4C_4ATOT	GTCATCTTAGTCATTACCTG	AACGAACACCACAGGTAATGACTAAGATG	15	14
RNF2_4C_4ATOG	GTCATCTTAGTCATTACCTG	AACGAACACCCCAGGTAATGACTAAGATG	15	14
RNF2_4C_5GTOT	GTCATCTTAGTCATTACCTG	AACGAACACATCAGGTAATGACTAAGATG	15	14
RNF2_4C_6GTOA	GTCATCTTAGTCATTACCTG	AACGAACATCTCAGGTAATGACTAAGATG	15	14
RNF2_4C_7TTOC	GTCATCTTAGTCATTACCTG	AACGAACGCCTCAGGTAATGACTAAGATG	15	14
FANCF_4D_1ATOG	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGGCGCTGCAGAAGGGAT	14	17
FANCF_4D_1ATOT	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGGAGCTGCAGAAGGGAT	14	17
FANCF_4D_2CTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGTTGCTGCAGAAGGGAT	14	17
FANCF_4D_3CTOG	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCACGTGCTGCAGAAGGGAT	14	17
FANCF_4D_3CTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAAGTGCTGCAGAAGGGAT	14	17
FANCF_4D_4TTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCTGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_4TTOG	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCCGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_5GTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCTAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_6GTOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATGCAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_7ATOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGAGCCAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_8TTOC	GGAATCCCTTCTGCAGCACC	GGAAAAGCGGTCCAGGTGCTGCAGAAGGGAT	14	17
FANCF_4D_10GTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGAGATCCAGGTGCTGCAGAAGGGAT	14	17
EMX1_4E_2ATOC	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTGCTTCTTCTGCTCGGA	14	16
EMX1_4E_2ATOT	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTACTTCTTCTGCTCGGA	14	16
EMX1_4E_3ATOG	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCCTCTTCTTCTGCTCGGA	14	16
EMX1_4E_4GTOC	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCGTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_5GTOA	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCTCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_5GTOT	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCACTTCTTCTTCTGCTCGGA	14	16

[1991]

EMX1_4E_7CTOA	GAGTCCGAGCAGAAGAAGAA	GTGATGGGATCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_8TTOA	GAGTCCGAGCAGAAGAAGAA	GTGATGGGTGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_8TTOC	GAGTCCGAGCAGAAGAAGAA	GTGATGGGGGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_8TTOG	GAGTCCGAGCAGAAGAAGAA	GTGATGGGCGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_9CTOG	GAGTCCGAGCAGAAGAAGAA	GTGATGGCAGCCCTTCTTCTTCTGCTCGGA	14	16
EMX1_4E_9CTOT	GAGTCCGAGCAGAAGAAGAA	GTGATGGAAGCCCTTCTTCTTCTGCTCGGA	14	16
RUNX1_4F_1CTOA	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCTCTTCCTCCTGAAAAT	15	15
RUNX1_4F_1CTOG	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCCCTTCCTCCTGAAAAT	15	15
RUNX1_4F_1CTOT	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCACTTCCTCCTGAAAAT	15	15
RUNX1_4F_2GTOA	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATTGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_3ATOC	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCAGCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_3ATOG	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCACCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_3ATOT	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCAACGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_4TTOA	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCTTCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_4TTOC	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCGTCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_4TTOG	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCCTCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_5GTOT	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCAATCGCTTCCTCCTGAAAAT	15	15
RUNX1_4F_6GTOC	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGGCATCGCTTCCTCCTGAAAAT	15	15
VEGFA_4G_1TTOA	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTCTTCTGGCCTGCAGA	13	22
VEGFA_4G_1TTOC	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTCGTCTGGCCTGCAGA	13	22
VEGFA_4G_1TTOG	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTCCTCTGGCCTGCAGA	13	22
VEGFA_4G_2GTOA	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTTATCTGGCCTGCAGA	13	22
VEGFA_4G_3ATOC	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCGCATCTGGCCTGCAGA	13	22
VEGFA_4G_3ATOG	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCCCATCTGGCCTGCAGA	13	22

[1992]

VEGFA_4G_3ATOT	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCACATCTGGCCTGCAGA	13	22
VEGFA_4G_5GTOT	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	22
VEGFA_4G_6GTOC	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGGCCTCATCTGGCCTGCAGA	13	22
VEGFA_4G_7CTOA	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGATCCCTCATCTGGCCTGCAGA	13	22
VEGFA_4G_7CTOT	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAACCCTCATCTGGCCTGCAGA	13	22
VEGFA_4G_9CTOG	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGCAGCCCTCATCTGGCCTGCAGA	13	22
DNMT1_4H_1ATOC	GATTCCTGGTGCCAGAAACA	GTCACCCCTGGTTCTGGCACCAGG	13	11
DNMT1_4H_1ATOG	GATTCCTGGTGCCAGAAACA	GTCACCCCTGCTTCTGGCACCAGG	13	11
DNMT1_4H_2CTOA	GATTCCTGGTGCCAGAAACA	GTCACCCCTTTTTCTGGCACCAGG	13	11
DNMT1_4H_2CTOG	GATTCCTGGTGCCAGAAACA	GTCACCCCTCTTTCTGGCACCAGG	13	11
DNMT1_4H_2CTOT	GATTCCTGGTGCCAGAAACA	GTCACCCCTATTTCTGGCACCAGG	13	11
DNMT1_4H_3ATOT	GATTCCTGGTGCCAGAAACA	GTCACCCCAGTTTCTGGCACCAGG	13	11
DNMT1_4H_4GTOA	GATTCCTGGTGCCAGAAACA	GTCACCCTTGTTTCTGGCACCAGG	13	11
DNMT1_4H_5GTOT	GATTCCTGGTGCCAGAAACA	GTCACCACTGTTTCTGGCACCAGG	13	11
DNMT1_4H_6GTOC	GATTCCTGGTGCCAGAAACA	GTCACGCCTGTTTCTGGCACCAGG	13	11
DNMT1_4H_8TTOA	GATTCCTGGTGCCAGAAACA	GCCCTCCCGTCTCCCCTGTTTCTGGCACCAGG	13	19
DNMT1_4H_8TTOC	GATTCCTGGTGCCAGAAACA	GCCCTCCCGTCGCCCCTGTTTCTGGCACCAGG	13	19
DNMT1_4H_8TTOG	GATTCCTGGTGCCAGAAACA	GCCCTCCCGTCCCCCCTGTTTCTGGCACCAGG	13	19
HEK3_4J_DEL1-5	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCGTGCTCAGTCTG	13	29
HEK3_4J_DEL1-10	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCCGTGCTCAGTCTG	13	24
HEK3_4J_DEL1-15	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCCGTGCTCAGTCTG	13	19
HEK3_4J_DEL1-25	GGCCCAGACTGAGCACGTGA	TGTCCTGCGACGCCCTCTGGAGGAAGCGTGCTCAGTCTG	13	26

[1993]

HEK3_4J_DEL1-30	GGCCCAGACTGAGCACGTGA	TGTCCTGCGACGCCCTCTGGACGTGCTCAGTCTG	13	21
HEK3_4J_DEL1-80	GGCCCAGACTGAGCACGTGA	AGTATCCCGGTGCAGGAGCTCGTGCTCAGTCTG	13	20
HEK3_4I_1AINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCATCGTGCTCAGTCTG	13	11
HEK3_4I_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	13
HEK3_4I_1TDEL	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	9
HEK3_4I_1-3TGADEL	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCACGTGCTCAGTCTG	13	31
RNF2_4I_1TINS	GTCATCTTAGTCATTACCTG	AACGAACACCTCAGAGTAATGACTAAGATG	15	15
RNF2_4I_1GTAINS	GTCATCTTAGTCATTACCTG	AACGAACACCTCAGTACGTAATGACTAAGATG	15	17
RNF2_4I_4ADEL	GTCATCTTAGTCATTACCTG	AACGAACACCCAGGTAATGACTAAGATG	15	13
RNF2_4I_3-5GAGDEL	GTCATCTTAGTCATTACCTG	AACGAACACAGGTAATGACTAAGATG	15	11
FANCF_4I_3CINS	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGGGTGCTGCAGAAGGGAT	14	18
FANCF_4I_4GATINS	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAATCGGTGCTGCAGAAGGGAT	14	20
FANCF_4I_6GDEL	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAGGTGCTGCAGAAGGGAT	14	16
FANCF_4I_5-7GGADEL	GGAATCCCTTCTGCAGCACC	GGAAAAGCGAAGGTGCTGCAGAAGGGAT	14	14
EMX1_4I_6TINS	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCACCTTCTTCTTCTGCTCGGA	14	17
EMX1_4I_1TGCINS	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTTCGCATTCTTCTGCTCGGA	14	19
EMX1_4I_5GDEL	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCTTCTTCTTCTGCTCGGA	14	15
EMX1_4I_4-6GGGDEL	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTCTTCTTCTGCTCGGA	14	13
RUNX1_4I_1CINS	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCGGCTTCCTCCTGAAAAT	15	16
RUNX1_4I_1ATGINS	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATCCATGCTTCCTCCTGAAAAT	15	18
RUNX1_4I_2GDEL	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCATGCTTCCTCCTGAAAAT	15	14
RUNX1_4I_2-4GATDEL	GCATTTTCAGGAGGAAGCGA	TGTCTGAAGCCGCTTCCTCCTGAAAAT	15	12

[1994]

VEGFA_4I_4CINS	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCGCTCATCTGGCCTGCAGA	13	23
VEGFA_4I_2ACAINS	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCTTGTCATCTGGCCTGCAGA	13	25
VEGFA_4I_3ADEL	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCCCATCTGGCCTGCAGA	13	21
VEGFA_4I_2-4GAGDEL	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCCATCTGGCCTGCAGA	13	19
DNMT1_4I_4CINS	GATTCCTGGTGCCAGAAACA	TCCCGTCACCCGCTGTTTCTGGCACCAGG	13	16
DNMT1_4I_1TCAINS	GATTCCTGGTGCCAGAAACA	TCCCGTCACCCCTGTGATTTCTGGCACCAGG	13	18
DNMT1_4I_3ADEL	GATTCCTGGTGCCAGAAACA	TCCCGTCACCCCGTTTCTGGCACCAGG	13	14
DNMT1_4I_3-5AGGDEL	GATTCCTGGTGCCAGAAACA	TCCCGTCACCGTTTCTGGCACCAGG	13	12
HEK3_4K_1CTTINS_5GDEL	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCATCAAAGCGTGCTCAGTCTG	13	36
HEK3_4K_1CTTINS_2GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATGAAAGCGTGCTCAGTCTG	13	37
HEK3_4K_3TDEL_5GTOC	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCGATCCGTGCTCAGTCTG	13	33
HEK3_4K_3GTOC_6GTOT	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGACATGACGTGCTCAGTCTG	13	34
RNF2_4K_2AAINS_3-4GADEL	GTCATCTTAGTCATTACCTG	AACGAACACCATTGGTAATGACTAAGATG	15	14
RNF2_4K_1AINS_5GTOC	GTCATCTTAGTCATTACCTG	AACGAACACGTCAGTGTAATGACTAAGATG	15	15
RNF2_4K_1-2CTDEL_6GTOT	GTCATCTTAGTCATTACCTG	AACGAACAACTCGTAATGACTAAGATG	15	12
RNF2_4K_1CTOA_5GTOT	GTCATCTTAGTCATTACCTG	AACGAACACATCATGTAATGACTAAGATG	15	14
FANCF_4K_1TINS_3-4TGDEL	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCGGTAGCTGCAGAAGGGAT	14	16

[1995]

FANCF_4K_1TINS_6GTOA	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATTCAGGTAGCTGCAGAAGGGAT	14	18
FANCF_4K_2CDEL_5GTOT	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCAAGTGCTGCAGAAGGGAT	14	16

[1996]

[1997] 표 3E: 도 41a-41k 닉킹 sgRNA 서열

닉킹 SGRNA	스페이서 서열 (SEQ ID NO: 3456-3463)
HEK3_4A_+90	GTCAACCAGTATCCCGGTGC
HEK3_4B_+90	GTCAACCAGTATCCCGGTGC
RNF2_4C_+41	GTCAACCATTAAGCAAAACAT
FANCF_4D_+48	GGGGTCCCAGGTGCTGACGT
EMX1_4E_+53	GACATCGATGTCCTCCCCAT
RUNX1_4F_+38	GATGAAGCACTGTGGGTACGA
VEGFA_4G_+57	GATGTACAGAGAGCCCAGGGC
DNMT1_4H_+49	GCCCTTCAGCTAAAATAAAGG

[1998]

[1999] 표 3F: 도 42a-42h PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 3464-3478)	3′ 연장부 (SEQ ID NO: 3479-3493)	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_5A_C3	GGCCCAGACTGAGCACGTGA	TCTGTCATCACGTGCTCAGTCTG	13	10
HEK3_5A_C4	GGCCCAGACTGAGCACGTGA	TCTGCTATCACGTGCTCAGTCTG	13	10
HEK3_5A_C7	GGCCCAGACTGAGCACGTGA	TCTGCCATTACGTGCTCAGTCTG	13	10
FANCF_5A_C3	GGAATCCCTTCTGCAGCACC	GGAAAAGTGATCCAGGTGCTGCAGAAGGGAT	14	17
FANCF_5A_C7	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATTCAGGTGCTGCAGAAGGGAT	14	17
FANCF_5A_C8	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCTAGGTGCTGCAGAAGGGAT	14	17
EMX1_5A_C5	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTCCTTCTTCTTCTGCTCGGA	14	16
EMX1_5A_C6	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCTCTTCTTCTTCTGCTCGGA	14	16
EMX1_5A_C7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCTTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C5_6	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTCTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C5_7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTCTTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C6_7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCTTTTCTTCTTCTGCTCGGA	14	16
EMX1_5C_C5_6_7	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTTTTCTTCTTCTGCTCGGA	14	16
HEK3_5D_A5	GGCCCAGACTGAGCACGTGA	TCTGCCGTCACGTGCTCAGTCTG	13	10
HEK3_5D_A8	GGCCCAGACTGAGCACGTGA	TCTGCCATCGCGTGCTCAGTCTG	13	10

[2000]

[2001] 표 3G: 도 42a-42h 닉킹 sgRNA

닉킹 SGRNA	가능한 스페이서 서열 (SEQ ID NO: 616-618)	가능한 스페이서 서열(SEQ ID NO: 619-621)
HEK3_5A-F_+90	GTCAACCAGTATCCCGGTGC	GTCAACCAGTATCCCGGTGC
FANCF_5A-F_+48	GGGGTCCCAGGTGCTGACGT	GATGTACAGAGAGCCCAGGGC
EMX1_5A-F_+57	GATGTACAGAGAGCCCAGGGC	GGGGTCCCAGGTGCTGACGT

[2002]

[2003] 표 3H: 도 42a-42h 염기 편집 sgRNA

염기 편집 SGRNA	스페이서 서열
HEK3_5A-F_BE	GTGCCATCACGTGCTCAGTCT (SEQ ID NO: 455)
FANCF_5A-F_BE	GAGCGATCCAGGTGCTGCAGA (SEQ ID NO: 456)
EMX1_5A-F_BE	GGAGCCCTTCTTCTTCTGCT (SEQ ID NO: 457)

[2004]

[2005] 표 3I: 도 42a-42h 온-타겟 sgRNA

온-타겟 SGRNA	스페이서 서열
HEK3_5G	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 510)
HEK4_5G	GGCACTGCGGCTGGAGGTGG (SEQ ID NO: 511)
EMX1_5G	GAGTCCGAGCAGAAGAAGAA (SEQ ID NO: 512)
FANCF_5G	GGAATCCCTTCTGCAGCACC (SEQ ID NO: 513)

[2006]

[2007] 표 3J: 도 42a-42h 온-타겟 PEgRNA

온-타겟 PEGRNA	스페이서 서열 (SEQ ID NO: 663-677)	3′ 연장부 (SEQ ID NO: 678-692)	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_5G-H_PEGRNA_1	GGCCCAGACTGAGCACGTGA	TCTGCCATCTCGTGCTCAGTCTG	13	10
HEK3_5G-H_PEGRNA_2	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	13
HEK3_5G-H_PEGRNA_3	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	9
HEK3_5G-H_PEGRNA_4	GGCCCAGACTGAGCACGTGA	TCTGCGATCACGTGCTCAGTCTG	13	10
HEK4_5G-H_PEGRNA_1	GGCACTGCGGCTGGAGGTGG	TTAACGCCCACCTCCAGCC	9	10
HEK4_5G-H_PEGRNA_2	GGCACTGCGGCTGGAGGTGG	TTAACCCCCCCCTCCAGCC	9	10
HEK4_5G-H_PEGRNA_3	GGCACTGCGGCTGGAGGTGG	TTAACCCCTTACACCTCCAGCC	9	13
HEK4_5G-H_PEGRNA_4	GGCACTGCGGCTGGAGGTGG	TTAACCCCCCCTCCAGCC	9	9
EMX1_5G-H_PEGRNA_1	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCACTTCTTCTTCTGCTCGGA	14	16
EMX1_5G-H_PEGRNA_2	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTGCTTCTTCTGCTCGGA	14	16
EMX1_5G-H_PEGRNA_3	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGCCCTTCGCATTCTTCTGCTCGGA	14	19
EMX1_5G-H_PEGRNA_4	GAGTCCGAGCAGAAGAAGAA	GTGATGGGAGTTCTTCTTCTGCTCGGA	14	13
FANCF_5G-H_PEGRNA_1	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATGCAGGTGCTGCAGAAGGGAT	14	17
FANCF_5G-H_PEGRNA_2	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAGGCGCTGCAGAAGGGAT	14	17
FANCF_5G-H_PEGRNA_3	GGAATCCCTTCTGCAGCACC	GGAAAAGCGATCCAATCGGTGCTGCAGAAGGGAT	14	20

[2008]

[2009] 표 3K: 도 49a-49b PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 3494-3521)	3′ 연장부 (SEQ ID NO: 3522-3540)	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_6A_2G TOC	GGCCCAGACTG AGCACGTGA	TCTGCCATGACGTGCTC AGTCTG	13	10
HEK3_6A_2G TOC	GGCCCAGACTG AGCACGTGA			
EMX1_6A_3G TOC	GAGTCCGAGCA GAAGAAGAA	ATGGGAGCCCTTGTTCT TCTGCTCGG	13	13
EMX1_6A_3G TOC	GAGTCCGAGCA GAAGAAGAA			
FANCF_6A_5 GTOT	GGAATCCCTTCT GCAGCACC	AAAAGCGATCAAGGTGC TGCAGAAGGGA	13	15
FANCF_6A_5 GTOT	GGAATCCCTTCT GCAGCACC			
HEK3_6A_1H IS6INS	GGCCCAGACTG AGCACGTGA	TGGAGGAAGCAGGGCTT CCTTTCCTCTGCCATCAA TGATGGTGATGATGGTG CGTGCTCAGTCTG	13	52
HEK3_6A_1H IS6INS	GGCCCAGACTG AGCACGTGA			
HEK3_6A_5G TOT	GGCCCAGACTG AGCACGTGA	TCTGCAATCACGTGCTC AGTCTG	13	10
HEK3_6A_5G TOT	GGCCCAGACTG AGCACGTGA			
HEK3_6A_1C TTINS	GGCCCAGACTG AGCACGTGA	TCTGCCATCAAAGCGTG CTCAGTCTG	13	10
HEK3_6A_1C TTINS	GGCCCAGACTG AGCACGTGA			
*HBB*_6B_INS ALL	GCATGGTGCAC CTGACTCCTG	AGACTTCTCCACAGGAG TCAGGTGCAC	13	14
*HBB*_6B_INS ALL	GCATGGTGCAC CTGACTCCTG			
*HBB*_6B_COR RECT	GCATGGTGCAC CTGACTCCTG	AGACTTCTCCTCAGGAG TCAGGTGCAC	13	14
*HBB*_6B_COR RECT	GCATGGTGCAC CTGACTCCTG			
*HBB*_6B_COR RECT_W_SIL ENT	GCATGGTGCAC CTGACTCCTG	AGACTTCTCTTCAGGAG TCAGGTGCAC	13	14
*HBB*_6B_COR RECT_W_SIL ENT	GCATGGTGCAC CTGACTCCTG			
*HEXA*_6B_IN STALL	GTACCTGAACC GTATATCCTA	AGTCAGGGCCATAGGAT AGATATACGGTTC	12	14

[2010]

*HEXA*_6B_CORRECT	GATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTG	10	21
*HEXA*_6B_CORRECT_W_SILENT	GATCCTTCCAGTCAGGGCCAT	GTACCTGAACCGTATATCTTATGGCCCTGACT	9	27
*PRNP*_6C	GCAGTGGTGGGGGGCCTTGG	ATGTAGACGCCAAGGCCCCCCACC	12	12
HEK3_6E-G_1TTOG	GGCCCAGACTGAGCACGTGA	TCTGCCATCCCGTGCTCAGTCTG	13	10
HEK3_6E-G_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	10
RNF2_6E-G_1CTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCACGTAATGACTAAGATG	15	14
*HBB*_6E-G_4ATOT	GCATGGTGCACCTGACTCCTG	AGACTTCTCCACAGGAGTCAGGTGCAC	13	14
HEK3_6H_1HIS6INS	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCAATGATGGTGATGATGGTGCGTGCTCAGTCTG	13	52
HEK3_6H_1FLAGINS	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCACTTATCGTCGTCATCCTTGTAATCCGTGCTCAGTCTG	13	58

[2011]

[2012] 표 3L: 도 47a-74d PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 3541-3547)	3′ 연장부 서열 (SEQ ID NO: 3549-3556)	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_ED4B_1TDEL	GGCCCAGACTGAGCACGTGA	TCTGCCATCCGTGCTCAGTCTG	13	9
HEK3_ED4B_1AINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCATCGTGCTCAGTCTG	13	11
HEK3_ED4B_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	13
HEK3_ED4C_2GTOC	GGCCCAGACTGAGCACGTGA	TCTGCCATGACGTGCTCAGTCTG	13	10
HEK3_ED4D_1FLAGINS	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGCCATCACTTATCGTCGTCATCCTTGTAATCCGTGCTCAGTCTG	13	58
RNF2_ED4E_1CTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
EMX1_ED4F_1GTOC	GAGTCCGAGCAGAAGAAGAA	ATGGGAGCCCTTGTTCTTCTGCTCGG	13	13
*HBB*_ED4G_2TTOA	GTAACGGCAGACTTCTCCTC	ATCTGACTCCTGTGGAGAAGTCTGCC	12	14

[2013]

[2014] 표 3M: 도 48a-48c PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 3557-3627)	3′ 연장부 서열 (SEQ ID NO: 3628-3698)	PBS 길이 (NT)	RT 주형 길이 (NT)
VEGFA_ED5A_31	GATGTCTGCAGGCCAGATGA	CCCTCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	31
VEGFA_ED5A_30	GATGTCTGCAGGCCAGATGA	CCTCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	30
VEGFA_ED5A_29	GATGTCTGCAGGCCAGATGA	CTCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	29
VEGFA_ED5A_28	GATGTCTGCAGGCCAGATGA	TCTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	28
VEGFA_ED5A_27	GATGTCTGCAGGCCAGATGA	CTGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	27
VEGFA_ED5A_26	GATGTCTGCAGGCCAGATGA	TGACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	26
VEGFA_ED5A_25	GATGTCTGCAGGCCAGATGA	GACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	25
VEGFA_ED5A_24	GATGTCTGCAGGCCAGATGA	ACAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	24
VEGFA_ED5A_23	GATGTCTGCAGGCCAGATGA	CAATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	23
VEGFA_ED5A_22	GATGTCTGCAGGCCAGATGA	AATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	22
VEGFA_ED5A_21	GATGTCTGCAGGCCAGATGA	ATGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	21
VEGFA_ED5A_20	GATGTCTGCAGGCCAGATGA	TGTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	20
VEGFA_ED5A_19	GATGTCTGCAGGCCAGATGA	GTGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	19
VEGFA_ED5A_18	GATGTCTGCAGGCCAGATGA	TGCCATCTGGAGCACTCATCTGGCCTGCAGA	13	18
VEGFA_ED5A_17	GATGTCTGCAGGCCAGATGA	GCCATCTGGAGCACTCATCTGGCCTGCAGA	13	17

[2015]

VEGFA_ED 5A_16	GATGTCTGCAGGC CAGATGA	CCATCTGGAGCACTCATCTG GCCTGCAGA	13	16
VEGFA_ED 5A_15	GATGTCTGCAGGC CAGATGA	CATCTGGAGCACTCATCTGG CCTGCAGA	13	15
VEGFA_ED 5A_14	GATGTCTGCAGGC CAGATGA	ATCTGGAGCACTCATCTGGC CTGCAGA	13	14
VEGFA_ED 5A_13	GATGTCTGCAGGC CAGATGA	TCTGGAGCACTCATCTGGCC TGCAGA	13	13
VEGFA_ED 5A_12	GATGTCTGCAGGC CAGATGA	CTGGAGCACTCATCTGGCCT GCAGA	13	12
VEGFA_ED 5A_11	GATGTCTGCAGGC CAGATGA	TGGAGCACTCATCTGGCCTG CAGA	13	11
VEGFA_ED 5A_10	GATGTCTGCAGGC CAGATGA	GGAGCACTCATCTGGCCTGC AGA	13	10
VEGFA_ED 5A_9	GATGTCTGCAGGC CAGATGA	GAGCACTCATCTGGCCTGCA GA	13	9
VEGFA_ED 5A_8	GATGTCTGCAGGC CAGATGA	AGCACTCATCTGGCCTGCAG A	13	8
DNMT1_ED 5B_31	GATTCCTGGTGCC AGAAACA	AGGACTAGTTCTGCCCTCCC GTCACCACTGTTTCTGGCAC CAGG	13	31
DNMT1_ED 5B_30	GATTCCTGGTGCC AGAAACA	GGACTAGTTCTGCCCTCCCG TCACCACTGTTTCTGGCACC AGG	13	30
DNMT1_ED 5B_29	GATTCCTGGTGCC AGAAACA	GACTAGTTCTGCCCTCCCGT CACCACTGTTTCTGGCACCA GG	13	29
DNMT1_ED 5B_28	GATTCCTGGTGCC AGAAACA	ACTAGTTCTGCCCTCCCGTC ACCACTGTTTCTGGCACCAG G	13	28
DNMT1_ED 5B_27	GATTCCTGGTGCC AGAAACA	CTAGTTCTGCCCTCCCGTCA CCACTGTTTCTGGCACCAGG	13	27
DNMT1_ED 5B_26	GATTCCTGGTGCC AGAAACA	TAGTTCTGCCCTCCCGTCAC CACTGTTTCTGGCACCAGG	13	26
DNMT1_ED 5B_25	GATTCCTGGTGCC AGAAACA	AGTTCTGCCCTCCCGTCACC ACTGTTTCTGGCACCAGG	13	25
DNMT1_ED 5B_24	GATTCCTGGTGCC AGAAACA	GTTCTGCCCTCCCGTCACCA CTGTTTCTGGCACCAGG	13	24

[2016]

DNMT1_ED 5B_23	GATTCCTGGTGCC AGAAACA	TTCTGCCCTCCCGTCACCAC TGTTTCTGGCACCAGG	13	23
DNMT1_ED 5B_22	GATTCCTGGTGCC AGAAACA	TCTGCCCTCCCGTCACCACT GTTTCTGGCACCAGG	13	22
DNMT1_ED 5B_21	GATTCCTGGTGCC AGAAACA	CTGCCCTCCCGTCACCACTG TTTCTGGCACCAGG	13	21
DNMT1_ED 5B_20	GATTCCTGGTGCC AGAAACA	TGCCCTCCCGTCACCACTGT TTCTGGCACCAGG	13	20
DNMT1_ED 5B_19	GATTCCTGGTGCC AGAAACA	GCCCTCCCGTCACCACTGTT TCTGGCACCAGG	13	19
DNMT1_ED 5B_18	GATTCCTGGTGCC AGAAACA	CCCTCCCGTCACCACTGTTT CTGGCACCAGG	13	18
DNMT1_ED 5B_17	GATTCCTGGTGCC AGAAACA	CCTCCCGTCACCACTGTTTC TGGCACCAGG	13	17
DNMT1_ED 5B_16	GATTCCTGGTGCC AGAAACA	CTCCCGTCACCACTGTTTCT GGCACCAGG	13	16
DNMT1_ED 5B_15	GATTCCTGGTGCC AGAAACA	TCCCGTCACCACTGTTTCTG GCACCAGG	13	15
DNMT1_ED 5B_14	GATTCCTGGTGCC AGAAACA	CCCGTCACCACTGTTTCTGG CACCAGG	13	14
DNMT1_ED 5B_13	GATTCCTGGTGCC AGAAACA	CCGTCACCACTGTTTCTGGC ACCAGG	13	13
DNMT1_ED 5B_12	GATTCCTGGTGCC AGAAACA	CGTCACCACTGTTTCTGGCA CCAGG	13	12
DNMT1_ED 5B_11	GATTCCTGGTGCC AGAAACA	GTCACCACTGTTTCTGGCAC CAGG	13	11
DNMT1_ED 5B_10	GATTCCTGGTGCC AGAAACA	TCACCACTGTTTCTGGCACC AGG	13	10
DNMT1_ED 5B_9	GATTCCTGGTGCC AGAAACA	CACCACTGTTTCTGGCACCA GG	13	9
DNMT1_ED 5B_8	GATTCCTGGTGCC AGAAACA	ACCACTGTTTCTGGCACCAG G	13	8
RUNX1_ED 5C_31	GCATTTTCAGGAG GAAGCGA	AATGACTCAAATATGCTGTC TGAAGCAATCGCTTCCTCCT GAAAAT	15	31
RUNX1_ED 5C_30	GCATTTTCAGGAG GAAGCGA	ATGACTCAAATATGCTGTCT GAAGCAATCGCTTCCTCCTG AAAAT	15	30

[2017]

RUNX1_ED 5C_29	GCATTTTCAGGAG GAAGCGA	TGACTCAAATATGCTGTCTG AAGCAATCGCTTCCTCCTGA AAAT	15	29
RUNX1_ED 5C_28	GCATTTTCAGGAG GAAGCGA	GACTCAAATATGCTGTCTGA AGCAATCGCTTCCTCCTGAA AAT	15	28
RUNX1_ED 5C_27	GCATTTTCAGGAG GAAGCGA	ACTCAAATATGCTGTCTGAA GCAATCGCTTCCTCCTGAAA AT	15	27
RUNX1_ED 5C_26	GCATTTTCAGGAG GAAGCGA	CTCAAATATGCTGTCTGAAG CAATCGCTTCCTCCTGAAAA T	15	26
RUNX1_ED 5C_25	GCATTTTCAGGAG GAAGCGA	TCAAATATGCTGTCTGAAGC AATCGCTTCCTCCTGAAAAT	15	25
RUNX1_ED 5C_24	GCATTTTCAGGAG GAAGCGA	CAAATATGCTGTCTGAAGCA ATCGCTTCCTCCTGAAAAT	15	24
RUNX1_ED 5C_23	GCATTTTCAGGAG GAAGCGA	AAATATGCTGTCTGAAGCAA TCGCTTCCTCCTGAAAAT	15	23
RUNX1_ED 5C_22	GCATTTTCAGGAG GAAGCGA	AATATGCTGTCTGAAGCAAT CGCTTCCTCCTGAAAAT	15	22
RUNX1_ED 5C_21	GCATTTTCAGGAG GAAGCGA	ATATGCTGTCTGAAGCAATC GCTTCCTCCTGAAAAT	15	21
RUNX1_ED 5C_20	GCATTTTCAGGAG GAAGCGA	TATGCTGTCTGAAGCAATCG CTTCCTCCTGAAAAT	15	20
RUNX1_ED 5C_19	GCATTTTCAGGAG GAAGCGA	ATGCTGTCTGAAGCAATCGC TTCCTCCTGAAAAT	15	19
RUNX1_ED 5C_18	GCATTTTCAGGAG GAAGCGA	TGCTGTCTGAAGCAATCGCT TCCTCCTGAAAAT	15	18
RUNX1_ED 5C_17	GCATTTTCAGGAG GAAGCGA	GCTGTCTGAAGCAATCGCTT CCTCCTGAAAAT	15	17
RUNX1_ED 5C_16	GCATTTTCAGGAG GAAGCGA	CTGTCTGAAGCAATCGCTTC CTCCTGAAAAT	15	16
RUNX1_ED 5C_15	GCATTTTCAGGAG GAAGCGA	TGTCTGAAGCAATCGCTTCC TCCTGAAAAT	15	15
RUNX1_ED 5C_14	GCATTTTCAGGAG GAAGCGA	GTCTGAAGCAATCGCTTCCT CCTGAAAAT	15	14
RUNX1_ED 5C_13	GCATTTTCAGGAG GAAGCGA	TCTGAAGCAATCGCTTCCTC CTGAAAAT	15	13

[2018]

RUNX1_ED 5C_12	GCATTTTCAGGAG GAAGCGA	CTGAAGCAATCGCTTCCTCC TGAAAAT	15	12
RUNX1_ED 5C_11	GCATTTTCAGGAG GAAGCGA	TGAAGCAATCGCTTCCTCCT GAAAAT	15	11
RUNX1_ED 5C_10	GCATTTTCAGGAG GAAGCGA	GAAGCAATCGCTTCCTCCTG AAAAT	15	10
RUNX1_ED 5C_9	GCATTTTCAGGAG GAAGCGA	AAGCAATCGCTTCCTCCTGA AAAT	15	9

[2019]

[2020] 표 3N: 도 48a-48c PEgRNA

PEGRNA	스페이서 서열	3′ 연장부 서열	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_ED6_5GTOA	GGCCCAGACTGAGCACGTGA (SEQ ID NO: 393)	TCTGCTATCACGTGCTCAGTCTG (SEQ ID NO: 394)	13	10

[2021]

[2022] 표 30: 도 48a-48c 닉킹 sgRNA

닉킹 SGRNA	스페이서 서열
HEK3_ED6_+63	GCACATACTAGCCCCTGTCT (SEQ ID NO: 395)

[2023]

[2024] 표 3P: 도 50a-50b PEgRNA

PEG RNA	스페이서 (SEQ ID NO: 3699-3754)	3′ 연장부 (5′에서 3′)(SEQ ID NO: 3755-3810)	PBS 길이 (NT)	RT 주형 길이 (NT)
HBB 3.5	GTAACGGCAGACTTCTCCAC	AGACTTCTCCTCAGGAGTCAGGTGCAC	12	14
HBB 3.7	GCATGGTGCACCTGACTCCTG	AGACTTCTCTTCAGGAGTCAGGTGCAC	13	14
HBB 5.2	GCATGGTGCACCTGACTCCTG	TAACGGCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	19
HBB 5.3	GCATGGTGCACCTGACTCCTG	ACGGCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	17
HBB 5.4	GCATGGTGCACCTGACTCCTG	GGCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	16
HBB 5.5	GCATGGTGCACCTGACTCCTG	GCAGACTTCTCCTCAGGAGTCAGGTGCAC	13	13
HBB 5.6	GCATGGTGCACCTGACTCCTG	GACTTCTCCTCAGGAGTCAGGTGCAC	13	12
HBB 5.7	GCATGGTGCACCTGACTCCTG	ACTTCTCCTCAGGAGTCAGGTGCAC	13	21
HBB 5.8	GCATGGTGCACCTGACTCCTG	TAACGGCAGACTTCTCCTCAGGAGTCAGGTGCA	12	19
HBB 5.9	GCATGGTGCACCTGACTCCTG	ACGGCAGACTTCTCCTCAGGAGTCAGGTGCA	12	17
HBB 5.10	GCATGGTGCACCTGACTCCTG	GGCAGACTTCTCCTCAGGAGTCAGGTGCA	12	16
HBB 5.11	GCATGGTGCACCTGACTCCTG	GCAGACTTCTCCTCAGGAGTCAGGTGCA	12	13
HBB 5.12	GCATGGTGCACCTGACTCCTG	GACTTCTCCTCAGGAGTCAGGTGCA	12	12
HBB 5.13	GCATGGTGCACCTGACTCCTG	ACTTCTCCTCAGGAGTCAGGTGCA	12	14
*HEXA*S 1	ATCCTTCCAGTCAGGGCCAT	ATATCTTATGGCCCTGACTGGAA	13	14
*HEXA*S 2	ATCCTTCCAGTCAGGGCCAT	TATATCTTATGGCCCTGACTGGAA	13	15
*HEXA*S 3	ATCCTTCCAGTCAGGGCCAT	GTATATCTTATGGCCCTGACTGGAA	13	16
*HEXA*S 4	ATCCTTCCAGTCAGGGCCAT	ACCGTATATCTTATGGCCCTGACTGGAA	13	19
*HEXA*S 5	ATCCTTCCAGTCAGGGCCAT	AACCGTATATCTTATGGCCCTGACTGGAA	13	20

[2025]

[2026]

HEX *AS* 6	ATCCTTCCAGTCAGG GCCAT	GAACCGTATATCTTATGGCCCTGACTG GAA	13	21
HEX *AS* 7	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT GGAA	13	22
HEX *AS* 8	ATCCTTCCAGTCAGG GCCAT	ATATCTTATGGCCCTGACT	9	14
HEX *AS* 9	ATCCTTCCAGTCAGG GCCAT	TATATCTTATGGCCCTGACT	9	15
HEX *AS* 10	ATCCTTCCAGTCAGG GCCAT	GTATATCTTATGGCCCTGACT	9	16
HEX *AS* 11	ATCCTTCCAGTCAGG GCCAT	ACCGTATATCTTATGGCCCTGACT	9	19
HEX *AS* 12	ATCCTTCCAGTCAGG GCCAT	AACCGTATATCTTATGGCCCTGACT	9	20
HEX *AS* 13	ATCCTTCCAGTCAGG GCCAT	GAACCGTATATCTTATGGCCCTGACT	9	21
HEX *AS* 14	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT	9	22
HEX *AS* 15	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGAC	8	22
HEX *AS* 16	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT G	10	22
HEX *AS* 17	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT GG	11	22
HEX *AS* 18	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT GGA	12	22
HEX *AS* 19	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT GGAA	13	22
HEX *AS* 20	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT GGAAG	14	22
HEX *AS* 21	ATCCTTCCAGTCAGG GCCAT	TGAACCGTATATCTTATGGCCCTGACT GGAAGG	15	22
HEX *AS* 22	ATCCTTCCAGTCAGG GCCAT	ACCTGAACCGTATATCTTATGGCCCTG ACT	9	25

*HEXA*S 23	ATCCTTCCAGTCAGGGCCAT	TACCTGAACCGTATATCTTATGGCCCTGACT	9	26
*HEXA*S 24	ATCCTTCCAGTCAGGGCCAT	GTACCTGAACCGTATATCTTATGGCCCTGACT	9	27
*HEXA*S 25	ATCCTTCCAGTCAGGGCCAT	GGTACCTGAACCGTATATCTTATGGCCCTGACT	9	28
*HEXA*S 26	ATCCTTCCAGTCAGGGCCAT	TGGTACCTGAACCGTATATCTTATGGCCCTGACT	9	29
HEXA 5	ATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTGGAA	13	21
HEXA 6	ATCCTTCCAGTCAGGGCCAT	ACCGTATATCCTATGGCCCTGACTGGAA	13	15
HEXA 7	ATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTGGAAGG	15	21
HEXA 8	ATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTGGAAG	14	21
HEXA 9	ATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTGGA	12	21
HEXA 10	ATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTGG	11	21
HEXA 11	ATCCTTCCAGTCAGGGCCAT	ACCTGAACCGTATATCCTATGGCCCTGACTG	10	21
HEXA 12	ATCCTTCCAGTCAGGGCCAT	AACCGTATATCCTATGGCCCTGACTGGAA	13	16
HEXA 13	ATCCTTCCAGTCAGGGCCAT	TGAACCGTATATCCTATGGCCCTGACTGGAA	13	18
HEXA 14	ATCCTTCCAGTCAGGGCCAT	TACCTGAACCGTATATCCTATGGCCCTGACTGGAA	13	22
HEXA 15	ATCCTTCCAGTCAGGGCCAT	TGGTACCTGAACCGTATATCCTATGGCCCTGACTGGAA	13	25
HEXA 16	ATCCTTCCAGTCAGGGCCAT	GTACCTGAACCGTATATCCTATGGCCCTGACTGGAA	13	23
HEXA 17	ATCCTTCCAGTCAGGGCCAT	AACCGTATATCCTATGGCCCTGACTG	10	16
HEXA 18	ATCCTTCCAGTCAGGGCCAT	TGAACCGTATATCCTATGGCCCTGACTG	10	18
HEXA 19	ATCCTTCCAGTCAGGGCCAT	TACCTGAACCGTATATCCTATGGCCCTGACTG	10	22
HEXA 20	ATCCTTCCAGTCAGGGCCAT	TGGTACCTGAACCGTATATCCTATGGCCCTGACTG	10	25

[2027]

[2028] 표 3Q: 도 50a-50b 닉킹 sgRNA

닉킹 SGRNA	스페이서 서열
*HBB*_ED7A_+72	GCCTTGATACCAACCTGCCCA (SEQ ID NO: 626)
*HEXA*_ED7B_+60	GCTGGAACTGGTCACCAAGGC(SEQ ID NO: 627)
*HEXA*_ED7B_CORRECT_WT_PE3B	GTACCTGAACCGTATATCCTA(SEQ ID NO: 628)
*HEXA*_ED7B_CORRECT_SILENT_PE3B	GTACCTGAACCGTATATCTTA(SEQ ID NO: 629)

[2029]

[2030] 표 3R: 도 51a-51g PEgRNA

PEGRNA	스페이서 서열 (SEQ ID NO: 632-640)	3′ 연장부 (SEQ ID NO: 641-649)	PBS 길이 (NT)	RT 주형 길이 (NT)
HEK3_ED8_1TTOG	GGCCCAGACTGAGCACGTGA	TCTGCCATCCCGTGCTCAGTCTG	13	10
HEK3_ED8_3ATOC	GGCCCAGACTGAGCACGTGA	TCTGCCAGCACGTGCTCAGTCTG	13	10
HEK3_ED8_3ATOT	GGCCCAGACTGAGCACGTGA	TCTGCCAACACGTGCTCAGTCTG	13	10
HEK3_ED8_3ATOT_5-6GGTOTT	GGCCCAGACTGAGCACGTGA	TGGAGGAAGCAGGGCTTCCTTTCCTCTGAAAACACGTGCTCAGTCTG	13	34
HEK3_ED8_1CTTINS	GGCCCAGACTGAGCACGTGA	TCTGCCATCAAAGCGTGCTCAGTCTG	13	10
RNF2_ED8_1CTOA	GTCATCTTAGTCATTACCTG	AACGAACACCTCATGTAATGACTAAGATG	15	14
RNF2_ED8_1CTOG	GTCATCTTAGTCATTACCTG	AACGAACACCTCACGTAATGACTAAGATG	15	14
RNF2_ED8_1GTAINS	GTCATCTTAGTCATTACCTG	AACGAACACCTCAGTACGTAATGACTAAGATG	15	17
*HBB*_ED8_4ATOT	GCATGGTGCACCTGACTCCTG	AGACTTCTCCACAGGAGTCAGGTGCAC	13	14

[2031]

[2032] 표 4: 포유동물 세포 게놈 DNA 증폭 및 HTS에 사용된 프라이머의 서열[181].

설명	서열 (SEQ ID NO: 3811-3863)
HEK3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
HEK3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
RNF2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACGTCTCATATGCCCCTTGG
RNF2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGTAGGAATTTTGGTGGGACA
HEK4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAACCCAGGTAGCCAGAGAC
HEK4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
EMX1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
EMX1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
FANCF FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
FANCF REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
HBB FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGGGTTGGCCAATCTACTCCC
HBB REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTTCTCTGTCTCCACATGCC
PRNP FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGTGGAACAAGCCGAGT
PRNP REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTACTTGGTTGGGGTAACGGTG
HEXA FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATACAGGTGTGGCGAGAGG
HEXA REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCCTCCTTTGGTTAGCA
RUNX1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCACAAACAAGACAGGGAACTG
RUNX1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGATGTAGGGCTAGAGGGGTG
VEGFA FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACTTGGTGCCAAATTCTTCTCC
VEGFA REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAAGAGGGAATGGGCTTTGGA
DNMT FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAACAGCTTCATGTCAGCC
DNMT REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGTTAATGTTTCCTGATGGTCC

[2033]

HEK3 오프-타겟 부위 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCC TGTTGACCTGGAGAA
HEK3 오프-타겟 부위 1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCC CTGACCA
HEK3 오프-타겟 부위 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGT GTTGACAGGGAGCAA
HEK3 오프-타겟 부위 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGC AGAAGGG
HEK3 오프-타겟 부위 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGA GGGAACAGAAGGGCT
HEK3 오프-타겟 부위 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCC AAGAACCT
HEK3 오프-타겟 부위 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTA GCACTTTGGAAGGTCG
HEK3 오프-타겟 부위 4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCT GCTCAGCC
HEK4 오프-타겟 부위 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCAT GGCTTCTGAGACTCA
HEK4 오프-타겟 부위 1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCAC TCCCTGTCTTT
HEK4 오프-타겟 부위 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGG CAATGGAGGCATTGG
HEK4 오프-타겟 부위 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGAGGCTGCC CATGAGAG
HEK4 오프-타겟 부위 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCT GAGGCTCGAATCCTG
HEK4 오프-타겟 부위 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCAT ATCCCTG
HEK4 오프-타겟 부위 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTCC ACCAGAACTCAGCCC
HEK4 오프-타겟 부위 4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCCTCC ACAACAC
EMX1 오프-타겟 부위 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGGG GAGATTTGCATCTGTGGAGG
EMX1 오프-타겟 부위 1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTTATACCATC TTGGGGTTACAG
EMX1 오프-타겟 부위 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAATG TGCTTCAACCCATCACGGC
EMX1 오프-타겟 부위 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCATGAATTTGTG ATGGATGCAGTCTG
EMX1 오프-타겟 부위 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGAA GGAGGTGCAGGAGCTAGAC

[2034]

EMX1 오프-타겟 부위 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCATCCCGACCTTC ATCCCTCCTGG
EMX1 오프-타겟 부위 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTAGT TCTGACATTCCTCCTGAGGG
EMX1 오프-타겟 부위 4 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCAAACAAGGTG CAGATACAGCA
FANCF 오프-타겟 부위 1 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCGGG CAGTGGCGTCTTAGTCG
FANCF 오프-타겟 부위 1 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCTGGGTTTGGT TGGCTGCTC
FANCF 오프-타겟 부위 2 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTCCT TGCCGCCCAGCCGGTC
FANCF 오프-타겟 부위 2 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGGGGAAGA GGCGAGGACAC
FANCF 오프-타겟 부위 3 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGT GTTTCCCATCCCCAACAC
FANCF 오프-타겟 부위 3 REV	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAATGGATCCCCC CCTAGAGCTC
FANCF 오프-타겟 부위 4 FWD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGGC CCACAGGTCCTTCTGGA

[2035]

[2036] 표 5: HDR 실험 및 HBB E6V HEK293T 세포주의 생성에 사용된 100-mer 단일 가닥 DNA 올리고뉴클레오티드 공여자 주형의 서열. 올리고뉴클레오티드는 길이가 100-103 nt이고, 상동성 아암이 대략 편집 부위의 중심이 된다. 올리고뉴클레오티드는 인티그레이티드 DNA 테크놀로지스로부터 입수하였고, PAGE에 의해 정제하였다.

HEK3 +3 A에서 T	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAG ACTGAGCACGTGTTGGCAGAGGAAAGGAAGCCCTGCTTC CTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 717)
HEK3 +3 A에서 T, +5,6 GG에서 TT	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAG ACTGAGCACGTGTTTTCAGAGGAAAGGAAGCCCTGCTTC CTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 718)
HEK3 +1 T에서 G	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAG ACTGAGCACGGGATGGCAGAGGAAAGGAAGCCCTGCTTC CTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 719)
HEK3 +3 A에서 C	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAG ACTGAGCACGTGCTGGCAGAGGAAAGGAAGCCCTGCTTC CTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 720)
HEK3 +1 CTT 삽입	GCTTCTCCAGCCCTGGCCTGGGTCAATCCTTGGGGCCCAG ACTGAGCACGCTTTGATGGCAGAGGAAAGGAAGCCCTGC TTCCTCCAGAGGGCGTCGCAGGAC (SEQ ID NO: 721)
RNF2 +1 C에서 A	CCCAGTTTACACGTCTCATATGCCCCTTGGCAGTCATCTTA GTCATTACATGAGGTGTTCGTTGTAACTCATATAAACTGAG TTCCCATGTTTTGCTTAA (SEQ ID NO: 722)
RNF2 +1 C에서 G	CCCAGTTTACACGTCTCATATGCCCCTTGGCAGTCATCTTA GTCATTACGTGAGGTGTTCGTTGTAACTCATATAAACTGAG TTCCCATGTTTTGCTTAA (SEQ ID NO: 723)
RNF2 +1 GTA 삽입	CAGTTTACACGTCTCATATGCCCCTTGGCAGTCATCTTAGT CATTACGTACTGAGGTGTTCGTTGTAACTCATATAAACTGA GTTCCCATGTTTTGCTTA (SEQ ID NO: 724)
HBB E6V 설치 (또한 *HBB* E6V HEK293T 세포주의 생성에 사용됨)	ACTTCATCCACGTTCACCTTGCCCCACAGGGCAGTAACGG CAGACTTCTCCACAGGAGTCAGATGCACCATGGTGTCTGT TTGAGGTTGCTAGTGAACAC (SEQ ID NO: 725)
HBB E6V 교정 프로토스페이서 A	ACTTCATCCACGTTCACCTTGCCCCACAGGGCAGTAACGG CAGACTTCTCCTCAGGAGTCAGGTGCACCATGGTGTCTGT TTGAGGTTGCTAGTGAACAC (SEQ ID NO: 726)
HBB E6V 교정 프로토스페이서 B	GTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACCT GACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGC AAGGTGAACGTGGATGAAGT (SEQ ID NO: 727)
HBB E6V 교정 프로토스페이서 B, 침묵 PAM 돌연변이	GTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACCT GACTCCTGATGAGAAGTCTGCCGTTACTGCCCTGTGGGGC AAGGTGAACGTGGATGAAGT (SEQ ID NO: 728)
PRNP G127V	CACATGGCTGGTGCTGCAGCAGCTGGGGCAGTGGTGGGG GGCCTTGGCGTCTACATGCTGGGAAGTGCCATGAGCAGGC CCATCATACATTTCGGCAGTG (SEQ ID NO: 729)

[2037]

[2038] 추가의 서열

[2039] 본원에 사용된 효모 이중 형광 리포터 플라스미드의 서열

[2040] p425-GFP_정지_mCherry:

ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGATGGTGA
TGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTACTTACGGTAAA
TTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCAC
TACTTTCGGTTATGGTGTTCAATGTTTTGCTAGATACCCAGATCATATGAAACAACATGACT
TTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTCAAAGATGA
CGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAGAATC
[2041] GAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTGGAATACA

ACTATAACTCTCACAATGTTTACATCATGGCTGACAAACAAAAGAATGGTATCAAAGTTAA
CTTCAAAATTAGACACAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAA
AATACTCCAATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATC
TGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTGTTACT
GCTGCTGGTATTACCCATGGTATGGATGAATTGTACAAAGCTAGCAACCTGGGTCAATCCTT
GG**GGCCCAGACTGAGCACGTGATGG**CAGAGCACAGGAGACGTC*ATGGTTTCAAAAGGTG*
AAGAAGATAATATGGCTATTATTAAAGAATTTATGAGATTTAAAGTTCATATGGAAGGTTCAGTTAAT
GGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGTAGACCATATGAAGGTACTCAAACTGCTAA
ATTGAAAGTTACTAAAGGTGGTCCATTACCATTTGCTTGGGATATTTTGTCACCACAATTTATGTAT
GGTTCAAAAGCTTATGTTAAACATCCAGCTGATATTCCAGATTATTTAAAATTGTCATTTCCAGAAG
GTTTTAAATGGGAAAGAGTTATGAATTTTGAAGATGGTGGTGTTGTTACTGTTACTCAAGATTCAT
CATTACAAGATGGTGAATTTATTTATAAAGTTAAATTGAGAGGTACTAATTTTCCATCAGATGGTCC
AGTTATGCAAAAAAAAACTATGGGTTGGGAAGCTTCATCAGAAAGAATGTATCCAGAAGATGGTG
CTTTAAAAGGTGAAATTAAACAAAGATTGAAATTAAAAGATGGTGGTCATTATGATGCTGAAGTTAA
AACTACTTATAAAGCTAAAAAACCAGTTCAATTACCAGGTGCTTATAATGTTAATATTAAATTGGATA
TTACTTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGAGCTGAAGGTAGACATTCAAC
*TGGTGGTATGGATGAATTATATAAAGGTACCGCTCGAGCAGCTGTGATTGATTGA*GTCGACTTG
GTTGAACACGTTGCCAAGGCTTAAGTGAATTTACTTTAAATCTTGCATTTAAATAAATTTTC
TTTTTATAGCTTTATGACTTAGTTTCAATTTATATACTATTTTAATGACATTTTCGATTCGGATC
CACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAG
GGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCG
CTCACAATTCCACACAACATAGGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGGTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAG
CGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG
TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT
CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGT
TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC
CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTG
GTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
GGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTT
TGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGT
AGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
CAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCT
AGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCG
TGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCG
CGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC
TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
ATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT
[2042] GCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTC

AATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA
GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAACGAAG
CATCTGTGCTTCATTTTGTAGAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAA
GAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACG
AAGAATCTGTGCTTCATTTTTGTAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAA
CAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTTACC
AACAAAGAATCTATACTTCTTTTTTGTTCTACAAAAATGCATCCCGAGAGCGCTATTTTTCT
AACAAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTCTTGATAA
CTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTTTGGTGTCTATTTTCTCTTC
CATAAAAAAAGCCTGACTCCACTTCCCGCGTTTACTGATTACTAGCGAAGCTGCGGGTGCA
TTTTTTCAAGATAAAGGCATCCCCGATTATATTCTATACCGATGTGGATTGCGCATACTTTGT
GAACAGAAAGTGATAGCGTTGATGATTCTTCATTGGTCAGAAAATTATGAACGGTTTCTTC
TATTTTGTCTCTATATACTACGTATAGGAAATGTTTACATTTTCGTATTGTTTTCGATTCACTC
TATGAATAGTTCTTACTACAATTTTTTTGTCTAAAGAGTAATACTAGAGATAAACATAAAAA
ATGTAGAGGTCGAGTTTAGATGCAAGTTCAAGGAGCGAAAGGTGGATGGGTAGGTTATATA
GGGATATAGCACAGAGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGT
ATTCGCAATATTTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTTGAAAGTGCGTC
TTCAGAGCGCTTTTGGTTTTCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGA
ACTTCGGAATAGGAACTTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAAAATGCAACGCG
AGCTGCGCACATACAGCTCACTGTTCACGTCGCACCTATATCTGCGTGTTGCCTGTATATAT
ATATACATGAGAAGAACGGCATAGTGCGTGTTTATGCTTAAATGCGTACTTATATGCGTCTAT
TTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTATCCCATTCCATGCGGGGTATC
GTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTATATGCTGCCACTCCTCAATTGGATTAG
TCTCATCCTTCAATGCTATCATTTCCTTTGATATTGGATCATACTAAGAAACCATTATTATCAT
GACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGAT
GACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCG
GATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGG
CTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATCGACTACGTC
GTAAGGCCGTTTCTGACAGAGTAAAATTCTTGAGGGAACTTTCACCATTATGGGAAATGCT
TCAAGAAGGTATTGACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTT
GTATTTTTTTTTTTTTAGAGAAAATCCTCCAATATCAAATTAGGAATCGTAGTTTCATGATTT
TCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCAATATTAATGTTAAAGTGCA
ATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTACCTGTATTCCTTTACTATC
CTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATATAGTTTCGTCTACCCTATGAA
CATATTCCATTTTGTAATTTCGTGTCGTTTCTATTATGAATTTCATTTATAAAGTTTATGTACA
AATATCATAAAAAAAGAGAATCTTTTTAAGCAAGGATTTTCTTAACTTCTTCGGCGACAGC
ATCACCGACTTCGGTGGTACTGTTGGAACCACCTAAATCACCAGTTCTGATACCTGCATCC
AAAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTT
CAACATCATTGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCT
GGAGCAGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAG
GCCAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATCGGA
GATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTGGGTTGGGTTCTTAA
CTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACCTTCAATGTAGGGAATTCGTT
CTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGCCAAAAGATTAGCTTTAT
CCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGGCCATTCTTG
TGATTCTTTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACCATCG
TCTTCCTTTCTCTTACCAAAGTAAATACCTCCCACTAATTCTCTGACAACAACGAAGTCAG
TACCTTTAGCAAATTGTGGCTTGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTT
ACATGGTCTTAAGTTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAG
GTCTAACACTACCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAA
CCTTCTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGC
ACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGGAACGAACATCAGAAATAGCT
TTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACGTGGTCACCTGGCAAAACGA
[2043] CGATCTTCTTAGGGGCAGACATAGGGGCAGACATTAGAATGGTATATCCTTGAAATATATAT

ATATATTGCTGAAATGTAAAAGGTAAGAAAAGTTAGAAAGTAAGACGATTGCTAACCACCT
ATTGGAAAAAACAATAGGTCCTTAAATAATATTGTCAACTTCAAGTATTGTGATGCAAGCAT
TTAGTCATGAACGCTTCTCTATTCTATATGAAAAGCCGGTTCCGGCCTCTCACCTTTCCTTT
TTCTCCCAATTTTTCAGTTGAAAAAGGTATATGCGTCAGGCGACCTCTGAAATTAACAAAA
AATTTCCAGTCATCGAATTTGATTCTGTGCGATAGCGCCCCTGTGTGTTCTCGTTATGTTGA
GGAAAAAAATAATGGTTGCTAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTC
CCACAGTTAACTGCGGTCAAGATATTTCTTGAATCAGGCGCCTTAGACCGCTCGGCCAAAC
AACCAATTACTTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTTGAACACA
CATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGATTTTGATGTAATTGT
TGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACTAGAAGTTCTCCTCG
ACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG
AAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTT
TTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGG
GTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGT
CAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATC
AAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCG
ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCG
AAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCAC
ACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCG
CAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGG
GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTG
TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACCGG
GCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGA
TCCGTTAGAATCATTTTGAATAAAAAACACGCTTTTTCAGTTCGAGTTTATCATTATCAATA
CTGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCTAACTTTATTTAGTCAA
AAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGTTACAC
AGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAA
TGGAGCCCGCTTTTTAAGCTGGCATCCAGAAAAAAAAAGAATCCCAGCACCAAAATATTG
TTTTCTTCACCAACCATCAGTTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGG
GCACAAACAGGCAAAAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTA
AATGATGACACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTTCTATTA
CCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAGTTCCCTGAA
ATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAAA
TCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACC
AAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAGAATTC (SEQ ID NO: 730)

[2044]

[2045] *p425-GFP_+1fs_mCherry*:

ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGATGGTGA
TGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTACTTACGGTAAA
TTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCAC
TACTTTCGGTTATGGTGTTCAATGTTTTGCTAGATACCCAGATCATATGAAACAACATGACT
TTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTCAAAGATGA
CGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAGAATC
GAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTGGAATACA
ACTATAACTCTCACAATGTTTACATCATGGCTGACAAACAAAAGAATGGTATCAAAGTTAA
CTTCAAAATTAGACACAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAA
AATACTCCAATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATC
TGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTGTTACT
GCTGCTGGTATTACCCATGGTATGGATGAATTGTACAAAGCTAGCAAACCTGGGTCAATCC
TTGG**GGCCCAGACTGAGCACGTGATGG**CAGAGCACAGGAGACGTC*ATGGTTTCAAAAGG*
TGAAGAAGATAATATGGCTATTATTAAAGAATTTATGAGATTTAAAGTTCATATGGAAGGTTCAGTTA
ATGGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGTAGACCATATGAAGGTACTCAAACTGCT
AAATTGAAAGTTACTAAAGGTGGTCCATTACCATTTGCTTGGGATATTTTGTCACCACAATTTATGT
ATGGTTCAAAAGCTTATGTTAAACATCCAGCTGATATTCCAGATTATTTAAAATTGTCATTTCCAGA

[2046]

AGGTTTTAAATGGGAAAGAGTTATGAATTTTGAAGATGGTGGTGTTGTTACTGTTACTCAAGATTC
ATCATTACAAGATGGTGAATTTATTTATAAAGTTAAATTGAGAGGTACTAATTTTCCATCAGATGGTC
CAGTTATGCAAAAAAAAACTATGGGTTGGGAAGCTTCATCAGAAAGAATGTATCCAGAAGATGGT
GCTTTAAAAGGTGAAATTAAACAAAGATTGAAATTAAAAGATGGTGGTCATTATGATGCTGAAGTT
AAAACTACTTATAAAGCTAAAAAACCAGTTCAATTACCAGGTGCTTATAATGTTAATATTAAATTGGA
TATTACTTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGAGCTGAAGGTAGACATTCA
*ACTGGTGGTATGGATGAATTATATAAAGGTACCGCTCGAGCAGCTGTGATTGATTGA*GTCGACTT
GGTTGAACACGTTGCCAAGGCTTAAGTGAATTTACTTTAAATCTTGCATTTAAATAAATTTT
CTTTTTATAGCTTTATGACTTAGTTTCAATTTATATACTATTTTAATGACATTTTCGATTCGGAT
CCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAG
GGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCG
CTCACAATTCCACACAACATAGGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGGTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAG
CGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG
TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT
CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGT
TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC
CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTG
GTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
GGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTT
TGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGT
AGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
CAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCT
AGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCG
TGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCG
CGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC
TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
ATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT
GCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTC
AATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA
GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAACGAAG
CATCTGTGCTTCATTTTGTAGAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAA
GAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACG
AAGAATCTGTGCTTCATTTTTGTAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAA
CAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTTACC
AACAAAGAATCTATACTTCTTTTTTGTTCTACAAAAATGCATCCCGAGAGCGCTATTTTTCT
AACAAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTCTTGATAA
CTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTTTGGTGTCTATTTTCTCTTC
[2047] CATAAAAAAAGCCTGACTCCACTTCCCGCGTTTACTGATTACTAGCGAAGCTGCGGGTGCA

TTTTTTCAAGATAAAGGCATCCCCGATTATATTCTATACCGATGTGGATTGCGCATACTTTGT
GAACAGAAAGTGATAGCGTTGATGATTCTTCATTGGTCAGAAAATTATGAACGGTTTCTTC
TATTTTGTCTCTATATACTACGTATAGGAAATGTTTACATTTTCGTATTGTTTTCGATTCACTC
TATGAATAGTTCTTACTACAATTTTTTTGTCTAAAGAGTAATACTAGAGATAAACATAAAAA
ATGTAGAGGTCGAGTTTAGATGCAAGTTCAAGGAGCGAAAGGTGGATGGGTAGGTTATATA
GGGATATAGCACAGAGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGT
ATTCGCAATATTTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTTGAAAGTGCGTC
TTCAGAGCGCTTTTGGTTTTCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGA
ACTTCGGAATAGGAACTTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAAAATGCAACGCG
AGCTGCGCACATACAGCTCACTGTTCACGTCGCACCTATATCTGCGTGTTGCCTGTATATAT
ATATACATGAGAAGAACGGCATAGTGCGTGTTTATGCTTAAATGCGTACTTATATGCGTCTAT
TTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTATCCCATTCCATGCGGGGTATC
GTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTATATGCTGCCACTCCTCAATTGGATTAG
TCTCATCCTTCAATGCTATCATTTCCTTTGATATTGGATCATACTAAGAAACCATTATTATCAT
GACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGAT
GACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCG
GATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGG
CTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATCGACTACGTC
GTAAGGCCGTTTCTGACAGAGTAAAATTCTTGAGGGAACTTTCACCATTATGGGAAATGCT
TCAAGAAGGTATTGACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTT
GTATTTTTTTTTTTTTAGAGAAAATCCTCCAATATCAAATTAGGAATCGTAGTTTCATGATTT
TCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCAATATTAATGTTAAAGTGCA
ATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTACCTGTATTCCTTTACTATC
CTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATATAGTTTCGTCTACCCTATGAA
CATATTCCATTTTGTAATTTCGTGTCGTTTCTATTATGAATTTCATTTATAAAGTTTATGTACA
AATATCATAAAAAAAGAGAATCTTTTTAAGCAAGGATTTTCTTAACTTCTTCGGCGACAGC
ATCACCGACTTCGGTGGTACTGTTGGAACCACCTAAATCACCAGTTCTGATACCTGCATCC
AAAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTT
CAACATCATTGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCT
GGAGCAGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAG
GCCAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATCGGA
GATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTGGGTTGGGTTCTTAA
CTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACCTTCAATGTAGGGAATTCGTT
CTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGCCAAAAGATTAGCTTTAT
CCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGGCCATTCTTG
TGATTCTTTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACCATCG
TCTTCCTTTCTCTTACCAAAGTAAATACCTCCCACTAATTCTCTGACAACAACGAAGTCAG
TACCTTTAGCAAATTGTGGCTTGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTT
ACATGGTCTTAAGTTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAG
GTCTAACACTACCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAA
CCTTCTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGC
ACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGGAACGAACATCAGAAATAGCT
TTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACGTGGTCACCTGGCAAAACGA
CGATCTTCTTAGGGGCAGACATAGGGGCAGACATTAGAATGGTATATCCTTGAAATATATAT
ATATATTGCTGAAATGTAAAAGGTAAGAAAAGTTAGAAAGTAAGACGATTGCTAACCACCT
ATTGGAAAAAACAATAGGTCCTTAAATAATATTGTCAACTTCAAGTATTGTGATGCAAGCAT
TTAGTCATGAACGCTTCTCTATTCTATATGAAAAGCCGGTTCCGGCCTCTCACCTTTCCTTT
TTCTCCCAATTTTTCAGTTGAAAAAGGTATATGCGTCAGGCGACCTCTGAAATTAACAAAA
AATTTCCAGTCATCGAATTTGATTCTGTGCGATAGCGCCCCTGTGTGTTCTCGTTATGTTGA
GGAAAAAAATAATGGTTGCTAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTC
CCACAGTTAACTGCGGTCAAGATATTTCTTGAATCAGGCGCCTTAGACCGCTCGGCCAAAC
AACCAATTACTTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTTGAACACA
CATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGATTTTGATGTAATTGT
[2048] TGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACTAGAAGTTCTCCTCG

ACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG
AAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTT
TTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGG
GTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGT
CAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATC
AAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCG
ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCG
AAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCAC
ACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCG
CAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGG
GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTG
TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACCGG
GCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGA
TCCGTTAGAATCATTTTGAATAAAAAACACGCTTTTTCAGTTCGAGTTTATCATTATCAATA
CTGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCTAACTTTATTTAGTCAA
AAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGTTACAC
AGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAA
TGGAGCCCGCTTTTTAAGCTGGCATCCAGAAAAAAAAAGAATCCCAGCACCAAAATATTG
TTTTCTTCACCAACCATCAGTTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGG
GCACAAACAGGCAAAAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTA
AATGATGACACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTTCTATTA
CCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAGTTCCCTGAA
ATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAAA
TCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACC
AAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAGAATTC (SEQ ID NO: 731)

[2049]

[2050] *p425-GFP_-1fs_mCherry*:

ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGATGGTGA
TGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTACTTACGGTAAA
TTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCAC
TACTTTCGGTTATGGTGTTCAATGTTTTGCTAGATACCCAGATCATATGAAACAACATGACT
TTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTCAAAGATGA
CGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAGAATC
GAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTGGAATACA
ACTATAACTCTCACAATGTTTACATCATGGCTGACAAACAAAAGAATGGTATCAAAGTTAA
CTTCAAAATTAGACACAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAA
AATACTCCAATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATC
TGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTGTTACT
GCTGCTGGTATTACCCATGGTATGGATGAATTGTACAAAGCTAGCAAACCTGGGTCAATCC
TTGG**GGCCCAGACTGAGCACGTGATGG**CAGAGCACAGGACGTC*ATGGTTTCAAAAGGTG*
AAGAAGATAATATGGCTATTATTAAAGAATTTATGAGATTTAAAGTTCATATGGAAGGTTCAGTTAAT
GGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGTAGACCATATGAAGGTACTCAAACTGCTAA
ATTGAAAGTTACTAAAGGTGGTCCATTACCATTTGCTTGGGATATTTTGTCACCACAATTTATGTAT
GGTTCAAAAGCTTATGTTAAACATCCAGCTGATATTCCAGATTATTTAAAATTGTCATTTCCAGAAG
GTTTTAAATGGGAAAGAGTTATGAATTTTGAAGATGGTGGTGTTGTTACTGTTACTCAAGATTCAT
CATTACAAGATGGTGAATTTATTTATAAAGTTAAATTGAGAGGTACTAATTTTCCATCAGATGGTCC
AGTTATGCAAAAAAAAACTATGGGTTGGGAAGCTTCATCAGAAAGAATGTATCCAGAAGATGGTG
CTTTAAAAGGTGAAATTAAACAAAGATTGAAATTAAAAGATGGTGGTCATTATGATGCTGAAGTTAA
AACTACTTATAAAGCTAAAAAACCAGTTCAATTACCAGGTGCTTATAATGTTAATATTAAATTGGATA
TTACTTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGAGCTGAAGGTAGACATTCAAC
*TGGTGGTATGGATGAATTATATAAAGGTACCGCTCGAGCAGCTGTGATTGATTGA*GTCGACTTG
GTTGAACACGTTGCCAAGGCTTAAGTGAATTTACTTTAAATCTTGCATTTAAATAAATTTTC
TTTTTATAGCTTTATGACTTAGTTTCAATTTATATACTATTTTAATGACATTTTCGATTCGGATC
CACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAG

[2051]

GGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCG
CTCACAATTCCACACAACATAGGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGGTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAG
CGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG
TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC
CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT
CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGT
TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC
CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTG
GTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
GGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTT
TGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGT
AGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
CAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCT
AGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCG
TGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCG
CGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC
TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
ATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT
GCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTC
AATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA
GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAACGAAG
CATCTGTGCTTCATTTTGTAGAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAA
GAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACG
AAGAATCTGTGCTTCATTTTTGTAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAA
CAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTTACC
AACAAAGAATCTATACTTCTTTTTTGTTCTACAAAAATGCATCCCGAGAGCGCTATTTTTCT
AACAAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTCTTGATAA
CTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTTTGGTGTCTATTTTCTCTTC
CATAAAAAAAGCCTGACTCCACTTCCCGCGTTTACTGATTACTAGCGAAGCTGCGGGTGCA
TTTTTTCAAGATAAAGGCATCCCCGATTATATTCTATACCGATGTGGATTGCGCATACTTTGT
GAACAGAAAGTGATAGCGTTGATGATTCTTCATTGGTCAGAAAATTATGAACGGTTTCTTC
TATTTTGTCTCTATATACTACGTATAGGAAATGTTTACATTTTCGTATTGTTTTCGATTCACTC
TATGAATAGTTCTTACTACAATTTTTTTGTCTAAAGAGTAATACTAGAGATAAACATAAAAA
ATGTAGAGGTCGAGTTTAGATGCAAGTTCAAGGAGCGAAAGGTGGATGGGTAGGTTATATA
GGGATATAGCACAGAGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGT
ATTCGCAATATTTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTTGAAAGTGCGTC
TTCAGAGCGCTTTTGGTTTTCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGA
ACTTCGGAATAGGAACTTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAAAATGCAACGCG
[2052] AGCTGCGCACATACAGCTCACTGTTCACGTCGCACCTATATCTGCGTGTTGCCTGTATATAT

ATATACATGAGAAGAACGGCATAGTGCGTGTTTATGCTTAAATGCGTACTTATATGCGTCTAT
TTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTATCCCATTCCATGCGGGGTATC
GTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTATATGCTGCCACTCCTCAATTGGATTAG
TCTCATCCTTCAATGCTATCATTTCCTTTGATATTGGATCATACTAAGAAACCATTATTATCAT
GACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGAT
GACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCG
GATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGG
CTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATCGACTACGTC
GTAAGGCCGTTTCTGACAGAGTAAAATTCTTGAGGGAACTTTCACCATTATGGGAAATGCT
TCAAGAAGGTATTGACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTT
GTATTTTTTTTTTTTTAGAGAAAATCCTCCAATATCAAATTAGGAATCGTAGTTTCATGATTT
TCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCAATATTAATGTTAAAGTGCA
ATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTACCTGTATTCCTTTACTATC
CTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATATAGTTTCGTCTACCCTATGAA
CATATTCCATTTTGTAATTTCGTGTCGTTTCTATTATGAATTTCATTTATAAAGTTTATGTACA
AATATCATAAAAAAAGAGAATCTTTTTAAGCAAGGATTTTCTTAACTTCTTCGGCGACAGC
ATCACCGACTTCGGTGGTACTGTTGGAACCACCTAAATCACCAGTTCTGATACCTGCATCC
AAAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTT
CAACATCATTGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCT
GGAGCAGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAG
GCCAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATCGGA
GATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTGGGTTGGGTTCTTAA
CTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACCTTCAATGTAGGGAATTCGTT
CTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGCCAAAAGATTAGCTTTAT
CCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGGCCATTCTTG
TGATTCTTTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACCATCG
TCTTCCTTTCTCTTACCAAAGTAAATACCTCCCACTAATTCTCTGACAACAACGAAGTCAG
TACCTTTAGCAAATTGTGGCTTGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTT
ACATGGTCTTAAGTTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAG
GTCTAACACTACCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAA
CCTTCTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGC
ACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGGAACGAACATCAGAAATAGCT
TTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACGTGGTCACCTGGCAAAACGA
CGATCTTCTTAGGGGCAGACATAGGGGCAGACATTAGAATGGTATATCCTTGAAATATATAT
ATATATTGCTGAAATGTAAAAGGTAAGAAAAGTTAGAAAGTAAGACGATTGCTAACCACCT
ATTGGAAAAAACAATAGGTCCTTAAATAATATTGTCAACTTCAAGTATTGTGATGCAAGCAT
TTAGTCATGAACGCTTCTCTATTCTATATGAAAAGCCGGTTCCGGCCTCTCACCTTTCCTTT
TTCTCCCAATTTTTCAGTTGAAAAAGGTATATGCGTCAGGCGACCTCTGAAATTAACAAAA
AATTTCCAGTCATCGAATTTGATTCTGTGCGATAGCGCCCCTGTGTGTTCTCGTTATGTTGA
GGAAAAAAATAATGGTTGCTAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTC
CCACAGTTAACTGCGGTCAAGATATTTCTTGAATCAGGCGCCTTAGACCGCTCGGCCAAAC
AACCAATTACTTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTTGAACACA
CATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGATTTTGATGTAATTGT
TGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACTAGAAGTTCTCCTCG
ACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG
AAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTT
TTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGG
GTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGT
CAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATC
AAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCG
ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCG
AAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCAC
ACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCG
CAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGG

[2053]

GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTG
TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACCGG
GCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGA
TCCGTTAGAATCATTTTGAATAAAAAACACGCTTTTTCAGTTCGAGTTTATCATTATCAATA
CTGCCATTTCAAAGAATACGTAAATAATTAATAGTAGTGATTTTCCTAACTTTATTTAGTCAA
AAAATTAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGGTTACAC
AGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTGGCATCCACTAAATATAA
TGGAGCCCGCTTTTTAAGCTGGCATCCAGAAAAAAAAAAGAATCCCAGCACCAAAATATTG
TTTTCTTCACCAACCATCAGTTCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGG
GCACAAACAGGCAAAAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTA
AATGATGACACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTTCTATTA
CCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGGTTGAAACCAGTTCCCTGAA
ATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAAA
TCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACC
AAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAGAATTC (SEQ ID NO: 732)

[2054]

[2055] 설명:

[2056] GFP 오픈 리딩 프레임

[2057] 정지 코돈 +1 프레임시프트, 또는 -1 프레임시프트를 함유하는 링커

[2058] *mCherry 오픈 리딩 프레임*

[2059] 플라스미드 백본 (GPD 프로모터, Leu2 마커, 및 AmpR 함유)

[2060] **프로토스페이서 (밑줄표시됨)**

[2061] **PAM (볼드체)**

[2062] 포유동물 프라임 편집제 플라스미드 및 예시적 PEgRNA 플라스미드의 DNA 서열

[2063] *pCMV-PE1:*

[2064]
ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCG
ACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCA
CCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCAC
AGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGC
CACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCT
ATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGAC
TGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC
AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG
AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC
ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC
ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAG
ACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAA
ACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
AGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTG
GCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCC
ATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCC
TCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTG
CGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTA
CGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCAT
CCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGC
TGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAG
CTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAA
AAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGA
AACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTT
CGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACT
TCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACT

[2065]

TCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCC
GCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCG
GAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATC
TGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGG
AAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGA
AAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACA
CCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGC
AAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTG
ATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCA
GGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGG
CATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGC
CCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAG
AACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGAT
CCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACT
ACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCG
ACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACA
AGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCC
AGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAG
GCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACA
GATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGT
GAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC
AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTG
GGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTAC
AAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTAC
CGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCC
AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGT
GTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAA
TATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAA
GAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCG
GCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCA
AGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGC
AGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAA
GGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAG
AATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAA
TGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGC
AGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGT
CCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACC
TGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGA
CCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC*TCTGGAGGAT*
CTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGC
*AGTGGCGGCAGCAGCGGCGGCAGCAGC***ACCCTAAATATAGAAGATGAGTATCGGCTACA**
TGAGACCTCAAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCC
TCAGGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCCTCTG
ATCATACCTCTGAAAGCAACCTCTACCCCCGTGTCCATAAAACAATACCCCATGTCAC
AAGAAGCCAGACTGGGGATCAAGCCCCACATACAGAGACTGTTGGACCAGGGAATA
CTGGTACCCTGCCAGTCCCCCTGGAACACGCCCCTGCTACCCGTTAAGAAACCAGG
GACTAATGATTATAGGCCTGTCCAGGATCTGAGAGAAGTCAACAAGCGGGTGGAAG
ACATCCACCCCACCGTGCCCAACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCC
ACCAGTGGTACACTGTGCTTGATTTAAAGGATGCCTTTTTCTGCCTGAGACTCCACC
CCACCAGTCAGCCTCTCTTCGCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA

[2066]

GGACAATTGACCTGGACCAGACTCCCACAGGGTTTCAAAAACAGTCCCACCCTGTT
TGATGAGGCACTGCACAGAGACCTAGCAGACTTCCGGATCCAGCACCCAGACTTGA
TCCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCCACTTCTGAGCTAGACTGCC
AACAAGGTACTCGGGCCCTGTTACAAACCCTAGGGAACCTCGGGTATCGGGCCTCG
GCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATCTTCTAAAA
GAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAGCCTAC
TCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGACGGCAGGCTTCTGTCGCC
TCTGGATCCCTGGGTTTGCAGAAATGGCAGCCCCCCTGTACCCTCTCACCAAAACGG
GGACTCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAATCAAGCAAG
CTCTTCTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTAAGCCCTTTGAACTCT
TTGTCGACGAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGACCT
TGGCGTCGGCCGGTGGCCTACCTGTCCAAAAAGCTAGACCCAGTAGCAGCTGGGTG
GCCCCCTTGCCTACGGATGGTAGCAGCCATTGCCGTACTGACAAAGGATGCAGGCA
AGCTAACCATGGGACAGCCACTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAG
TCAAACAACCCCCCGACCGCTGGCTTTCCAACGCCCGGATGACTCACTATCAGGCCT
TGCTTTTGGACACGGACCGGGTCCAGTTCGGACCGGTGGTAGCCCTGAACCCGGCT
ACGCTGCTCCCACTGCCTGAGGAAGGGCTGCAACACAACTGCCTTGATATCCTGGC
CGAAGCCCACGGAACCCGACCCGACCTAACGGACCAGCCGCTCCCAGACGCCGACC
ACACCTGGTACACGGATGGAAGCAGTCTCTTACAAGAGGGACAGCGTAAGGCGGGA
GCTGCGGTGACCACCGAGACCGAGGTAATCTGGGCTAAAGCCCTGCCAGCCGGGAC
ATCCGCTCAGCGGGCTGAACTGATAGCACTCACCCAGGCCCTAAAGATGGCAGAAG
GTAAGAAGCTAAATGTTTATACTGATAGCCGTTATGCTTTTGCTACTGCCCATATCCAT
GGAGAAATATACAGAAGGCGTGGGTTGCTCACATCAGAAGGCAAAGAGATCAAAAA
TAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTTTCTGCCCAAAAGACTTAGCAT
AATCCATTGTCCAGGACATCAAAAGGGACACAGCGCCGAGGCTAGAGGCAACCGGA
TGGCTGACCAAGCGGCCCGAAAGGCAGCCATCACAGAGACTCCAGACACCTCTACC
CTCCTCATAGAAAATTCATCACCCTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGC
GAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCATTG
AGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGC
CCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAA
ATGAGAAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGG
GCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCT
AGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATT
CCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGC
TAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT
CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG
CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTT
TTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG
CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAA
GCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTAT
CGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAAC
TACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCG
GAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCT
TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA
GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC
TAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT
CTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA

CGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCT
CACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGT
GGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAA
GTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTC
ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACA
TGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAA
GTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACA
TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAG
GATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCA
GCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCA
AAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATT
ATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGAT
CGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGC
ATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGC
AAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGG
GTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATT
GACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCG
CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTG
ACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATG
GGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGT
ACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA
CCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA
TGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG
TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG
GTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTA
ATACGACTCACTATAGGGAGAGCCGCCACC (SEQ ID NO: 733)

[2067]

[2068] *pCMV-PE2*:

ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCG
ACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCA
CCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCAC
AGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGC
CACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCT
ATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGAC
TGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC
AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG
AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC
ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC
ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAG
ACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAA
ACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
AGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTG
GCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCC
ATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCC
TCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTG
CGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTA
CGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCAT
CCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGC
TGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAG

[2069]

CTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAA
AAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGA
AACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTT
CGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACT
TCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACT
TCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCC
GCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCG
GAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATC
TGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGG
AAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGA
AAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACA
CCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGC
AAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTG
ATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCA
GGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGG
CATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGC
CCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAG
AACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGAT
CCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACT
ACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCG
ACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACA
AGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGA
GGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCC
AGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAG
GCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACA
GATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGT
GAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC
AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTG
GGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTAC
AAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTAC
CGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCC
AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGT
GTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAA
TATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAA
GAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCG
GCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCA
AGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGC
AGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAA
GGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAG
AATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAA
TGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGC
AGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGT
CCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACC
TGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGA
CCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCA
TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACTCTGGAGGAT
CTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAG
AGCAGTGGCGGCAGCAGCGGCGGCAGCAGCACCCTAAATATAGAAGATGAGTATCGGCTA
CATGAGACCTCAAAAGAGCCAGATGTTTCTCTAGGGTCCACATGGCTGTCTGATTTTCCTC
AGGCCTGGGCGGAAACCGGGGGCATGGGACTGGCAGTTCGCCAAGCTCCTCTGATCATAC
CTCTGAAAGCAACCTCTACCCCCGTGTCCATAAAACAATACCCCATGTCACAAGAAGCCA
[2070] GACTGGGGATCAAGCCCCACATACAGAGACTGTTGGACCAGGGAATACTGGTACCCTGCC

AGTCCCCCTGGAACACGCCCCTGCTACCCGTTAAGAAACCAGGGACTAATGATTATAGGCC
TGTCCAGGATCTGAGAGAAGTCAACAAGCGGGTGGAAGACATCCACCCCACCGTGCCCA
ACCCTTACAACCTCTTGAGCGGGCTCCCACCGTCCCACCAGTGGTACACTGTGCTTGATTT
AAAGGATGCCTTTTTCTGCCTGAGACTCCACCCCACCAGTCAGCCTCTCTTCGCCTTTGAG
TGGAGAGATCCAGAGATGGGAATCTCAGGACAATTGACCTGGACCAGACTCCCACAGGGT
TTCAAAAACAGTCCCACCCTGTTTAATGAGGCACTGCACAGAGACCTAGCAGACTTCCGG
ATCCAGCACCCAGACTTGATCCTGCTACAGTACGTGGATGACTTACTGCTGGCCGCCACTT
CTGAGCTAGACTGCCAACAAGGTACTCGGGCCCTGTTACAAACCCTAGGGAACCTCGGGT
ATCGGGCCTCGGCCAAGAAAGCCCAAATTTGCCAGAAACAGGTCAAGTATCTGGGGTATC
TTCTAAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAAGAGACTGTGATGGGGCAG
CCTACTCCGAAGACCCCTCGACAACTAAGGGAGTTCCTAGGGAAGGCAGGCTTCTGTCGC
CTCTTCATCCCTGGGTTTGCAGAAATGGCAGCCCCCCTGTACCCTCTCACCAAACCGGGG
ACTCTGTTTAATTGGGGCCCAGACCAACAAAAGGCCTATCAAGAAATCAAGCAAGCTCTT
CTAACTGCCCCAGCCCTGGGGTTGCCAGATTTGACTAAGCCCTTTGAACTCTTTGTCGACG
AGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACTGGGACCTTGGCGTCGGCCG
GTGGCCTACCTGTCCAAAAAGCTAGACCCAGTAGCAGCTGGGTGGCCCCCTTGCCTACGG
ATGGTAGCAGCCATTGCCGTACTGACAAAGGATGCAGGCAAGCTAACCATGGGACAGCCA
CTAGTCATTCTGGCCCCCCATGCAGTAGAGGCACTAGTCAAACAACCCCCCGACCGCTGG
CTTTCCAACGCCCGGATGACTCACTATCAGGCCTTGCTTTTGGACACGGACCGGGTCCAGT
TCGGACCGGTGGTAGCCCTGAACCCGGCTACGCTGCTCCCACTGCCTGAGGAAGGGCTGC
AACACAACTGCCTTGATATCCTGGCCGAAGCCCACGGAACCCGACCCGACCTAACGGACC
AGCCGCTCCCAGACGCCGACCACACCTGGTACACGGATGGAAGCAGTCTCTTACAAGAG
GGACAGCGTAAGGCGGGAGCTGCGGTGACCACCGAGACCGAGGTAATCTGGGCTAAAGC
CCTGCCAGCCGGGACATCCGCTCAGCGGGCTGAACTGATAGCACTCACCCAGGCCCTAAA
GATGGCAGAAGGTAAGAAGCTAAATGTTTATACTGATAGCCGTTATGCTTTTGCTACTGCCC
ATATCCATGGAGAAATATACAGAAGGCGTGGGTGGCTCACATCAGAAGGCAAAGAGATCA
AAAATAAAGACGAGATCTTGGCCCTACTAAAAGCCCTCTTTCTGCCCAAAAGACTTAGCAT
AATCCATTGTCCAGGACATCAAAAGGGACACAGCGCCGAGGCTAGAGGCAACCGGATGG
CTGACCAAGCGGCCCGAAAGGCAGCCATCACAGAGACTCCAGACACCTCTACCCTCCTCA
TAGAAAATTCATCACCCTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGC
CCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCATTGAGTTTAAACCCG
CTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGAAAATTG
CATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCA
AGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTT
CTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGT
AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATA
CGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTA
ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAT
GAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGC
TCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAG
GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCC
GCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGA
CCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCA
TAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTG
CACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT
[2071] GACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG

ATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTC
TATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGC
TTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATT
TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTAT
CCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAAT
AGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTG
CAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGT
GTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGAT
GCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACC
GAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAA
GTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGA
GATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGC
GACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG
GTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTT
CCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCC
GATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCT
GCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCT
GCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTA
ATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGT
AAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTAT
GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGT
CAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTA
CTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC
ATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG
TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGC
TGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGG
GAGAGCCGCCACC (SEQ ID NO: 734)

[2072]

[2073] *N-말단 NLS + Cas9 H840A*

[2074] *가요성 링커*

[2075] **M-MLV 리버스 트랜스크립타제 + C-말단 NLS**

[2076] 플라스미드 백본 (CMV 프로모터 및 AmpR 함유)

[2077] *pU6-HEK3_PEgRNA_CTTins:*

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGAT
AATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAG
TAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTT
ACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAC
C*GGCCCAGACTGAGCACGTGA*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
CGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC**TCTGCCATCAAAGCGTGCTCAGT**
CTGTTTTTTTAAGCTTGGGCCGCTCGAGGTACCTCTCTACATATGACATGTGAGCAAAAGG
CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG
CCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAT
AGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAA
CCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC

[2078]

GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG
TAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGA
TCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGT
AAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCT
ATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCT
TACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTT
ATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTAT
GGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC
AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTG
TTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGC
TTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGA
GTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAG
ATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCA
GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCG
ACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGT
TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC
GCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGCTAGCTGTACAAAAAAGCAGGCTTT
AAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGGAA (SEQ ID NO:
735)

[2079]

[2080] U6 프로모터 서열

[2081] *스페이서 서열*

[2082] sgRNA 스캐폴드

[2083] **3' 연장부 (PBS 및 RT 주형 함유)**

[2084] 백본 (AmpR 함유)

[2085] *pLenti-hSyn-N-PE2-NpuN-P2A-GFP-KASH_U6-DNMT1-PEgRNA:*

GTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCC
GCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCA
AAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAG
GCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTT
ATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAAC
TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGAC
GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAAT
GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCA
GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAA
TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTTTGCCTGTACTGGGTCTC
TCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCC
TCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAAC
TAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGG
ACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
CGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGG
CTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGCGATG
GGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAATATAAATTAAAACATATAGTATGGGC

[2086]

AAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAG
ACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATAT
AATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTT
TAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCTGAT
CTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTA
GTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGA
AAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTAT
GGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAG
CAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGC
ATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTG
GGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGA
GTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTA
ACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATG
AACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGG
CTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCT
GTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCC
AACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGA
GACAGATCCATTCGATTAGTGAACGGATCGGCACTGCGTGCGCCAATTCTGCAGACAAATGGCA
GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATA
GTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAA
ATTTTCGGGTTTATTACAGGGACAGCAGAGATCCAGTTTGGTTAATTAAGGTACCAAGGTCGGG
CAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGA
GATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGT
AATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGT
AACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACC**GCGGGC**
TGGAGCTGTTCGCGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTA
TCAACTTGAAAAAGTGGCACCGAGTCGGTGCAAGATGCAAGCGCGAACAGCTCCAG<u>TTT</u>
<u>TTTTGAATTC</u>*AGTGCAAGTGGGTTTTAGGACCAGGATGAGGCGGGGTGGGGGTGCCTACCTGACG*
ACCGACCCCGACCCACTGGACAAGCACCCAACCCCCATTCCCCAAATTGCGCATCCCCTATCAGAGA
GGGGGAGGGGAAACAGGATGCGGCGAGGCGCGTGCGCACTGCCAGCTTCAGCACCGCGGACAGT
GCCTTCGCCCCCGCCTGGCGGCGCGCGCCACCGCCGCCTCAGCACTGAAGGCGCGCTGACGTCA
CTCGCCGGTCCCCCGCAAACTCCCCTTCCCGGCCACCTTGGTCGCGTCCGCGCCGCCGGCCC
AGCCGGACCGCACCACGCGAGGCGCGAGATAGGGGGGCACGGGCGCGACCATCTGCGCTGCGGC
GCCGGCGACTCAGCGCTGCCTCAGTCTGCGGTGGGCAGCGGAGGAGTCGTGTCGTGCCTGAGAGC
*GCAG**TCGAGAATCTAGAGCGCTGCCACC***ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTC
ACCAAAGAAGAAGCGGAAAGTCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACT
CTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGG
GCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGC
GAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGA
ACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCT
TCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCT
TCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAA
AGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACA
TGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGG
ACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACG
CCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAA
AATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTG
AGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAG
CTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTAC
GCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGA
GTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCAC
CACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGA
GATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGA
AGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGT

[2087]

GAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCC
ACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCT
GAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCC
TCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCC
CCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGA
CCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGT
ACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCG
CCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAA
GTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAG*TGCCTGTCCTACGAGACA*
GAGATCCTGACAGTGGAGTATGGCCTGCTGCCAATCGGCAAGATCGTGGAGAAGAGGATCGAGTGTA
CCGTGTACTCTGTGGATAACAATGGCAACATCTATACACAGCCCGTGGCACAGTGGCACGATAGGGG
AGAGCAGGAGGTGTTCGAGTATTGCCTGGAGGACGGCAGCCTGATCAGGGCAACCAAGGACCACAA
GTTCATGACAGTGGATGGCCAGATGCTGCCCATCGACGAGATTTTCGAGCGGGAGCTGGACCTGAT
*GAGAGTGGATAACCTGCCTAAT**AGCGGAGGCAGTAAAAGAACAGCAGACGGGAGTGAGTTTGAGC***
CCAAGAAAAAGAGAAAGGTGGGATCCGGCGCAACAAACTTCTCTCTGCTGAAACAAGCCGGAGA
TGTCGAAGAGAATCCTGGACCGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT
CCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGG
CGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT
CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGT
GAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGT
GAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCG
CCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGAGGAGGAGGAG
GAGACAGACAGCAGGATGCCCCACCTCGACAGCCCCGGCAGCTCCCAGCCGAGACGCTCCTTCC
TCTCAAGGGTGATCAGGGCAGCGCTACCGTTGCAGCTGCTTCTGCTGCTGCTGCTGCTCCTGGC
CTGCCTGCTACCTGCCTCTGAAGATGACTACAGCTGCACCCAGGCCAACAACTTTGCCCGATCCT
TCTACCCCATGCTGCGGTACACCAACGGGCCACCTCCCACCTAAACGCGTTAAGTCGACAATCA
ACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCT
ATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCC
TCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGG
CGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAG
CTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCT
TGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAA
ATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT
GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCG
GCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCG
CGTCGACTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAA
GGGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGG
GTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTG
GTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGGGCCCGTTTAAA
CCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATC
GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGA
AAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGC
GGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC
GCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCT
CCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
[2088] GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT

TCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTT
GATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATT
TAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAG
CAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAG
GCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCC
CCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGAC
TAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGA
GGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCG
GATCTGATCAGCACGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAA
GGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGT
CGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGA
CTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGT
GCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGT
CGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGC
AGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCC
GAGGAGCAGGACTGACACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTG
GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGG
AGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATC
ACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAA
TGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCT
GTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAG
TGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCG
CTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAG
GCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG
GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG
TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCT
CGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAA
GCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAG
CTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACA
CTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG
TAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAG
ATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTC
AGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA
CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGT
TGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCT
GCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCG
GAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA
GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAA
GGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGT
TGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTA
CTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGA
ATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATA
GCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTT
ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTA
CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATA
AGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAG
GGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGAC (SEQ ID NO: 736)

[2089]

[2090] U6프로모터

[2091] PEgRNA

[2092] *hSyn프로모터*

[2093] N-termPE2

[2094] *N-termNpu*

[2095] *P2A-GFP-KASH*

[2096] *pLenti-hSyn-C-PE2-NpuC:*

GTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCC
GCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCA
AAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAG
GCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTT
ATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAAC
TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGAC
GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAAT
GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCA
GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAA
TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTTTGCCTGTACTGGGTCTC
TCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCC
TCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAAC
TAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGG
ACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCG
CGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGG
CTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGCGATG
GGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAATATAAATTAAAACATATAGTATGGGC
AAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAG
ACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATAT
AATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTT
TAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCTGAT
CTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTA
GTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGA
AAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTAT
GGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAG
CAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGC
ATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTG
GGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGA
GTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTA
ACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATG
AACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGG
CTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCT
GTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCC
AACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGA
GACAGATCCATTCGATTAGTGAACGGATCGGCACTGCGTGCGCCAATTCTGCAGACAAATGGCA
GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATA
GTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAA
ATTTTCGGGTTTATTACAGGGACAGCAGAGATCCAGTTTGGTTAATTAAGGTACCAGTGCAAGT
GGGTTTTAGGACCAGGATGAGGCGGGGTGGGGGTGCCTACCTGACGACCGACCCCGACCCACT
GGACAAGCACCCAACCCCCATTCCCCAAATTGCGCATCCCCTATCAGAGAGGGGGAGGGGAAA
CAGGATGCGGCGAGGCGCGTGCGCACTGCCAGCTTCAGCACCGCGGACAGTGCCTTCGCCCCC
GCCTGGCGGCGCGCGCCACCGCCGCCTCAGCACTGAAGGCGCGCTGACGTCACTCGCCGGTCC

[2097]

CCCGCAAACTCCCCTTCCCGGCCACCTTGGTCGCGTCCGCGCCGCCGCCGGCCCAGCCGGACC
GCACCACGCGAGGCGCGAGATAGGGGGGCACGGGCGCGACCATCTGCGCTGCGGCGCCGGCG
ACTCAGCGCTGCCTCAGTCTGCGGTGGGCAGCGGAGGAGTCGTGTCGTGCCTGAGAGCGCAG*T*
CGAGAATCTAGAGCGCTGCCACCATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGA
*AGAAGCGGAAAGTC*ATCAAGATTGCTACACGGAAATACCTGGGAAAGCAGAACGTGTACGACATCGG
*CGTGGAGCGGGATCACAACTTCGCCCTGAAGAATGGCTTTATCGCCAGCAAT***TGCTTCGACTCCG**
TGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGC
TGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAA
GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAA
AACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACA
CCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGG
CAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCT
GATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCC
AGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAG
GGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACA
AGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAG
AAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCC
AGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTG
TACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTG
TCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGAC
AACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCG
AAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATT
ACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGG
ATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTG
GCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCG
GGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCC
AGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACG
CCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTAC
GGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGG
CAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGAT
TACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCG
GGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATG
CCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTC
TATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTA
AGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAA
GTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCAT
CATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAA
ACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGC
CCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGG
CTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGG
ACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATC
TGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCC
GAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTAC
TTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCAC
CCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGG
GAGGTGACTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAA
GAAGAGGAAAGTCTAAACGCGTTAAGTCGACAATCAACCTCTGGATTACAAAATTTGTGAAA
GATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTT
GTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCT
CTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACG
CAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCC
CCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
[2098] CTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCG

CCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCC
AGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGC
CCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCGTCGACTTTAAGACCAATGACTTA
CAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCA
CTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGGGTCTCTCTGGTTAGACCAGATCTGA
GCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAG
TGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTT
TTAGTCAGTGTGGAAAATCTCTAGCAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCA
CTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCT
ATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCA
TGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGG
GTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGT
GACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCA
CGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCT
TTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCT
GATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA
ACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCG
GCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTG
TGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCAT
CTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAA
AGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAAC
TCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCG
AGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTT
TTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGTTGACAATTA
ATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGT
TGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCG
ACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACG
TGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGT
GGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGG
GACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCT
GCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTGCTAC
GAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGC
CGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTT
ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTT
TTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTC
GACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC
TCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAG
TGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG
CCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACT
CAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA
AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCC
GCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGC
CGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGC
TGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG
TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGA
CTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT
ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCG
CTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA
CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTC
AAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGG
GATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTT

[2099]

TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG
GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCT
CACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCG
CCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTT
TGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGT
GCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTT
ATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTC
TGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCT
TGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTA
ACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCA
AAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACT
CATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTG
CCACCTGAC (SEQ ID NO: 737)

[2100]

[2101] hSyn프로모터

[2102] *C-termNpu*

[2103] C-termwtCas9

[2104] *pLenti-U6-DNMT1_닉킹_sgRNA:*

TAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGA
ATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAG
ACCCACCTCCCAACCCCGAGGGGACCCAGA*GAGGGCCTATTTCCCATGATTCCTTCATATTTGCAT*
ATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAA
TACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATC
ATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAC
CGCCGCGCGCGCGAAAAAGCCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTAAGCTTGGCGTAACTAGATCTTGAGA
CAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGG
GAAAGAATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACA
AAAATTCAAAATTTTCGGGTTTATTACAGGGACAGCAGAGATCCACTTTGGCGCCGGCTCGAGG
GGGCCCGGGTGCAAAGATGGATAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGACCT
TCTAGGTCTTGAAAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGC
CCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGATCCGGTGCCTAGAGAAGGTGG
CGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGA
GAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAA
CACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGT
GCCTTGAATTACTTCCACCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGT
GGGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCC
TGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTT
TCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGAT
AGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCG
ACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCG
AGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCG
TGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGA
TGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCG
GGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGA
CTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCG
TCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTG
AAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGATCT
TGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCAGGTGTCGTGACG
TACGGCCACCATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCAG

[2105]

GGCCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCC
GGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGCTCGA
CATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCCGTGGCGGTCTGGACCACGCCGGAGA
GCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCC
GGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCG
TGGTTCCTGGCCACCGTCGGAGTCTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTC
GTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCCTGGAGACCTCCGC
GCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTGCC
CGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGAACGCGTTAAGTCGACA
ATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTA
CGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAAC
GTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCT
GTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCC
TGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCG
GGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGT
CCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGC
TCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCC
TCCCCGCGTCGACTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAA
AGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTG
TACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCA
CTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTG
ACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTACGTAT
AGTAGTTCATGTCATCTTATTATTCAGTATTTATAACTTGCAAAGAAATGAATATCAGAGAGTGAG
AGGAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAA
TAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGT
CTGGCTCTAGCTATCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCA
TTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTG
AGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGGACGTACCCAATTCGCCCTAT
AGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTG
GCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGA
GGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTG
TAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
CGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCC
GTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCC
CAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGC
CCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCA
ACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAA
ATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGTG
GCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGT
ATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGT
ATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATC
GAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGA
TGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCA
ACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAG
CATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACA
CTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACA
ACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAA
CGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGG
CGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCA
GGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTG
AGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGT
[2106] TATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGT

GCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTA
AAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATC
CCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTT
GAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGT
GGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCG
CAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGG
GGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGC
GTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGC
GGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTA
TAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
TTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGT
GAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG
GAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGC
ACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCA
CTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGC
GGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTC
ACTAAAGGGAACAAAAGCTGGAGCTGCAAGCTTAATGTAGTCTTATGCAATACTCTTGTAGTCT
TGCAACATGGTAACGATGAGTTAGCAACATGCCTTACAAGGAGAGAAAAAGCACCGTGCATGC
CGATTGGTGGAAGTAAGGTGGTACGATCGTGCCTTATTAGGAAGGCAACAGACGGGTCTGACAT
GGATTGGACGAACCACTGAATTGCCGCATTGCAGAGATATTGTATTTAAGTGCCTAGCTCGATAC
ATAAACGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACC
CACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTG
TGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGC
GCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGG
CTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTT
GACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAA
TTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAAC
ATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATC
AGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACT
TAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACA
CCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAA
GCGGCCGCTGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATAT
AAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTG
GTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGC
AGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGG
TATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTC
ACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGAT
CAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGA
ATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGA
CAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAA
GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAA
CA (SEQ ID NO: 738)

[2107]

[2108] *U6프로모터*

[2109] ***sgRNA***

[2110] 본원에 사용된 말로니 뮤린 백혈병 바이러스 리버스 트랜스크립타제 (M-MLV RT) 변이체의 아미노산 서열.

[2111] PE1 M-MLV RT:

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY
PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP
TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQ
GFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYR
ASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP
GFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK
GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV
EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA
HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI
ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLP
KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 739)

[2112]

[2113] M3 M-MLV RT (D200N, T330P, L603W) (문헌 [Baranauskas et al.][182] 참조):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY
PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP
TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQ
GFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYR
ASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP
GFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK
GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV
EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA
HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI
ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP
KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 740)

[2114]

[2115] PE2 M-MLV RT (D200N, T306K, W313F, T330P, L603W):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY
PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP
TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQ
GFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYR
ASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIP
GFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK
GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV
EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA
HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI
ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP
KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 741)

[2116]

[2117] M3-기능상실RT M-MLV RT:

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY
PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKLPGTNDYSPVQDLREVNKRVEDIHP
TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQ
GFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYR
ASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIP
GFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAK
GVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV
EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA
HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI
ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLP
KRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 742)

[2118]

[2119] 실시예 2에 대한 참고문헌

[2120] 각각의 하기 참고문헌이 실시예 12에 인용되며, 이들 각각은 본원에 참조로 포함된다.

1. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).

2. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

3. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819– 823 (2013).

4. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).

5. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **154**, 1370–1379 (2013).

6. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).

7. Orlando, S. J. *et al.* Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.* **38**, e152–e152 (2010).

8. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPRCas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).

9. Suzuki, K. *et al. In vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).

10. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).

11. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).

12. Ihry, R. J. *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).

13. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci.* **91**, 6064–6068 (1994).

14. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell* **47**, 497–510 (2012).

15. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).

16. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).

17. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).

18. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).

19. Rees, H. A., Yeh, W.-H. & Liu, D. R. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* **10**, 1–12 (2019).

20. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).

21. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770 (2018).

[2121]

22. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

23. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

24. Gao, X. *et al.* Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217–221 (2018).

25. Ingram, V. M. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* **178**, 792–794 (1956).

26. Myerowitz, R. & Costigan, F. C. The major defect in Ashkenazi Jews with Tay-Sachs disease is an insertion in the gene for the alpha-chain of beta-hexosaminidase. *J. Biol. Chem.* **263**, 18587–18589 (1988).

27. Zielenski, J. Genotype and Phenotype in Cystic Fibrosis. *Respiration* **67**, 117–133 (2000).

28. Mead, S. *et al.* A Novel Protective Prion Protein Variant that Colocalizes with Kuru Exposure. *N. Engl. J. Med.* **361**, 2056–2065 (2009).

29. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).

30. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).

31. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).

32. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).

33. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).

34. Liu, Y., Kao, H.-I. & Bambara, R. A. Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem.* **73**, 589–615 (2004).

35. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

36. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

37. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12**, 664–670 (2015).

38. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 15939 (2017).

39. Jinek, M. *et al.* Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**, 1247997 (2014).

40. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**, 935–949 (2014).

41. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).

42. Baranauskas, A. *et al.* Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng. Des. Sel.* **25**, 657–668 (2012).

[2122]

43. Gerard, G. F. *et al.* The role of template-primer in protection of reverse transcriptase from thermal inactivation. *Nucleic Acids Res.* **30**, 3118–3129 (2002).

44. Arezi, B. & Hogrefe, H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res.* **37**, 473–481 (2009).

45. Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M. & Gerard, G. F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res.* **16**, 265–277 (1988).

46. Thuronyi, B. W. *et al.* Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0193-0

47. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).

48. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4172

49. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off target effects. *Nature* **529**, 490–495 (2016).

50. Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).

51. Jin, S. *et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).

52. Kim, D., Kim, D., Lee, G., Cho, S.-I. & Kim, J.-S. Genome-wide target specificity of CRISPR RNA guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).

53. Grunewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).

54. Zhou, C. *et al.* Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).

55. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **5**, eaax5717 (2019).

56. Ostertag, E. M. & Kazazian Jr, H. H. Biology of Mammalian L1 Retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).

57. Griffiths, D. J. Endogenous retroviruses in the human genome sequence. *Genome Biol.* **2**, REVIEWS1017 (2001).

58. Berkhout, B., Jebbink, M. & Zsiros, J. Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus. *J. Virol.* **73**, 2365–2375 (1999).

59. Halvas, E. K., Svarovskaia, E. S. & Pathak, V. K. Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity. *J. Virol.* **74**, 10349–10358 (2000).

60. Dever, D. P. *et al.* CRISPR/Cas9 Beta-globin Gene Targeting in Human Hematopoietic Stem Cells. *Nature* **539**, 384–389 (2016).

61. Park, S. H. *et al.* Highly efficient editing of the β-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease. *Nucleic Acids Res.* doi:10.1093/nar/gkz475

62. Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. *Annu. Rev. Neurosci.* **24**, 519–550 (2001).

63. Asante, E. A. *et al.* A naturally occurring variant of the human prion protein completely prevents prion disease. *Nature* **522**, 478–481 (2015).

[2123]

64. Zettler, J., Schutz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).

65. Kugler, S., Kilic, E. & Bahr, M. Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. *Gene Ther.* **10**, 337–347 (2003).

66. de Felipe, P., Hughes, L. E., Ryan, M. D. & Brown, J. D. Co-translational, intraribosomal cleavageof polypeptides by the foot-and-mouth disease virus 2A peptide. *J. Biol. Chem.* **278**, 11441–11448 (2003).

[2124]

[2125] 1. 돌연변이유발 DNA 가닥의 가이드 RNA-주형화된 역전사를 확립한다.

[2126] 배경 및 근거

[2127] 제안된 게놈 편집 전략에서, Cas9-닉킹된 비-표적 DNA 가닥 (R-루프를 형성하는 PAM-함유 가닥)은 DNA 합성을 위한 프라이머로서 작용한다. 이는 생화학적 및 구조적 데이터의 여러 조각에 기초하여 가능한 것으로 가설화된다. 뉴클레아제 보호 실험[32], 결정학적 연구[33], 및 염기 편집 윈도우[4,24]는 Cas9-결합된 복합체의 소위 R-루프

내에서 비-표적 가닥 뉴클레오티드 -20 내지 -10에 대한 큰 정도의 유연성 및 불규칙성을 입증하였다 (넘버링은 제1 PAM 뉴클레오티드로부터 5' 거리를 나타냄). 또한, 절단된 비-표적 가닥의 PAM-원위 부분은 상보적 ssDNA가 트랜스로 부가될 경우 치밀하게 결합된 3원 복합체로부터 대체될 수 있다[20]. 이들 연구는 비-표적 가닥이 고도로 유연하고, 효소에 접근가능하고, 닉킹 후에 PAM-원위 단편의 3' 말단이 Cas9 해리 전에 방출된다는 것을 지지한다. 또한, gRNA는 DNA 합성의 주형화를 위해 연장될 수 있는 것으로 가설화된다. 선행 연구는 SpCas9, SaCas9, 및 LbCas12a (이전에 Cpf1)에 대한 gRNA가 RNA 압타머[34], 리간드-유도성 자기-절단 리보자임[35], 및 긴 비-코딩 RNA[36]를 갖는 gRNA 연장부를 용인한다는 것을 제시하였다. 이 문헌은 이용될 2가지 주요 특색에 대한 선례를 확립한다. 이 전략을 평가하는데 있어서, 여러 CRISPR-Cas 시스템은 시험관내 및 세포 검정의 조합을 사용하여 5' 및 3' 연장된 gRNA 설계와 함께 평가될 것이다 (도 2a-2c).

[2128] TRT 편집을 위한 조작된 gRNA에 대한 설계는 도 3a-3b에 제시된다. DNA 합성은 5'에서 3'으로 진행되고, 따라서 RNA 주형을 3'에서 5' 방향으로 카피한다. 5' 연장부를 위한 설계는 링커 영역, 닉킹된 DNA 가닥이 어닐링되는 프라이머 결합 부위, 및 역전사에 의한 DNA 합성을 위한 주형을 함유한다. 3' 연장된 gRNA는 프라이머 결합 부위 및 역전사 주형을 함유한다. 일부 경우에, gRNA 코어의 3' RNA 헤어핀은 DNA 표적 서열과 매칭되도록 변형되는데, 이는 시험관내 실험이 역전사가 3' 연장된 gRNA 구축물의 경우에 gRNA 코어 내로 ~3개의 뉴클레오티드를 연장시킨다는 것을 보여주었기 때문이다 (헤어핀 RNA 구조를 유지하는 보상적 변화가 이루어지는 한, 헤어핀 서열의 변형은 잘 용인되는 것으로 보임). DNA 합성은 5'에서 3'으로 진행되고, 뉴클레오티드는 성장하는 DNA 가닥의 3' OH에 부가된다.

[2129] 예비 결과

[2130] Cas9 닉킹된 DNA는 gRNA 주형의 역전사를 프라이밍한다. 닉킹된 비-표적 DNA 가닥의 접근성을 평가하기 위해, 에스. 피오게네스로부터의 Cas9 뉴클레아제 (SpCas9) 및 Cy5 형광 표지된 듀플렉스 DNA 기질 (51개 염기 쌍)을 사용하여 시험관내 생화학적 검정을 수행하였다. 먼저, 다양한 합성 주형 길이를 갖는 5' 연장부를 함유하는 일련의 gRNA를 시험관내 전사에 의해 제조하였다 (전체 설계는 도 2b에 제시됨). 뉴클레아제 기능상실 Cas9 (dCas9)를 사용한 전기영동 이동성 변화 검정 (EMSA)은 5' 연장된 gRNA가 표적 결합 친화도를 유지한다는 것을 확립하였다 (데이터는 제시되지 않음). 다음으로, dCas9, 5'-연장된 gRNA, 및 몰로니-뮤린 백혈병 바이러스 (M-MLV) 리버스 트랜스크립타제 (슈퍼스크립트 III)를 사용하여 사전-닉킹된 Cy5-표지된 듀플렉스 DNA 기질 상에서 TPRT 활성을 시험하였다. 37℃에서 1시간 인큐베이션 후에, 변성 폴리아크릴아미드 겔 전기영동 (PAGE)에 의해 생성물을 평가하고, Cy5 형광을 사용하여 영상화하였다 (도 4a). 각각의 5'-연장된 gRNA 변이체는 유의한 생성물 형성으로 이어졌고, 관찰된 DNA 생성물 크기는 연장부 주형의 길이와 일치하였다 (도 4b). 중요한 것으로, dCas9의 부재 하에, 사전-닉킹된 기질이 DNA 기질의 전장 51-bp 길이까지 연장되었으며, 이는 gRNA가 아닌 상보적 DNA 가닥이 dCas9가 존재하지 않는 경우에 DNA 합성을 위한 주형으로서 사용되었음을 강하게 시사한다 (도 4c). 시스템은 새로 합성된 DNA 가닥이 표적 부위 편집에 요구될 생성물 (단일 뉴클레오티드 변화를 갖는 상동 가닥)을 반영하도록 설계되었음에 주목한다. 이 결과는 Cas9:gRNA 결합이 닉킹된 비-표적 가닥의 3' 단부를 노출시키고, 비-표적 가닥이 역전사에 접근가능하다는 것을 확립한다.

[2131] 다음으로, 비-표적 DNA 가닥을 닉킹하는 Cas9(H840A) 돌연변이체를 사용하여 비-닉킹된 dsDNA 기질을 평가하였다. 먼저, 5'-연장된 gRNA와 함께 Cas9(H840A) 닉카제 활성을 시험하기 위해, 이전에 기재된 바와 같이 시험관내 절단 검정을 수행하였다[37]. 닉킹은 표준 gRNA와 비교하여 손상되었지만, 상당한 절단 생성물이 형성되었다 (도 4d). 중요한 것으로, 5'-연장된 gRNA 및 Cas9(H840A)를 이용하여 TPRT 반응을 수행한 경우에도, 아마도 감소된 닉킹 활성에 의해 설명될 비록 보다 낮은 수율이기는 하지만, RT 생성물이 관찰되었다 (도 4d). 이 결과는 5'-연장된 gRNA:Cas9(H840A) 복합체가 DNA를 닉킹하고 역전사의 주형이 될 수 있다는 것을 확립한다.

[2132] 마지막으로, 3' gRNA 연장부를 Cas9(H840A) 닉킹 및 TPRT에 대해 평가하였다. 5'-연장된 gRNA와 비교하여, 3'-연장된 gRNA에 의한 DNA 절단은 표준 gRNA와 비교하여 어떠한 검출가능한 정도로 손상되지 않았다. 유의하게, 3'-연장된 gRNA 주형은 또한 M-MLV RT가 트랜스로 공급된 경우에 사전-닉킹된 것 및 무손상 듀플렉스 DNA 기질 둘 다에 의한 효율적인 역전사를 지지하였다 (도 4e). 놀랍게도, 3'-연장된 주형에 대해 단일 생성물만이 관찰되었으며, 이는 역전사가 gRNA 스캐폴드를 따라 특정 위치에서 종결된다는 것을 나타낸다. 말단 트랜스퍼라제에 의한 생성물의 단독중합체 테일링에 이은 클레나우 연장 및 생어 서열분석은 전장 gRNA 합성 주형이 gRNA 코어의 말단 3개 뉴클레오티드에 더하여 카피되었음을 밝혀내었다. 향후, 플랩 말단은 말단 gRNA 서열을 변형시키는 것에 의해 재프로그램화될 것이다[38,39]. 이러한 결과는 3'-연장된 gRNA가 효율적인 뉴클레아제 표적

화 가이드로서의 역할을 할 수 있고, 역전사의 주형이 될 수 있다는 것을 입증한다.

[2133] Cas9-TPRT는 닉킹된 DNA 및 gRNA를 시스로 사용한다. 이중 색상 실험을 사용하여 RT 반응이 gRNA와 시스로 (동일한 복합체에서 결합됨) 우선적으로 발생하는지 결정하였다 (도 8 참조). 5'-연장된 gRNA 및 3'-연장된 gRNA에 대해 2개의 별개의 실험을 수행하였다. 주어진 실험을 위해, dCas9, gRNA, 및 DNA 기질의 3원 복합체를 별개의 튜브에서 형성하였다. 하나의 튜브에서, gRNA는 긴 RT 생성물을 코딩하고, DNA 기질은 Cy3 (적색)으로 표지되고; 다른 튜브에서, gRNA는 짧은 RT 생성물을 코딩하고, DNA 기질은 Cy5 (청색)로 표지된다. 짧은 인큐베이션 후에, 복합체를 혼합한 다음, RT 효소 및 dNTP로 처리하였다. 생성물을 우레아-변성 PAGE에 의해 분리하고, Cy3 및 Cy5 채널에서 형광에 의해 가시화하였다. 반응 생성물은 DNA 기질과 사전-복합체화된 gRNA 주형을 사용하여 우선적으로 형성되는 것으로 확인되었고, 이는 RT 반응이 시스로 발생할 가능성이 있음을 나타낸다. 이러한 결과는 단일 Cas9:gRNA 복합체가 DNA 부위를 표적화하고, 돌연변이유발 DNA 가닥의 역전사의 주형이 될 수 있음을 지지한다.

[2134] 다른 Cas 시스템을 사용한 TPRT 시험

[2135] 에스. 아우레우스로부터의 Cas9 및 엘. 박테리움(L. bacterium)으로부터의 Cas12a를 포함한 다른 Cas 시스템을 사용하여 이전 섹션에 제시된 것과 유사한 실험을 수행할 것이다 (도 2a-2c 참조). TRPT가 또한 이들 Cas 변이체에 대해 입증될 수 있는 경우에, 세포에서의 잠재적 편집 범주 및 전반적 성공 가능성은 증가할 것이다.

[2136] RT-Cas9 융합 단백질을 사용한 TPRT 시험

[2137] 일련의 상업적으로 입수가능하거나 정제가능한 RT 효소는 먼저 TPRT 활성에 대해 트랜스로 평가될 것이다. M-MLV로부터의 이미 시험된 RT에 더하여, 조류 골수모구증 바이러스 (AMV), 게오바실루스 스테아로써모필루스 군 II 인트론 (GsI-IIC)[41,42], 및 유박테리움 렉탈레 군 II 인트론 (Eu.re.I2)[43,44]으로부터의 RT가 평가될 것이다. 유의하게, 후자의 2개의 RT는 그의 천연 생물학적 상황에서 TPRT를 수행한다. 관련된 경우에, RNAse 불활성화 돌연변이 및 다른 잠재적으로 유익한 RT 효소 변형이 시험될 것이다. 기능적 RT가 트랜스로 공급된 경우에 확인되면, 각각은 Cas9 변이체에 대한 융합 단백질로서 평가될 것이다. N-말단 및 C-말단 융합 배향 둘 다가 다양한 링커 길이 및 아키텍처와 함께 시험될 것이다. RT 효소를 시스로 사용하여 TPRT가 발생할 수 있는지를 동역학적 시간 경과 실험을 사용하여 결정할 것이다. 효율적인 TPRT 화학을 가능하게 하는 RT-Cas9 융합 아키텍처가 구축될 수 있는 경우에, 이는 세포와 관련하여 기능적 편집의 가능성을 크게 증가시킬 것이다.

[2138] 세포에서 조작된 gRNA를 사용한 Cas9 표적화

[2139] 이전의 하위-목표에서 개발된 후보 조작된 gRNA를 인간 세포 배양 실험 (HEK293)에서 평가하여 Cas9 표적화 효율을 확인할 것이다. 야생형 SpCas9를 사용하는 확립된 indel 형성 검정을 사용하여[45], 조작된 gRNA를 인간 게놈 내의 5개 이상의 부위에 걸쳐 표준 gRNA와 나란히 비교할 것이다. 게놈 편집 효율은 실험실에 수용된 일루미나 MiSeq 플랫폼을 사용하여 멀티플렉스에서의 앰플리콘 서열분석에 의해 특징화될 것이다. 본 섹션 및 선행 섹션으로부터의 결과는 설계-구축-시험 주기의 후속 반복에 대한 정보를 제공하는 통찰을 생성할 것으로 예상되며, 여기서 gRNA는 세포에서 역전사의 주형이 되는 것 및 효율적인 Cas9 표적화 둘 다에 대해 최적화될 수 있다.

[2140] 시험관내 검증의 결과를 도 5-7에 제시한다. 시험관내 실험은 닉킹된 비-표적 DNA 가닥이 유연하고, DNA 합성을 프라이밍하는 데 이용가능하며, gRNA 연장부가 역전사를 위한 주형으로서의 역할을 할 수 있음을 입증하였다 (도 5 참조). 이 실험 세트는 다양한 길이의 합성 주형 (좌측에 열거됨)을 갖는 5'-연장된 gRNA (도 3a-3b에 제시된 바와 같이 설계됨)를 사용하였다. 형광 표지된 (Cy5) DNA 표적을 기질로서 사용하였고, 이 실험 세트에서 사전-닉킹하였다. 이들 실험에 사용된 Cas9는 촉매 기능상실 Cas9 (dCas9)이므로, DNA를 절단할 수 없지만, 여전히 효율적으로 결합할 수 있다. 몰로니-뮤린 백혈병 바이러스 (M-MLV)로부터 유래된 상업적 RT인 슈퍼스크립트 III은 트랜스로 공급되었다. 먼저, 정제된 성분으로부터 dCas9:gRNA 복합체를 형성하였다. 이어서, 형광 표지된 DNA 기질을 dNTP 및 RT 효소와 함께 첨가하였다. 37℃에서 1시간 인큐베이션 후에, 변성 우레아-폴리아크릴아미드 겔 전기영동 (PAGE)에 의해 반응 생성물을 분석하였다. 겔 영상은 역전사 주형의 길이와 일치하는 길이까지의 원래의 DNA 가닥의 연장을 보여준다. dCas9의 부재 하에 수행된 반응은 사용된 gRNA와 상관없이 길이 51개 뉴클레오티드의 DNA 생성물을 생산하였음에 주목한다. 이 생성물은 RNA가 아닌 DNA 합성을 위한 주형으로서의 상보적 DNA 가닥의 사용에 상응한다 (데이터는 제시되지 않음). 따라서, Cas9 결합은 DNA 합성을 RNA 주형으로 지시하는데 요구된다. 이러한 세트의 시험관내 실험은, DNA 기질이 사전-닉킹되지 않고 Cas9 닉카제 (SpyCas9 H840A 돌연변이체)가 사용된 것을 제외하고, 도 5에 제시된 것과 밀접하게 대응한다. 겔에 제시된 바

와 같이, 닉카제는 표준 gRNA가 사용된 경우에 (gRNA_0, 레인 3) DNA 가닥을 효율적으로 절단한다. 다중 절단 생성물이 관찰되며, 이는 SpyCas9의 이전의 생화학적 연구와 일치한다. 5' 연장부는 닉킹 활성을 손상시키지만 (레인 4-8), 일부 RT 생성물은 여전히 관찰된다. 도 7은 3' 연장부가 DNA 합성을 지지하고 Cas9 닉카제 활성에 유의하게 영향을 미치지 않는다는 것을 보여준다. 사전-닉킹된 기질 (흑색 화살표)은 dCas9 또는 Cas9 닉카제가 사용된 경우 RT 생성물로 거의 정량적으로 전환된다 (레인 4 및 5). RT 생성물 (적색 화살표)로의 50% 초과의 전환이 전장 기질에 의해 관찰된다 (레인 3). RT 생성물의 길이 및 서열을 결정하기 위해, 겔로부터 생성물 밴드를 절제하고, 추출하고, 서열분석하였다. 이는 RT가 gRNA 코어의 3' 말단 헤어핀 내로 3개의 뉴클레오티드를 연장시켰음을 밝혀내었다. 후속 실험 (제시되지 않음)은 헤어핀 RNA 구조를 유지하는 상보적 변화가 이루어지는 한, 이들 3개의 뉴클레오티드는 표적 DNA 서열과 매칭되도록 변화될 수 있다는 것을 입증하였다.

[2141] 잠재적 어려움 및 대안

[2142] (1) RT는 융합체로서 기능하지 않는다: 분자 밀집 및/또는 불리한 기하구조는 Cas9-융합된 RT 효소에 의한 폴리머라제 연장을 방해할 수 있다. 먼저, 링커 최적화를 시험할 수 있다. DNA 프라이머, gRNA, 및 RT 효소 사이의 공간적 관계를 재배향시킬 수 있는 Cas9의 원형 순열 변이체를 평가할 것이다. 목표 2에 상술된 바와 같은 비-공유 RT 동원 전략을 시험할 수 있다. (2) 연장된 gRNA 변이체에 의한 감소된 Cas 표적화 효율: 이는 5'-연장된 gRNA에 대한 이슈일 가능성이 가장 크다. 구조적 데이터에 기초하여[24], Cas9 돌연변이체를 설계하고 스크리닝하여 gRNA 연장부를 보다 크게 용인하는 변이체를 확인할 수 있다. 또한, gRNA 라이브러리는 표적화 활성을 개선시키는 링커에 대해 세포에서 스크리닝될 수 있다.

[2143] 유의성

[2144] 이들 예비 결과는 Cas9 닉카제 및 연장된 gRNA가 트랜스로 공급된 리버스 트랜스크립타제를 사용하여 결합된 DNA 표적 상에서 표적-프라이밍된 역전사를 개시할 수 있다는 것을 확립한다. 중요한 것으로, Cas9 결합은 생성물 형성에 결정적으로 중요한 것으로 밝혀졌다. 아마도 세포에서의 게놈 편집을 위한 절대 요건은 아니지만, RT 효소 기능을 시스로 혼입시키는 시스템의 추가의 개발은 세포-기반 적용에서 성공 가능성을 유의하게 증가시킬 것이다. 이러한 목표의 나머지 측면의 달성은 인간 게놈과 관련하여 정밀 게놈 편집을 수행하기 위한 분자적 토대를 제공할 것이다.

[2145] 2. 인간 세포에서 프라임 편집을 확립한다.

[2146] 배경 및 근거

[2147] 제안된 전략에서, 조작된 RT-Cas9:gRNA 복합체는 게놈 표적 부위에 돌연변이유발 3' DNA 플랩을 도입할 것이다. 단일 미스매치를 함유하는 돌연변이유발 3' 플랩은 우선적으로 제거될 인접한 5' 플랩과의 에너지적으로 접근가능한 평형을 통해 DNA 복구 기구에 의해 혼입될 것으로 가설화된다 (도 1c-1d). DNA 복제 및 복구 기구는 오카자키 단편을 프로세싱할 때[46] 및 긴-패치 염기 절제 복구 (LP-BER) 동안[47] 5' ssDNA 플랩과 마주친다. 5' 플랩은 동종삼량체 슬라이딩 클램프 복합체 PCNA에 의해 DNA 복구 부위로 동원되는, 광범위하게 발현되는 플랩 엔도뉴클레아제 FEN1에 대한 바람직한 기질이다[48]. PCNA는 또한 DNA 리가제 Lig1을 포함한 다른 복구 인자의 동시 동원을 위한 스캐폴드로서의 역할을 한다[49]. '툴벨트'로서 작용하여, PCNA는 모든 세포 분열 동안 생성된 수백만개의 오카자키 단편을 프로세싱하는데 필수적인 연속 플랩 절단 및 라이게이션을 가속화한다[50,51]. 이들 천연 DNA 중간체에 대한 유사성에 기초하여, 돌연변이유발 가닥은 5' 플랩과의 평형을 통해 혼입된 다음, 배위된 5' 플랩 절제 및 라이게이션이 이어질 것으로 가설화된다. 이어서, 미스매치 복구 (MMR)가 어느 한 가닥 상에서 동등한 확률로 발생하여, 편집 또는 복귀로 이어질 것이다 (도 1c-1d). 대안적으로, DNA 복제가 먼저 일어날 수 있고, 새롭게 합성된 딸 가닥에서의 편집의 혼입으로 직접 이어질 수 있다. 이 과정으로부터의 최고 예상 수율은 50%이지만, 다중 기질 편집 시도는 편집 복구의 비가역성으로 인해 반응을 완료쪽으로 구동할 수 있다.

[2148] 예비 결과

[2149] DNA 플랩은 효모 및 HEK 세포에서 플라스미드 모델 기질에서 부위-특이적 돌연변이유발을 유도한다. 제안된 편집 전략을 시험하기 위해, TPRT의 생성물과 유사한 돌연변이유발 3' 플랩을 함유하는 모델 플라스미드 기질을 사용하여 연구를 개시하였다. GFP와 mCherry 사이의 정지 코돈을 코딩하는 이중 형광 단백질 리포터를 생성하였다. 돌연변이유발 플랩은 정지 코돈에 대한 교정을 코딩하여 (도 9a), mCherry 합성을 가능하게 한다. 따라서, 돌연변이유발 효율은 GFP:mCherry 비에 의해 정량화될 수 있다. 플라스미드 기질을 시험관내에서

제조하고, 효모 (에스. 세레비지아에) 또는 인간 세포 (HEK293) 내로 도입하였다. 둘 다의 시스템에서 높은 빈도의 돌연변이유발이 관찰되었고 (도 9b), 단리된 효모 콜로니는 복귀된 염기, 돌연변이된 염기, 또는 둘 다의 생성물의 혼합물을 함유하였다 (도 9c). 후자의 검출은 이들 경우에 플라스미드 복제가 MMR 전에 일어났음을 시사하고, 추가로 플랩 절제 및 라이게이션이 MMR에 선행함을 시사한다. 이 결과는 3' 돌연변이유발 가닥을 사용한 DNA 편집의 실행가능성을 확립한다.

[2150] • 모델 플랩 기질을 이용한 체계적인 연구

[2151] 상기 기재된 예비 결과에 기초하여, 효율적인 편집의 원리를 추론하기 위해 보다 넓은 스펙트럼의 플랩 기질이 HEK 세포에서 평가될 것이다. 3' ssDNA 플랩은 미스매치 쌍 형성의 영향, 플랩을 따른 돌연변이유발 뉴클레오티드의 위치, 및 말단 뉴클레오티드의 정체를 결정하기 위해 체계적으로 달라질 것이다 (도 9d). 단일 뉴클레오티드 삽입 및 결실이 또한 시험될 것이다. 앰플리콘 서열분석은 편집 정밀도를 분석하기 위해 사용될 것이다. 이들 결과는 gRNA 역전사 주형의 설계에 대한 정보를 제공하는 것을 도울 것이다.

[2152] 플라스미드 기질 상의 시험관내 TPRT는 효율적인 편집 결과로 이어진다. 목표 1에서 개발된 TPRT 반응을 사용하여 플라스미드 기질 내에서 돌연변이유발을 유도하였다. 반응을 원형 DNA 플라스미드 기질 상에서 수행하였다 (도 10 참조). 이는 이전의 시험관내 실험에서 RT 연장을 위한 메카니즘으로서 DNA 가닥 해리의 가능성을 배제한다. 이는 또한 세포에서 플랩 기질의 DNA 복구의 시험을 가능하게 하였다. 효모 (에스. 세레비지아에) 발현을 위해 이중-형광 리포터 플라스미드를 구축하였다. 이 플라스미드는 개재 정지 코돈 (TGA)을 갖는 GFP (녹색 형광 단백질) 및 mCherry (적색 형광 단백질)를 코딩한다. 효모에서의 이러한 구축물의 발현은 GFP만을 생산한다. 플라스미드를 시험관내 TRT [Cas9(H840A) 닉카제, 조작된 gRNA, MLV RT 효소, dNTPS]에 대한 기질로서 사용하였다. gRNA 연장부는 정지 코돈에 대한 돌연변이를 코딩한다. 플랩 가닥은 정지 코돈의 복구에 사용되고, 융합 단백질로서 GFP 및 mCherry 둘 다를 발현하는 플라스미드를 생산할 것으로 예상된다. 효모 이중-FP 플라스미드 형질전환체를 도 10에 제시한다. 모 플라스미드 또는 시험관내 Cas9(H840A) 닉킹된 플라스미드의 형질전환은 녹색 GFP 발현 콜로니만을 생성한다. 5'-연장된 또는 3'-연장된 gRNA와의 TRT 반응은 녹색 및 황색 콜로니의 혼합물을 생산한다. 후자는 GFP 및 mCherry 둘 다를 발현한다. 보다 많은 황색 콜로니가 3'-연장된 gRNA에 의해 관찰된다. 정지 코돈을 함유하지 않는 양성 대조군이 또한 제시된다.

[2153] 이 결과는 긴 이중 가닥 기질이 TPRT를 겪을 수 있고, TPRT 생성물이 진핵 세포에서 편집을 유도한다는 것을 확립한다.

[2154] 상기 프라임 편집 실험과 유사한 또 다른 실험을 수행하였지만, 정지 코돈에 점 돌연변이를 설치하는 대신, TRT 편집은 프레임시프트 돌연변이를 복구하고 하류 mCherry의 합성을 가능하게 하는 단일 뉴클레오티드 삽입 (좌측) 또는 결실 (우측)을 설치한다 (도 11 참조). 둘 다의 실험은 3' 연장된 gRNA를 사용하였다. TRT 형질전환으로부터의 개별 콜로니를 선택하고, 생어 서열분석에 의해 분석하였다 (도 12 참조). 녹색 콜로니는 원래 DNA 서열을 갖는 플라스미드를 함유한 반면, 황색 콜로니는 TRT 편집 gRNA에 의해 설계된 정확한 돌연변이를 함유하였다. 다른 점 돌연변이 또는 indel은 관찰되지 않았다.

[2155] RT-Cas9 아키텍처를 사용한 HEK 세포에서의 프라임 편집의 확립

[2156] 이전의 목표로부터 최적화된 구축물은 인간 게놈 내의 표적화된 부위에서 포유동물 발현 및 편집을 위해 적합화될 것이다. 2차 gRNA (닉킹을 방지하기 위해 말단절단됨)로의 인접한 표적화에 더하여, 다중 RT 효소 및 융합 아키텍처가 시험될 것이다. 비-공유 RT 동원이 또한 선-태그 시스템[52] 및 MS2 압타머 시스템[53]을 사용하여 평가될 것이다. 표준 gRNA 및 RT-Cas9 융합체 (상기와 같음)를 사용한 표적화 효율을 평가하기 위해 Indel 형성 검정이 사용될 것이다. 이어서, 각각의 게놈 부위에 대해, 연장된 gRNA 및 RT-Cas9 쌍을 단일 뉴클레오티드 편집에 대해 검정할 것이다. 편집 결과는 MiSeq로 평가될 것이다.

[2157] HEK 세포에서의 초기 실험을 Cas9-RT 융합체를 사용하여 수행하였다. 세포 내에서 발현된 성분에 의한 편집은 Cas9(H840A) 닉카제, 리버스 트랜스크립타제 (융합체로서 발현되거나 또는 트랜스로 공급됨), 및 3' 연장부를 갖는 조작된 gRNA를 필요로 한다 (도 14 참조). 예비 연구는 gRNA 연장부 내의 프라이머 결합 부위의 길이가 인간 세포에서 편집 효율을 증가시키는데 중요함을 나타내었다 (도 15 참조).

[2158] HEK 세포에서의 프라임 편집 파라미터의 최적화

[2159] 세포에서 프라임 편집을 수행할 수 있는 Cas9-RT 아키텍처를 확인한 후에, 고효율 편집을 달성하도록 성분 및 설계가 최적화될 것이다. 코딩된 점 돌연변이의 위치 및 뉴클레오티드 정체, 및 새로 합성된 DNA 가닥의 총 길

이는 편집 범주 및 잠재적 제한을 평가하기 위해 달라질 것이다. 짧은 삽입 및 결실 돌연변이가 또한 평가될 것이다. 단백질 발현 구축물은 코돈 최적화될 것이다. 성공적인 경우에, 이는 포유동물 세포에서의 효율적인 프라임 편집을 확립할 것이다.

[2160] 예비 결과. 융합된 RT 효소에 의한 분자내 역전사가 가능하지 않은 경우에, 편집 유전자좌에서 RT 효소를 보다 높은 국부 농도로 만들기 위해 추가의 gRNA를 설계하였다. 이들 보조 가이드는 5' 단부 (14-15 nt 스페이서)에서 말단절단되며, 이는 이전에 Cas9 절단을 방지하지만 결합은 보유하는 것으로 제시된 바 있다 (도 16 참조). HEK3 유전자좌를 선택하여 이 전략을 연구하였다.

[2161] 잠재적 어려움 및 대안

[2162] 1) 세포에서의 gRNA 분해: 연장된 gRNA 말단이 세포에서 말단절단된 경우에, 안정화 2차 구조가 설치될 수 있거나, 또는 안정화 변형을 갖는 합성 gRNA가 시험될 수 있다. (2) 인간 세포에서 관찰된 편집 없음: 인접한 게놈 부위에 대한 RT-Cas9 융합체의 2차 표적화를 포함한 추가의 전략이 연구될 것이다[54]. 또한, 이. 콜라이 또는 에스. 세레비지아에에서 잠재적 방향적 진화 전략이 연구될 수 있다.

[2163] 유의성

[2164] 프라임 편집이 실험 세포주에서 확립될 수 있다면, 이는 인간 유전자에서 다수의 점 돌연변이의 신속한 생성 및 특징화를 가능하게 함으로써 기초 생의학 연구에 즉각적인 영향을 미칠 것이다. 방법의 일반성, 및 염기 편집제에 관한 그의 직교 편집 윈도우는 많은 현재의 접근불가능한 돌연변이를 설치하기 위한 접근법을 제공할 것이다. 더욱이, 프라임 편집이 높은 효율 및 생성물 순도를 위해 최적화될 수 있다면, 다른 인간 세포 유형에서 질환 돌연변이를 교정하는 그의 잠재적 적용가능성은 유의할 것이다.

[2165] 3. 배양된 인간 세포에서 병원성 돌연변이의 부위-특이적 편집을 달성한다.

[2166] 배경 및 근거.

[2167] 다수의 병원성 돌연변이는 PAM 제한, 또는 전환 또는 indel 돌연변이 교정에 대한 필요로 인해 현재의 염기 편집제에 의해 교정될 수 없다. 프라임 편집에 의하면, 모든 전이 및 전환은 작은 삽입 및 결실일 수 있기 때문에 이론적으로 가능하다. 더욱이, PAM과 관련하여, 프라임 편집 윈도우 (예상 -3 내지 +4)는 염기 편집제의 것 (-18 내지 -12)과 구별된다 (도 13). 염기 편집제에 의해 현재 교정가능하지 않은 멘델 상태는 (1) 헤모글로빈베타에서의 겸상 적혈구 질환 Glu6Val 창시자 돌연변이 (A·T에서 T·A로의 전환을 필요로 함); (2) ATP7B에서의 가장 흔한 윌슨병 변이체 His1069Gln (G·C에서 T·A로의 전환을 필요로 함); 및 (3) 낭성 섬유증을 유발하는 CFTR에서의 ΔPhe508 돌연변이 (3-뉴클레오티드 삽입을 필요로 함)를 포함한다. 이들 표적 각각은 SpCas9 표적화 및 프라임 편집을 위해 적절하게 위치한 PAM을 함유한다.

[2168] 예비 결과.

[2169] • HEK3 세포에서의 T에서 A로의 편집은 현행 염기 편집에 의해 달성가능하지 않지만, TRPT 편집에 의해서는 달성가능하다 (도 17a-17c 참조).

[2170] 도 17a는 인간 배아 신장 (HEK) 세포에서의 성분의 형질감염 후 표적 뉴클레오티드에서의 % T에서 A로의 전환을 나타내는 그래프를 보여준다. 이 데이터는 Cas9(H840A) 닉카제 (32-아미노산 링커)에 대한 야생형 MLV 리버스 트랜스크립타제의 N-말단 융합을 사용한 결과를 제시한다. 편집 효율은 프라이머 결합 부위의 길이가 7개 뉴클레오티드에서 11 또는 12개 뉴클레오티드로 연장된 경우에 현저하게 개선되었다. 추가적으로, 편집 유전자좌의 바로 상류에 위치하는 보조 가이드 A (도 16 참조)는, 특히 보다 짧은 길이의 프라이머 결합 부위에 대해 편집 활성을 유의하게 개선시킨다. 편집 효율을 일루미나 MiSeq 플랫폼을 사용하여 앰플리콘 서열분석에 의해 정량화하였다. 도 17b도 또한 인간 배아 신장 (HEK) 세포에서의 성분의 형질감염 후 표적 뉴클레오티드에서의 % T에서 A로의 전환을 보여주지만, 이 데이터는 RT 효소의 C-말단 융합을 사용한 결과를 제시한다. 여기서, 보조 가이드 A는 그만큼의 효과를 갖지 않고, 편집 효율이 전체적으로 더 높다. 도 17c는 도 17a에서 사용된 것과 유사한 Cas9(H840A) 닉카제에 대한 야생형 MLV 리버스 트랜스크립타제의 N-말단 융합을 사용한 결과를 제시하는 데이터를 보여주지만; MLV RT와 Cas9 사이의 링커는 32개 아미노산 대신 60개 아미노산 길이이다.

[2171] • TRPT 편집 결과에 의한 HEK3 부위에서의 T에서 A로의 편집은 높은 순도를 나타낸다.

[2172] 도 18은 고처리량 앰플리콘 서열분석에 의한 서열 분석의 산출을 보여준다. 산출은 편집된 세포의 가장 풍부한

유전자형을 나타낸다. 어떠한 주요 indel 생성물도 수득되지 않고, 목적하는 점 돌연변이 (T에서 A로)는 방관자 편집 없이 깔끔하게 설치됨에 주목한다. 제1 서열은 참조 유전자형을 보여준다. 상단의 2개의 생성물은 내인성 다형성 (G 또는 A)을 함유하는 출발 유전자형이다. 하단의 2개의 생성물은 정확하게 편집된 유전자형을 나타낸다.

[2173] • MLV RT 돌연변이체는 편집을 개선시킨다.

[2174] 문헌 [Baranauskas, et al., (doi:10.1093/protein/gzs034)]에 기재된 돌연변이체 리버스 트랜스크립타제를 인간 배아 신장 (HEK) 세포에서 표적 뉴클레오티드 편집을 위한 Cas9(H840A) 닉카제에 대한 C-말단 융합체로서 시험하였다. Cas9-RT 편집제 플라스미드를 역전사의 주형이 되는 3'-프라임 편집 가이드 RNA를 코딩하는 플라스미드와 함께 공동-형질감염시켰다. 표적 뉴클레오티드에서의 편집 효율 (청색 막대)을 도 19에서 indel 비율 (오렌지색 막대)과 함께 제시한다. WT는 야생형 MLV RT 효소를 지칭한다. 돌연변이체 효소 (M1 내지 M4)는 우측에 열거된 돌연변이를 함유한다. 편집 속도를 게놈 DNA 앰플리콘의 고처리량 서열분석에 의해 정량화하였다.

[2175] 제2 gRNA에 의한 상보적 가닥 닉킹은 편집을 개선시킨다.

[2176] 이 실험은 단일 가닥 닉이 표적 뉴클레오티드에 근접한 상보적 DNA 가닥에 도입되는 경우의 표적 뉴클레오티드의 편집 효율을 평가하며, 이는 미스매치 복구를 지시하여 원래의 뉴클레오티드를 우선적으로 제거하고 염기 쌍을 목적하는 편집으로 전환시킬 것으로 가설화된다. Cas9(H840A)-RT 편집 구축물을 2개의 가이드 RNA 코딩 플라스미드와 함께 공동-형질감염시켰고, 그 중 하나는 역전사 반응의 주형이 되고, 다른 것은 닉킹을 위한 상보적 DNA 가닥을 표적화한다. 표적 뉴클레오티드로부터 다양한 거리에서 닉킹을 시험하였다 (오렌지색 삼각형) (도 20 참조). 표적 염기 쌍 (청색 막대)에서의 편집 효율은 indel 형성 속도 (오렌지색 막대)와 함께 제시된다. "부재" 예는 상보적 가닥 닉킹 가이드 RNA를 함유하지 않는다. 편집 속도를 게놈 DNA 앰플리콘의 고처리량 서열분석에 의해 정량화하였다.

[2177] 도 21은 목적하는 T에서 A로의 전환 돌연변이 및 다른 주요 게놈 편집 부산물의 일반적 부재를 보여주는 프로세싱된 고처리량 서열분석 데이터를 보여준다.

[2178] 범주. 새로운 편집 기술에 대한 잠재적 범주를 도 13에 제시하며, 이를 데아미나제-매개된 염기 편집제 기술과 비교한다. 이전에 개발된 염기 편집제는 PAM의 ~15 ± 2 bp 상류 영역을 표적화한다. 표적 C 또는 A 뉴클레오티드를 각각 T 또는 G로 전환시킴으로써, 이전에 개발된 염기 편집제는 모든 전이 돌연변이 (A:T에서 G:C로의 전환)를 가능하게 한다. 그러나, 이전에 개발된 염기 편집제는 전환 돌연변이 (A에서 T로, A에서 C로, G에서 T로, G에서 C로, T에서 A로, T에서 G로, C에서 A로, C에서 G로)를 설치할 수 없다. 더욱이, 편집 윈도우에 다중 표적 뉴클레오티드가 존재하는 경우에, 추가의 바람직하지 않은 편집이 발생할 수 있다.

[2179] 새로운 프라임 편집 기술은 이론적으로 임의의 뉴클레오티드 및 염기 쌍 전환, 및 또한 잠재적으로 작은 삽입 및 결실 편집을 설치할 수 있다. PAM과 관련하여, 프라임 편집 윈도우는 DNA 닉킹 부위 (PAM의 3개 염기 상류)에서 시작하고, PAM의 아직 결정되지 않은 위치 하류에서 끝난다. 이러한 편집 윈도우는 데아미나제 염기 편집제의 편집 윈도우와 구별됨에 주목한다. TPRT 시스템은 DNA 폴리머라제 효소를 사용하여 편집을 수행하기 때문에, 이는 잠재적으로 일반성, 정밀도, 및 충실도를 포함한 모든 그의 이익을 갖는다.

[2180] 환자-유래 세포주에서 병원성 돌연변이를 교정한다.

[2181] 관련 돌연변이를 보유하는 세포주 (겸상 적혈구 질환: CD34+ 조혈 줄기 세포; 윌슨병: 배양된 섬유모세포; 낭성 섬유증: 배양된 기관지 상피)는 ATCC, 코리엘 바이오뱅크(Coriell Biobank), 또는 협력 하버드/브로드 (Harvard/Broad) 제휴 실험실로부터 입수할 것이다. 편집 효율은 고처리량 서열분석에 의해 평가될 것이고, 교정된 유전자형의 효능은 표현형 검정 (헤모글로빈 HPLC, ATP7B 면역염색, 및 CFTR 막 전위 검정)을 사용하여 시험될 것이다.

[2182] 오프-타겟 편집 활성을 특징화한다.

[2183] 잠재적 오프-타겟 편집은 야생형 Cas9와 쌍형성된 표적 gRNA를 사용하여 확립된 방법, 예컨대 가이드-seq(GUIDE-seq)[55] 및 서클-seq(CIRCLE-seq)[56]에 의해 스크리닝될 것이다. 잠재적 오프-타겟이 확인되는 경우에, 이들 유전자좌는 진정한 오프-타겟 편집 사건을 확인하기 위해 TPRT 편집된 세포에서 프로빙될 것이다.

[2184] 잠재적 어려움 및 대안.

[2185] (1) 낮은 편집 효율: 프라임 편집제 (PE)는 각각의 표적에 대한 최적화를 필요로 할 수 있다. 이러한 경우에,

gRNA 라이브러리는 특정한 적용을 위한 가장 높은 기능적 변이체를 확인하기 위해 시험될 수 있다. RT-Cas 융합체 발현 및 핵 국재화가 최적화될 수 있다. 리포솜 RNP 전달은 오프-타겟 편집을 제한하는 데 사용될 수 있다.

[2186] 앞으로의 실험.

[2187] gRNA 설계의 최적화는 프라이머 결합 부위 길이 및 합성 주형의 연장의 추가의 연구에 의해 달성될 수 있다. 시험 범주 및 일반성은 상이한 뉴클레오티드 전환, 작은 삽입 및 결실, 뿐만 아니라 PAM에 관한 상이한 편집 위치, 및 인간 게놈 내의 다중 부위를 포함할 것이다. RT 성분의 최적화는 활성을 증진시키기 위한 MLV RT에서의 돌연변이 (Rnase H 불활성화, 프라이머-주형 결합 친화도 증가, 프로세싱성에 대한 조정), 및 새로운 RT 효소 (군 II 인트로 RT, 다른 레트로바이러스 RT)를 연구하는 것을 포함할 것이다.

[2188] 유의성.

[2189] 수많은 유전 장애는 개별 유전자에서의 단일 뉴클레오티드 변화로부터 발생한다. 본원에 기재된 게놈 편집 기술을 개발하고, 이를 질환-관련 세포 유형에 적용하는 것은 임상으로의 번역을 위한 토대를 확립할 것이다. 일부 질환, 예컨대 겸상 적혈구 질환의 경우, 단일 점 돌연변이는 집단 전반에 걸쳐 우성 유전자형을 나타낸다. 그러나, 많은 다른 유전 장애의 경우, 단일 유전자 내의 상이한 점 돌연변이의 큰 이질성이 환자 집단 전반에 걸쳐 관찰되고, 이들 각각은 유사한 질환 표현형을 발생시킨다. 따라서, 이론상 다수의 이러한 돌연변이를 표적화할 수 있는 일반적 게놈 편집 방법으로서, 본 기술은 많은 이들 환자 및 그의 가족에게 막대한 잠재적 이익을 제공할 수 있다. 이들 적용에 대한 원리의 증명이 세포에서 확립될 수 있다면, 이는 질환의 동물 모델에서의 연구에 대한 기반을 확립할 것이다.

[2190] 이점

[2191] 정밀도: 목적하는 편집은 핵산 서열에서 코딩되고 지시된다. 일반성: 이론상, 전환 편집 뿐만 아니라 작은 삽입 또는 결실을 포함한 임의의 염기 쌍 전환을 만드는 것이 가능할 수 있다. Cas9 프로토스페이서 인접 모티프 (PAM) 서열과 관련하여 염기 편집제의 것과 별개의 편집 윈도우가 존재한다. 이 방법은 상동성-지정 복구 (HDR)의 편집 능력 중 많은 것을 달성하지만, HDR의 주요 제한 (대부분의 세포 유형에서 비효율적이고, 통상적으로 과량의 바람직하지 않은 부산물 예컨대 indel을 동반함)은 받지 않는다. 또한, 이는 이중 가닥 DNA 파괴 (DSB)를 일으키지 않고, 따라서 indel, 전위, 큰 결실, p53 활성화 등을 거의 일으키지 않는다..

[2192] 실시예 3 - PE에 의한 펩티드 태그부착

[2193] 본원에 기재된 프라임 편집 시스템 (즉, PE 시스템)은 또한 다양한 펩티드 태그를 단백질 코딩 유전자 내로 도입하는 데 사용될 수 있다. 이러한 태그는 헥사히스티딘 태그, FLAG-태그, V5-태그, GCN4-태그, HA-태그, Myc-태그 등을 포함할 수 있다. 이러한 접근법은 단백질 형광 표지, 면역침전, 이뮤노블롯팅, 면역조직화학, 단백질 동원, 유도성 단백질 데그론, 및 게놈-전반 스크리닝과 같은 적용에 유용할 수 있다. 실시양태는 도 25 및 26에 도시된다.

[2194] 도 25는 내인성 게놈 유전자좌에서의 펩티드 태그부착 유전자 및 TPRT 게놈 편집 (즉, 프라임 편집)에 의한 펩티드 태그부착에 대한 gRNA 설계를 보여주는 개략도이다. FlAsH 및 ReAsH 태그부착 시스템은 2개의 부분을 포함한다: (1) 형광단-비아르세니칼 프로브, 및 (2) 서열 FLNCCPGCCMEP (서열식별번호: 1)로 예시되는, 테트라시스테인 모티프를 함유하는 유전자 코딩된 펩티드. 세포 내에서 발현되는 경우에, 테트라시스테인 모티프를 함유하는 단백질은 형광단-비소 프로브로 형광 표지될 수 있다 (참고문헌 [J. Am. Chem. Soc., 2002, 124 (21), pp 6063-6076. DOI: 10.1021/ja017687n] 참조). "소르태깅" 시스템은 표지된 펩티드 프로브를 적합한 펩티드 기질을 함유하는 단백질에 공유 접합시키는 박테리아 소르타제 효소를 사용한다 (참고문헌 [Nat. Chem. Biol. 2007 Nov;3(11):707-8. DOI: 10.1038/nchembio.2007.31] 참조). FLAG-태그 (DYKDDDDK (서열식별번호: 2)), V5-태그 (GKPIPNPLLGLDST (서열식별번호: 3)), GCN4-태그 (EELLSKNYHLENEVARLKK (서열식별번호: 4)), HA-태그 (YPYDVPDYA (서열식별번호: 5)), 및 Myc-태그 (EQKLISEEDL (서열식별번호: 6))가 면역검정을 위한 에피토프 태그로서 통상적으로 사용된다. 파이-클램프는 펜타플루오로-방향족 기질로 표지될 수 있는 펩티드 서열 (FCPF) (서열식별번호: 622)을 코딩한다 (참고문헌 [Nat. Chem. 2016 Feb;8(2):120-8. doi: 10.1038/nchem.2413]).

[2195] 도 26은 게놈 DNA 내로의 His6-태그 및 FLAG-태그의 정확한 설치를 보여준다. 18-nt His-태그 삽입 또는 24-nt FLAG-태그 삽입을 코딩하는 역전사 주형을 사용하여 HEK3 유전자좌를 표적화하는 가이드 RNA를 설계하였다. 형질감염된 HEK 세포에서의 편집 효율을 앰플리콘 서열분석을 사용하여 평가하였다. FLAG-태그의 전장 24-nt

서열은 보이는 프레임의 외부에 있다는 것을 주목한다 (서열분석으로 완전하고 정확한 삽입을 확인함).

[2196] 실시예 4 - PE를 사용한 RNA 태그부착 및 조작

[2197] 포유동물, 진핵, 및 박테리아 세포 내에서 RNA를 태그부착하거나 또는 달리 조작하는 유전자 서열 내로 모티프를 삽입하기 위한 새로운 방법이 본원에 기재된다. 인간 게놈의 1%만이 단백질을 코딩하는 것으로 추정되지만, 실질적으로 모든 게놈이 일부 수준으로 전사된다. 대부분의 이들 추정 RNA의 역할이 무엇인지는 물론, 생성된 비-코딩 RNA (ncRNA)가 얼마나 많이 기능적 역할을 하는지는 해결되지 않은 의문이다. 유용한 특성을 갖는 신규 RNA-코딩 서열의 관심 유전자 내로의 삽입을 통한 이들 RNA 분자의 "태그부착"은 세포에서 RNA 분자의 생물학적 기능을 연구하기 위한 유용한 방법이다. 이는 또한 mRNA 변형이 어떻게 세포 기능에 영향을 미칠 수 있는지를 교란하고, 이에 따라 더 잘 이해하기 위한 수단으로서 단백질-코딩 mRNA 상에 태그를 설치하는데 유용할 수 있다. 예를 들어, 편재성 천연 RNA 태그 - 폴리아데닐화 - 는 세포에 의해 mRNA의 세포질 내로의 수송에 영향을 미치는데 사용된다. 상이한 유형의 폴리아데닐화 신호는 상이한 수송 속도 및 상이한 mRNA 수명 및 - 이에 따라 - 코딩된 단백질이 발현되는 수준의 차이를 발생시킨다.

[2198] 세포 내에서 태그부착된 RNA를 발현시키기 위한 통상적인 접근법은 (i) 종종 초생리학적 수준에서 유전자 발현의 단기 버스트를 생성하는 일시적 플라스미드 형질감염; 또는 (ii) 연장된 발현을 가능하게 하는 렌티바이러스 통합 또는 트랜스포손을 사용한 게놈 내로의 태그부착된 RNA 유전자의 영구적 통합 (무작위 부위)을 사용하여 합성 구축물을 외인적으로 도입하는 것이다. 이들 접근법 둘 다는 변경된 발현 수준의 생산에 의해, 및 유전자의 발현 또는 활성을 조절하는 천연 메카니즘의 부재에 의해 제한된다. 대안적 전략은 Cas9 또는 다른 표적화된 DNA 뉴클레아제에 의해 유도된 이중 가닥 DNA 파괴의 상동성-지정 복구 (HDR)를 사용하여 그의 내인성 유전자좌에서 관심 유전자를 직접 태그부착하는 것이다. 이러한 접근법은 광범위한 내인성으로 태그부착된 유전자의 생성을 가능하게 하지만, HDR은 현저하게 비효율적이고, 따라서 성공적으로 태그부착된 목적하는 세포의 클론 집단을 확인하기 위해 유의한 스크리닝을 필요로 한다. 더욱이, HDR은 전형적으로 다수의 세포 유형, 가장 특히 유사분열후 세포에서 매우 비효율적이거나 또는 전적으로 불활성이다. HDR의 낮은 효율은 바람직하지 않은 indel 생성물의 생성에 의해 추가로 복잡해지고, RNA 유전자의 경우에 특히 문제가 될 수 있는데, 이는 그의 활성이 정상 대립유전자의 기능을 방해하는 RNA의 생산으로 이어질 수 있기 때문이다. 마지막으로, 연구원들은 종종 최적의 성능을 달성하기 위해 RNA 분자 내의 다양한 태그부착 위치를 스크리닝할 필요가 있다. 이들 결점의 합은 HDR을 RNA에 태그를 설치하는데 덜 바람직한 방법으로 만든다.

[2199] 프라임 편집은 RNA로부터 DNA로의 유전자 정보의 전달을 통해 게놈 유전자좌의 표적화된 편집을 가능하게 하는 새로운 게놈 편집 기술이다. 프라임 편집을 사용하여, RNA 유전자는 다양한 성분, 예컨대 RNA 압타머, 리보자임, 또는 다른 RNA 모티프로 태그부착될 수 있다. 프라임 편집은 HDR 전략과 비교하여 보다 빠르고, 보다 저렴하고, 보다 다양한 세포 유형에서 효과적인 잠재력을 갖는다. 따라서, 기재된 발명은 건강 및 질환에서 RNA 유전자의 생물학을 조사하기 위한 신규하고 유용하며 비-자명한 도구를 나타낸다. 프라임 편집제 (PE)를 사용하여 RNA를 태그부착하거나 또는 달리 조작하는 유전자 서열 내로의 RNA 모티프의 삽입을 위한 새로운 방법이 본원에 기재된다. PE는 CRISPR/Cas 시스템에 의해 표적화가능한 목적하는 게놈 유전자좌에서 다중 뉴클레오티드를 부위-특이적으로 삽입, 돌연변이, 및/또는 결실시킬 수 있다. PE는 Cas9 뉴클레아제 도메인과 리버스 트랜스크립타제 도메인 사이의 융합체로 구성된다. 이들은 DNA 표적화를 위한 가이드 스페이서 부분, 뿐만 아니라 목적하는 게놈 편집을 코딩하는 역전사를 위한 주형을 함유하는 조작된 PEgRNA (프라임 편집 가이드 RNA)에 의해 그의 게놈 표적으로 가이드된다 (도 28a 참조). PE는 비-코딩 RNA 또는 mRNA를 태그부착하거나 또는 달리 조작하기 위해 RNA 수준에서 기능적인 모티프 (이하 RNA 모티프)를 삽입하는 데 사용될 수 있는 것으로 고려된다. 이들 모티프는 유전자 발현을 증가시키거나, 유전자 발현을 감소시키거나, 스플라이싱을 변경하거나, 전사후 변형을 변화시키거나, RNA의 세포하 위치에 영향을 미치거나, RNA의 단리 또는 세포내 또는 세포외 위치의 결정을 가능하게 하거나 (예를 들어, 형광 RNA 압타머, 예컨대 스피니치, 스피니치2, 베이비 스피니치, 또는 브로콜리를 사용함), 내인성 또는 외인성 단백질 또는 RNA 결합제를 동원하거나, sgRNA를 도입하거나, 또는 자기-절단 또는 RNAse에 의한 RNA의 프로세싱을 유도하는 역할을 할 수 있다 (도 28b 참조). 프라임 편집의 유연성으로 인해, 게놈 내에 설치될 수 있는 RNA 모티프의 포괄적인 목록을 제공하는 것은 가능하지 않다. RNA 유전자를 태그부착하는 데 사용될 수 있는 PE-설치된 RNA 모티프의 예측 범주를 광범위하게 예시하는 일련의 예가 본원에 제시된다. PE 이외의 임의의 다른 보고된 게놈 편집 방법을 사용하여 대부분의 유형의 세포 (HDR을 지지하지 않는 많은 것 포함)에서 이들 변화를 효율적이고 상당히 깔끔하게 만드는 것은 현재 가능하지 않다.

[2200] 유전자 발현은 핵 수송 또는 체류 또는 mRNA 수명에서 변화를 발생시키는 3' 비번역 영역 (UTR)을 코딩하는 것에 의해 영향을 받을 수 있다. 예를 들어, 폴리오마바이러스 원숭이 바이러스 40 (SV40)으로부터의 폴리A 테일

은 효율적인 전사 종결을 가능하게 하고 다른 3' UTR에 비해 유전자 발현을 증가시킬 수 있는 추가의 헬퍼 서열을 갖는다[57, 58]. SV40 폴리A 테일의 예시적 서열:

SV40 폴리A 테일	AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGC ATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT TGTGGTTTGTCCAAACTCATCAATGTATCTTA (SEQ ID NO: 331)

[2201]

[2202] 폴리아데닐화 신호 이외의 번역후 변형 신호가 또한 PE에 의해 코딩될 수 있다. 이들은 N6-메틸아데노신, N1-메틸아데노신, 5-메틸시토신, 및 슈도우리딘 변형을 포함하는 신호를 포함한다[59]. PE를 사용하여 RNA 전사체 내에 이들 변형을 기록 또는 제거하는 효소에 의해 결합된 서열을 포함시킴으로써, 그의 기록 또는 소거를 유도하는 것이 가능할 것이다. 이는 이들 마커의 효과를 연구하거나, 세포 분화를 유도하거나, 스트레스 반응에 영향을 미치거나, 또는 이들 마커의 기능이 아직 조사 중이라면 다른 방식으로 표적화된 세포에 영향을 미치기 위한 도구로서 사용될 수 있다.

[2203] PE는 세포하 국재화에 영향을 미치는 돌연변이를 코딩할 수 있다. 예를 들어, mRNA 내의 tRNA-Lys의 혼입은 이론적으로 미토콘드리아로의 수송을 발생시킬 수 있는 한편[60], 다양한 3' UTR은 핵 체류 또는 수송을 발생시킬 수 있다[61].

[2204] 예:

[2205] SV40 폴리A 신호는 수송을 발생시킨다.

SV40 폴리A 테일	AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGC ATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT TGTGGTTTGTCCAAACTCATCAATGTATCTTA (SEQ ID NO: 331)

[2206]

[2207] U1 snRNA 3' 박스는 체류를 발생시킨다.

U1 SNRNA 3′ 박스	TTCATTCAGCAAGTTCAGAGAAATCTGAACTTGCTGGATTTTTGGAGC AGGGAGATGGAATAGGAGCTTGCTCCGTCCACTCCACGCATCGACCTG GTATTGCAGTACCTCCAGGAACGGTGCACCCACTTTCTGGAGTTTCAA AAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTT GGTTTGTGTCTTGGTTGGCGTCTTAA (SEQ ID NO: 625)

[2208]

[2209] 내인성 RNA의 세포하 국재화를 결정하는 것은 도전과제일 수 있고, 이는 FISH의 경우에서와 같이 외인성, 형광-태그부착된 뉴클레오티드 프로브의 부가, 또는 시간-소모적이고 잠재적으로 부정확한 세포 분획화에 이은 RNA 검출을 필요로 한다. 내인성 RNA 내에 프로브를 코딩하는 것은 이들 이슈 중 많은 것을 제거할 것이다. 한 예는 내인성 RNA 내에 형광 RNA 압타머, 예컨대 스피니치[62] 또는 브로콜리를 코딩함으로써, 소분자 원-형광단의 부가를 통해 그러한 RNA의 존재를 가시화하는 것일 것이다.

[2210] 브로콜리 압타머:

브로콜리 압타머	GAGACGGTCGGGTCCAGATATTCGTATCTGTCGAGTAGAGTGTG GGCTC (SEQ ID NO: 357)

[2211]

[2212] PE는 RNA 결합 단백질에 의해 인식되는 RNA를 코딩하는 서열을 삽입하거나 제거하여, RNA 안정성, 발현, 국재화, 또는 변형에 영향을 미칠 수 있다 (예를 들어, 열거된 단백질 참조[63]).

[2213] PE는 바이러스 또는 암 방어 메카니즘으로서 게놈 내에 sgRNA를 코딩하는 서열을 삽입할 수 있다. 유사하게, 이는 마이크로RNA (예를 들어, 프리-마이크로RNA)를 삽입하여 표적화된 유전자의 침묵을 지시하는 데 사용될 수 있다.

[2214] PE는 RNA의 다양한 부분의 기능을 연구하기 위한 요법 또는 도구로서, RNA를 그 자체로 또는 외부 인자에 의해 프로세싱하는 서열을 삽입할 수 있다. 예를 들어, HDV 리보자임[64]은 RNA 서열 내에 포함되는 경우에 리보자임의 바로 5'의 RNA의 프로세싱을 발생시키는 한편, 해머헤드 리보자임은 리보자임 내의 제3 스템 전에 절단한다[65]. 다른 자기-절단 리보자임은 피스톨[66], 해체트[66], 헤어핀[67], 뉴로포라 바르쿠드(Neuropora Varkud) 위성[68], glmS[69],

트위스터[70], 및 트위스터 시스터[66]를 포함한다. 이들 서열은 리보자임의 야생형 또는 조작된 또는 진화된 버전을 포함할 수 있다. 대다수의 이들 리보자임은 리보자임 절단 부위가 위치하는 곳에 따라, 이들이 회합된 RNA의 영역에 따라 상이한 서열을 가질 수 있다. 외부 인자에 의한 RNA의 프로세싱을 지시할 서열, 예컨대 서열 특이적 RNAse[71], 특이적 구조를 인식하는 RNAse[72] - 예컨대 다이서[73] 및 드로샤[74]가 또한 달성될 수 있다.

[2215] HDV 리보자임:

[2216]

HDV 리보자임	GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAA CATGCTTCGGCATGGCGAATGGGAC (SEQ ID NO: 365)

[2217] 실시예 4에 대한 참고문헌

[2218] 하기 참고문헌은 그 전문이 본원에 참조로 포함된다.

1. Schek N, Cooke C, Alwine JC. *Molecular and Cellular Biology*. 1992.
2. Gil A, Proudfoot NJ. *Cell*. 1987.
3. Zhao, B.S., Roundtree, I.A., He, C. Nat Rev Mol Cell Biol. 2017.
4. Rubio, M.A.T., Hopper, A.K. Wiley Interdiscip Rev RNA 2011.
5. Shechner, D.M., Hacisuleyman E., Younger, S.T., Rinn, J.L. Nat Methods. 2015.
6. Paige, J.S., Wu, K.Y., Jaffrey, S.R. *Science* 2011.
7. Ray D., … Hughes TR. Nature 2013.
8. Chadalavada, D. M., Cerrone-Szakal, A. L., Bevilacqua, P.C. RNA 2007.
9. Forster AC, Symons RH. Cell. 1987.
10. Weinberg Z, Kim PB, Chen TH, Li S, Harris KA, Lünse CE, Breaker RR. Nat. Chem. Biol. 2015.
11. Feldstein PA, Buzayan JM, Bruening G. Gene 1989.
12. Saville BJ, Collins RA. Cell. 1990.
13. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. Nature 2004.
14. Roth A, Weinberg Z, Chen AG, Kim PG, Ames TD, Breaker RR. Nat Chem Biol. 2013.
15. Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z. Nat Commun. 2012.
16. MacRae IJ, Doudna JA. Curr Opin Struct Biol. 2007.
17. Bernstein E, Caudy AA, Hammond SM, Hannon GJ Nature 2001.
18. Filippov V, Solovyev V, Filippova M, Gill SS. Gene 2000.

[2219]

[2220] 실시예 5 - PE에 의한 이뮤노에피토프 삽입

[2221] CRISPR/Cas 시스템을 사용하는 정확한 게놈 표적화 기술은 최근, 조작된 DNA 서열의 표적화된 게놈 유전자좌 내로의 삽입을 포함한 광범위한 적용에서 연구되었다. 이전에는 상동성-지정 복구 (HDR)가 이러한 적용에 사용되어 왔으며, 이는 ssDNA 공여자 주형 및 이중 가닥 DNA 파괴 (DSB)에 의한 복구 개시를 필요로 하였다. 이러한 전략은 세포에서 이루어지는 가장 넓은 범위의 가능한 변화를 제공하고, 대형 DNA 서열을 포유동물 세포 내로 삽입하는데 이용가능한 유일한 방법이다. 그러나, HDR은 그의 개시 DSB로부터 초래되는 바람직하지 않은 세포 부작용, 예컨대 높은 수준의 indel 형성, DNA 전위, 대형 결실, 및 P53 활성화에 의해 방해받는다. 이들 결점에 더하여, HDR은 많은 세포 유형에서의 낮은 효율에 의해 제한된다 (T 세포는 이러한 관찰에 대한 주목할 만한 예외임). 이들 결점을 극복하기 위한 최근의 노력은 인간 Rad51 돌연변이체를 Cas9 D10A 닉카제에 융합시켜 (RDN), 개선된 HDR 생성물:indel 비 및 보다 낮은 오프 타겟 편집을 특색으로 하는 DSB-부재 HDR 시스템을 생성하는 것을 포함하지만, 이는 여전히 세포-유형 의존성 및 단지 보통의 HDR 편집 효율에 의해 방해받는다.

[2222] PEgRNA와 커플링된 리버스 트랜스크립타제에 대한 Cas9의 최근에 개발된 융합체 ("프라임 편집제")는 부위-특이적 방식으로 임의의 단일 뉴클레오티드 치환을 설치하고 임의의 짧은 스트레치의 뉴클레오티드 (최대 적어도 수십개의 염기)를 삽입 또는 결실시키는 능력을 포함하여, 기존의 게놈 편집 방법에 비해 다수의 이점을 제공하는 신규한 게놈 편집 기술을 나타낸다. 특히, PE 편집은 일반적으로 낮은 비율의 비의도된 indel로 달성된다. 따라서, PE는 이전에 불가능하였거나 비실용적이었던 표적화된 삽입-기반 편집 적용을 가능하게 한다.

[2223] 이러한 특정한 발명은 공지된 면역원성 에피토프를 내인성 또는 외래 게놈 DNA에 삽입하여, 치료 또는 생명공학

적 적용을 위한 상응하는 단백질의 변형을 발생시키는 수단으로서 프라임 편집을 사용하는 방법을 기재한다 (도 31 및 32 참조). 프라임 편집의 발명 전에, 이러한 삽입은 단지 비효율적으로, DSB로부터의 높은 비율의 indel 형성으로 달성될 수 있다. 프라임 편집은 일반적으로 HDR보다 더 높은 효율을 제공하면서 삽입 편집으로부터의 높은 indel 형성의 문제를 해결한다. 이러한 보다 낮은 비율의 indel 형성은, 특히 면역원성 에피토프를 삽입하는 기재된 적용에서 표적화된 DNA 삽입을 위한 방법으로서 HDR에 비해 프라임 편집의 주요 이점을 제공한다. 에피토프의 길이는 수개의 염기 내지 수백개의 염기의 범위이다. 프라임 편집제는 포유동물 세포에서 이러한 표적화된 삽입을 달성하기 위한 가장 효율적이고 가장 흠없는 기술이다.

[2224] 본 발명의 주요 개념은 이전에 기재된 면역원성 에피토프를 함유하는 뉴클레오티드 서열을 내인성 또는 외래 게놈 DNA 내에 그의 단백질 생성물 및/또는 발현 세포 유형의 하향조절 및/또는 파괴를 위해 삽입하기 위한 프라임 편집제의 사용이다. 면역원성 에피토프 삽입을 위한 뉴클레오티드 서열은 표적화된 유전자의 코딩된 단백질 및 삽입된 면역원성 에피토프의 상응하는 단백질 번역의 융합 단백질을 생산하는 방식으로 유전자에 표적화될 것이다. 환자의 면역계는 예를 들어 파상풍 또는 디프테리아 또는 홍역에 대한 상용 백신접종으로부터의 표준 사전 면역화의 결과로서 이들 에피토프를 인식하도록 이전에 훈련되었을 것이다. 융합된 에피토프의 면역원성 성질의 결과로서, 환자의 면역계는 프라임 편집된 단백질 (삽입된 에피토프만이 아님) 및 잠재적으로 그것이 발현되는 세포를 인식하고 이를 무력화할 것으로 예상될 것이다.

[2225] 표적화된 유전자에 대한 융합은 삽입된 에피토프 단백질 번역이 면역계 인식을 위해 노출되는 것을 보장하기 위해 필요에 따라 조작될 것이다. 이는 표적화된 유전자에 대한 면역원성 에피토프의 C-말단 융합체, 표적화된 유전자에 대한 면역원성 에피토프의 N-말단 융합체, 또는 면역원성 에피토프가 단백질 구조의 표면-노출 영역 내에 코딩되도록 하는 유전자 내로의 뉴클레오티드의 삽입을 생성하는 단백질 번역을 발생시키는 표적화된 뉴클레오티드 삽입을 포함할 수 있다.

[2226] 표적 유전자 서열과 삽입된 면역원성 에피토프 뉴클레오티드 서열 사이에 삽입된 뉴클레오티드로서 코딩된 단백질 링커는 표적화된 유전자의 면역계 인식, 세포 트래픽킹, 단백질 기능, 또는 단백질 폴딩을 용이하게 하기 위해 본 발명의 일부로서 조작될 필요가 있을 수 있다. 이들 삽입된 뉴클레오티드-코딩된 단백질 링커는 XTEN 링커의 가변 길이 및 서열 또는 글리신-세린 링커의 가변 길이 및 서열을 포함할 수 있다 (그러나 이에 제한되지는 않음). 이들 조작된 링커는 단백질 융합을 성공적으로 용이하게 하기 위해 이전에 사용되었다.

[2227] 본 발명의 구별되는 특색은 특정 아미노산 서열에 대해 이전에 획득된 면역 반응을 달리 비-면역원성 단백질에 대한 면역 반응을 유도하는 수단으로서 사용하는 능력을 포함한다. 또 다른 구별되는 특색은 부산물 편집으로서 원치 않는 indel을 높은 수준으로 유도하지 않고 그의 삽입에 있어서 효율적인 방식으로 이들 면역원성 에피토프의 뉴클레오티드 서열을 표적에 삽입하는 능력이다. 본원에 논의된 본 발명은 또한 세포 유형-특이적 전달 방법 (예컨대 AAV 혈청형)을 조합하여 관심 세포 유형 내에 에피토프를 삽입하여 그에 대한 면역 반응을 촉발하는 능력을 갖는다.

[2228] 면역원성 에피토프를 병원성 유전자 내로 삽입하는 수단으로서의 프라임 편집은 환자의 면역계가 매우 다양한 질환 (면역-종양학 전략에 의하는 것과 같이 암으로 제한되지는 않음)과 싸우도록 프로그램화하는 데 사용될 수 있다. 이러한 기술의 즉각적인 관련 사용은 관련 종양유전자 예컨대 HER2 또는 성장 인자 예컨대 EGFR에 대한 면역 반응을 유발함으로써 종양의 면역 회피 메카니즘을 손상시킬 수 있기 때문에, 암 치료제로서의 사용일 것이다. 이러한 접근법은 T-세포 조작과 유사한 것으로 보일 수 있지만, 이러한 접근법의 한가지 신규 진전은 이것이 조작된 T-세포를 생성하여 환자 내로 도입할 필요 없이 많은 세포 유형에서, 그리고 암 이외의 질환에 대해 이용될 수 있다는 것이다.

[2229] PE를 사용하여 대부분의 사람들이 이미 백신접종된 면역원성 에피토프 (파상풍, 백일해, 디프테리아, 홍역, 볼거리, 풍진 등)를 질환을 구동하는 외래 또는 내인성 유전자 내로 삽입하여, 환자의 면역계가 그러한 단백질을 무력화하도록 학습시킨다.

[2230] 상기 언급된 전략으로부터 잠재적 치료 이익을 갖는 것으로 여겨지는 질환은, 예컨대 치명적 가족성 불면증에서 독성 단백질의 응집에 의해 유발되는 것을 포함한다. 유익할 수 있는 다른 질환은 달리 비독성인 내인성 단백질의 병원성 과다발현에 의해 유발되는 것, 및 외래 병원체에 의해 유발되는 것을 포함한다.

[2231] 1차 치료 적응증은 상기 언급된 것, 예컨대 암, 프리온 및 다른 신경변성 질환, 감염성 질환을 위한 치료, 및 예방적 의학을 포함한다. 2차 치료 적응증은 후기-발병 유전 질환을 갖는 환자에 대한 예방적 관리를 포함할 수 있다. 현행 표준 관리 의학은 일부 질환, 예컨대 특히 공격성 암에 대해, 또는 의약이 질환 증상의 완화를

돕는 경우에 질환이 완전히 치유될 때까지 프라임 편집과 함께 사용될 수 있을 것으로 예상된다.

[2232] 하기는 프라임 편집에 의해 삽입될 수 있는 면역원성 에피토프의 예이며, 이는 하기를 달성하는 데 사용될 수 있다:

백신	질환	에피토프 아미노산 서열	예시적인 핵산 서열 (8)
1	파상풍 톡소이드	QYIKANSKFIGITEL (SEQ ID NO: 396)	CATGATATAAAAGCAAATTCTAAATTTATAGGTATAACTGAACTA (SEQ ID NO: 397)
2	디프테리아 독소 돌연변이체 CRM197	GADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDWKEFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYINNWEQAKALSVELEINFETRGKRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDWDVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVGELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPFLHDGYAVSWNTVEDSIIRTGFQGESGHDIKITAENTPLPIAGVLLPTIPGKLDVNKSKTHISVNGRKIRMRCRAIDGDVTFCRPKSPVYVGNGVHANLHVAFHRSSSEKIHSNEISS	GGCGCCGACGACGTGGTGGACAGCAGCAAGAGCTTCGTGATGGAGAACTTCAGCAGCTACCACGGCACCAAGCCCGGCTACGTGGACAGCATCCAGAAGGGCATCCAGAAGCCCAAGAGCGGCACCCAGGGCAACTACGACGACGACTGGAAGGAGTTCTACAGCACCGACAACAAGTACGACGCCGCCGGCTACAGCGTGGACAACGAGAACCCCCTGAGCGGCAAGGCCGGCGGCGTGGTGAAGGTGACCTACCCCGGCCTGACCAAGGTGCTGGCCCTGAAGGTGGACAACGCCGAGACCATCAAGAAGGAGCTGGGCCTGAGCCTGACCGAGCCCCTGATGGAGCAGGTGGGCACCGAGGAGTTCATCAAGAGGTTCGGCGACGGCGCCAGCAGGGTGGTGCTGAGCCTGCCCTTCGCCGAGGGCAGCAGCAGCGTGGAGTACATCAACAACTGGGAGCAGGCCAAGGCCCTGAGCGTGGAGCTGGAGATCAACTTCGAGACCAGGGGCAAGAGGGGCCAGGACGCCATGTACGAGTACATGGCCCAGGCCTGCGCCGGCAACAGGGTGAGGAGGAGCGTGGGCAGCAGCCTGAGCTGCATCAACCTGGACTGGGACGTGATCAGGGACAAGACCAAGACCAAGATCGAGAGCCTGAAGGAGCACGGCCCCATCAAGAACAAGATGAGCGAGAGCCCCAACAAGACCGTGAGCGAGGAGAAGGCCAAGCAGTACCTGGAGGAGTTCCACCAGACCGCCCTGGAGCACCCCGAGCTGAGCGAGCTGAAGACCGTGACCGGCACCAACCCCGTGTTCGCCGGCGCCAACTACGCCGCCTGGGCCGTGAACGTGGCCCAGGTGATCGACAGCGAGACCGCCGACAACCTGGAGAAGACCACCGCCGCCCTGAGCATCCTGCCCGGCATCGGCAGCGTGATGGGCATCGCCGACGGCGCCGTGCACCACAACACCGAGGAGATCGTGGCCCAGAGC

[2233]

		DSIGVLGYQKTV DHTKVNSKLSLFF EIKS (SEQ ID NO: 630)	ATCGCCCTGAGCAGCCTGATGGTGG CCCAGGCCATCCCCCTGGTGGGCGA GCTGGTGGACATCGGCTTCGCCGCC TACAACTTCGTGGAGAGCATCATCA ACCTGTTCCAGGTGGTGCACAACAG CTACAACAGGCCCGCCTACAGCCCC GGCCACAAGACCCAGCCCTTCCTGC ACGACGGCTACGCCGTGAGCTGGAA CACCGTGGAGGACAGCATCATCAGG ACCGGCTTCCAGGGCGAGAGCGGCC ACGACATCAAGATCACCGCCGAGAA CACCCCCCTGCCCATCGCCGGCGTG CTGCTGCCCACCATCCCCGGCAAGC TGGACGTGAACAAGAGCAAGACCCA CATCAGCGTGAACGGCAGGAAGATC AGGATGAGGTGCAGGGCCATCGACG GCGACGTGACCTTCTGCAGGCCCAA GAGCCCCGTGTACGTGGGCAACGGC GTGCACGCCAACCTGCACGTGGCCT TCCACAGGAGCAGCAGCGAGAAGAT CCACAGCAACGAGATCAGCAGCGAC AGCATCGGCGTGCTGGGCTACCAGA AGACCGTGGACCACACCAAGGTGAA CAGCAAGCTGAGCCTGTTCTTCGAG ATCAAGAGC (SEQ ID NO: 399)
3	볼거리	GTYRLIPNARANL TA (SEQ ID NO: 400)	GGCACCTACAGGCTGATCCCCAACG CCAGGGCCAACCTGACCGCC (SEQ ID NO: 401)
4	볼거리	PSKFFTISDSATFA PGPVSNA (SEQ ID NO: 402) PSKLFIMLDNATF APGPVVNA (SEQ ID NO: 404)	Ccgagcaaatttttaccattagcgatagcgcgacctttgcg ccgggcccggtgagcaacgcg (SEQ ID NO: 403) Ccgagcaaactgtttattatgctggataacgcgacctttgcg ccgggcccggtggtgaacgcg (SEQ ID NO: 405) 발생 균주 (표1) 분기 중에서 백신 균주 JL5와 발생 균주 (표2) 사이의 헤마글루티닌-뉴라미니다제 (HN) 다양성으로부터 선택된 예
5	풍진 바이러스 (RV)	TPPPYQVSCGGES DRASARVIDPAA QS (SEQ ID NO: 406)	ACCCCCCCCCCCTACCAGGTGAGCT GCGGCGGCGAGAGCGACAGGGCCA GCGCCAGGGTGATCGACCCCGCCGC CCAGAGC (SEQ ID NO: 407)

[2234]

[2235]

6	헤마글루티닌	PEYAYKIVKNKK MEDGFLQGMVD GWYGHHSNEQGS GLMENERTLDKA NPNNDLCSWSDH EASSNNTNQEDLL QRESRRKKRIGTS TLNQRGNCNTKC QTEEARLKREEVS LVKSDQCSNGSL QCRANNSTEQVD (SEQ ID NO: 408)	CCCGAGTACGCCTACAAGATCGTGA AGAACAAGAAGATGGAGGACGGCT TCCTGCAGGGCATGGTGGACGGCTG GTACGGCCACCACAGCAACGAGCAG GGCAGCGGCCTGATGGAGAACGAG AGGACCCTGGACAAGGCCAACCCCA ACAACGACCTGTGCAGCTGGAGCGA CCACGAGGCCAGCAGCAACAACACC AACCAGGAGGACCTGCTGCAGAGGG AGAGCAGGAGGAAGAAGAGGATCG GCACCAGCACCCTGAACCAGAGGGG CAACTGCAACACCAAGTGCCAGACC GAGGAGGCCAGGCTGAAGAGGGAG GAGGTGAGCCTGGTGAAGAGCGACC AGTGCAGCAACGGCAGCCTGCAGTG CAGGGCCAACAACAGCACCGAGCA GGTGGAC (SEQ ID NO: 409)
7	뉴라미니다제	TKSTNSRSGGISG PDNEAPVGEAPSP YGDNPRPNDGNN IRIGSKGYNGIITD TIEESCSCYPDAK VVKSVELDSTIWT SGSSPNQKIITIGW DPNGWTGTPMSP NGAYGTDGPSNG QANQHQAESISA GNSSLCPIRDNWH GSNRSWSWPDGA E (SEQ ID NO: 410)	ACCAAGAGCACCAACAGCAGGAGC GGCGGCATCAGCGGCCCCGACAACG AGGCCCCCGTGGGCGAGGCCCCCAG CCCCTACGGCGACAACCCCAGGCCC AACGACGGCAACAACATCAGGATCG GCAGCAAGGGCTACAACGGCATCAT CACCGACACCATCGAGGAGAGCTGC AGCTGCTACCCCGACGCCAAGGTGG TGAAGAGCGTGGAGCTGGACAGCAC CATCTGGACCAGCGGCAGCAGCCCC AACCAGAAGATCATCACCATCGGCT GGGACCCCAACGGCTGGACCGGCAC CCCCATGAGCCCCAACGGCGCCTAC GGCACCGACGGCCCCAGCAACGGCC AGGCCAACCAGCACCAGGCCGAGA GCATCAGCGCCGGCAACAGCAGCCT GTGCCCCATCAGGGACAACTGGCAC GGCAGCAACAGGAGCTGGAGCTGGC CCGACGGCGCCGAG (SEQ ID NO: 411)
8	H5N1 바이러스 헤마글루티닌 상의 TAP (항원 제시 수송)	EKIVLLLAMMEKI VLLLAKCQTPMG AIKAVDGVTNKC PYLGSPSF (SEQ ID NO: 412)	GAGAAGATCGTGCTGCTGCTGGCCA TGATGGAGAAGATCGTGCTGCTGCT GGCCAAGTGCCAGACCCCCATGGGC GCCATCAAGGCCGTGGACGGCGTGA CCAACAAGTGCCCCTACCTGGGCAG CCCCAGCTTC (SEQ ID NO: 413)
9	h5n1 바이러스	IRPCFWVELNPNQ KIITIRPCFWVELI	ATCAGGCCCTGCTTCTGGGTGGAGC TGAACCCCAACCAGAAGATCATCAC

	뉴라미니다제 상의 TAP (항원 제시 수송)	CYPDAGEIT (SEQ ID NO: 414)	CATCAGGCCCTGCTTCTGGGTGGAG CTGATCTGCTACCCCGACGCCGGCG AGATCACC (SEQ ID NO: 415)
10	부류 I HLA에 대한 헤마글루티닌 에피토프	MEKIVLLLAEKIV LLLAMCPYLGSPS FKCQTPMGAIKA VDGVTNK (SEQ ID NO: 416)	ATGGAGAAGATCGTGCTGCTGCTGG CCGAGAAGATCGTGCTGCTGCTGGC CATGTGCCCCTACCTGGGCAGCCCC AGCTTCAAGTGCCAGACCCCCATGG GCGCCATCAAGGCCGTGGACGGCGT GACCAACAAG (SEQ ID NO: 417)
11	부류 I HLA에 대한 뉴라미니다제 에피토프	NPNQKIITICYPDAG EITIRPCFWVELRPC FWVELI (SEQ ID NO: 418)	AACCCCAACCAGAAGATCATCACCAT CTGCTACCCCGACGCCGGCGAGATCA CCATCAGGCCCTGCTTCTGGGTGGAG CTGAGGCCCTGCTTCTGGGTGGAGCT GATC (SEQ ID NO: 419)
12	부류 II HLA에 대한 헤마글루티닌 에피토프	MVSLVKSDQIGTST LNQR (SEQ ID NO: 420)	ATGGTGAGCCTGGTGAAGAGCGACCA GATCGGCACCAGCACCCTGAACCAGA GG (SEQ ID NO: 421)
13	부류 II HLA에 대한 뉴라미니다제 에피토프	YNGIITDTI (SEQ ID NO: 422)	TACAACGGCATCATCACCGACACCAT C (SEQ ID NO: 423)
14	헤마글루티닌 에피토프 H5N1-결합된 부류 I 및 부류 II HLA	MEKIVLLLAEKIVL LLAMMVSLVKSDQ CPYLGSPSFIGTSTL NQRKCQTPMGAIK AVDGVTNK (SEQ ID NO: 424)	ATGGAGAAGATCGTGCTGCTGCTGGC CGAGAAGATCGTGCTGCTGCTGGCCA TGATGGTGAGCCTGGTGAAGAGCGAC CAGTGCCCCTACCTGGGCAGCCCCAG CTTCATCGGCACCAGCACCCTGAACC AGAGG (SEQ ID NO: 425)
15	뉴라미니다제 에피토프 H5N1-결합된 부류 I 및 부류 II HLA	NPNQKIITIYNGIIT DTICYPDAGEITIRP CFWVELRPCFWVE LI (SEQ ID NO: 426)	AACCCCAACCAGAAGATCATCACCAT CTACAACGGCATCATCACCGACACCA TCTGCTACCCCGACGCCGGCGAGATC ACCATCAGGCCCTGCTTCTGGGTGGA GCTGAGGCCCTGCTTCTGGGTGGAGC TGATC (SEQ ID NO: 427)

[2236]

[2237] 하기는 이뮤노에피토프 태그부착을 위한 표적 유전자 내로 통합될 수 있는 에피토프의 추가의 예이다:

[2238] 실시예 5에 인용된 참고문헌

[2239] 하기 참고문헌은 그 전문이 참조로 포함된다.

1. X. Wen, K. Wen, D. Cao, G. Li, R. W. Jones, J. Li, S. Szu, Y. Hoshino, L. Yuan, Inclusion of a universal tetanus toxoid CD4(+) T cell epitope P2 significantly enhanced the immunogenicity of recombinant rotavirus ΔVP8* subunit parenteral vaccines. *Vaccine* 32, 4420-4427 (2014).

2. G. Ada, D. Isaacs, Carbohydrate-protein conjugate vaccines. *Clin Microbiol Infect* 9, 79-85 (2003).

3. E. Malito, B. Bursulaya, C. Chen, P. L. Surdo, M. Picchianti, E. Balducci, M. Biancucci, A. Brock, F. Berti, M. J. Bottomley, M. Nissum, P. Costantino, R. Rappuoli, G. Spraggon, Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197. Proceedings of the National Academy of Sciences 109, 5229 (2012).

4. J. de Wit, M. E. Emmelot, M. C. M. Poelen, J. Lanfermeijer, W. G. H. Han, C. van Els, P. Kaaijk, The Human CD4(+) T Cell Response against Mumps Virus Targets a Broadly Recognized Nucleoprotein Epitope. *J Virol* 93, (2019).

5. M. May, C. A. Rieder, R. J. Rowe, Emergent lineages of mumps virus suggest the need for a polyvalent vaccine. *Int J Infect Dis* 66, 1-4 (2018).

6. M. Ramamurthy, P. Rajendiran, N. Saravanan, S. Sankar, S. Gopalan, B. Nandagopal, Identification of immunogenic B-cell epitope peptides of rubella virus E1 glycoprotein towards development of highly specific immunoassays and/or vaccine. *Conference Abstract*, (2019).

7. U. S. F. Tambunan, F. R. P. Sipahutar, A. A. Parikesit, D. Kerami, Vaccine Design for H5N1 Based on B- and T-cell Epitope Predictions. Bioinform Biol Insights 10, 27-35 (2016).

[2240]

[2241] 실시예 6 - 생체내 표적 단백질의 화학적-유도된 이량체화를 가능하게 하기 위한 PE의 사용

[2242] 본원에 기재된 프라임 편집제는 또한 프라임 편집을 사용한 소분자 결합 단백질을 코딩하는 유전자의 게놈 통합에 의해 편리한 소분자 약물의 제어 하에 이량체화-유도된 생물학적 과정, 예컨대 수용체 신호전달을 배치하는 데 사용될 수 있고, 이는 본원에 기재된다. 본원에 기재된 프라임 편집제를 사용하여, 소분자 결합 단백질을 코딩하는 유전자 서열은 살아있는 세포 또는 환자에서 관심 표적 단백질을 코딩하는 유전자 내에 삽입될 수 있다. 이러한 편집 단독은 생리학적 효과를 갖지 않아야 한다. 전형적으로 각각이 표적 단백질의 카피에 융합된 2개의 약물-결합 단백질 도메인에 동시에 결합할 수 있는 이량체 소분자인 소분자 약물의 투여 시, 소분자는 표적화된 단백질의 이량체화를 유도한다. 이어서, 이러한 표적 단백질 이량체화 사건은 생물학적 신호전달 사건, 예컨대 적혈구생성 또는 인슐린 신호전달을 유도한다.

[2243] 실시예 7 - DNA 서열 삽입, 결실 및 이중 프라임 편집제를 사용한 대체

[2244] 본 실시예는 새로운 DNA 서열의 정확한 삽입, 내인성 게놈 DNA 서열의 정확한 결실 또는 내인성 게놈 DNA 서열의 새로운 DNA 서열로의 대체를 위한 다중-플랩 프라임 편집의 사용, 구체적으로 이중 프라임 편집제의 사용을 기재한다. 도 90을 참조한다. 본 발명은 게놈 조작에서의 오랜 목적: 살아있는 세포의 게놈에서의 DNA 서열의 프로그램가능한 삽입, 결실 또는 대체를 다룬다. 이 기술은 연구, 치료 또는 산업 적용을 위해 임의의 살아있는 세포의 게놈을 변형시키거나 편집하는 데 적용될 수 있다.

[2245] 이전에, 뉴클레아제-유도된 이중 가닥 DNA 파괴 (DSB)의 상동성 지정 복구 (HDR)는 살아있는 세포의 게놈에서 DNA 서열을 삽입, 결실 또는 대체하는 데 사용되었다. 다재다능하지만, HDR의 낮은 효율은 많은 포유동물 세포 유형, 특히 치료상 적절한 인간 세포 유형에서 그의 적용을 제한한다[1-4]. 더욱이, DSB의 생성은 오류-유발 단부-연결 메카니즘에 의한 우선적 복구로 이어지며, 이는 삽입 및 결실 (indel) 복구 생성물을 생산하고, 이는 후속적으로 목적하는 편집된 DNA 서열로 전환될 수 없다[5-10]. 그 결과, indel은 전형적으로 뉴클레아제가 포유동물 세포에서 사용되는 경우에 대다수의 변형된 게놈 DNA 생성물을 나타낸다. 더욱이, 뉴클레아제에 의해 생성된 DSB는 바람직하지 않은 게놈 변화, 예컨대 큰 결실 및 염색체 전위[11] 및 p53의 활성화[12,13]로 이어질 수 있다. 구조적 변이체 (SV)는 또한 상동 재조합에 의존하지 않으면서 프로그램가능한 뉴클레아제를 사용하여 생성되거나 교정될 수 있으며, 다만 뉴클레아제-기반 전략은 이들 적용에 대해 유사한 결점을 겪는다. 예를 들어, 이중-절단 전략을 사용하여 표적화된 게놈 DNA 서열을 역전시킬 수 있지만, 대부분의 생성물 대립유전자는 절단 부위 사이에 목적하는 역전된 서열 대신 큰 결실을 함유하거나 또는 개별 절단 부위에 indel을 함유한다.

[2246] 프라임 편집은 오류-유발 이중 가닥 DNA 파괴를 요구하지 않으면서 게놈 DNA 서열의 삽입, 결실 또는 대체를 가능하게 하는 최근에 보고된 게놈 편집 기술이다[14]. 프라임 편집은 Cas9의 표적 게놈 부위로의 지시 및 목적하는 편집을 설치하기 위한 정보의 코딩 둘 다를 수행하는 조작된 프라임 편집 가이드 RNA (pegRNA)와 쌍형성된 조작된 Cas9 닉카제-리버스 트랜스크립타제 융합 단백질 (PE1 또는 PE2)을 사용한다. 프라임 편집은 다단계 편집 과정을 통해 진행된다: 1) Cas9 도메인은 pegRNA의 스페이서 서열에 의해 특정된 표적 게놈 DNA 부위에 결합하여 닉킹하고; 2) 리버스 트랜스크립타제 도메인은 닉킹된 게놈 DNA를 프라이머로서 사용하여, 역전사를 위한 주형으로서 pegRNA 상의 조작된 연장부를 사용하여 편집된 DNA 가닥의 합성을 개시하고-이는 편집된 DNA 서열을 함유하는 단일 가닥 3' 플랩을 생성하고; 3) 세포 DNA 복구는 5' 플랩 종의 변위에 의해 3' 플랩 중간체를 분해하고, 이는 편집된 3' 플랩에 의한 침입, 원래 DNA 서열을 함유하는 5' 플랩의 절제 및 새로운 3' 플랩의 라이게이션을 통해 발생하여 편집된 DNA 가닥을 혼입시킴으로써, 1개의 편집된 가닥 및 1개의 비편집된 가닥의 헤테로듀플렉스를 형성하고; 4) 세포 DNA 복구는 복구를 위한 주형으로서 편집된 가닥을 사용하여 헤테로듀플렉스 내의 비편집된 가닥을 대체함으로써, 편집 과정을 완료한다.

[2247] 목적하는 편집의 효율적인 혼입은 새로 합성된 3' 플랩이 게놈 DNA 부위에 상동인 서열의 부분을 함유하는 것을 요구한다. 이러한 상동성은 편집된 3' 플랩이 DNA 듀플렉스 내로의 혼입을 위해 내인성 DNA 가닥 (상응하는 5' 플랩)과 경쟁하는 것을 가능하게 한다. 편집된 3' 플랩은 내인성 5' 플랩보다 더 적은 서열 상동성을 함유할 것이기 때문에, 경쟁은 5' 플랩 가닥에 유리할 것으로 예상된다. 따라서, 편집을 함유하는 3' 플랩은 5' 플랩 가닥을 효과적으로 침입 및 대체하지 못할 수 있다. 더욱이, 성공적인 3' 플랩 침입 및 5' 플랩의 제거는 이중 가닥 DNA 게놈의 한 가닥 상에만 편집을 혼입시킨다. 편집의 영구적 설치는 주형으로서 편집된 가닥을 사용하여 비편집된 상보적 DNA 가닥을 대체하기 위한 세포 DNA 복구를 요구한다. 세포는 2차 sgRNA (PE3시스템)를 사용하여 편집에 인접한 비편집된 가닥에 닉을 도입함으로써 편집된 가닥에 비해 비편집된 가닥의 대체를 유리하게 할 수 있지만 (상기 단계 4), 이 과정은 여전히 DNA 복구의 제2 단계에 의존한다. 이들 DNA 복구 단계는 긴 5' 및 3' 플랩 중간체의 평형을 요구하거나 긴 비-상동 영역, 예컨대 긴 삽입 또는 긴 결실을 함유하는 편집에 특히 비효율적일 수 있다.

[2248] 본 실시예는 두 DNA 가닥을 동시에 편집함으로써 플랩 평형 및 비-편집된 상보적 게놈 DNA 가닥 내로의 편집의 후속 혼입과 연관된 도전과제를 다루는 이중 프라임 편집 시스템 (또는 이중-플랩 프라임 편집 시스템)을 기재한다. 이중-플랩 프라임 편집 시스템에서, 2개의 pegRNA는 게놈 부위의 대향 가닥을 표적화하고 편집된 DNA 서열을 함유하는 2개의 상보적 3' 플랩의 합성을 지시하는 데 사용된다 (도 91). 전형적 프라임 편집과는 달리, 상보적인 편집된 가닥이 대신 혼성화에 이용가능하기 때문에, 편집된 DNA 가닥의 쌍 (3' 플랩)이 내인성 게놈 DNA에서 5' 플랩과 직접 경쟁하는 것이 요구되지 않는다. 듀플렉스의 두 가닥이 편집된 DNA로서 합성되기 때문에, 이중-플랩 프라임 편집 시스템은 전형적 프라임 편집에 의해 요구되는 비-편집된 상보적 DNA 가닥의 대체에 대한 필요를 제거한다. 대신, 세포 DNA 복구 기구는 단지 쌍형성된 5' 플랩 (원래 게놈 DNA)만을 절제하고, 쌍형성된 3' 플랩 (편집된 DNA)을 유전자좌 내로 라이게이션시킬 필요가 있다. 따라서, 또한 새로 합성된 DNA 가닥에 게놈 DNA에 상동인 서열을 포함할 필요가 없으므로, 새로운 가닥의 선택적 혼성화를 허용하고 최소 게놈 상동성을 함유하는 편집을 용이하게 한다. DNA의 두 가닥을 절단하는 프라임 편집제의 뉴클레아제-활성 버전은 또한 원래 DNA 서열의 제거를 가속화하는 데 사용될 수 있다.

[2249] 이중 프라임 편집제에 의해 생성된 엇갈린 DNA 닉의 배향에 따라, 시스템은 2개의 닉 사이의 서열을 새로운 목적하는 서열로 대체하거나 (5' 오버행 엇갈린 배향) 또는 새로운 DNA 서열을 수반되는 표적 부위 중복과 함께 삽입할 수 있다 (3' 오버행 엇갈린 배향). 도 90은 닉 사이의 서열이 대체된 배향 (5' 대체된 닉)을 보여준다. 중요하게는, 새로운 3' 플랩은 새로운 DNA 서열의 길이를 따라 완전히 중첩될 필요는 없지만, 대신에 새로운 서열의 단지 한 부분에 상보적일 수 있다. 이러한 방식으로, 각각의 플랩은 새로운 DNA 서열의 전장보다 더 짧을 수 있다.

[2250] HEK293 부위 3 유전자좌에서의 이중-플랩 프라임 편집

[2251] 인간 세포 (HEK293T) 내의 HEK293 부위 3 유전자좌에서의 이중-플랩 프라임 편집을 사용하여 이종 DNA 서열의 고도로 효율적인 도입을 달성하였다. 예를 들어, 유전자좌 내의 90-bp 서열을 6xHis 태그를 코딩하는 새로운 22-bp 서열로 대체하는 것이 80% 초과의 효율로 발생하였다 (도 92 대립유전자 표). 유사하게, 이러한 동일한 90-bp 서열을 GFP-11 펩티드 서열 (52 bp)로 대체하는 것에 대해 >60% 수율을 달성하였다. 또한 GFP-11 태그 삽입을 MYC 유전자의 C-말단에서 >20%의 효율로 수행하였으며, 이에 수반하여 Myc 3' UTR 서열의 한 부분의 프로그램화된 결실이 발생하였다.

[2252] 이중-플랩 프라임 편집 pegRNA

[2253] 이중-플랩 프라임 편집에 사용된 pegRNA의 일반적 설계가 도 93에 제시된다. 이중-플랩 프라임 편집에 사용된 pegRNA는 전형적 프라임 편집에 사용된 것과 유사한 설계를 갖지만, RT 주형 영역이 표적 유전자좌에 대한 임의의 상동성을 코딩할 필요는 없다. 대신, 2개의 pegRNA는 3' 단부가 서로의 역 상보체 서열인 3' 플랩의 합성을 코딩하는 RT 주형을 함유할 수 있다. 3' 플랩 사이의 이러한 상보성은 그의 어닐링 및 내인성 DNA 서열의 의도된 새로운 DNA 서열로의 대체를 촉진한다. 이는 2개의 pegRNA 내의 RT 주형의 5' 영역이 서로에 대해 역 상보체 서열일 것을 필요로 하고, 이러한 상보성의 양은 달라질 수 있다 (도 93).

[2254] 재조합 부위를 설치하기 위한 이중-플랩 프라임 편집의 사용

[2255] 이중-플랩 프라임 편집은 펩티드 태그, RNA 태그, 이뮤노에피토프, 이량체화 도메인 및 레콤비나제 표적 부위를 설치하는 것과 같은 많은 잠재적 적용을 갖는다.

[2256] 하나의 이러한 적용은 게놈 내 사용자-명시된 위치에서의 레콤비나제 또는 인테그라제 서열의 설치이다. 도 94는 Bxb1 레콤비나제 attB (38 bp) 및 attP (50 bp) 부위를 인간 게놈의 표적화된 영역 (HEK293T 부위 3 또는 HEK3) 내로 설치하고, 동시에 2개의 닉 부위 사이에 90 bp의 개재 서열을 결실시키는 것을 예시한다. 3' 플랩 사이의 다양한 정도의 상보성이 성공적인 편집을 허용하며, 다만 보다 긴 상보성 서열이 보다 유리한 목적하는 편집 대 indel 비를 생성한다. 다른 이중-플랩 프라임 편집 적용은 펩티드 또는 단백질 서열로의 유전자의 내인성 태그부착, 또는 돌연변이가 엑손 서열 내에 속하는 다수의 변이체를 치환시키는 잠재력을 갖는 새로운 DNA 서열로의 엑손의 대체를 포함한다.

[2257] 이중-플랩 프라임 편집을 사용하여 인간 게놈 내 표적화된 위치에 1 또는 2개의 레콤비나제 부위를 도입할 수 있다. 단일 레콤비나제 부위가 삽입되는 경우에, 이들은 게놈 레콤비나제 부위와 외인적으로 공급된 DNA, 예컨대 플라스미드 내의 제2 레콤비나제 부위 사이의 레콤비나제-매개된 반응을 위한 랜딩 부위로서 사용될 수 있다. 이는 DNA 카고의 표적화된 통합을 가능하게 한다. 2개의 레콤비나제 부위가 DNA의 인접한 영역에 삽입되는 경우에, 레콤비나제 부위의 배향에 따라, 이들은 레콤비나제-매개된 절제 또는 개재 서열의 역위를 위해, 또는 카고 통합을 위한 외인성 DNA와의 레콤비나제-매개된 카세트 교환을 위해 사용될 수 있다. 상이한 염색체 상의 상용성 레콤비나제 부위의 통합은 표적화된 및 방향적인 염색체 전위를 가능하게 한다. 이중-플랩 프라임 편집을 사용하여 인간 게놈 내 다수의 유전자좌에 레콤비나제 부위를 효율적으로 도입할 수 있다 (도 94). 따라서, 이중-플랩 프라임 편집에 의한 레콤비나제 부위 통합과 DNA 레콤비나제 효소의 쌍형성은 많은 유형의 SV 편집 및 DNA 카고의 표적 통합을 달성하기 위한 강력한 접근법을 나타낸다.

[2258] 실시예 8 - 사중-플랩 프라임 편집을 사용한 DNA 서열 삽입, 결실, 역위, 전위 및 통합

[2259] 본 실시예는 내인성 게놈 DNA 서열의 표적화된 삽입, 결실, 중복 및 대체 및 표적화된 게놈 DNA 서열 역위, 표적화된 염색체 전위 및 외인성 DNA의 표적화된 통합을 위한 다중-플랩 프라임 편집, 구체적으로 사중-플랩 프라임 편집의 사용을 기재한다. 이 기술은 게놈 조작에서의 오랜 목적: 구조적 변이체 (일반적으로 >50 bp의 게놈 서열 변화로서 정의되며, 종종 큰 결실, 삽입/중복, 역위 및 전위에 의해 나타내어짐)의 프로그램가능한 설치 또는 교정, 및 인간 게놈 내 특정 위치에서의 DNA 카고의 표적화된 통합을 다룰 잠재력을 갖는다. 후자는 유전자 증대 치료 전략 및 다른 생명공학 적용을 가능하게 할 수 있다. 이 기술은 연구, 치료 또는 산업 적용을 위해 임의의 살아있는 세포의 게놈을 변형시키거나 편집하는 데 잠재적으로 적용될 수 있다.

[2260] 본 발명은 두 DNA 가닥을 동시에 편집함으로써 플랩 평형 및 비-편집된 상보적 게놈 DNA 가닥 내로의 편집의 후속 혼입과 연관된 도전과제를 다루는 다중-플랩 프라임 편집 시스템을 기재한다. 이는 또한 새로 합성된 3' 플랩이 보다 큰 서열 재배열을 지시하도록 함으로써 전통적인 프라임 편집의 능력을 확장시킨다.

[2261] 레콤비나제 효소에 대한 필요 없이 (이중-플랩 프라임 편집에서와 같이) 그러나 유사한 다재다능성으로 목적하는 서열 변경을 직접 수행할 수 있는 사중-플랩 프라임 편집 접근법을 개발하였다. 사중-플랩 프라임 편집에서, 4개의 상이한 pegRNA 서열이 PE2 편집제와 함께 세포에 전달된다. 한 쌍의 pegRNA는 이중-플랩 프라임 편집에서와 같이 상보적 DNA 플랩의 합성의 주형이 되는 반면, 다른 쌍의 pegRNA는 직교 쌍의 상보적 DNA 플랩의 합성의 주형이 된다. 이들 플랩의 위치, 그의 상대적 배향 및 상보성 쌍형성은 어느 부류의 재배열이 발생할 것인지를 좌우한다. 재배열된 DNA의 접합부는 재배열의 배향을 지시하고 하류 과정을 위한 유용한 DNA 서열로서의 역할을 할 수 있는 pegRNA-주형화된 3' 플랩에 의해 코딩되는 서열을 함유한다. 표적화된 게놈 서열 역위를 사중-플랩 프라임 편집 전략으로 수행하였다 (도 95). HEK293T 세포 내 AAVS1 유전자좌에서의 2.7-kb 서열의 역위를, 역위 접합부에 Bxb1 attB 및 attP 서열을 배치하는 4개의 pegRNA를 사용하여 달성하였다. 앰플

리콘 서열분석을 사용하여 예상 접합부의 확인을 수행하였으며, 이는 서열분석 판독물의 ≥90%에서 attB 또는 attP 부위 서열을 함유하는 정확한 접합부를 입증하였다 (도 96).

[2262] 외인성 DNA 플라스미드의 표적화된 통합을 또한 사중-플랩 프라임 편집으로 달성하였다. 이러한 접근법에서, 2개의 pegRNA는 내인성 게놈 DNA 부위를 표적화하는 한편, 2개의 pegRNA는 카고 플라스미드 상의 영역을 표적화한다. 3' DNA 플랩은 플라스미드 및 게놈 DNA가 혼성화된 3' 플랩에 의해 가교되도록 게놈 DNA 및 플라스미드 DNA 둘 다 상에서 합성된다 (도 97a). 5' 오버행의 절제 및 닉의 라이게이션은 통합된 플라스미드 생성물을 생성한다. 통합의 배향은 pegRNA 주형 서열의 쌍형성에 의해 제어될 수 있다.

[2263] H2B-GFP 융합 단백질의 발현을 코딩하는 플라스미드를 사중-플랩 프라임 편집을 사용하여 HEK293T 세포 내 AAVS1 유전자좌에서 통합시켰다. 이로써 유동 세포측정법에 의해 평가 시 1.6%의 세포가 장기간의 핵 국재화된 GFP 형광을 수득하였다. 예측 접합부의 PCR은 pegRNA-주형화된 BxB1 attP 서열을 함유하는 예상 생성물을 나타냈다 (도 97b).

[2264] 사중 플랩 프라임 편집을 사용하여 염색체 전위를 또한 수행하여 (도 98a) 염색체 8 상의 MYC 유전자좌와 염색체 19 상의 TIMM44 유전자좌 사이의 염색체 전위를 지시하였다 (도 98b).

[2265] IDS 유전자좌에서의 이중 플랩 프라임 편집을 12쌍의 pegRNA 및 PE2로 수행하였다. HEK293T 세포를 PE2 및 상이한 쌍의 pegRNA (예를 들어, 제1 칼럼에서 pegRNA A1_a 및 pegRNA B2_a (attP 부위를 정방향으로 설치하기 위한 주형을 가짐))로 형질감염시켰다. 효율을 HTS에 의해 측정하였다. 이 데이터는 이중-플랩 편집이 관심 서열을 최대 ~80%의 효율로 IDS 유전자좌에 성공적으로 삽입할 수 있다는 것을 보여준다 (도 99).

[2266] 이중-플랩-매개된 중복을 또한 293T 세포 내 AAVS1 유전자좌에서 수행하였다 (도 100a). 이중-플랩 pegRNA과 PE2를 사용하여 AAVS1에서의 유전자 서열의 중복을 유도하였다 (도 100b).

[2267] MYC-CCR5의 다중-플랩 프라임 편집 유도된 전위를 또한 수행하였다 (도 101). MYC-CCR5 전위를 사중-플랩 pegRNA 및 PE2에 의해 유도하였다. MYC-CCR5 전위 사건을 사중-pegRNA에 의해 유도하였다. 4개의 상이한 세트의 pegRNA를 HEK293T 세포에서 시험하였다. 유도된 chr8과 chr3 사이의 전위 접합부 생성물을 접합부 프라이머에 의해 증폭시켰다. 예상 접합부 대립유전자에 대해 정렬된 판독물의 퍼센트가 그래프에 제시된다. 이는 사중-플랩 프라임 편집이 MYC 및 CCR5 유전자의 전위를 매개할 수 있으며, 생성물 순도는 접합부 1에서 거의 100% 및 접합부 2에서 ~50%라는 것을 보여준다. 대표적인 대립유전자 플롯은 접합부 1에서 예상 대립유전자 서열에 대해 정렬된 서열을 보여준다.

[2268] 이중-플랩 및 사중-플랩 프라임 편집을 또한 다른 인간 세포주에서 수행하였다 (도 102a-102b). 이중-플랩 프라임 편집을 4종의 상이한 인간 세포주에서 수행하였다 (도 102a). 3종의 상이한 게놈 유전자좌 (IDS, MYC 및 TIMM44)를 편집하기 위해 HEK293T 및 HeLa 세포를 이중 pegRNA 및 PE2로 형질감염시켰다. U2OS 및 K562 세포를 동일한 성분으로 뉴클레오펙션시켰다. 이중-플랩 프라임 편집은 표적화된 유전자좌에서, 특히 HEK293T 및 K562 세포에서 모든 4종의 인간 세포에 걸쳐 강건한 편집 효율을 나타냈다. 이중-플랩 프라임 편집을 가능하게 하기 위한 세포 메카니즘은 많은 인간 세포 유형에 걸쳐 보존된다. 2.7 kb 서열의 다중-플랩 (사중-pegRNA) 지시된 역위를 또한 HeLa 세포 내 AAVS1 유전자좌에서 수행하였다 (도 102b).

[2269] 역위 효율을 또한 CCR5 유전자좌에서 HTS에 의해 측정하였다 (도 103a-103b). 예상 역위 편집 대립유전자의 백분율을 HTS에 의해 측정하였다. PE2의 4개의 사중-pegRNA 세트를 각각 HEK293T 세포에 형질감염시켰다. HEK293T 세포 내 CCR5 유전자좌에서 이중-플랩 매개된 서열 중복 (~100nt)을 수행하였다 (도 103a). 편집 효율은 HTS 300-사이클 쌍-단부 서열 분석을 통해 ~1.5%를 달성한다. 사중-플랩-매개된 서열 역위 (~95-117nt)를 또한 HeLa 세포 내 CCR5 유전자좌에서 수행하였다 (도 103b). 편집 효율은 HTS 300-사이클 쌍-단부 서열 분석을 통해 ~1.2%를 달성한다. 이러한 결과는 다중-플랩 프라임 편집이 정확하게 CCR5 유전자좌에서의 중복 및 역위를 성공적으로 매개할 수 있다는 것을 보여준다. 편집 특이성은 표적화된 서열이 중복되는 경우에 높다 (indel의 백분율 < 2%).

[2270] 다음으로, HEK293T 세포를 pegRNA 및 PE2와 함께 Exo1, Fen1, 적색 형광 단백질 (대조군), DNA2, Mlh1 neg 및 P53 억제제를 발현하는 플라스미드로 형질감염시켰다 (도 104a-104d). 편집 효율을 HTS에 의해 측정하였다. 편집 효율을 후보와 RFP 대조군 사이에 비교하였다. HEK293T 세포를 또한 각각의 표적 유전자좌에 대해 siRNA 플라스미드 및 pegRNA 및 PE2로 형질감염시켰다 (도 104e). 비-표적화 siRNA (siNT)를 대조군으로서 사용하였다. 편집 효율을 HTS에 의해 측정하였다. 편집 효율을 각각의 표적 유전자좌에서의 각각의 siRNA 녹다운과 siNT ctrl 사이에 비교하였다. HEK293T 세포를 이중-pegRNA, PE2 및 각각 Exo1, Fen1, 적색 형광 단백질

(ctrl), DNA2, M1h1 neg 및 P53 억제제를 발현하는 플라스미드로 형질감염시켰다. 편집 효율을 HTS에 의해 측정하였다. 편집 효율을 후보와 RFP 대조군 사이에 비교하였다. FEN1의 과다발현은 모든 4종의 표적화된 유전자좌 (MYC, TIMM44, IDS, CCR5)에서 이중 플랩 편집 효율을 개선시키는 것으로 밝혀졌다.

[2271] 이중-플랩-매개된 서열 중복을 또한 AAVS1 유전자좌에서 수행하였다 (도 105a). 2개의 고유한 3' 플랩 구조를 생성하는 이중 pegRNA를 사용함으로써, 293T 세포 내 AAVS1 유전자좌에서 ~300bp 서열 중복이 유도되었다 (도 105b). 예상 대립유전자를 특이적 프라이머로 증폭시키고, HTS에 적용하였다. 대략 94%의 판독물을 중복을 갖는 예상 대립유전자에 대해 정렬하였다. 중복 생성물은 비처리된 샘플에서는 관찰되지 않았다.

[2272] 표적화된 IDS 게놈 서열 역위를 또한 사중-플랩 프라임 편집으로 수행하였다 (도 106). 헌터 증후군 환자의 대략 13%는 IDS 유전자 서열의 역위를 갖는 것으로 나타났다 (Bondeson et al., Human Molecular Genetics, 1995). 사중-플랩 프라임 편집을 적용하여 HEK293T 세포에서 대략 40kb IDS 게놈 서열의 이러한 병원성 역위를 유도하였다. 6 세트의 사중 pegRNA를, HEK293T 세포를 pegRNA 및 PE2로 형질감염시킴으로써 시험하였다. 프라이머를 사용하여 접합부 "ab" 및 접합부 "cd"에서 역전된 서열을 특이적으로 증폭시켰다. 예상된 역전된 대립유전자 서열의 대략 95%가 IDS_QF1의 두 접합부에서 관찰되었다. 다른 세트의 pegRNA는 또한 두 접합부에서 높은 백분율의 예상 대립유전자 서열을 생성한다. 역전된 접합부 생성물은 비처리된 샘플에서는 관찰되지 않았다.

[2273] 또한, PegRNA 3' 모티프 변형이 IDS 유전자좌에서 이중-플랩 편집 효율을 개선시키는 것으로 밝혀졌다 (도 107). 이중 플랩 편집 효율을 추가로 개선시키기 위해, 유사매듭형 evoPreQ1 모티프를 도입하여 pegRNA 3' 단부를 보호하였다. 비변형된 및 evoPreQ1-변형된 이중 pegRNA에 의해 생성된 편집 효율을 비교하면, 표적화된 IDS 유전자좌에서 변형된 pegRNA에 의한 편집 효율의 전체 증가가 존재하였다. 이중-플랩 편집 효율의 개선은 최대 5.3배까지 도달할 수 있다.

[2274] 참고문헌 (실시예 7 및 8):

1. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. PNAS 91, 6064–6068 (1994).

2. Heyer, W.-D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. Annu. Rev. Genet. 44, 113–139 (2010).

3. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. Mol. Cell 47, 497–510 (2012).

4. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. Nature Medicine 21, 121–131 (2015).

5. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol 14, 8096–8106 (1994).

6. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161, 1169–1175 (2002).

7. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11, 636–646 (2010).

8. Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).

9. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013).

10. Jinek, M. et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013).

11. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nature Biotechnology 36, 765–771 (2018).

12. Ihry, R. J. et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. Nature Medicine 24, 939–946 (2018).

13. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. Nature Medicine 24, 927–930 (2018).

14. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

[2275]

[2276] 실시예 9 - 트윈 프라임 편집 및 부위-특이적 레콤비나제를 사용한 프로그램가능한 큰 DNA 결실, 대체, 통합 및 역위

[2277] 게놈 내 명시된 위치에서의 DNA 서열의 표적화된 결실, 대체, 통합 또는 역위는 원칙적으로 많은 인간 유전 질환을 연구 또는 치료하는 데 사용될 수 있다. 그러나, 포유동물 세포 게놈과 관련하여 이들 DNA 서열 형질전환을 정확하고 효율적으로 수행하는 것은 도전과제로 남아있다. 이중 가닥 DNA 파괴를 요구하지 않으면서 내인성 인간 게놈 부위에서 DNA 서열의 프로그램가능한 대체 또는 절제 방법인 트윈 프라임 편집 (트윈PE)이 본원에 제시된다. 트윈PE는 프라임 편집제 (PE) 단백질 및 게놈 DNA의 대향 가닥 상의 상보적 DNA 플랩의 합성의 주형이 되는 2개의 프라임 편집 가이드 RNA (pegRNA)를 사용하여, PE-유도된 닉 부위 사이의 내인성 DNA 서열을 pegRNA-코딩된 서열로 대체한다. 트윈PE는 최대 780 bp의 정확한 결실 및 최대 108 bp의 새로운 서열로의 게놈 DNA 서열의 정확한 대체를 수행할 수 있는 것으로 본원에 제시된다. 단일 또는 멀티플렉스화 트윈PE를 부위-특이적 세린 레콤비나제와 조합함으로써, AAVS1, CCR5 및 ALB를 포함한 세이프-하버 유전자좌에서의 유전자-크기 조정된 DNA 플라스미드의 표적화된 통합, 및 추가로, 공통 헌터 증후군 대립유전자를 교정하는 데 사용될 수 있는 IDS에서의 39-kb 역위가 또한 입증되는 것으로 제시된다. 트윈PE는 이중 가닥 DNA 파괴를 요구하지 않으면서 게놈 편집 능력을 실질적으로 확장시키고, 다른 게놈 편집 도구와 조합되어, 인간 세포에서 복합 병원성 대립유전자 변이체의 교정 또는 보완을 가능하게 한다.

[2278] 질환-연관 인간 유전자 변이체는 단일-염기 쌍 치환에서 메가 염기-규모 결실 및 재배열에 이르는 다양한 서열 변화를 통해 발생한다. 인간 세포에서 이들 병원성 변이체를 설치, 교정 또는 보완할 수 있는 게놈 편집 접근

법은 인간 유전 질환의 이해를 진보시킬 잠재력을 갖고, 신규 치료제의 개발로 이어질 수 있다. 최근에, 뉴클레아제, 염기 편집제 및 프라임 편집제를 포함한, CRISPR-Cas 시스템에 기초한 여러 게놈 편집 접근법이 게놈 편집을 위해 개발되었다. CRISPR-Cas 뉴클레아제는, 하나의 가이드 RNA와 함께 사용되는 경우에, 확률적 indel로 이어지는 이중 가닥 DNA 파괴 (DSB)의 생성을 통해 유전자를 파괴하는 데 사용될 수 있다. 염기 편집제는 C·G에서 T·A 또는 A·T에서 G·C 전이 염기 쌍 편집 또는 C·G에서 G·C 염기 쌍 편집을 정확하게 설치하는 데 사용될 수 있다. 프라임 편집제는 12개의 염기 쌍 치환 편집뿐만 아니라 작은 내지 중간-크기의 삽입 및 결실 중 어느 하나를 정확하게 설치할 수 있다. 그러나, 이들 전략 단독은 보다 실질적인 게놈 편집, 예컨대 큰 결실, 삽입, 대체 또는 역위를 만드는 데 직접적으로 적용가능하지 않다.

[2279] 이전에, ~50 내지 >100,000개 염기 쌍 범위의 길이인 게놈 DNA 서열의 표적화된 결실을 위한 쌍형성된 Cas9 뉴클레아제 전략이 개발되었다. 더욱이, 선형 공여자 DNA 서열을 제공함으로써, 새로운 DNA 서열의 표적화된 삽입이 단부-연결 또는 상동성-지정 DNA 복구 과정을 통해 단일 절단 부위에서 또는 쌍형성된 절단 부위 사이에서 수행될 수 있다. 많은 방식으로 다재다능하지만, 단일-뉴클레아제 및 쌍형성된-뉴클레아제 접근법은 여러 결점을 동반한다. 먼저, 결실을 위한 쌍형성된 뉴클레아제의 사용은 indel을 갖는 목적하는 결실, 목적하는 결실이 없는 개별 DSB 부위에서의 indel, DSB 부위 사이의 DNA 서열의 바람직하지 않은 역위 및 DSB 부위에서의 외인성 DNA 서열의 의도하지 않은 통합을 포함한 다수의 온-타겟 부산물을 생성할 수 있다. 추가로, 결실의 정확한 위치는 PAM 이용가능성 및 뉴클레아제에 의해 생성된 상응하는 절단 부위에 의해 제한된다. 유사한 제한 및 부산물이 DNA 공여자 녹-인에 대해 존재하며, 이는 종종 상동성-비의존성 접근법이 사용되는 경우에 서열 배향의 제어 없이 발생한다. 마지막으로, 온-타겟 및 오프-타겟 부위에서의 다수의 DSB의 동시 생성은, 추가의 가이드 RNA에 의해 보다 가능해지며, 오프-타겟 게놈 변형, 염색체 재배열 또는 염색체파열로 이어질 수 있다. 따라서, 보다 큰 DNA 서열의 절제, 대체 또는 삽입을 위해 뉴클레아제-기반 편집 전략을 적용하는 경우에 생성물 제어 및 순도와 관련하여 상당한 도전과제가 존재한다.

[2280] 프라임 편집은 높은 목적하는 생성물 대 부산물 비로 인간 세포에서 최대 ~40 bp의 정확한 삽입 및 최대 ~80bp의 결실을 만들 수 있는 것으로 밝혀졌다. 원칙적으로, 프라임 편집은 DSB의 생성을 피함으로써 뉴클레아제와 연관된 많은 도전과제를 회피하는 데 사용될 수 있다. 그러나, 프라임 편집은 전형적인 유전자 코딩 서열의 크기의 보다 큰 삽입 및 결실을 매개하는 것으로 아직 밝혀지지 않았고, 간단한 프라임 편집 반응의 추정된 메카니즘은 긴 pegRNA 역전사 주형을 요구함으로써 이들 보다 큰 DNA 변화를 어렵게 만든다. 부위-특이적 DNA 레콤비나제 효소는 큰 DNA 서열의 절제, 역위, 통합 또는 교환을 포함한 여러 DNA 변환을 수행하는 능력을 갖는다. 그러나, 부위-특이적 레콤비나제를 재프로그램화하는 문제로 인해 현재 그것을 게놈 조작 및 편집 적용에 적용하는 것은 어렵다.

[2281] 인간 세포에서 높은 효율로 내인성 게놈 부위에서의 DNA 서열의 결실, 치환 또는 삽입을 가능하게 하는 트윈 프라임 편집 (트윈PE)의 개발이 본원에 제시된다. 이들 변화에 추가로, 트윈PE를 사용하여 인간 게놈 내 표적화된 부위에 높은 효율로 1개 이상의 레콤비나제 인식 부위를 통합시킬 수 있고, 이 부위는 후속적으로 유전자 카고의 표적화된 통합 또는 게놈 DNA 서열의 표적화된 역위를 가능하게 하는 부위-특이적 세린 레콤비나제 효소에 대한 기질로서 사용될 수 있다.

[2282] 결과

[2283] 프라임 편집은 촉매적으로 손상된 Cas9 닉카제 및 야생형 (PE1) 또는 조작된 (PE2) MMLV 리버스 트랜스크립타제 효소의 융합체를 포함하는 프라임 편집제 단백질, 및 표적 게놈 부위의 특정 및 목적하는 편집의 코딩 둘 다를 수행하는 프라임 편집 가이드 RNA (pegRNA)를 사용한다. 표적 부위 인식 시에, PE·pegRNA 복합체는 PAM-함유 DNA 가닥을 닉킹하고, 닉킹된 가닥을 프라이머로서 사용하여 pegRNA의 RT 주형을 게놈 DNA로 직접 역전사시킨다. 역전사 후에, 새로 합성된 3' 플랩은 그에 대해 대체로 상동인 인접한 DNA를 침입하여 중복 5' 플랩 서열을 대체한다. 이어서, 편집된 DNA 가닥을 주형으로서 사용하여 대향 비편집된 가닥을 복구시킨다. 따라서, 이러한 제안된 편집 경로는 세포 DNA 복구가 목적하는 편집을 거부하고 DNA 서열을 그의 원래 형태로 복귀시킬 수 있는 2가지 기회를 제공한다.

[2284] DNA 복구에서 잠재적으로 불리한 단계를 우회하는 것은 프라임 편집이 증가된 효율로 발생하게 하고 게놈 편집의 별개의 부류를 가능하게 할 수 있는 것으로 예측되었다. 추가로, 각각 하나 또는 다른 DNA 가닥을 표적화하고 다른 pegRNA에 의해 주형화된 서열에 상보적인 3' 플랩 DNA 서열의 합성의 주형이 되는 pegRNA의 쌍을 사용하는 트윈 프라임 편집 (트윈PE) 전략이 고려되었다 (도 108a). 새로 합성된 DNA 가닥이 내인성 표적 부위와 고도로 상이한 경우에, 3' 플랩이 우선적으로 혼성화하여 새로운 DNA 서열의 어닐링된 3' 오버행 및 원래 DNA

서열의 어닐링된 5' 오버행을 보유하는 중간체를 생성할 것으로 가정되었다 (도 108a). 두 편집된 가닥은 프라임 편집제 복합체에 의해 합성되기 때문에, 편집-함유 플랩 가닥에 의한 표적 부위의 침입 또는 상보적 DNA 가닥 내로의 편집의 카피가 요구되지 않는다. 원래 DNA 서열 (어닐링된 5' 오버행)의 절제, 폴리머라제에 의한 갭의 충전 및 닉 쌍의 라이게이션은 닉 부위 사이의 내인성 서열을 쌍형성된 3' 플랩 서열로 대체할 것이다 (도 108a). 주형 설계의 유연성으로 인해, 편집은 원칙적으로 새로운 DNA 서열을 도입하거나, DNA 서열의 부분을 대체하거나 또는 DNA 서열의 부분을 결실시킬 수 있다.

[2285] 트윈PE 전략을 평가하기 위해, HEK293T 내 HEK 부위 3 유전자좌 (이전에 HEK3으로 지칭됨)를 표적화하여 90 bp의 내인성 서열을 38-bp Bxb1 attB 부착 서열로 대체하였다 (도 108b). 각각의 프로토스페이서에 대해, 30, 34 또는 38 nt의 38-bp attB 서열을 함유하는 RT 주형으로 3개의 pegRNA를 설계하였다 (도 108c). 이들 pegRNA의 쌍별 조합은 22 내지 38 bp 범위의 중첩 상보성을 갖는 3' 플랩을 생성할 것으로 예측된다 (도 108c). 흥미롭게도, 두 pegRNA가 플라스미드 DNA 형질감염에서 PE2와 함께 세포에 전달되었을 때, attB 부위 삽입의 높은 효율이 관찰되었고, pegRNA의 일부 조합은 앰플리콘 서열분석에 기초하여 목적하는 생성물로의 >80% 전환을 생성하였다 (도 108c). 50-bp Bxb1 attP 부착 서열의 삽입을 위한 유사한 전략은 최대 58%에 도달하는 효율로 편집을 달성하였다 (도 108c). 특히, 부분적으로 중첩되는 상보적 플랩이 전장 attB 또는 attP 서열 혼입을 가능하게 하였기 때문에, 각각의 pegRNA가 전체 삽입 서열을 코딩할 필요는 없었다. 그러나, 더 큰 중첩을 갖는 3' 플랩은 HEK 부위 3에서 Bxb1 attB 및 attP 서열의 삽입에 대해 약간 더 높은 편집 효율 및 더 적은 indel로 이어졌다.

[2286] 다음으로, 트윈PE가 PE2 또는 PE3 시스템을 사용하여 입증된 것보다 더 큰 DNA 서열의 삽입을 지지할 수 있는지를 결정하기 위해 그를 시험하였다. 먼저, 12 내지 321 bp 범위의 FKBP12 코딩 서열 단편의 PE3-매개된 삽입을 검사하였다. PE3으로 HEK 부위 3 유전자좌를 표적화하였을 때, 보다 짧은 12-bp 및 36-bp 삽입에 대해서는 보통의 효율이 달성되었지만 (각각 31.8% 및 17.3%), 108-bp 삽입에 대해서는 목적하는 생성물이 거의 관찰되지 않았고 (0.8%), 321-bp 삽입에 대해서는 전장 생성물이 관찰되지 않았다 (도 109a). 대조적으로, 트윈PE는 108-bp 단편에 대해 16% 삽입 효율을 가능하게 하였으며, 이는 PE3에 비해 20배 개선에 이른다. 추가로, 쌍형성된 Bxb1 레콤비나제 부위를 함유하는 113-bp 및 103-bp 삽입이 CCR5 유전자좌에서 유사한 효율 (각각 10.7% 및 9.7%)로 달성되었다 (데이터는 제시되지 않음). 이들 결과는 PE3과 비교하여 트윈PE를 사용하는 경우 더 큰 서열 삽입에 대한 잠재력을 예시한다.

[2287] 트윈PE 전략의 하나의 잠재적 적용은 엑손 코딩 서열을, 단백질 서열을 유지하고 표적 트윈PE-유도된 닉 부위 사이의 임의의 돌연변이를 교정할 잠재력을 갖는 재코딩된 DNA 서열로 대체하는 것이다. 이러한 접근법을 시험하기 위해, 유전적 대사 장애 페닐케톤뇨 (PKU)를 유발하는 돌연변이인 PAH 유전자를 표적화하였다. 엑손 4 및 엑손 7의 부분의 재코딩을 야생형 HEK293T 세포의 PAH에서 시험하였으며, 여기서 교정 편집이 검출가능하다. 상이한 플랩 중첩 길이를 시험하고 pegRNA의 3' 단부에 evoPreQ1 모티프를 부가함으로써, 엑손 4에서 64-bp 재코딩에 대해 최대 9.4% 평균 효율, 엑손 7에서 46-bp 재코딩에 대해 최대 22.7% 평균 효율 및 엑손 7에서 64-bp 재코딩에 대해 최대 27.4% 평균 효율에 도달하는 보통의 효율로 목적하는 서열 재코딩이 달성되었다 (도 109b). 바이오PKU 데이터베이스에 주석달린 돌연변이에 기초하여, 이들 영역의 트윈PE 편집은 집합적으로 PKU에 연루된 별개의 돌연변이의 교정을 허용한다. 검사된 추가의 엑손이 또한 보다 낮은 효율일지라도 재코딩될 수 있다 (도 112). 이들 결과는 트윈PE가 원칙적으로 단일 쌍의 pegRNA로 다중 돌연변이를 교정하기 위해 사용될 수 있다는 것을 입증한다.

[2288] DNA 서열의 삽입 및 대체에 추가로, 트윈PE는 또한 정확한 결실 편집을 할 수 있어야 한다. 쌍형성된-뉴클레아제 결실 전략은 2개의 DSB 부위에 걸쳐 있는 결실을 생성하고, 따라서 PAM 이용가능성에 의해 제한되며, 목적하는 결실은 종종 바람직하지 않은 indel 부산물을 동반한다. 대조적으로, 트윈PE는 DSB 생성의 결여 및 트윈PE-유도된 닉 부위에서 추가의 DNA 서열을 기록하는 능력으로 인해 보다 큰 유연성 및 정밀도로 결실을 만드는 잠재력을 갖는다. 정확한 결실에 대해 트윈PE를 평가하기 위해, 쌍형성된 pegRNA를 사용하는 3가지 전략을 비교하였다: "단일-앵커" 트윈PE 전략, "하이브리드-앵커" 트윈PE 전략 및 최근에 보고된 "프라임델" 전략 (도 109c). 각각의 전략은 상보적 플랩에서 코딩되는 서열이 상이하여, 닉 부위에 대한 결실의 배치에 있어서 유연성을 허용한다.

[2289] 단일-앵커 전략을 사용하여, 13-nt 상보적 플랩을 사용하여 pegRNA-유도된 닉 부위 중 하나에 인접한 77 bp의 서열을 14.8% 효율로 결실시키고, 34-nt 상보적 플랩을 사용하여 56 bp의 서열을 18.8% 효율로 정확하게 절제하였다 (도 109d). 하이브리드-앵커 전략을 사용하여, pegRNA-유도된 닉 부위 사이에 64 bp를 11.7% 효율로 결실시켜 생성물이 각각의 닉의 3'에 13 bp의 서열을 보유하도록 하였다 (도 109d). 마지막으로, 프라임델 전략을

pegRNA-유도된 닉 부위 사이의 90 bp의 결실에 대해 시험하였으며, 이는 40.4% 효율로 발생하였다 (도 109d). 중요하게는, 프라임델 전략은 두 가닥 상의 PAM 서열을 파괴하며, 이는 아마도 효율을 증가시키면서 또한 indel을 감소시킨다. 편집 효율은 pegRNA의 3' 단부에 evoPreQ1 모티프를 부가함으로써 1.5배 내지 2.5배 개선될 수 있었지만, 일부 경우에는 indel의 증가가 동반되었다 (도 109d). 종합하면, 이들 데이터는 트윈PE가 완벽하게 위치한 뉴클레아제 절단 부위의 이용가능성에 의존하지 않는 높은 유연성으로 표적화된 결실을 수행하기 위한 전략을 제공한다는 것을 보여준다.

[2290] 마지막으로, 보다 큰 결실을 위해 트윈PE를 적용하기 위해 치료상 적절한 유전자좌인 DMD를 표적화하였다. 뒤시엔느 근육 이영양증을 담당하는 병원성 DMD 대립유전자는 통상적으로 엑손을 함유하는 영역에 큰 결실을 함유하여 프레임시프트된 mRNA 전사체를 생성한다. 결실된 엑손의 대체가 없는 전장 디스트로핀 단백질의 생산은 단백질 기능의 부분적인 구출로 이어질 수 있기 때문에, 리딩 프레임을 복구하는 제2 엑손의 파괴가 잠재적 치료 전략으로서 제안되었다. DMD에서 엑손 51을 절제하기 위해 이전에 보고된 Cas9 뉴클레아제 결실 전략과 함께 3가지 트윈PE 결실 전략을 검사하였다. 단일-앵커 트윈PE 결실 접근법을 사용하여, DMD에서 엑손 51을 함유하는 780-bp 서열의 결실에 대해 14.4% 내지 24% 효율이 관찰되었다 (도 109e). 쌍형성된 Cas9 뉴클레아제 전략은 트윈PE 전략 (평균 14% 내지 24%)과 비교하여 더 높은 결실 효율 (평균 49.8%)을 달성하였지만, 쌍형성된 Cas9 뉴클레아제-매개된 결실은 또한 트윈PE (1% 내지 12% 총 indel)와 비교하여 훨씬 더 높은 indel 수준 (33%의 목적하는 결실을 갖는 indel, 11%의 목적하는 결실을 갖지 않는 indel)을 동반하였다 (도 109e). DMD의 엑손 51은 또한 프라임델 전략으로 627-bp 결실 (28.8% 평균 효율) 또는 단일-앵커 트윈PE 전략으로 590-bp 결실 (26.4% 평균 효율)을 생성하는 대안적 스페이서 쌍을 사용하여, 또는 589-bp 서열을 38-bp Bxb1 attB 서열로 대체하는 트윈PE (최대 40.4% 평균 효율)를 사용하여 절제될 수 있다 (도 109e). 이들 실험은 트윈PE 및 프라임델이, 비록 더 낮은 평균 효율일지라도, 쌍형성된 Cas9 뉴클레아제 전략과 비교하여 우수한 생성물 순도로 인간 세포 내 치료상 적절한 유전자좌에서 큰 결실을 생성할 수 있다는 것을 보여준다.

[2291] 트윈PE 및 Bxb1 인테그라제를 사용한 세이프 하버 유전자좌에서의 표적화된 DNA 통합

[2292] 트윈PE가 PE3보다 더 큰 삽입 편집을 수행할 수 있지만, 서열 삽입 크기의 상한은 유전자-크기조정된 DNA 단편의 통합을 지지하기에 여전히 충분하지 않았다. 따라서, 트윈PE를 사용하여 높은 효율로 내인성 인간 게놈 부위 내로 Bxb1 부착 서열을 성공적으로 삽입한 후, DNA 카고의 부위-특이적 통합을 위해 트윈PE를 세린 레콤비나제와 조합하였다. DNA 카고 통합을 위한 위치를 확인하기 위해, HEK293T 세포 내 확립된 세이프 하버 유전자좌에서 Bxb1 attB 및 attP 부착 서열의 트윈PE-매개된 삽입을 수행하였다. AAVS1 유전자좌를 표적화하는 32개의 스페이서 쌍을 50-bp Bxb1 attP 서열의 삽입에 대해 스크리닝하였다 (데이터는 제시되지 않음). 평가된 스페이서 쌍 중에서, 32개의 스페이서 조합 중 18개에 대한 최적의 pegRNA는 앰플리콘 서열분석에 의해 >50%의 정확한 삽입 효율을 달성하였다 (도 110a). 추가적으로, CCR5 유전자좌를 표적화하는 19개의 스페이서 쌍을 38-bp Bxb1 attB 서열의 삽입에 대해 스크리닝하였으며 (데이터는 제시되지 않음), 이들 중 6개의 스페이서 쌍은 앰플리콘 서열분석에 의해 >50%의 목적하는 편집 효율을 달성하였다 (도 110b). 이들 결과는 트윈PE가 인간 세포 내 세이프 하버 유전자좌에 레콤비나제 부착 서열을 높은 효율로 삽입하는 데 사용될 수 있다는 것을 입증한다.

[2293] 다음으로, 트윈PE-혼입된 Bxb1 attB 및 attP 서열을 검사하여, 이들이 파트너 Bxb1 부착 부위를 보유하는 플라스미드 DNA 카고의 통합을 위한 표적 부위로서의 역할을 할 수 있는지를 결정하였다. 먼저, 트윈PE를 사용하여 CCR5 유전자좌에 동형접합 attB 부위 삽입을 보유하는 단일-세포 클론을 생성하였다 (데이터는 제시되지 않음). 이러한 클론 세포주를 Bxb1 레콤비나제를 발현하는 플라스미드 및 5.6-kB attP-함유 공여자 DNA 플라스미드로 형질감염시키면, ddPCR에 의해 측정 시 표적 부위에서 12-17% 녹-인이 생성되었으며 (도 113), 이는 이전에 보고된 Bxb1 통합 효율과 일치한다. 이들 결과에 의해 고무되어, 트윈PE-매개된 부착 부위 삽입 및 BxB1-매개된 DNA 공여자 통합을 조사하여 이들이 단일 형질감염 단계에서 달성될 수 있는지를 결정하였다. 흥미롭게도, HEK293T 세포를 PE2, 둘 다의 pegRNA, BxB1 및 공여자 DNA를 코딩하는 플라스미드로 형질감염시키면, ddPCR에 의해 측정 시 1.4 - 6.8% 녹-인 효율이 발생되었다 (도 110c). 예상 attL 및 attR 재조합 생성물을 함유하는 예상 접합부 서열을 앰플리콘 서열분석에 의해 71-95% 범위의 생성물 순도로 확인하였다 (도 114).

[2294] "원-포트" 녹-인 효율을 개선시키기 위한 노력으로, attB 또는 attP 부착 부위, 정규 (GT) Bxb1 코어 서열, 대안적 (GA) Bxb1 코어 서열 및 다양한 플랩 중첩 길이의 트윈PE-매개된 혼입을 시험하였다. 트윈PE에 의한 attB 부위의 혼입이 원-포트 녹-인 형질감염에 대한 attP 부위의 혼입보다 더 우수하게 수행된 것으로 밝혀졌다 (3.3% 대 0.5% 녹-인, 도 110d). 추가로, 정규 GT 코어 서열을 갖는 부착 부위의 삽입은 일반적으로 대안적 GA 서열을 갖는 부착 부위의 삽입보다 더 높은 녹-인 효율을 유도하는 것으로 관찰되었다 (attB 맥락에서 3.3% 대 0%, 도 110d). 그러나, 중요하게는, 정규 및 대안적 코어 서열은 상응하는 attP 부착 부위와의 직교 재조합을

허용한다. pegRNA-코딩된 레콤비나제 서열 사이의 중첩을 감소시키는 것은 또한, 아마도 pegRNA-발현 플라스미드 DNA 서열 내의 부착 서열의 감소된 정도로 인해, attB 삽입의 경우 녹-인 효율을 38개 염기 쌍 중첩에서의 3.3%로부터 20개 염기 쌍 중첩에서의 5.5%로 개선시켰다 (도 110d). 유사한 트윈PE 효율이 중첩 길이에 걸쳐 관찰되었지만, 공여자 DNA와 pegRNA 플라스미드 사이의 재조합은 감소되었다 (도 115a-115b).

[2295] 다음으로, 공여자 통합을 ALB 유전자좌에서 평가하였으며, 이는 간세포로부터의 치료 단백질 분비에 대해 수집될 수 있다. 알부민은 간에 의해 활발하게 분비되고, 인간 혈장 단백질의 대략 60%를 구성한다. 따라서, 비교적 작은 백분율의 간세포 내 ALB 유전자좌에서의 치료적 트랜스진 통합은 많은 질환에 대해 치료적 수준의 단백질 발현을 달성할 수 있다. ALB의 인트론 1을 표적화하는 전략을 고안하였으며, 여기서 Bxb1은 스플라이스 수용자 서열을 함유하는 원형 DNA에 이어서 관심 단백질에 대한 cDNA의 통합을 매개할 것이다. 통합은 원칙적으로 엑손 1에 코딩된 알부민 분비 신호의 치료적 트랜스진 코딩 서열로의 스플라이싱을 가능하게 할 것이다. 먼저, pegRNA 쌍을 ALB의 인트론 1 내의 Bxb1 attB 서열의 삽입에 대해 스크리닝하였다. pegRNA 쌍을 스크리닝한 후, 앰플리콘 서열분석에 의해 attB 서열의 43% 정확한 삽입을 달성한 스페이서 조합이 확인되었다 (도 110e). 이 유전자좌에서의 단일 형질감염 녹-인은 HEK293T에서 1.3% 효율로 달성되었다 (도 110f). Huh7 세포에서, attB 서열의 34% 정확한 삽입과 2.6%의 단일 형질감염 녹-인 효율이 관찰되었다. 흥미롭게도, HEK293T 세포에서 CCR5 녹-인이 ALB 녹-인보다 더 효율적이었음에도 불구하고, Huh7에서는 ALB에서의 녹-인이 CCR5에서의 녹-인보다 더 효율적이었다 (2.6% 대 1.1%).

[2296] 표 14. 본 실시예에서 사용된 DNA 공여자 서열

DNA 공여자	서열	SEQ ID NO:
attB_Puro_GA 공여자 DNA:	gatgccagctcattcctcccactcatgatctatagatcccccgggctgcaggaatttctacccactctgtcgatacccccaccgagaccc cattggggccaatacgcccgcgtttcttccttttccccacccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggcttgtcgacgacggcggactccgtcgtcaggatcatccgtgagcaagggcgaggga gctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcga gggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgacc accctgacctacggcgtgcagtgcttcagccgctacccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaag gctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccc tggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaaca gccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggca gcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagca cccagtccgccctgagcaaagacccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactct cggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgccaccatgggcga tcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccgccaagcttatg catgtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagttgggggggagggtcggcaattgaaccggtgcct agagaaggtggcgcggggtaaactgggaaagtgatgtcgtgtactggctccgccttttcccgagggtgggggagaaccgtatata agtgcagtagtcgccgtgaacgttcttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcccgcgggcct ggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagcttcgggttgga agtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgggcgctggggcc gccgcgtgcgaatctggtggcaccttcgcgcctgtctcgctgctttcgataagtctctagccatttaaaatttttgatgacctgctgcga cgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggtttttgggccgcgggcggcgacggg gcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggggtagtctcaag ctggccggcctgctctggtgcctggcctcgcgccgccgtgtatcgccccgccctgggcggcaaggctggcccggtcggcaccag ttgcgtgagcggaaagatggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgggagagcgggc gggtgagtcacccacacaaaggaaaagggcctttccgtcctcagccgtcgcttcatgtgactccacggagtaccgggcgccgtcc aggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggggagggttttatgcgatggagtttccccacactga gtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttggttcattctcaa gcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccaccatgaccgagtacaagcccacggtgcgcct cgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcg acccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtggg tcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcc cgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaagga gcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagt ggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctcccttctacgagcggctcggctt caccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccggatcgggagag ggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccaccagcgagctgattaaggagaac atgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagccctacgag ggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccctcgccttcgacatcctggctactagcttcctctacggc agcaagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggagagagtcac cacatacgaagacggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtcaagatcag aggggtgaacttcacatccaacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgtaccccgct gacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatcaagaccacatat agatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaaggaggccaaca acgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagcttaattaaga attctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtg aaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcag ggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcctcgtcgccgcgg tttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatgggatcggccattg aacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacgatcggctg ctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaact gcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcggg aagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggc tgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgta ctcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggct caaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgc ttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagc	4282

[2297]

	ttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgcccccgattcgcagcgcatcgccttctatcgccttcttgac gagttcttctgagcgggactctgggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaattcggtaccaataa aagagctttattttcatgatctgtgtgttggtttttggcccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaa aatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctc tcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggta tctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaac tatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgta ggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccag ttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttttgtttgcaagcagcagatt acgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg	
attB_Puro 공여자 DNA:	gatgccagctcattcctcccactcatgatctatagatccccgggctgcaggaattctacccactctgtcgatacccccaccgagaccc cattgggccaatacgcccgcgtttcttccttttccccacccacccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggcttgtcgacgacggcggtctccgtcgtcaggatcatccgtgagcaagggcgaggga gctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcga gggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgacc accctgacctacggcgtgcagtgcttcagccgctacccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaag gctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccc tggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaaca gccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggca gcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagca cccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactct cggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgccaccatgggcga tcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccgccaagcttatg catgtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagttgggggggagggtcggcaattgaaccggtgcct agagaaggtggcgcgggtaaactgggaaagtgatgtcgtgtactggctccgcctttttcccgagggtgggggagaaccgtatata agtgcagtagtcgccgtgaacgttcttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcccgcgggcct ggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagcttcgggttgga agtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgggcgctggggcc gccgcgtgcgaatctggtggcaccttcgcgcctgtctcgctgctttcgataagtctctagccatttaaaatttttgatgacctgctgcga cgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggttttgggccgcgggcggcgacgggg gcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggggtagtctcaag ctggccggcctgctctggtgcctggcctcgcgccgccgtgtatcgccccgccctgggcggcaaggctggcccggtcggcaccag ttgcgtgagcggaaagatggccgcttcccgtgccctgctgcagggagctcaaaatggaggacgcggcgctcgggagagcgggc gggtgagtcacccacacaaaggaaaagggcctttccgtcctcagccgtcgcttcatgtgactccacggagtaccgggcgccgtcc aggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggggagggttttatgcgatggagtttccccacactga gtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttggttcattctcaa gcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccaccatgaccgagtacaagcccacggtgcgcct cgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcg acccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtggg tcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcc cgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaagga gcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagt ggaggcggccgagcgcgccggggtgcccgccttcctggagacctccgcgccccgcaacctcccctttctacgagcggctcggctt caccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccggatcgggagag ggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccaccagcgagctgattaaggagaac atgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagccctacgag ggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccccttcgccttcgacatcctggctactagcttcctctacggc agcaagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggagagagtcac cacatacgaagacgggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtcaagatcag aggggtgaacttcacatccaacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgtaccccgct gacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatcaagaccacatat agatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaaggaggccaaca acgagaccctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagcttaattaaga attctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtg aaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtttcaggttcag ggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcctcgtcgccgcgg tttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatgggatcggccattg aacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacgatcggctg ctctgatgccgccgtgttccggctgtcagcgcagggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaact	4283

[2298]

	gcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcggg aagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggc tgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgta ctcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggct caaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgc ttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagc ttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgccccgattcgcagcgcatcgccttctatcgccttcttgac gagttcttctgagcgggactctggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaattcggtaccaataa aagagctttattttcatgatctgtgtgttggtttttggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaa aatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctc tcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggta tctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaac tatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgta ggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccag ttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagatt acgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg	
attP_Puro_GA 공여자 DNA:	gatgccagctcattcctcccactcatgatctatagatcccccgggctgcaggaattctacccactctgtcgatacccccaccgagaccc cattggggccaatacgcccgcgtttcttccttttccccaccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggtttgtctggtcaaccaccgcggactcagtggtgtacggtacaaaccccgtgagcaa gggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccgg cgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccac cctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgcca tgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagg gcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtac aactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcg aggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccac tacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgcc gggatcactctcggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgcca ccatgggcgatcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccg ccaagcttatgcatgtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagttgggggaggggtcggcaattg aaccggtgcctagagaaggtggcgcggggtaaactgggaaagtgatgtcgtgtactggctccgcctttttcccgagggtgggggga gaaccgtatataagtgcagtagtcgccgtgaacgttcttttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggtt cccgcgggcctggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgag cttcgggttggaagtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgg gcgctggggccgccgcgtgcgaatctggtggcaccttcgcgcctgtctcgctgctttcgataagtctctagccatttaaaatttttgatg acctgctgcgacgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggtttttgggccgcggg cggcgacgggggcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggg gtagtctcaagctggccggcctgctctggtgcctggcctcgcgccgccgtgtatcgccccgccctgggcggcaaggctggcccg gtcggcaccagttgcgtgagcggaaagatggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgg gagagcgggcgggtgagtcacccacacaaaggaaaagggcctttccgtcctcagccgtcgcttcatgtgactccacggagtaccg ggcgccgtccaggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggaggggttttatgcgatggagtttc cccacactgagtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttg gttcattctcaagcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccaccatgaccgagtacaagccca cggtgcgcctcgccacccgcgacgacgtccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgc cacaccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggc aaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgcc gagatcggccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccg gcccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgc tccccggagtggaggcggccgagcgcgccgggggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgag cggctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccgg atcgggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccacagcgagctga ttaaggagaacatgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaa gccctacgagggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccctcgccttcgacatcctggctactagct tcctctacggcagcaagaccttcatcaaccacaccccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatggg agagagtcaccacatacgaagacgggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacg tcaagatcagagggggtgaacttcacatccaacggccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgct gtaccccgctgacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatca agaccacatatagatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaag gaggccaacaacgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaa	4284

[2299]

	gcttaattaagaattctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaa atgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttat gtttcaggttcaggggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcc tcgtcgccgcggtttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatg ggatcggccattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaaca gacgatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgc cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtc actgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagt atccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcga gcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaact gttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtg gaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgat attgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgccccgattcgcagcgcatcgccttct atcgccttcttgacgagttcttctgagcgggactctggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaat tcggtaccaataaaagagctttattttcatgatctgtgttggttttggccgcgttgctggcgtttttccataggctccgccccctgac gagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagct ccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgccttttctcccttcgggaagcgtggcgctttctatagct cacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcg ccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagca gagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcg ctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgttt gcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg	
attP_Puro 공여자 DNA:	gatgccagctcattcctcccactcatgatctatagatcccccgggctgcaggaattctacccactctgtcgatacccccaccgagaccc cattggggccaatacgcccgcgtttcttccttttccccaccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtc ggggcggcaagcttacatcgagatcccggtttgtctggtcaaccaccgcggtctcagtggtgtacggtacaaaccccgtgagcaag ggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggc gagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccacc ctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccat gcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggc gacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgggcacaagctggagtacaa ctacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgag gacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccacta cctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgg gatcactctcggcatggacgagctgtacaagagcggcctgaggagcagagcccaggcgagcaacagcgccgtggacgccacc atgggcgatcgcccgggaattgactagtgcggccgcctaggatccatgccgatagcgttggttgagtggataaccgtattaccgcc aagcttatgcatgtcccgtcagtgggcagagcgcacatcgcccacagtcccccgagaagttgggggggagggtcggcaattgaa ccggtgcctagagaaggtggcgcggggtaaactgggaaagtgatgtcgtgtactggctccgccttttcccgagggtgggggaga accgtatataagtgcagtagtcgccgtgaacgttcttttcgcaacgggtttgccgccagaacacaggtaagtgccgtgtgtggttcc cgcgggcctggcctctttacgggttatggcccttgcgtgccttgaattacttccacctggctgcagtacgtgattcttgatcccgagctt cgggttggaagtgggtgggagagttcgaggccttgcgcttaaggagccccttcgcctcgtgcttgagttgaggcctggcctgggcg ctgggcccgccgcgtgcgaatctggtggcacccttcgcgcctgtctcgctgctttcgataagtctctagccatttaaaatttttgatgacc tgctgcgacgctttttttctggcaagatagtcttgtaaatgcgggccaagatctgcacactggtatttcggtttttgggccgcgggcg gcgacggggcccgtgcgtcccagcgcacatgttcggcgaggcggggcctgcgagcgcggccaccgagaatcggacgggggt agtctcaagctggccggcctgctctggtgcctggcctcgccgccgccgtgtatcgccccgccctgggcggcaaggctggcccggt cggcaccagttgcgtgagcggaaagatggccgcttcccggccctgctgcagggagctcaaaatggaggacgcggcgctcgggga gagcgggcgggtgagtcacccacacaaaggaaaagggccttttccgtcctcagccgtcgcttcatgtgactccacggagtaccggg cgccgtccaggcacctcgattagttctcgcgcttttggagtacgtcgtctttaggttgggggggagggggttttatgcgatggagtttccc cacactgagtgggtggagactgaagttaggccagcttggcacttgatgtaattctccttggaatttgccctttttgagtttggatcttggtt cattctcaagcctcagacagtggttcaaagttttttttcttccatttcaggtgtcgtgagctagcccaccatgaccgagtacaagcccacg gtgcgcctcgccacccgcgacgacgtcccccgggccgtacgcaccctcgccgccgcgttcgccgactaccccgccacgcgcca caccgtcgacccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaa ggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccga gatcggcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggc ccaaggagcccgcgtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctcc ccggagtggaggcggccgagcgcgccgggggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgagcgg ctcggcttcaccgtcaccgccgacgtcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgccggatc gggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggcccaccggtcgccaccagcgagctgatta aggagaacatgcacatgaagctgtacatggagggcaccgtggacaaccatcacttcaagtgcacatccgagggcgaaggcaagc cctacgagggcacccagaccatgagaatcaaggtggtcgagggcggccctctcccctccgccttcgacatcctggctactagcttc	4285

[2300]

	ctctacggcagcaagaccttcatcaaccacacccagggcatccccgacttcttcaagcagtccttccctgagggcttcacatgggag agagtcaccacatacgaagacgggggcgtgctgaccgctacccaggacaccagcctccaggacggctgcctcatctacaacgtc aagatcagagggtgaacttcacatccaacggcccctgtgatgcagaagaaaacactcggctgggaggccttcaccgagacgctgt accccgctgacggcggcctggaaggcagaaacgacatggccctgaagctcgtgggcgggagccatctgatcgcaaacatcaag accacatatagatccaagaaacccgctaagaacctcaagatgcctggcgtctactatgtggactacagactggaaagaatcaagga ggccaacaacgagacctacgtcgagcagcacgaggtggcagtggccagatactgcgacctccctagcaaactggggcacaagc ttaattaagaattctctagaggatccagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatg ctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaattgcattcattttatgtt tcaggttcaggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgatcctgcaagcctcg tcgccgcggtttattctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatggga tcggccattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagac gatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccct gaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcact gaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatc catcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagc gagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgtt cgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaa aatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattg ctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgccccgattcgcagcgcatcgccttctatc gccttcttgacgagttcttctgagcgggactctggggttcgaataaagaccgaccaagcgacgtctgagagctccctggcgaattcg gtaccaataaaagagctttatttcatgatctgtgtgttggttttggccgcgttgctggcgttttccataggctccgccccccctgacgag catcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccct cgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcac gctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgcctt atccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagag cgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctct gctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttttgtttgca agcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctagtgtgcg	

[2301]

[2302] 멀티플렉스 편집 및 레콤비나제-매개된 역위

[2303] 큰 구조적 변이체는 많은 인간 병원성 대립유전자, 예컨대 중증 A형 혈우병 사례의 ~50%를 유발하는 F8 유전자좌에서 큰 0.6-Mb 역위로 발견된다. 트윈PE를 사용한 레콤비나제 부착 부위 삽입의 높은 효율에 의해 영감을 받아, attB 및 attP Bxb1 부착 부위 둘 다의 삽입을 멀티플렉스화하는 것이 개재 DNA 서열의 단방향성 결실 또는 역위에 의해 보다 복잡한 유전자 변이체를 교정하는 데 사용될 수 있는 것으로 추론되었다. 치료상 적절한 유전자좌에 대한 이러한 접근법을 시험하기 위해, IDS와 그의 위유전자 IDS2 사이의 39-kb 역위를 시도하였다. 이들 부위 사이의 역위는 헌터 증후군 사례의 ~13%에서 관찰되었고, 병원성 대립유전자에서의 절결점의 특징화는 역위가 IDS 및 IDS2 둘 다에 존재하는 인트론 7의 재조합 핫스폿 내에서 종종 발생한다는 것을 밝혀내었다. 따라서, 이들 재조합 핫스폿에 플랭킹된 영역을 표적화하여 Bxb1 attP 및 attB 부착 서열을 삽입하였으며, 이는 반대 방향으로 배향되는 경우에, 환자 세포에서 병원성 대립유전자 서열의 교정을 위한 Bxb1에 의한 단방향성 역위를 위한 기질로서 사용될 수 있다 (도 111a). 12개의 pegRNA 스페이서 쌍을 스크리닝하고 pegRNA 서열 특징부를 최적화한 후, pegRNA 쌍은 각각 73.7% 및 69.8% 효율로 좌측 및 우측 표적화 부위에 attP 또는 attB 부착 부위를 삽입할 수 있는 것으로 밝혀졌다 (도 111b).

[2304] 다음으로, attP 및 attB 부착 부위 둘 다의 멀티플렉스화 트윈PE-매개된 삽입을 IDS 및 IDS2에서 39-kb 서열의 Bxb1 레콤비나제-매개된 역위로 조사하였다. 먼저, 트윈PE 성분에 이어서 Bxb1 레콤비나제로 순차적 DNA 형질감염을 조사하였다. IDS2의 인트론 3에 정방향-배향된 attP 서열 및 IDS의 인트론 7에 역방향-배향된 attB 서열을 설치하는 pegRNA의 제1 세트 (세트 1)를 시험하였다. 추가로, pegRNA의 제2 세트 (세트 2)를 사용하여 제1 형질감염에서 IDS2에 역방향-배향된 attP 서열 및 IDS에 정방향-배향된 attB 서열을 설치하였다. 각각 pegRNA 세트 1 및 세트 2를 사용한 멀티플렉스 편집에 대해 52.3% 또는 54.5% attP 서열 삽입 및 27.6% 또는 30.9% attB 서열 삽입이 관찰되었다 (도 111c). 후속적으로 편집된 세포를 Bxb1 레콤비나제로 형질감염시켰을 때, 모의 형질감염 대조군과 비교하여 attP 및 attB 서열을 함유하는 증폭된 대립유전자의 유의하게 감소된 백분율이 관찰되었으며 (p-값 < 0.05), 이는 일부 편집된 대립유전자가 Bxb1-매개된 재조합에 대한 기질로서 사용되고 있다는 것을 시사한다 (도 111c). 예상 역위 접합부의 증폭에 이은 앰플리콘 서열분석은 Bxb1-매개된 재조합의 attL 및 attR 서열 시그너쳐 둘 다를 보여주었고, 생성물 순도는 두 접합부에서 80% 이상이었다 (도 111d).

[2305] 원-포트 트윈PE 및 Bxb1-매개된 역위를 수행하고 pegRNA 플라스미드 DNA 사이의 원치않는 재조합을 피하기 위해, PE2 mRNA, 세트 1로부터의 합성 pegRNA 및 Bxb1 mRNA를 포함하는 모든-RNA 성분을 뉴클레오펙션시켰다. 앰플리콘 서열분석을 사용하여, attR 및 attL 서열을 함유하는 예상된 역전된 대립유전자 접합부를 포획하였다. 추가로, 역위 효율을 정량화하기 위해, 비-역전 및 역전된 대립유전자 둘 다에서 동일한 서열에 결합할 수 있고 따라서 동일한 프라이머 쌍을 사용하여 두 편집을 증폭시킬 수 있는 역방향 프라이머를 설계하였다 (도 116). "순차적" 및 "원-포트" 전략에 대한 ~4.6-7.2% 및 ~1.5-2.3% 역위 효율이 각각 관찰되었다 (도 111e). 종합하

면, 이들 데이터는 트윈PE를 부위-특이적 세린 레콤비나제와 조합하는 것이 헌터 증후군 대립유전자에서 발견되는 통상의 39-kb 역위를 교정하는 데 사용될 수 있고, 다른 복잡한 병원성 변이체를 표적화하기 위한 유망한 치료적 플랫폼으로서의 역할을 할 수 있다는 것을 지지한다.

[2306] 논의

[2307] 본원에 제시된 바와 같이, 인간 세포의 게놈 내 표적화된 위치에서 DNA 서열을 결실, 대체 또는 삽입하는 데 사용될 수 있는 트윈 프라임 편집 접근법이 개발되었다. 부위-특이적 세린 레콤비나제 효소와 조합되는 경우에, 트윈PE는 또한 DNA 카고의 통합 또는 DNA 서열의 역위를 지지할 수 있다.

[2308] 방법

[2309] 일반적 방법

[2310] 달리 나타내지 않는 한, 퓨전 U 그린 멀티플렉스(Phusion U Green Multiplex) PCR 마스터 믹스 (써모피셔 사이언티픽) 또는 Q5 핫 스타트 하이-피델리티(Q5 Hot Start High-Fidelity) 2x 마스터 믹스 (뉴 잉글랜드 바이오랩스)를 사용한 PCR에 의해 DNA 증폭을 수행하였다. DNA 올리고뉴클레오티드를 인티그레이티드 DNA 테크놀로지스로부터 입수하였다. 어닐링된 올리고뉴클레오티드를 BsmBI-소화된 수용자 벡터 내로 라이게이션하여 sgRNA를 발현하는 플라스미드를 구축하였다. 이전에 기재된 바와 같이 깁슨 조립 또는 골든 게이트 조립에 의해 pegRNA를 발현하는 플라스미드를 구축하였다. 본 실시예에 사용된 sgRNA 및 pegRNA 구축물의 서열이 표 11-12에 열거된다.

[2311] 표 11. 본 실시예에 사용된 sgRNA의 서열

[2312]

sgRNA	**스페이서 서열**	**SEQ ID NO:**
HEK3_3b_+90_닉킹	GTCAACCAGTATCCCGGTGC	616
pU6-Sp-sgRNA-DMDExo51-A1	GATTGGCTTTGATTTCCCTA	3971
pU6-Sp-sgRNA-DMDExo51-A3	GTATATGATTGTTACTGAGA	3972
pU6-Sp-sgRNA-DMDExo51-B1	Gcagttgcctaagaactggt	3973

[2313] 표 12. 본 실시예에 사용된 pegRNA의 서열

pegRNA	스페이서 서열	스페이서 SEQ ID NO:	3' 연장부	연장부 SEQ ID NO:
HEK3_attB_A_38	ggcccagactgagcacgtga	383	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCGTGCTCAGTCTG	4029
HEK3_attB_A_34	ggcccagactgagcacgtga	383	TCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCGTGCTCAGTCTG	4030
HEK3_attB_A_30	ggcccagactgagcacgtga	383	GACGACGGAGACCGCCGTCGTCGACAAGCCCGTGCTCAGTCTG	4031
HEK3_attB_B_38	gtcaaccagtatcccggtgc	616	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGATACTGG	4032
HEK3_attB_B_34	gtcaaccagtatcccggtgc	616	TGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGATACTGG	4033
HEK3_attB_B_30	gtcaaccagtatcccggtgc	616	GACGACGGCGGTCTCCGTCGTCAGGATCATCCGGGATACTGG	4034
HEK3_attP_A_43	ggcccagactgagcacgtga	383	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGCTCAGTCTG	4035
HEK3_attP_A_39	ggcccagactgagcacgtga	383	GTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGCTCAGTCTG	4036
HEK3_attP_A_35	ggcccagactgagcacgtga	383	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGCTCAGTCTG	4037
HEK3_attP_B_44	gtcaaccagtatcccggtgc	616	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCGGGATACTGG	4038
HEK3_attP_B_40	gtcaaccagtatcccggtgc	616	GGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCGGGATACTGG	4039
HEK3_attP_B_36	gtcaaccagtatcccggtgc	616	AACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCGGGATACTGG	4040
PE3_HEK3_FKBP_ins12bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcacacctgcactcccgtgctcagtctg	4041
PE3_HEK3_FKBP_ins36bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcacccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4042
PE3_HEK3_FKBP_ins108bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcaaaatttctttccatcttcaagcatcccggtgtagtgcaccacgcaggtctggccgcgcttggggaaggtgcgcccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4043
PE3_HEK3_FKBP_ins321bp	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcattccagttttagaagctccacatcgaagacgagagtggcatgtggtgggatgatgcctgggtgcccagtggcaccataggcataatctggagatatagtcagtttggctctctgacccacactcatctgggcaacccctttcttcccagcctcggatcacctcctgcttgcctagcataaacttaaagggcttgtttctgtcccgggaggaatcaaatttctttccatcttcaagcatcccggtgtagtgcaccacgcaggtctggccgcgcttggggaaggtgcgcccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4044
트윈PE_HEK3_FKBP_ins108bp_32bp_오버행_A	ggcccagactgagcacgtga	383	accacgcaggtctggccgcgcttggggaaggtgcgcccgtctcctggggagatggtttccacctgcactcccgtgctcagtctg	4045
트윈PE_HEK3_FKBP_ins108bp_32bp_오버행_B	gtcaaccagtatcccggtgc	616	accttccccaagcgcggccagacctgcgtggtgcactacaccgggatgcttgaagatggaaagaaatttccgggatactgg	4046
pU6-Sp-gRNA-HEK3_del_DF_A_13nt TEMP_베이직	ggcccagactgagcacgtga	383	tcctctgccatcacgtgctcagtctg	4047

[2314]

[2315]

pU6-Sp-gRNA-HEK3_del_DF_B_13nt TEMP_베이직	gtcaaccagtatcccggtgc	616	tgatggcagaggaccgggatactgg	4048
pU6-Sp-gRNA-HEK3_del_DF_A_34nt TEMP_베이직	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcacgtgctcagtctg	4049
pU6-Sp-gRNA-HEK3_del_DF_B_34nt TEMP_베이직	gtcaaccagtatcccggtgc	616	tgatggcagaggaaaggaagccctgcttcctccaccgggatactgg	4050
pU6-Sp-gRNA-HEK3_del_DF_A_13-13TEMP_하이브리드	ggcccagactgagcacgtga	383	tgcaggagctgcatcctctgccatcacgtgctcagtctg	4051
pU6-Sp-gRNA-HEK3_del_DF_B_13-13TEMP_하이브리드	gtcaaccagtatcccggtgc	616	tgatggcagaggatgcagctcctgcaccgggatactgg	4052
pU6-Sp-gRNA-HEK3_del_DF_A_31nt TEMP_센듀어	ggcccagactgagcacgtga	383	gcccagccaaacttgtcaaccagtatcccggcgtgctcagtctg	4053
pU6-Sp-gRNA-HEK3_del_DF_B_31nt TEMP_센듀어	gtcaaccagtatcccggtgc	616	gggtcaatccttggggcccagactgagcacgccgggatactgg	4054
pU6-Sp-gRNA-HEK3_del_DF_A_13nt TEMP_베이직_EvoPreQ1	ggcccagactgagcacgtga	383	tcctctgccatcacgtgctcagtctgTTAAATAACGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4055
pU6-Sp-gRNA-HEK3_del_DF_B_13nt TEMP_베이직_EvoPreQ1	gtcaaccagtatcccggtgc	616	tgatggcagaggaccgggatactggAAAATATACGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4056
pU6-Sp-gRNA-HEK3_del_DF_A_34nt TEMP_베이직_EvoPreQ1	ggcccagactgagcacgtga	383	tggaggaagcagggcttcctttcctctgccatcacgtgctcagtctgTTAATAATCGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4057
pU6-Sp-gRNA-HEK3_del_DF_B_34nt TEMP_베이직_EvoPreQ1	gtcaaccagtatcccggtgc	616	tgatggcagaggaaaggaagccctgcttcctccaccgggatactggAAAAAAAAACGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4058
pU6-Sp-gRNA-HEK3_del_DF_A_13-13TEMP_하이브리드_EvoPreQ1	ggcccagactgagcacgtga	383	tgcaggagctgcatcctctgccatcacgtgctcagtctgATAAATAACGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4059
pU6-Sp-gRNA-HEK3_del_DF_B_13-13TEMP_하이브리드_EvoPreQ1	gtcaaccagtatcccggtgc	616	tgatggcagaggatgcagctcctgcaccgggatactggAAAAAAGGCGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4060
pU6-Sp-gRNA-HEK3_del_DF_A_31nt TEMP_센듀어_EvoPreQ1	ggcccagactgagcacgtga	383	gcccagccaaacttgtcaaccagtatcccggcgtgctcagtctgTTAAATACCGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4061
pU6-Sp-gRNA-HEK3_del_DF_B_31nt TEMP_센듀어_EvoPreQ1	gtcaaccagtatcccggtgc	616	gggtcaatccttggggcccagactgagcacgccgggatactggAAAATAATCGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA	4062
pU6-Sp-pegRNA-DMDExo51-A1_a_베이직	GATTGGCTTTGATTTCCCTA	3971	ttgaataggaagtaaattaatttgaagctggaccctagGGAAATCAA	4063

pU6-Sp-pegRNA-DMDExo51-A1_b_베이직	GATTGGCTTTGATTTCCCTA	3971	ttgaataggaagtaaattaatttgaagctggaccctagGGAAATCAAAG	4064
pU6-Sp-pegRNA-DMDExo51-A1_c_베이직	GATTGGCTTTGATTTCCCTA	3971	ttgaataggaagtaaattaatttgaagctggaccctagGGAAATCAAAGCC	4065
pU6-Sp-pegRNA-DMDExo51-B1_A1_b_베이직	Gcagttgcctaagaactggt	3973	tagggtccagcttcaaattaatttacttcctattcaaagttcttaggc	4066
pU6-Sp-pegRNA-DMDExo51-B1_A1_c_베이직	Gcagttgcctaagaactggt	3973	tagggtccagcttcaaattaatttacttcctattcaaagttcttaggcaa	4067
pU6-Sp-pegRNA-DMDExo51-A3_c_베이직	GTATATGATTGTTACTGAGA	3972	atgcctggacaagtaacttaagttaaataagccttctCAGTAACAATCAT	4068
pU6-Sp-pegRNA-DMDExo51-B1_A3_a_베이직	Gcagttgcctaagaactggt	3973	agaaggcttatttaacttaagttacttgtccaggcatagttctta g	4069
pu6-sp-pegrna-dmdexo51-a3-b1_a_프라임델	GTATATGATTGTTACTGAGA	3972	tgctgagagagaaacagttgcctaagaactCAGTAACAAT	4070
pu6-sp-pegrna-dmdexo51-b1-a3_a-프라임델	Gcagttgcctaagaactggt	3973	accacttccacaatgtatatgattgttactgagttcttagg	4071
pu6-sp-pegrna-dmdexo51-b1-a3_b-프라임델	Gcagttgcctaagaactggt	3973	accacttccacaatgtatatgattgttactgagttcttaggcaa	4072
pU6-Sp-pegRNA-DMDExo51-A3_b_attB_fwd	GTATATGATTGTTACTGAGA	3972	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCAGTAACAATC	4073
pU6-Sp-pegRNA-DMDExo51-A3_c_attB_fwd	GTATATGATTGTTACTGAGA	3972	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCAGTAACAATCATA	4074
pU6-Sp-pegRNA-DMDExo51-B1_b_attB_fwd	Gcagttgcctaagaactggt	3973	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATagttcttaggc	4075
pU6-Sp-pegRNA-DMDExo51-B1_c_attB_fwd	Gcagttgcctaagaactggt	3973	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATagttcttaggcaac	4076
pU6-Sp-pegRNA-DMDExo51-B2_b_attB_fwd	Gaggagagtaaagtgattgg	3974	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATatcactttactc	4077
pU6-Sp-pegRNA-DMDExo51-B2_c_attB_fwd	Gaggagagtaaagtgattgg	3974	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATatcactttactctcc	4078
A_1077a	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCT	4079
A_1077b	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCTGG	4080
A_1077c	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCTGGCT	4081
A_1098a	GGGAAGGGGCAGGAGAGCCA	3976	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCTCCTGC	4082

[2316]

A_1098b	GGGAAGGGGCAGGAGAGCCA	3976	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCTCCTGCCC	4083
A_1098c	GGGAAGGGGCAGGAGAGCCA	3976	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCTCCTGCCCCT	4084
A_1246a	GAATATGTCCCAGATAGCAC	3977	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTATCTGGG	4085
A_1246b	GAATATGTCCCAGATAGCAC	3977	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTATCTGGGAC	4086
A_1246c	GAATATGTCCCAGATAGCAC	3977	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTATCTGGGACAT	4087
A_1267a	GGGGACTCTTTAAGGAAAGA	3978	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTTCCTTAAA	4088
A_1267b	GGGGACTCTTTAAGGAAAGA	3978	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTTCCTTAAAGA	4089
A_1267c	GGGGACTCTTTAAGGAAAGA	3978	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTTCCTTAAAGAGT	4090
A_1293a	GAGAAAGAGAAAGGGAGTAG	3979	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCCCTTTC	4091
A_1293b	GAGAAAGAGAAAGGGAGTAG	3979	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCCCTTTCTC	4092
A_1293c	GAGAAAGAGAAAGGGAGTAG	3979	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCTCCCTTTCTCTT	4093
A_1307a	GAGTAGAGGCGGCCACGACC	3980	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGGCCGC	4094
A_1307b	GAGTAGAGGCGGCCACGACC	3980	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGGCCGCCT	4095
A_1307c	GAGTAGAGGCGGCCACGACC	3980	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGTGGCCGCCTCT	4096
A_1582a	GATCAGTGAAACGCACCAGA	3981	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGTGCGTTT	4097
A_1582b	GATCAGTGAAACGCACCAGA	3981	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGTGCGTTTCA	4098
A_1582c	GATCAGTGAAACGCACCAGA	3981	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGTGCGTTTCACT	4099
A_1615a	GCAGCTCAGGTTCTGGGAGA	3982	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCCAGAACC	4100
A_1615b	GCAGCTCAGGTTCTGGGAGA	3982	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCCAGAACCTG	4101

[2317]

[2318]

A_1615c	GCAGCTCAGGTTCTGGGAGA	3982	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCCAGAACCTGAG	4102
A_1647a	GTGGCCACTGAGAACCGGGC	3983	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGGTTCTCA	4103
A_1647b	GTGGCCACTGAGAACCGGGC	3983	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGGTTCTCAGT	4104
A_1647c	GTGGCCACTGAGAACCGGGC	3983	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCGGTTCTCAGTGG	4105
A_1810a	GAATCTGCCTAACAGGAGGT	3984	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTCCTGTTAG	4106
A_1810b	GAATCTGCCTAACAGGAGGT	3984	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTCCTGTTAGGC	4107
A_1810c	GAATCTGCCTAACAGGAGGT	3984	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTTCCTGTTAGGCAG	4108
A_1890a	GTCACCAATCCTGTCCCTAG	3985	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGACAGGA	4109
A_1890b	GTCACCAATCCTGTCCCTAG	3985	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGACAGGATT	4110
A_1890c	GTCACCAATCCTGTCCCTAG	3985	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGACAGGATTGG	4111
A_3786a	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGG	4112
A_3786b	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGGGG	4113
A_3786c	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGGGGGC	4114
A_3835a	GACGTCACGGCGCTGCCCCA	3987	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCAGCGCC	4115
A_3835b	GACGTCACGGCGCTGCCCCA	3987	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCAGCGCCGT	4116
A_3835c	GACGTCACGGCGCTGCCCCA	3987	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCAGCGCCGTGA	4117
A_3856a	GGTGTGCTGGGCAGGTCGCG	3988	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGACCTGCCC	4118
A_3856b	GGTGTGCTGGGCAGGTCGCG	3988	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGACCTGCCCAG	4119
A_3856c	GGTGTGCTGGGCAGGTCGCG	3988	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGACCTGCCCAGCA	4120

[2319]

B_1154a	GTCCTTGGCAAGCCCAGGAG	3989	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTG	4121
B_1154b	GTCCTTGGCAAGCCCAGGAG	3989	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTGCC	4122
B_1154c	GTCCTTGGCAAGCCCAGGAG	3989	cTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTGCCAA	4123
B_1314a	GTGCGTCCTAGGTGTTCACC	3990	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAACACCTA	4124
B_1314b	GTGCGTCCTAGGTGTTCACC	3990	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAACACCTAGG	4125
B_1314c	GTGCGTCCTAGGTGTTCACC	3990	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAACACCTAGGAC	4126
B_1376a	GTCCTGGCAGGGCTGTGGTG	3991	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACAGCCCT	4127
B_1376b	GTCCTGGCAGGGCTGTGGTG	3991	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACAGCCCTGC	4128
B_1376c	GTCCTGGCAGGGCTGTGGTG	3991	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACAGCCCTGCCA	4129
B_1640a	GTGACCTGCCCGGTTCTCAG	3992	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGAACCGGG	4130
B_1640b	GTGACCTGCCCGGTTCTCAG	3992	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGAACCGGGCA	4131
B_1640c	GTGACCTGCCCGGTTCTCAG	3992	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGAACCGGGCAGG	4132
B_1676a	GAGCTTGGCAGGGGGTGGGA	3993	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACCCCCTG	4133
B_1676b	GAGCTTGGCAGGGGGTGGGA	3993	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACCCCCTGCC	4134
B_1676c	GAGCTTGGCAGGGGGTGGGA	3993	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCACCCCCTGCCAA	4135
B_1701a	GAGCCAGAGAGGATCCTGGG	3994	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGGATCCTC	4136
B_1701b	GAGCCAGAGAGGATCCTGGG	3994	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGGATCCTCTC	4137
B_1701c	GAGCCAGAGAGGATCCTGGG	3994	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTAGGATCCTCTCTG	4138
B_1705a	GATGGAGCCAGAGAGGATCC	3995	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTCCTCTCTG	4139

B_1705b	GATGGAGCCAGAGAGGATCC	3995	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTCCTCTCTGGC	4140
B_1705c	GATGGAGCCAGAGAGGATCC	3995	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTCCTCTCTGGCTC	4141
B_1883a	GGGGCCACTAGGGACAGGAT	3996	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGTCCCTA	4142
B_1883b	GGGGCCACTAGGGACAGGAT	3996	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGTCCCTAGT	4143
B_1883c	GGGGCCACTAGGGACAGGAT	3996	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGTCCCTAGTGG	4144
B_1902a	GTCCCCTCCACCCCACAGTG	3997	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGTGGGGTG	4145
B_1902b	GTCCCCTCCACCCCACAGTG	3997	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGTGGGGTGGA	4146
B_1902c	GTCCCCTCCACCCCACAGTG	3997	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGTGGGGTGGAGG	4147
B_1962a	GGGACCACCTTATATTCCCA	3998	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAATATAAG	4148
B_1962b	GGGACCACCTTATATTCCCA	3998	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAATATAAGGT	4149
B_1962c	GGGACCACCTTATATTCCCA	3998	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGAATATAAGGTGG	4150
B_3839a	GACCTGCCCAGCACACCCTG	3999	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGGTGTGCTG	4151
B_3839b	GACCTGCCCAGCACACCCTG	3999	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGGTGTGCTGGG	4152
B_3839c	GACCTGCCCAGCACACCCTG	3999	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGGTGTGCTGGGCA	4153
B_3903a	GCGACTCCTGGAAGTGGCCA	4000	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCA	4154
B_3903b	GCGACTCCTGGAAGTGGCCA	4000	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCAGG	4155
B_3903c	GCGACTCCTGGAAGTGGCCA	4000	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCAGGAG	4156
B_3930a	GGACTTCCCAGTGTGCATCG	4001	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTG	4157
B_3930b	GGACTTCCCAGTGTGCATCG	4001	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTGGG	4158

[2320]

B_3930c	GGACTTCCCAGTGTGCATCG	4001	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTGGGAA	4159
A_1_223_a	TTTGCAGTTTATCAGGATGA	4002	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGCAGTTTATC	4160
A_1_223_b	TTTGCAGTTTATCAGGATGA	4002	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGCAGTTTATCAG	4161
A_1_223_c	TTTGCAGTTTATCAGGATGA	4002	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGCAGTTTATCAGG	4162
A_2_260_a	GGTTGAGCAGGTAGATGTCA	4003	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCTGAGCAGGT	4163
A_2_260_b	GGTTGAGCAGGTAGATGTCA	4003	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCTGAGCAGGTAG	4164
A_2_260_c	GGTTGAGCAGGTAGATGTCA	4003	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCTGAGCAGGTAGAT	4165
A_3_325_a	CTGGGCGGCAGCATAGTGAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGCAGC	4166
A_3_325_b	CTGGGCGGCAGCATAGTGAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGCAGCAT	4167
A_3_325_c	CTGGGCGGCAGCATAGTGAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGCAGCATAG	4168
A_4_360_a	GTCAAGAGTTGACACATTGT	4005	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGAGTTGAC	4169
A_4_360_b	GTCAAGAGTTGACACATTGT	4005	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGAGTTGACAC	4170
A_4_360_c	GTCAAGAGTTGACACATTGT	4005	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGAGTTGACACA	4171
A_5_507_a	CCCAAGTGATCACACTTGTC	4006	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGTGATCAC	4172
A_5_507_b	CCCAAGTGATCACACTTGTC	4006	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGTGATCACAC	4173
A_5_507_c	CCCAAGTGATCACACTTGTC	4006	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGTGATCACACTT	4174
A_6_510_a	CCACCCAAGTGATCACACTT	4007	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCCCCAAGTGA	4175
A_6_510_b	CCACCCAAGTGATCACACTT	4007	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCCCCAAGTGATC	4176
A_6_510_c	CCACCCAAGTGATCACACTT	4007	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCCCCAAGTGATCAC	4177

[2321]

A_7_532_a	GGGAGAGACGCAAACACAGC	4008	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAGAGACGCA	4178
A_7_532_b	GGGAGAGACGCAAACACAGC	4008	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAGAGACGCAAA	4179
A_7_532_c	GGGAGAGACGCAAACACAGC	4008	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAGAGACGCAAACA	4180
B_1_272_a	TCTCTGACCTGTTTTTCCTT	4009	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCTGACCTG	4181
B_1_272_b	TCTCTGACCTGTTTTTCCTT	4009	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCTGACCTGT	4182
B_1_272_c	TCTCTGACCTGTTTTTCCTT	4009	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCTGACCTGTTT	4183
B_2_291_a	TCTTACTGTCCCCTTCTGGG	4010	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATTACTGTCCC	4184
B_2_291_b	TCTTACTGTCCCCTTCTGGG	4010	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATTACTGTCCCCT	4185
B_2_291_c	TCTTACTGTCCCCTTCTGGG	4010	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATTACTGTCCCCTTC	4186
B_3_305_a	TCTGGGCTCACTATGCTGCC	4011	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGGGCTCACT	4187
B_3_305_b	TCTGGGCTCACTATGCTGCC	4011	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGGGCTCACTAT	4188
B_3_305_c	TCTGGGCTCACTATGCTGCC	4011	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGGGCTCACTATGC	4189
B_4_326_a	CCCAGTGGGACTTTGGAAAT	4012	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTGGGACT	4190
B_4_326_b	CCCAGTGGGACTTTGGAAAT	4012	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTGGGACTTTG	4191
B_4_326_c	CCCAGTGGGACTTTGGAAAT	4012	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTGGGACTTTGG	4192
B_5_330_a	GTGGGACTTTGGAAATACAA	4013	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGGACTTTGG	4193
B_5_330_b	GTGGGACTTTGGAAATACAA	4013	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGGACTTTGGAA	4194
B_5_330_c	GTGGGACTTTGGAAATACAA	4013	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGGACTTTGGAAAT	4195
B_6_414_a	CCTGACAATCGATAGGTACC	4014	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGACAATCGA	4196
B_6_414_b	CCTGACAATCGATAGGTACC	4014	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGACAATCGATA	4197
B_6_414_c	CCTGACAATCGATAGGTACC	4014	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGACAATCGATAGG	4198
B_7_538_a	GGAATCATCTTTACCAGATC	4015	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATATCATCTTTA	4199

[2322]

B_7_538_b	GGAATCATCTTTACCAGATC	4015	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATATCATCTTTACC	4200
B_7_538_c	GGAATCATCTTTACCAGATC	4015	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATATCATCTTTACCAG	4201
B_8_584_a	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATT	4202
B_8_584_b	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTC	4203
B_8_584_c	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTCCA	4204
B_9_601_a	TACAGTCAGTATCAATTCTG	4017	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTCAGTAT	4205
B_9_601_b	TACAGTCAGTATCAATTCTG	4017	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTCAGTATCA	4206
B_9_601_c	TACAGTCAGTATCAATTCTG	4017	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATAGTCAGTATCAAT	4207
A_3_325_a	CTGGGCGGCAGCATAGTGAG	4004	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGGCGGCAGC	4166
B_6_414_b	CCTGACAATCGATAGGTACC	4014	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGACAATCGATA	4197
A_5_507_c	CCCAAGTGATCACACTTGTC	4006	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAAGTGATCACACTT	4174
A_6_510_b	CCACCCAAGTGATCACACTT	4007	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCCCAAGTGATC	4176
A_7_532_b	GGGAGAGACGCAAACACAGC	4008	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCAGAGACGCAAA	4179
B_8_584_b	GCAGCTCTCATTTTCCATAC	4016	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGCTCTCATTTTC	4203
A_1077c	GCAGAGCCAGGAACCCCTGT	3975	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGGGTTCCTGGCT	4081
B_1154c	GTCCTTGGCAAGCCCAGGAG	3989	cTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCTGGGCTTGCCAA	4123
A_3786c	GCTGGCCCCCCACCGCCCCA	3986	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTGGCGGTGGGGGGC	4114
B_3903c	GCGACTCCTGGAAGTGGCCA	4000	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCACTTCCAGGAG	4156
B_3930c	GGACTTCCCAGTGTGCATCG	4001	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTTGCACACTGGGAA	4159

[2323]

A_7_532_b	GGGAGAGAC GCAAACACA GC	4008	ATGATCCTGACGACGGAGACCGCCGT CGTCGACAAGCCAGAGACGCAAA	4179
B_8_584_b	GCAGCTCTC ATTTTCCATA C	4016	GGCTTGTCGACGACGGCGGTCTCCGT CGTCAGGATCATGCTCTCATTTTC	4203
A7_attB_30	GGGAGAGAC GCAAACACA GC	4008	TCCTGACGACGGAGACCGCCGTCGTC GACAAGCCAGAGACGCAAA	4208
B8_attB_30	GCAGCTCTC ATTTTCCATA C	4016	TGTCGACGACGGCGGTCTCCGTCGTC AGGATCATGCTCTCATTTTC	4209
A7_attB_20	GGGAGAGAC GCAAACACA GC	4008	ACGACGGAGACCGCCGTCGTCGACA AGCCAGAGACGCAAA	4210
B8_attB_20	GCAGCTCTC ATTTTCCATA C	4016	ACGACGGCGGTCTCCGTCGTCAGGAT CATGCTCTCATTTTC	4211
A7_attB_GA_38	GGGAGAGAC GCAAACACA GC	4008	ATGATCCTGACGACGGAGTCCGCCGT CGTCGACAAGCCAGAGACGCAAA	4212
B8_attB_GA_38	GCAGCTCTC ATTTTCCATA C	4016	GGCTTGTCGACGACGGCGGACTCCGT CGTCAGGATCATGCTCTCATTTTC	4213
A7_attB_GA_30	GGGAGAGAC GCAAACACA GC	4008	TCCTGACGACGGAGTCCGCCGTCGTC GACAAGCCAGAGACGCAAA	4214
B8_attB_GA_30	GCAGCTCTC ATTTTCCATA C	4016	TGTCGACGACGGCGGACTCCGTCGTC AGGATCATGCTCTCATTTTC	4215
A7_attB_GA_20	GGGAGAGAC GCAAACACA GC	4008	ACGACGGAGTCCGCCGTCGTCGACAA GCCAGAGACGCAAA	4216
B8_attB_GA_20	GCAGCTCTC ATTTTCCATA C	4016	ACGACGGCGGACTCCGTCGTCAGGAT CATGCTCTCATTTTC	4217
A7_attP_50	GGGAGAGAC GCAAACACA GC	4008	GGGTTTGTACCGTACACCACTGAGAC CGCGGTGGTTGACCAGACAAACCAA GAGACGCAAA	4218
B8_attP_50	GCAGCTCTC ATTTTCCATA C	4016	TGGTTTGTCTGGTCAACCACCGCGGT CTCAGTGGTGTACGGTACAAACCCGC TCTCATTTTC	4219
A7_attP_40	GGGAGAGAC GCAAACACA GC	4008	TGTACCGTACACCACTGAGACCGCGG TGGTTGACCAGACAAACCAAGAGAC GCAAA	4220
B8_attP_40	GCAGCTCTC ATTTTCCATA C	4016	TGTCTGGTCAACCACCGCGGTCTCAG TGGTGTACGGTACAAACCCGCTCTCA TTTTC	4221
A7_attP_30	GGGAGAGAC GCAAACACA GC	4008	CGTACACCACTGAGACCGCGGTGGTT GACCAGACAAACCAAGAGACGCAAA	4222
B8_attP_30	GCAGCTCTC ATTTTCCATA C	4016	GGTCAACCACCGCGGTCTCAGTGGTG TACGGTACAAACCCGCTCTCATTTTC	4223
A7_attP_GA_50	GGGAGAGAC GCAAACACA GC	4008	GGGTTTGTACCGTACACCACTGAGTC CGCGGTGGTTGACCAGACAAACCAA GAGACGCAAA	4224

[2324]

B8_attP_GA_50	GCAGCTCTC ATTTTCCATA C	4016	TGGTTTGTCTGGTCAACCACCGCGGA CTCAGTGGTGTACGGTACAAACCCGC TCTCATTTTC	4225
A7_attP_GA_40	GGGAGAGAC GCAAACACA GC	4008	TGTACCGTACACCACTGAGTCCGCGG TGGTTGACCAGACAAACCAAGAGAC GCAAA	4226
B8_attP_GA_40	GCAGCTCTC ATTTTCCATA C	4016	TGTCTGGTCAACCACCGCGGACTCAG TGGTGTACGGTACAAACCCGCTCTCA TTTTC	4227
A7_attP_GA_30	GGGAGAGAC GCAAACACA GC	4008	CGTACACCACTGAGTCCGCGGTGGTT GACCAGACAAACCAAGAGACGCAAA	4228
B8_attP_GA_30	GCAGCTCTC ATTTTCCATA C	4016	GGTCAACCACCGCGGACTCAGTGGTG TACGGTACAAACCCGCTCTCATTTTC	4229
A277_b	GACTGAAAC TTCACAGAA TA	4018	AACTTCACAGAGGCTTGTCGACGACG GCGGTCTCCGTCGTCAGGATCAT	4230
A277_c	GACTGAAAC TTCACAGAA TA	4018	GAAACTTCACAGAGGCTTGTCGACGA CGGCGGTCTCCGTCGTCAGGATCAT	4231
B358_b	GATTTATGA GATCAACAG CAC	4019	GGCTTGTCGACGACGGCGGTCTCCGT CGTCAGGATCATCTGTTGATCTC	4232
B358_c	GATTTATGA GATCAACAG CAC	4019	GGCTTGTCGACGACGGCGGTCTCCGT CGTCAGGATCATCTGTTGATCTCAT	4233
A7_attB_20	GGGAGAGAC GCAAACACA GC	4008	ACGACGGAGACCGCCGTCGTCGACA AGCCAGAGACGCAAA	4210
B8_attB_20	GCAGCTCTC ATTTTCCATA C	4016	ACGACGGCGGTCTCCGTCGTCAGGAT CATGCTCTCATTTTC	4211
A277_c	GACTGAAAC TTCACAGAA TA	4018	GAAACTTCACAGAGGCTTGTCGACGA CGGCGGTCTCCGTCGTCAGGATCAT	4231
B358_c	GATTTATGA GATCAACAG CAC	4019	GGCTTGTCGACGACGGCGGTCTCCGT CGTCAGGATCATCTGTTGATCTCAT	4233
pU6-Sp-gRNA-IDS_DF_A1_a_attP_fwd.dna	GACACCAAA AACTGCCAC AGG	4020	TACCGTACACCACTGAGACCGCGGTG GTTGACCAGACAAACCTGTGGCAGTT A	4234
pU6-Sp-gRNA-IDS_DF_A1_a_attP_rev.dna	GACACCAAA AACTGCCAC AGG	4020	GTCTGGTCAACCACCGCGGTCTCAGT GGTGTACGGTACAAACCTGTGGCAGT TA	4235
pU6-Sp-gRNA-IDS_DF_A1_c_attP_fwd.dna	GACACCAAA AACTGCCAC AGG	4020	TACCGTACACCACTGAGACCGCGGTG GTTGACCAGACAAACCTGTGGCAGTT TTA	4236
pU6-Sp-gRNA-IDS_DF_A1_c_attP_rev.dna	GACACCAAA AACTGCCAC AGG	4020	GTCTGGTCAACCACCGCGGTCTCAGT GGTGTACGGTACAAACCTGTGGCAGT TTTA	4237
pU6-Sp-gRNA-IDS_DF_A4_b_attP_fwd.dna	GCACTCATTT CCTCCAAGCT C	4021	TACCGTACACCACTGAGACCGCGGTG GTTGACCAGACAAACCTCTTGGAGGA AA	4238
pU6-Sp-gRNA-IDS_DF_A4_b_attP_rev.dna	GCACTCATTT CCTCCAAGCT C	4021	GTCTGGTCAACCACCGCGGTCTCAGT GGTGTACGGTACAAACCTCTTGGAGG AAA	4239

[2325]

pU6-Sp-gRNA-IDS_DF_B2_a_attP_fwd.dna	GTAGGTACAGGACAGGGCAG	4022	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCCTGTCC	4240
pU6-Sp-gRNA-IDS_DF_B2_a_attP_rev.dna	GTAGGTACAGGACAGGGCAG	4022	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCCTGTCC	4241
pU6-Sp-gRNA-IDS_DF_B3_a_attP_fwd.dna	GAGATAGGTAGGTACAGGACA	4023	TCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCTGTACCT	4242
pU6-Sp-gRNA-IDS_DF_B3_a_attP_rev.dna	GAGATAGGTAGGTACAGGACA	4023	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCTGTACCT	4243
pU6-Sp-gRNA-IDS_DF_B4_b_attP_fwd.dna	GTGAAAAGATAGGTAGGTAC	4024	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTCCTACCTATCTA	4244
pU6-Sp-gRNA-IDS_DF_B4_b_attP_rev.dna	GTGAAAAGATAGGTAGGTAC	4024	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTCCTACCTATCTA	4245
pU6-Sp-gRNA-IDS_DF_B7_b_attP_fwd.dna	GTTATGGTTTACTCCATCTA	4025	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTATGGAGTAAACCT	4246
pU6-Sp-gRNA-IDS_DF_B7_b_attP_rev.dna	GTTATGGTTTACTCCATCTA	4025	TACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCTATGGAGTAAACC	4247
pU6-Sp-gRNA-IDS_DF_C2_c_attB_fwd.dna	GTTTTGGTTTACCCTATCTA	4026	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCATAGGGTAAACCA	4248
pU6-Sp-gRNA-IDS_DF_C2_c_attB_rev.dna	GTTTTGGTTTACCCTATCTA	4026	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATATAGGGTAAACCA	4249
pU6-Sp-gRNA-IDS_DF_D1_b_attB_fwd.dna	GCTGTGGAACTGCAACACACT	4027	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATGTGTTGCAGT	4250
pU6-Sp-gRNA-IDS_DF_D1_b_attB_rev.dna	GCTGTGGAACTGCAACACACT	4027	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCGTGTTGCAGT	4251
pU6-Sp-gRNA-IDS_DF_D2_c_attB_fwd.dna	GTGCCACCTAACAGTGAGCTG	4028	GGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCATCTCACTGTTAGGT	4252
pU6-Sp-gRNA-IDS_DF_D2_c_attB_rev.dna	GTGCCACCTAACAGTGAGCTG	4028	ATGATCCTGACGACGGAGACCGCCGTCGTCGACAAGCCCTCACTGTTAGGT	4253

[2326]

[2327] 포유동물 세포 실험을 위한 모든 벡터는 플라스미드 플러스 미디프렙(Plasmid Plus Midiprep) 키트 (퀴아젠) 또는 퓨어일드(PureYield) 플라스미드 미니프렙 키트 (프로메가) 또는 퀴아프렙 스핀 미니프렙(QIAprep Spin Miniprep) 키트를 사용하여 정제하였다.

[2328] 일반적 포유동물 세포 배양 조건

[2329] HEK293T (ATCC CRL-3216), U2OS (ATTC HTB-96), K562 (CCL-243) 및 HeLa (CCL-2) 세포를 ATCC로부터 구입하고, 각각 10% (v/v) 태아 소 혈청 (깁코(Gibco), 검증됨) 및 1x 페니실린 스트렙토마이신 (코닝(Corning))으로 보충된 둘베코 변형 이글 배지 (DMEM) 플러스 글루타맥스 (써모피셔 사이언티픽), 맥코이 5A 배지 (깁코), RPMI 배지 1640 플러스 글루타맥스 (깁코) 또는 이글 최소 필수 배지 (EMEM, ATCC)에서 각각 배양 및 계대배양하였다. 모든 세포 유형을 인큐베이션하고, 유지시키고, 5% CO_2 하에 37℃에서 배양하였다. 세포주는 그의 각각의 공급업체에 의해 인증되었고, 미코플라스마에 대해 음성으로 시험되었다.

[2330] HEK293T, HeLa 및 Huh7 조직 배양 형질감염 프로토콜 및 게놈 DNA 제조

[2331] 성장시킨 HEK293T 세포를 48-웰 폴리-D-리신 코팅된 플레이트 (코닝) 상에 시딩하였다. 시딩 16-24시간 후, 세포를 제조업체의 프로토콜에 따라 1 μL의 리포펙타민 2000 (써모피셔 사이언티픽) 및 750 ng의 PE2 플라스미드, 125 ng의 pegRNA 1 및 125 ng의 pegRNA 2 (트윈PE 형질감염의 경우); 750 ng의 PE2 플라스미드, 250 ng의 pegRNA 플라스미드 및 83 ng의 sgRNA 플라스미드 (PE3 형질감염의 경우); 또는 750 ng의 Cas9 및 125 ng의 sgRNA 1 및 125 ng의 sgRNA 2 (쌍형성된 Cas9 뉴클레아제 형질감염의 경우)로 대략 60% 전면생장률로 형질감염시켰다. 달리 언급되지 않는 한, 세포를 형질감염 후 3일 배양하였고, 그 후에 배지를 제거하고, 세포를 1x PBS 용액 (써모피셔 사이언티픽)으로 세척하고, 150 μL의 새로 제조된 용해 완충제 (10 mM 트리스-HCl, pH 7.5; 0.05% SDS; 25 μg/mL 프로테이나제 K (써모피셔 사이언티픽))를 조직 배양 플레이트의 각각의 웰에 직접 첨가함으로써 게놈 DNA를 추출하였다. 게놈 DNA 혼합물을 37℃에서 1-2시간 동안 인큐베이션한 후, 80℃ 효소 불활성화 단계가 30분 동안 이어졌다. 포유동물 세포 게놈 DNA 증폭에 사용된 프라이머가 표 13에 열거된다.

[2332] 표 13. 본 실시예에 사용된 프라이머의 서열

프라이머	서열	SEQ ID NO:
HEK3_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG	3811
HEK3_rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC	3812
HEK3_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG	3811
HEK3_rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC	3812
DMD_UMI_fwd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNtgctggccagtttactaacaat	4254
DMD_UMI_fwd2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNcagaaagaagatcttatcccatcttg	4255
DMD_rev0	TGGAGTTCAGACGTGTGCTCTTCCGATCTggctacttttgttatttgcatt	4256
AAVS1_AVA1713	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGAGCAGGGTCCCGCTTC	4257
AAVS1_AVA1717	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACGGGGCTCAGTCTGAAGAG	4258
AAVS1_AVA1651	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCAAGGACTCAAACCCAGA	4259
AAVS1_AVA1652	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCGTGCGTCAGTTTTACCT	4260
AAVS1_AVA1653	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACTGCTTCTCCTCTTGGGAA	4261
AAVS1_AVA1715	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTGCCAGAACCTCTAAGGT	4262
AAVS1_AVA1655	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATCCTCTCTGGCTCCATCGTA	4263
AAVS1_AVA1656	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCACTTCAGGACAGCATGTTT	4264
AAVS1_AVA1707	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCCGGGAACTGCCGCTGGC	4265
AAVS1_AVA1710	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAGGAGGCCCTCATCTGGCG	4266
CCR5_AVA1678	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAATCAATGTGAAGCAAATCGCAGC	4267
CCR5_AVA1679	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCGATTGTCAGGAGGATGATGAA	4268
CCR5_AVA1680	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTTCTTACTGTCCCCTTCTGGGC	4269
CCR5_AVA1681	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAAACACAGCATGGACGAC	4270
CCR5_AVA1682	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACAATCGATAGGTACCTGGCTGTC	4271
CCR5_AVA1683	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCAGCCCCAAGATGACTAT	4272
IDS_AVA1763	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGAAAACCTGAGCTTGGAGG	4273
IDS_AVA1764	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTACTCCAGCTTAATGGAAGTGG	4274
IDS_AVA1765	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGAAGATGTGGAAATGCCTCAC	4275
IDS_AVA1766	TGGAGTTCAGACGTGTGCTCTTCCGATCTAATCAACATGAAGGGTTGTGTTGT	4276

[2333]

IDS_AVA1769	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTCCCACACATGCGTTCCTC	4277
IDS_AVA1770	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCATGAAGGGTTGTTTTTAATTGA	4278
IDS_UMI_junc1_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNGCCCTGTCCTGTACCTACCT	4279
IDS_junc2_fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTCAGTTTGGGGTATGTGCCA	4280
IDS_범용_rev	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAAATTTACCCGTGGCAGC	4281

[2334]

[2335] HeLa 세포의 경우, 이들을 성장시키고, HEK293T 세포와 유사한 밀도로 시딩하고, 세포의 각각의 웰을 0.75 uL의 HeLa 몬스터 시약 (미루스(Mirus))과 함께 190ng PE2_P2A_블라스트 (블라스티시딘 저항성 유전자), 31.5ng의 pegRNA1 및 31.5 ng의 pegRNA 2 (트윈PE 형질감염의 경우)로 형질감염시켰다. 이어서, 세포를 형질감염 24시간 후 선택 동안 10 ug/mL로 처리하였다. 나머지 게놈 DNA 수집 단계는 HEK293T 세포와 동일하다. Huh7 세포를 24-웰 폴리-D-리신 코팅된 플레이트 (코닝) 상에 150k/웰로 시딩하였다. 시딩 16-24시간 후, 세포를 제조업체의 프로토콜에 따라 2 μL의 리포펙타민 2000 (써모피셔 사이언티픽) 및 최대 800 ng의 플라스미드 DNA를 사용

하여 높은 전면생장률로 형질감염시켰다. 게놈 DNA 수집을 상기 HEK293T에 대해서와 같이 진행하였다.

[2336] 게놈 DNA 샘플의 고처리량 DNA 서열분석

[2337] 관심 게놈 부위를 게놈 DNA 샘플로부터 증폭시키고, 하기 변형과 더불어 이전에 기재된 바와 같이 일루미나 MiSeq 상에서 서열분석하였다. 간략하게, 일루미나 정방향 및 역방향 어댑터를 함유하는 증폭 프라이머 (표 13)를 관심 게놈 영역을 증폭시키는 제1 라운드의 PCR (PCR 1)에 사용하였다. 25-μL PCR 1 반응을 0.5 μM의 각각의 정방향 및 역방향 프라이머, 1 μL의 게놈 DNA 추출물 및 12.5 μL의 퓨전 U 그린 멀티플렉스 PCR 마스터 믹스를 사용하여 수행하였다. PCR 반응을 하기와 같이 수행하였다: 98℃에서 2분, 이어서 [98℃에서 10초, 61℃에서 20초 및 72℃에서 30초]의 30 사이클, 이어서 최종 72℃에서 2분 연장. 고유 일루미나 바코딩 프라이머 쌍을 2차 PCR 반응 (PCR 2)에서 각각의 샘플에 첨가하였다. 구체적으로, 25 μL의 주어진 PCR 2 반응물은 0.5 μM의 각각의 고유한 정방향 및 역방향 일루미나 바코딩 프라이머 쌍, 1 μL의 비정제 PCR 1 반응 혼합물, 및 12.5 μL의 퓨전 U 그린 멀티플렉스 PCR 2x 마스터 믹스를 함유하였다. 바코딩 PCR 2 반응을 하기와 같이 수행하였다: 98℃에서 2분, 이어서 [98℃에서 10초, 61℃에서 20초 및 72℃에서 30초]의 12 사이클, 이어서 최종 72℃에서 2분 연장. PCR 생성물을 1.5% 아가로스 겔에서 전기영동에 의해 분석적으로 평가하였다. PCR 2 생성물 (공통 앰플리콘에 의해 풀링됨)을 퀴아퀵 겔 추출 키트 (퀴아젠)를 사용하여 1.5% 아가로스 겔에 의한 전기영동에 의해, 40 μL의 물로 용리시켜 정제하였다. DNA 농도를 형광측정 정량화 (큐비트(Qubit), 써모피셔 사이언티픽) 또는 qPCR (카파 라이브러리 정량화 키트-일루미나, 카파 바이오시스템즈)에 의해 측정하고, 일루미나 MiSeq 기기 상에서 제조업체의 프로토콜에 따라 서열분석하였다.

[2338] 서열분석 판독물을 MiSeq 리포터 (일루미나)를 사용하여 탈멀티플렉스화하였다. 참조 서열에 대한 앰플리콘 서열의 정렬을 CRISPResso2를 사용하여 수행하였다. 모든 프라임 편집 수율 정량화에 대해, 프라임 편집 효율을 다음과 같이 계산하였다: [indel을 함유하지 않는 목적하는 편집을 갖는 판독물의 #] ÷ [총 판독물의 #]의 %. 점 돌연변이 편집의 정량화에 대해, CRISPResso2를 표준 모드로 "폐기_indel_판독물"을 온으로 하여 실행하였다. 이어서, 점 돌연변이의 설치를 위한 프라임 편집을 다음과 같이 명백하게 계산하였다: [비-폐기된 판독물에서의 명시된 점 돌연변이의 빈도] x [비-폐기된 판독물의 #] ÷ [총 판독물]. 편집의 정량화를 위해, CRISPResso2를 HDR 모드로 예상된 대립유전자로서 목적하는 대립유전자 (e flag)를 사용하고 "폐기_indel_판독물"을 온으로 하여 실행하였다. 편집 수율을 다음과 같이 계산하였다: [비-폐기된 HDR 정렬된 판독물의 #] ÷ [총 판독물]. indel 수율을 다음과 같이 계산하였다: [indel-함유 폐기 판독물의 #] ÷ [총 판독물]. 고유 분자 식별자 (UMI)를 적용하여 결실 효율을 정량화하고, 이러한 3-단계 PCR 프로토콜에서 PCR 편향을 평가하였다. 간략하게, 선형 증폭의 제1 단계에서, 1 uL의 게놈 DNA 추출물을 25-uL 반응물에서 15-bp 또는 16-bp 고유 분자 식별자를 함유하는 0.1 uM의 단일 정방향 UMI 프라이머 및 12.5 uL 퓨전 U 그린 멀티플렉스 PCR 마스터 믹스에 의해 증폭시켰다. 이어서, PCR 생성물을 1.6X 암퓨어 비드 (베크만 쿨터)에 의해 정제하고, 20uL 퀴아젠 용리 완충제 중에서 용리시켰다. 이어서, 제2 단계에서, 1 또는 2 uL의 선형으로 증폭된 PCR 용리액을 25-uL 반응물에서 0.5 uM의 각각의 정방향 및 역방향 프라이머, 퓨전 U 그린 멀티플렉스 PCR 믹스로 30 사이클 동안 증폭시켰다. 이 경우에, 정방향 프라이머는 UMI 프라이머의 5' 및 UMI의 상류에 위치하는 P5 일루미나 어댑터 서열에 어닐링한다. PCR 생성물을 1X 암퓨어 비드에 의해 정제하고, 25 uL 용리 완충제 중에서 용리시켰다. 제3 단계에서, 1 uL 용리액을 고유 일루미나 바코딩을 부가하기 위해 상기 언급된 바와 같이 12 사이클 동안 증폭시켰다. DMD 유전자좌에서의 큰 결실을 평가하기 위해, 상단 밴드 (비편집된 큰 앰플리콘) 및 하단 밴드 (결실을 갖는 편집된 앰플리콘)를 1.5% 아가로스로부터 개별적으로 절제하고, 최종 라이브러리 (상단 밴드의 60 pM 대 하부 밴드의 20 pM)를 2개의 별개의 Miseq 키트에 로딩하여 짧은 앰플리콘 (~300bp)과 비교하여 큰 앰플리콘 (~1 kb)의 약한 클러스터링을 피하였다.

[2339] 미가공 서열분석 판독물을 AmpUMI를 사용하여 UMI 중복제거하였다. 쌍형성된-단부 판독물에 대해, SeqKit를 사용하여 R1을 R2의 역 상보체와 연결시켰다 (중첩 없이 병합함). 연결된 R1+R2를 R1의 5' 단부에서 UMI를 사용하여 UMI 중복제거하였다. UMI-중복제거된 R1 또는 연결된 R1+R2를 CRISPResso2를 사용하여 분석하였다. 연결된 R1+R2를 분석하기 위해, 적절한 연결된 참조 앰플리콘 서열을 제공하여 연결로 인한 서열분석 정렬 인공물을 최소화하였다.

[2340] U2OS 및 K562의 뉴클레오펙션

[2341] K562 및 U2OS 세포를 사용하여 모든 실험에서 뉴클레오펙션을 수행하였다. 뉴클레오펙션당 200,000개의 세포를 사용하였다. 카운팅된 세포를 회전 침강시키고, PBS로 세척한 다음, 론자 SE 세포주 4D-뉴클레오펙터 키트의 권장사항에 따라 뉴클레오펙션 용액 중에 재현탁시켰다. 장치에서 세포의 뉴클레오펙션 후에, 세포를 큐벳 내

에서 실온에서 10분 동안 인큐베이션하였다. 큐벳의 내용물을 미리 가온된 배지로 채워진 48 웰 플레이트로 옮겼다. 게놈 DNA를 추출하고, 상기 언급된 바와 같이 일루미나 Miseq 제조를 위해 준비하였다.

[2342] 원-포트 트윈PE 및 BxbI-매개된 카고 녹-인 및 큰 역위:

[2343] 원-포트 매개된 큰 역위 실험을 위해, 순차적 플라스미드 형질감염 및 원-포트 mRNA 뉴클레오펙션을 수행하였다. 순차적 플라스미드 형질감염 실험에서, HEK293T 세포를 상기 언급된 바와 같이 리포펙타민 2000을 사용하여 750 ng PE2, 62.5 ng의 각각의 pegRNA (총 4개)로 형질감염시켰다. 3일 후, 세포를 트립신처리하고, 24-웰 플레이트에 플레이팅하고, 약 7일 동안 계대배양하였다. 이어서, 20K 세포를 다시 시딩하고, 500 ng BxBI 플라스미드로 형질감염시켰다. 게놈 DNA를 추출하고, 상기 언급된 바와 같이 일루미나 Miseq 제조를 위해 준비하였다. 원-포트 mRNA 뉴클레오펙션 실험에서, 200,000개의 세포를 1ug PE2 mRNA, 30 pmol의 각각의 합성 pegRNA (총 4개의 pegRNA) 및 750ng BxBI mRNA로 뉴클레오펙션시켰다. 합성 pegRNA를 IDT로부터 주문하고, mRNA 전사체를 공개된 프로토콜에 따라 시험관내에서 전사시켰다. 론자 완충제와 세포의 20 uL 혼합물을 프로그램 CM-130으로 전기천공하고, 80 uL 가온된 배지로 5분 동안 회수하였다. 이어서, 큐벳으로부터의 25 uL 샘플을 48-웰 플레이트의 각각의 웰에 첨가하고, 37℃에서 72시간 동안 인큐베이션하였다.

[2344] 본 개시내용 전반에 걸쳐 언급된 다른 참고문헌

[2345] 하기 참고문헌은 각각 그 전문이 본원에 참조로 포함된다.

1. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).

3. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20–36 (2017).

4. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

5. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).

6. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

7. ClinVar, July 2019.

[2346]

8. Dunbar, C. E. *et al.* Gene therapy comes of age. *Science* **359**, eaan4672 (2018).

9. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).

10. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).

11. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).

12. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

13. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

14. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).

15. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* **5**, a012740 (2013).

16. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).

17. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).

18. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).

19. Ihry, R. J. *et al.* p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).

20. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

21. Srivastava, M. *et al.* An Inhibitor of Nonhomologous End-Joining Abrogates Double-Strand Break Repair and Impedes Cancer Progression. *Cell* **151**, 1474–1487 (2012).

[2347]

22. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).

23. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).

24. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).

25. Li, X. *et al.* Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).

26. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4199

27. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 1 (2018). doi:10.1038/s41576-018-0059-1.

28. Ostertag, E. M. & Kazazian Jr, H. H. Biology of Mammalian L1 Retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).

29. Zimmerly, S., Guo, H., Perlman, P. S. & Lambowltz, A. M. Group II intron mobility occurs by target DNA- primed reverse transcription. *Cell* **82**, 545–554 (1995).

30. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).

31. Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).

32. Jinek, M. *et al.* Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**, 1247997 (2014).

33. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* aad8282 (2016). doi:10.1126/science.aad8282

34. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

35. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-

[2348]

responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 15939 (2017).

36. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12**, 664–670 (2015).

37. Anders, C. & Jinek, M. Chapter One - *In vitro* Enzymology of Cas9. in *Methods in Enzymology* (eds. Doudna, J. A. & Sontheimer, E. J.) **546**, 1–20 (Academic Press, 2014).

38. Briner, A. E. *et al.* Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Mol. Cell* **56**, 333–339 (2014).

39. Nowak, C. M., Lawson, S., Zerez, M. & Bleris, L. Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Res.* **44**, 9555–9564 (2016).

40. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).

41. Mohr, S. *et al.* Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 (2013).

42. Stamos, J. L., Lentzsch, A. M. & Lambowitz, A. M. Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications. *Mol. Cell* **68**, 926- 939.e4 (2017).

43. Zhao, C. & Pyle, A. M. Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat. Struct. Mol. Biol.* **23**, 558–565 (2016).

44. Zhao, C., Liu, F. & Pyle, A. M. An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* **24**, 183–195 (2018).

45. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

46. Liu, Y., Kao, H.-I. & Bambara, R. A. Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem.* **73**, 589–615 (2004).

47. Krokan, H. E. & Bjørås, M. Base Excision Repair. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).

[2349] 48. Kelman, Z. PCNA: structure, functions and interactions. *Oncogene* **14**, 629–640 (1997).

49. Choe, K. N. & Moldovan, G.-L. Forging Ahead through Darkness: PCNA, Still the Principal Conductor at the Replication Fork. *Mol. Cell* **65**, 380–392 (2017).

50. Li, X., Li, J., Harrington, J., Lieber, M. R. & Burgers, P. M. Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *J. Biol. Chem.* **270**, 22109–22112 (1995).

51. Tom, S., Henricksen, L. A. & Bambara, R. A. Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. *J. Biol. Chem.* **275**, 10498–10505 (2000).

52. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).

53. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).

54. Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33**, 1159–1161 (2015).

55. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).

56. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).

57. Schek N, Cooke C, Alwine JC. Molecular and Cellular Biology. (1992).

58. Gil A, Proudfoot NJ. Cell. (1987).

59. Zhao, B.S., Roundtree, I.A., He, C. Nat Rev Mol Cell Biol. (2017).

60. Rubio, M.A.T., Hopper, A.K. Wiley Interdiscip Rev RNA (2011).

61. Shechner, D.M., Hacisuleyman E., Younger, S.T., Rinn, J.L. Nat Methods. (2015).

62. Paige, J.S., Wu, K.Y., Jaffrey, S.R. Science (2011).

63. Ray D., … Hughes TR. Nature (2013).

64. Chadalavada, D. M., Cerrone-Szakal, A. L., Bevilacqua, P.C. RNA (2007).

[2350]

65. Forster AC, Symons RH. Cell. (1987).

66. Weinberg Z, Kim PB, Chen TH, Li S, Harris KA, Lünse CE, Breaker RR. Nat. Chem. Biol. (2015).

67. Feldstein PA, Buzayan JM, Bruening G. Gene (1989).

68. Saville BJ, Collins RA. Cell. (1990).

69. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. Nature (2004).

70. Roth A, Weinberg Z, Chen AG, Kim PG, Ames TD, Breaker RR. Nat Chem Biol. (2013).

71. Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z. Nat Commun. (2012).

72. MacRae IJ, Doudna JA. Curr Opin Struct Biol. (2007).

73. Bernstein E, Caudy AA, Hammond SM, Hannon GJ Nature (2001).

74. Filippov V, Solovyev V, Filippova M, Gill SS. Gene (2000).

75. Cadwell RC and Joyce GF. PCR Methods Appl. (1992).

76. McInerney P, Adams P, and Hadi MZ. Mol Biol Int. (2014).

77. Esvelt KM, Carlson JC, and Liu DR. Nature. (2011).

78. Naorem SS, Hin J, Wang S, Lee WR, Heng X, Miller JF, Guo H. Proc Natl Acad Sci USA (2017).

79. Martinez MA, Vartanian JP, Wain-Hobson S. Proc Natl Acad Sci USA (1994).

80. Meyer AJ, Ellefson JW, Ellington AD. Curr Protoc Mol Biol. (2014).

81. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. Nature. (2009).

82. Nyerges Á et al. Proc Natl Acad Sci USA. (2016).

83. Mascola JR, Haynes BF. Immunol Rev. (2013).

84. X. Wen, K. Wen, D. Cao, G. Li, R. W. Jones, J. Li, S. Szu, Y. Hoshino, L. Yuan, Inclusion of a universal tetanus toxoid CD4(+) T cell epitope P2 significantly enhanced the [2351]

immunogenicity of recombinant rotavirus ΔVP8* subunit parenteral vaccines. Vaccine 32, 4420-4427 (2014).

85. G. Ada, D. Isaacs, Carbohydrate-protein conjugate vaccines. Clin Microbiol Infect 9, 79-85 (2003).

86. E. Malito, B. Bursulaya, C. Chen, P. L. Surdo, M. Picchianti, E. Balducci, M. Biancucci, A. Brock, F. Berti, M. J. Bottomley, M. Nissum, P. Costantino, R. Rappuoli, G. Spraggon, Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197. Proceedings of the National Academy of Sciences 109, 5229 (2012).

87. J. de Wit, M. E. Emmelot, M. C. M. Poelen, J. Lanfermeijer, W. G. H. Han, C. van Els, P. Kaaijk, The Human CD4(+) T Cell Response against Mumps Virus Targets a Broadly Recognized Nucleoprotein Epitope. J Virol 93, (2019).

88. M. May, C. A. Rieder, R. J. Rowe, Emergent lineages of mumps virus suggest the need for a polyvalent vaccine. Int J Infect Dis 66, 1-4 (2018).

89. M. Ramamurthy, P. Rajendiran, N. Saravanan, S. Sankar, S. Gopalan, B. Nandagopal, Identification of immunogenic B-cell epitope peptides of rubella virus E1 glycoprotein towards development of highly specific immunoassays and/or vaccine. Conference Abstract, (2019).

90. U. S. F. Tambunan, F. R. P. Sipahutar, A. A. Parikesit, D. Kerami, Vaccine Design for H5N1 Based on B- and T-cell Epitope Predictions. Bioinform Biol Insights 10, 27-35 (2016).

91. Asante, EA. et. al. "A naturally occurring variant of the human prion protein completely prevents prion disease". Nature. (2015).

92. Crabtree, G. R. & Schreiber, S. L. Three-part inventions: intracellular signaling and induced proximity. *Trends Biochem. Sci.* **21**, 418–22 (1996).

93. Liu, J. *et al.* Calcineurin Is a Common Target of A and FKBP-FK506 Complexes. *Cell* **66**, 807–815 (1991).

94. Keith, C. T. *et al.* A mammalian protein targeted by G1-arresting rapamycin–receptor complex. *Nature* **369**, 756–758 (2003).

95. Spencer, D. M., Wandless, T. J., Schreiber, S. L. S. & Crabtree, G. R. Controlling signal transduction with synthetic ligands. *Science* **262**, 1019–24 (1993).

96. Pruschy, M. N. *et al.* Mechanistic studies of a signaling pathway activated by the organic dimerizer FK1012. *Chem. Biol.* **1**, 163–172 (1994).

[2352]

97. Spencer, D. M. *et al.* Functional analysis of Fas signaling *in vivo* using synthetic inducers of dimerization. *Curr. Biol.* **6**, 839–847 (1996).
98. Belshaw, P. J., Spencer, D. M., Crabtree, G. R. & Schreiber, S. L. Controlling programmed cell death with a cyclophilin-cyclosporin-based chemical inducer of dimerization. *Chem. Biol.* **3**, 731–738 (1996).
99. Yang, J. X., Symes, K., Mercola, M. & Schreiber, S. L. Small-molecule control of insulin and PDGF receptor signaling and the role of membrane attachment. *Curr. Biol.* **8**, 11–18 (1998).
100. Belshaw, P. J., Ho, S. N., Crabtree, G. R. & Schreiber, S. L. Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins. *Proc. Natl. Acad. Sci.* **93**, 4604–4607 (2002).
101. Stockwell, B. R. & Schreiber, S. L. Probing the role of homomeric and heteromeric receptor interactions in TGF-β signaling using small molecule dimerizers. *Curr. Biol.* **8**, 761–773 (2004).
102. Spencer, D. M., Graef, I., Austin, D. J., Schreiber, S. L. & Crabtree, G. R. A general strategy for producing conditional alleles of Src-like tyrosine kinases. *Proc. Natl. Acad. Sci.* **92**, 9805–9809 (2006).
103. Holsinger, L. J., Spencer, D. M., Austin, D. J., Schreiber, S. L. & Crabtree, G. R. Signal transduction in T lymphocytes using a conditional allele of Sos. *Proc. Natl. Acad. Sci.* **92**, 9810–9814 (2006).
104. Myers, M. G. Insulin Signal Transduction and the IRS Proteins. *Annu. Rev. Pharmacol. Toxicol.* **36**, 615–658 (1996).
105. Watowich, S. S. The erythropoietin receptor: Molecular structure and hematopoietic signaling pathways. *J. Investig. Med.* **59**, 1067–1072 (2011).
106. Blau, C. A., Peterson, K. R., Drachman, J. G. & Spencer, D. M. A proliferation switch for genetically modified cells. *Proc. Natl. Acad. Sci.* **94**, 3076–3081 (2002).
107. Clackson, T. *et al.* Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci.* **95**, 10437–10442 (1998).
108. Diver, S. T. & Schreiber, S. L. Single-step synthesis of cell-permeable protein dimerizers that activate signal transduction and gene expression. *J. Am. Chem. Soc.* **119**, 5106–5109 (1997).
109. Guo, Z. F., Zhang, R. & Liang, F. Sen. Facile functionalization of FK506 for biological studies by the thiol-ene 'click' reaction. *RSC Adv.* **4**, 11400–11403 (2014).
110. Robinson, D. R., Wu, Y.-M. & Lin, S.-F. The protein tyrosine kinase family of the human genome. *Oncogene* **19**, 5548–5557 (2000).
111. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 44, D862–D868 (2016).

[2353]

112. Jinek, M. et al. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337, 816–821 (2012).
113. Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819–823 (2013).
114. Mali, P. et al. RNA-Guided Human Genome Engineering via Cas9. Science 339, 823– 826 (2013).
115. Yang, H. et al. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell 154, 1370–1379 (2013).
116. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 24, 1012–1019 (2014).
117. Orlando, S. J. et al. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic Acids Res. 38, e152–e152 (2010).
118. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197 (2015).
119. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology- independent targeted integration. Nature 540, 144–149 (2016).
120. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat. Biotechnol. 36, 765–771 (2018).
121. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. Nat. Med. 24, 927–930 (2018).
122. Ihry, R. J. et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. Nat. Med. 24, 939–946 (2018).
123. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double- strand break repair pathway choice. Mol. Cell 47, 497–510 (2012).
124. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. Nat. Med. 21, 121–131 (2015).
125. Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature 533, 125–129 (2016).
126. Chu, V. T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. 33, 543–548 (2015).
127. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR- Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. 33, 538–542 (2015).
128. Rees, H. A., Yeh, W.-H. & Liu, D. R. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. Nat. Commun. 10, 1–12 (2019).

[2354]

129. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420– 424 (2016).
130. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551, 464–471 (2017).
131. Gao, X. et al. Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. Nature 553, 217–221 (2018).
132. Ingram, V. M. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. Nature 178, 792–794 (1956).
133. Myerowitz, R. & Costigan, F. C. The major defect in Ashkenazi Jews with Tay-Sachs disease is an insertion in the gene for the alpha-chain of beta-hexosaminidase. J. Biol. Chem. 263, 18587–18589 (1988).
134. Zielenski, J. Genotype and Phenotype in Cystic Fibrosis. Respiration 67, 117–133 (2000).
135. Mead, S. et al. A Novel Protective Prion Protein Variant that Colocalizes with Kuru Exposure. N. Engl. J. Med. 361, 2056–2065 (2009).
136. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843–1845 (2008).
137. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712 (2007).
138. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. Annu. Rev. Biophys. 46, 505–529 (2017).
139. Hille, F. et al. The Biology of CRISPR-Cas: Backward and Forward. Cell 172, 1239–1259 (2018).
140. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non- LTR retrotransposition. Cell 72, 595–605 (1993).
141. Liu, Y., Kao, H.-I. & Bambara, R. A. Flap endonuclease 1: a central component of DNA metabolism. Annu. Rev. Biochem. 73, 589–615 (2004).
142. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. 19, 770 (2018).
143. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat. Biotechnol. 34, 339–344 (2016).
144. Qi, L. S. et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell 152, 1173–1183 (2013).

[2355]

145. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat. Methods 12, 664–670 (2015).
146. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand- responsive genome editing and transcriptional activation. Nat. Commun. 8, 15939 (2017).
147. Jinek, M. et al. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science 343, 1247997 (2014).
148. Nishimasu, H. et al. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell 156, 935–949 (2014).
149. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. Science 348, 1477–1481 (2015).
150. Baranauskas, A. et al. Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. Protein Eng. Des. Sel. 25, 657–668 (2012).
151. Gerard, G. F. et al. The role of template-primer in protection of reverse transcriptase from thermal inactivation. Nucleic Acids Res. 30, 3118–3129 (2002).
152. Arezi, B. & Hogrefe, H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. Nucleic Acids Res. 37, 473–481 (2009).
153. Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M. & Gerard, G. F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. Nucleic Acids Res. 16, 265–277 (1988).
154. Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. Nature 563, 646–651 (2018).
155. Thuronyi, B. W. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. Nat. Biotechnol. (2019). doi:10.1038/s41587-019- 0193-0
156. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. 35, 371–376 (2017).
157. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. (2018). doi:10.1038/nbt.4172
158. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Sci. Adv. 3, eaao4774 (2017).
159. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490–495 (2016).
160. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science 364, 289–292 (2019).

[2356]

161. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science 364, 292–295 (2019).
162. Kim, D., Kim, D., Lee, G., Cho, S.-I. & Kim, J.-S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. Nat. Biotechnol. 37, 430–435 (2019).
163. Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR- guided DNA base editors. Nature 569, 433–437 (2019).
164. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature 571, 275–278 (2019).
165. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. Sci. Adv. 5, eaax5717 (2019).
166. Ostertag, E. M. & Kazazian Jr, H. H. Biology of Mammalian L1 Retrotransposons. Annu. Rev. Genet. 35, 501–538 (2001).
167. Griffiths, D. J. Endogenous retroviruses in the human genome sequence. Genome Biol. 2, REVIEWS1017 (2001).
168. Berkhout, B., Jebbink, M. & Zsíros, J. Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus. J. Virol. 73, 2365–2375 (1999).
169. Halvas, E. K., Svarovskaia, E. S. & Pathak, V. K. Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity. J. Virol. 74, 10349–10358 (2000).
170. Dever, D. P. et al. CRISPR/Cas9 Beta-globin Gene Targeting in Human Hematopoietic Stem Cells. Nature 539, 384–389 (2016).
171. Park, S. H. et al. Highly efficient editing of the β-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease. Nucleic Acids Res. doi:10.1093/nar/gkz475
172. Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. Annu. Rev. Neurosci. 24, 519–550 (2001).
173. Asante, E. A. et al. A naturally occurring variant of the human prion protein completely prevents prion disease. Nature 522, 478–481 (2015).
174. Anzalone, A. V., Lin, A. J., Zairis, S., Rabadan, R. & Cornish, V. W. Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches. Nat. Methods 13, 453–458 (2016).
175. Badran, A. H. et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature 533, 58–63 (2016).
176. Anders, C. & Jinek, M. Chapter One - In Vitro Enzymology of Cas9. in Methods in Enzymology (eds. Doudna, J. A. & Sontheimer, E. J.) 546, 1–20 (Academic Press, 2014).

[2357]

177. Pirakitikulr, N., Ostrov, N., Peralta-Yahya, P. & Cornish, V. W. PCRless library mutagenesis via oligonucleotide recombination in yeast. Protein Sci. Publ. Protein Soc. 19, 2336–2346 (2010).
178. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. 37, 224–226 (2019).
179. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197 (2015).
180. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off- target effects. Nature 529, 490–495 (2016).
181. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. (2018). doi:10.1038/nbt.4172
182. Baranauskas, A. et al. Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. Protein Eng. Des. Sel. 25, 657–668 (2012).
183. Schechner, DM, Hacisuleyman E., Younger ST, Rinn JL. Nat Methods 664-70 (2015).
184. Brown JA, et al. Nat Struct Mol Biol 633-40 (2014).
185. Conrad NA and Steitz JA. EMBO J 1831-41 (2005).
186. Bartlett JS, et al. Proc Natl Acad Sci USA 8852-7 (1996).
187. Mitton-Fry RM, DeGregorio SJ, Wang J, Steitz TA, Steitz JA. Science 1244-7 (2010).
188. Forster AC, Symons RH. Cell. 1987.
189. Weinberg Z, Kim PB, Chen TH, Li S, Harris KA, Lünse CE, Breaker RR. Nat. Chem. Biol. 2015.
190. Feldstein PA, Buzayan JM, Bruening G. Gene 1989.
191. Saville BJ, Collins RA. Cell. 1990.
192. Roth A, Weinberg Z, Chen AG, Kim PG, Ames TD, Breaker RR. Nat Chem Biol. 2013.
193. Borchardt EK, et al. RNA 1921-30 (2015).
194. Zhang Y, et al. Mol Cell 792-806 (2013).
195. Dang Y, et al. Genome Biol 280 (2015).
196. Schaefer M, Kapoor U, and Jantsch MF. Open Biol 170077 (2017).
197. Nahar S, et al. Chem Comm 2377-80 (2018).
198. Gao Y and Zhao Y. J Integr Plant Biol 343-9 (2014).
199. Dubois N, Marquet R, Paillart J, Bernacchi S. Front Microbiol 527 (2018).
200. Costa M and Michel F. EMBO J 1276-85 (1995).
201. Hu JH, et al. Nature 57-63 (2018).
202. Furukawa K, Gu H, Breaker RR. Methods Mol Biol 209-20 (2014).

[2358]

203. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Lett. 583, 909–914 (2009).
204. Kügler, S., Kilic, E. & Bähr, M. Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. Gene Ther. 10, 337–347 (2003).
205. de Felipe, P., Hughes, L. E., Ryan, M. D. & Brown, J. D. Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide. J. Biol. Chem. 278, 11441–11448 (2003).
206. Levy, J. M. & Nicoll, R. A. Membrane-associated guanylate kinase dynamics reveal regional and developmental specificity of synapse stability. J. Physiol. 595, 1699–1709 (2017).
207. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011).
208. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47–e47 (2015).

[2359]

[2360] 등가물 및 범주

[2361] 단수형은 달리 나타내지 않는 한 또는 달리 문맥으로부터 명백하지 않는 한 하나 또는 하나 초과를 의미할 수 있다. 군의 1개 이상의 구성원들 사이에 "또는"을 포함하는 실시양태 또는 기재는 달리 나타내지 않는 한 또는 달리 문맥으로부터 명백하지 않는 한 군 구성원 중 하나, 하나 초과 또는 모두가 주어진 생성물 또는 방법에 존재하거나, 그에 사용되거나 또는 달리 그와 관련되는 경우에 충족된 것으로 간주된다. 본 발명은 군의 정확히 하나의 구성원이 주어진 생성물 또는 방법에 존재하거나, 그에 사용되거나 또는 달리 그와 관련되는 실시양태를 포함한다. 본 발명은 군 구성원 중 하나 초과 또는 모두가 주어진 생성물 또는 방법에 존재하거나, 그에 사용되거나 또는 달리 그와 관련되는 실시양태를 포함한다.

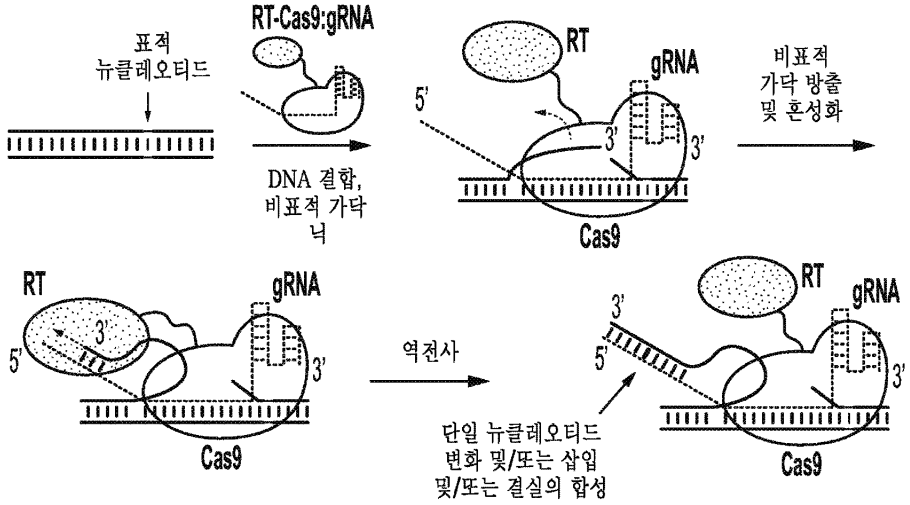
[2362] 추가로, 본 개시내용은 열거된 청구항 중 하나 이상으로부터의 하나 이상의 제한, 요소, 조항 및 서술적 용어를 또 다른 청구항에 도입하는 모든 변형, 조합 및 순열을 포괄한다. 예를 들어, 또 다른 청구항에 종속하는 임의의 청구항은 동일한 기본 청구항에 종속하는 임의의 다른 청구항에서 발견된 하나 이상의 제한을 포함하도록 변형될 수 있다. 요소가 목록으로서, 예를 들어 마쿠쉬 군 포맷으로 제시되는 경우에, 요소의 각각의 하위군이 또한 개시되고, 임의의 요소(들)는 군으로부터 제거될 수 있다. 일반적으로, 본 발명 또는 본 발명의 측면이 특정한 요소 및/또는 특색을 포함하는 것으로 언급되는 경우에, 본 개시내용의 특정 실시양태 또는 본 개시내용의 특정 측면은 이러한 요소 및/또는 특색으로 이루어지거나 또는 본질적으로 이루어지는 것으로 이해되어야 한다. 단순성의 목적을 위해, 이들 실시양태는 본원에 구체적으로 제시되지 않았다. 또한, 용어 "포함하는" 및 "함유하는"은 개방적인 것으로 의도되고, 추가의 요소 또는 단계의 포함을 허용한다는 것이 주목된다. 범위가 주어지는 경우, 종점이 포함된다. 추가로, 달리 나타내지 않는 한 또는 달리 문맥 및 관련 기술분야의 통상의 기술자의 이해로부터 명백하지 않는 한, 범위로서 표현된 값은 본 발명의 상이한 실시양태에서 언급된 범위 내의 임의의 구체적 값 또는 하위-범위를, 문맥이 달리 명확하게 지시하지 않는 한 그러한 범위의 하한치 단위의 1/10까지 가정할 수 있다.

[2363] 본 출원은 다양한 허여된 특허, 공개된 특허 출원, 학술지 논문 및 다른 공개물을 언급하며, 이들 모두는 본원에 참조로 포함된다. 임의의 포함된 참고문헌과 본 명세서 사이에 상충이 존재하는 경우, 본 명세서가 우선할 것이다. 추가로, 선행 기술에 속하는 본 발명의 임의의 특정한 실시양태는 실시양태 중 어느 하나 이상으로부터 명백하게 배제될 수 있다. 이러한 실시양태는 관련 기술분야의 통상의 기술자에게 공지된 것으로 간주되기 때문에, 이들은 배제가 본원에 명백하게 제시되지 않더라도 배제될 수 있다. 본 발명의 임의의 특정한 실시양태는 선행 기술의 존재와 관련되든 관련되지 않든 임의의 이유로 임의의 실시양태로부터 배제될 수 있다.

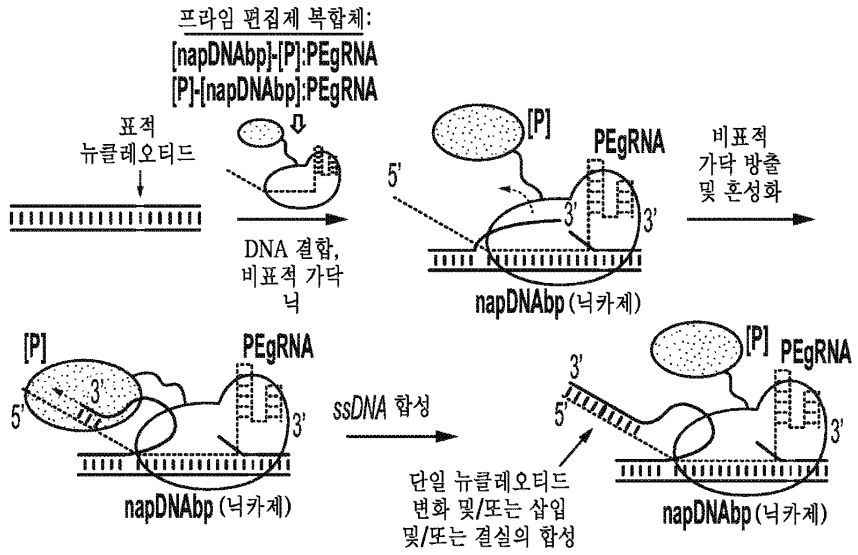
[2364] 관련 기술분야의 통상의 기술자는 상용 실험만을 사용하여 본원에 기재된 구체적 실시양태에 대한 많은 등가물을 인식하거나 확인할 수 있을 것이다. 본원에 기재된 본 실시양태의 범주는 상기 설명에 제한되는 것으로 의도되지 않으며, 오히려 첨부된 실시양태에 제시된 바와 같다. 관련 기술분야의 통상의 기술자는 하기 실시양태에 정의된 바와 같은 본 발명의 취지 또는 범주로부터 벗어나지 않으면서 본 설명에 대한 다양한 변화 및 변형이 이루어질 수 있음을 인지할 것이다.

도면

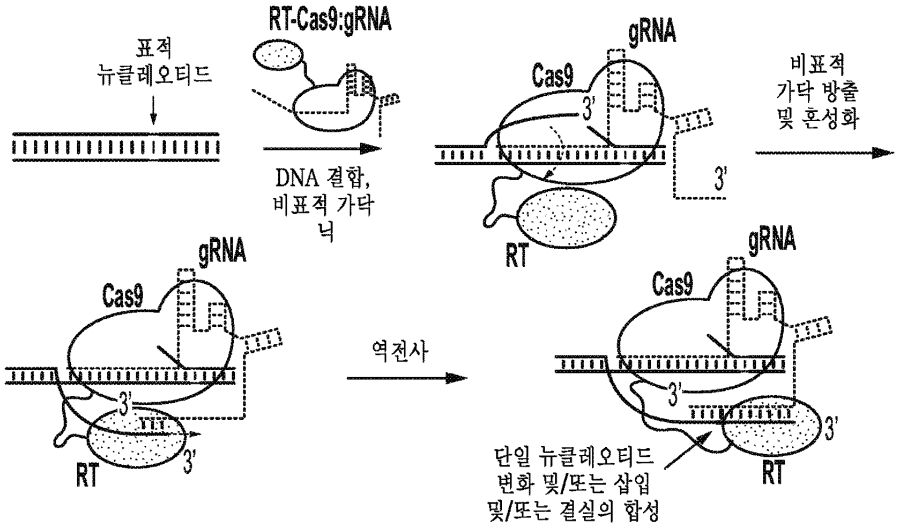
도면1a


표적
뉴클레오티드
RT-Cas9:gRNA
DNA 결합,
비표적 가닥
닉
RT
gRNA
5'
3'
3'
Cas9
비표적
가닥 방출
및 혼성화
RT
gRNA
5'
3'
3'
Cas9
역전사
3'
5'
RT
gRNA
3'
Cas9
단일 뉴클레오티드
변화 및/또는 삽입
및/또는 결실의 합성

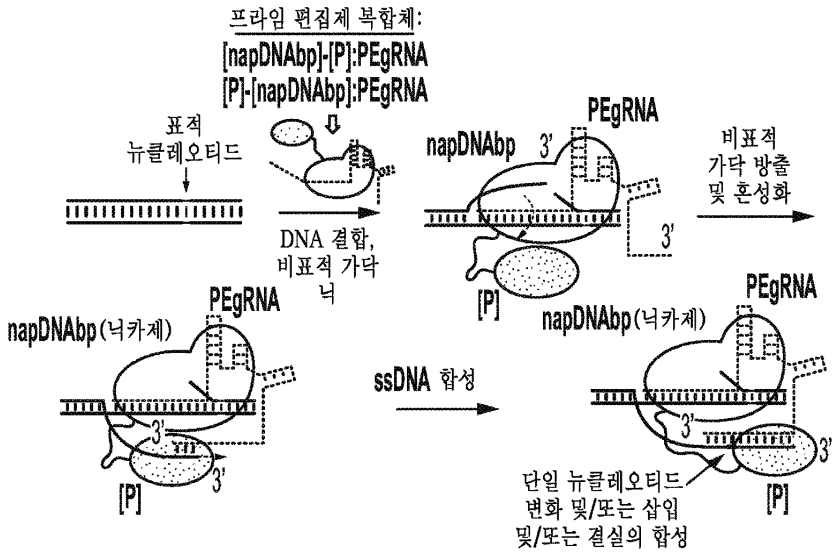
도면1b


프라임 편집제 복합체:
[napDNAbp]-[P]:PEgRNA
[P]-[napDNAbp]:PEgRNA
표적
뉴클레오티드
DNA 결합,
비표적 가닥
닉
[P]
PEgRNA
5'
3'
3'
napDNAbp (닉카제)
비표적
가닥 방출
및 혼성화
[P]
PEgRNA
5'
3'
3'
napDNAbp (닉카제)
ssDNA 합성
3'
5'
[P]
PEgRNA
3'
napDNAbp (닉카제)
단일 뉴클레오티드
변화 및/또는 삽입
및/또는 결실의 합성

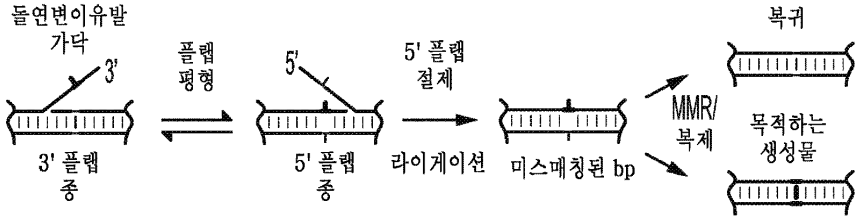
도면1c



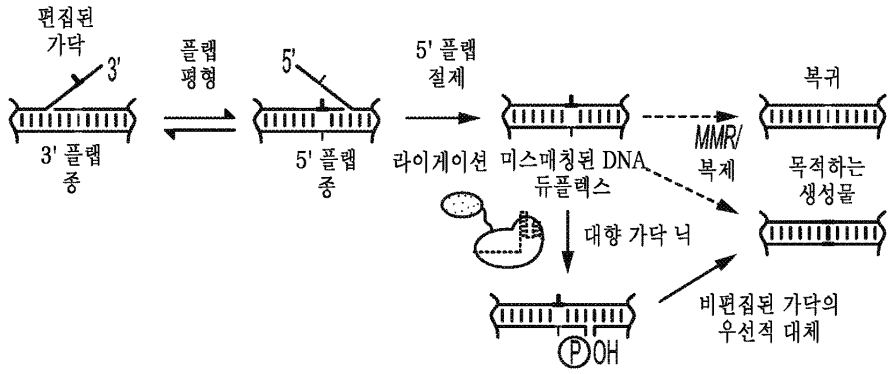
도면1d



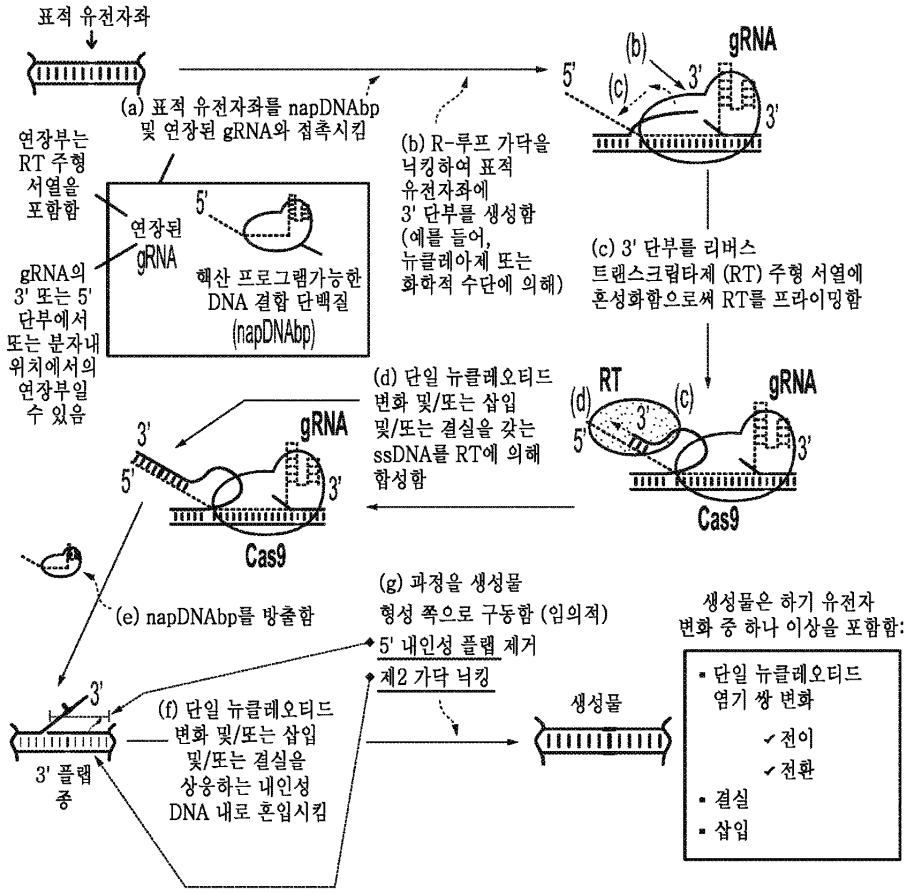
도면1e



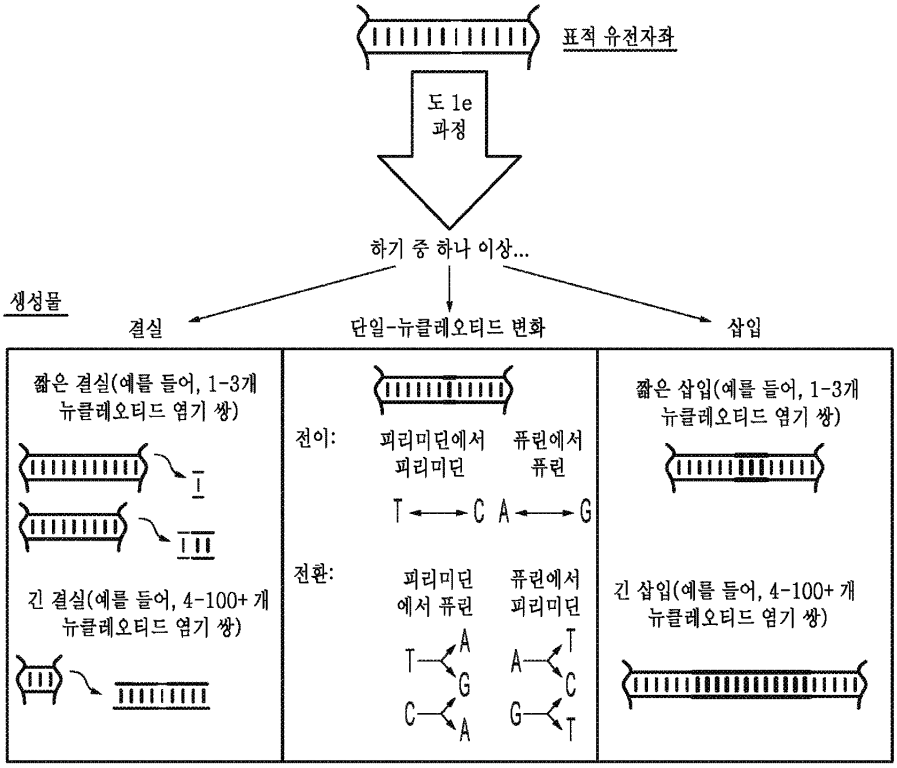
도면1f


편집된
가닥
3'
3' 플랩
종
플랩
평형
5'
5' 플랩
종
5' 플랩
절제
라이게이션
미스매칭된 DNA
듀플렉스
MMR/
복제
복귀
목적하는
생성물
대향 가닥 닉
비편집된 가닥의
우선적 대체
POH

도면1g


표적 유전자좌
(a) 표적 유전자좌를 napDNAbp
및 연장된 gRNA와 접촉시킴
연장부는
RT 주형
서열을
포함함
gRNA의
3' 또는 5'
단부에서
또는 분자내
위치에서의
연장부일
수 있음
연장된
gRNA
5'
핵산 프로그램가능한
DNA 결합 단백질
(napDNAbp)
(b) R-루프 가닥을
닉킹하여 표적
유전자좌에
3' 단부를 생성함
(예를 들어,
뉴클레아제 또는
화학적 수단에 의해)
(b)
gRNA
(c)
3'
(c) 3' 단부를 리버스
트랜스크립타제 (RT) 주형 서열에
혼성화함으로써 RT를 프라이밍함
RT
(d)
Cas9
(d) 단일 뉴클레오티드
변화 및/또는 삽입
및/또는 결실을 갖는
ssDNA를 RT에 의해
합성함
(e) napDNAbp를 방출함
(g) 과정을 생성물
형성 쪽으로 구동함 (임의적)
5' 내인성 플랩 제거
제2 가닥 닉킹
생성물
생성물은 하기 유전자
변화 중 하나 이상을 포함함:
▪ 단일 뉴클레오티드
염기 쌍 변화
✓전이
✓전환
▪ 결실
▪ 삽입
3' 플랩
종
(f) 단일 뉴클레오티드
변화 및/또는 삽입
및/또는 결실을
상응하는 내인성
DNA 내로 혼입시킴

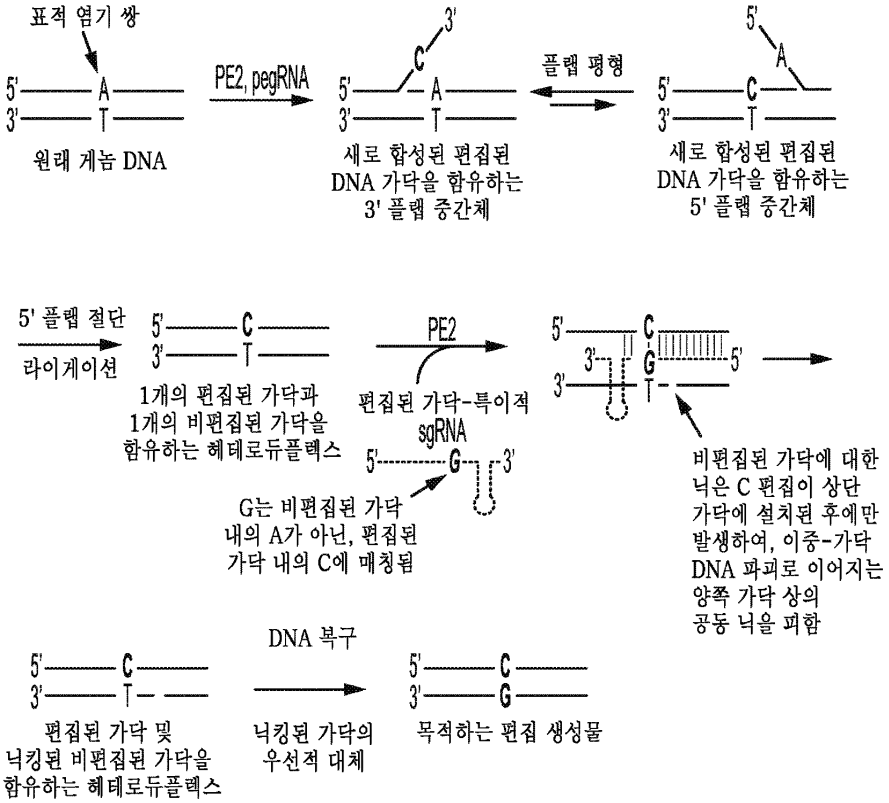
도면1h


표적 유전자좌
도 1e
과정
하기 중 하나 이상...
생성물
결실
단일-뉴클레오티드 변화
삽입
짧은 결실(예를 들어, 1-3개 뉴클레오티드 염기 쌍)
긴 결실(예를 들어, 4-100+ 개 뉴클레오티드 염기 쌍)
전이:
피리미딘에서 피리미딘
퓨린에서 퓨린
T ↔ C A ↔ G
전환:
피리미딘에서 퓨린
퓨린에서 피리미딘
T A G C
A T C G
짧은 삽입(예를 들어, 1-3개 뉴클레오티드 염기 쌍)
긴 삽입(예를 들어, 4-100+ 개 뉴클레오티드 염기 쌍)

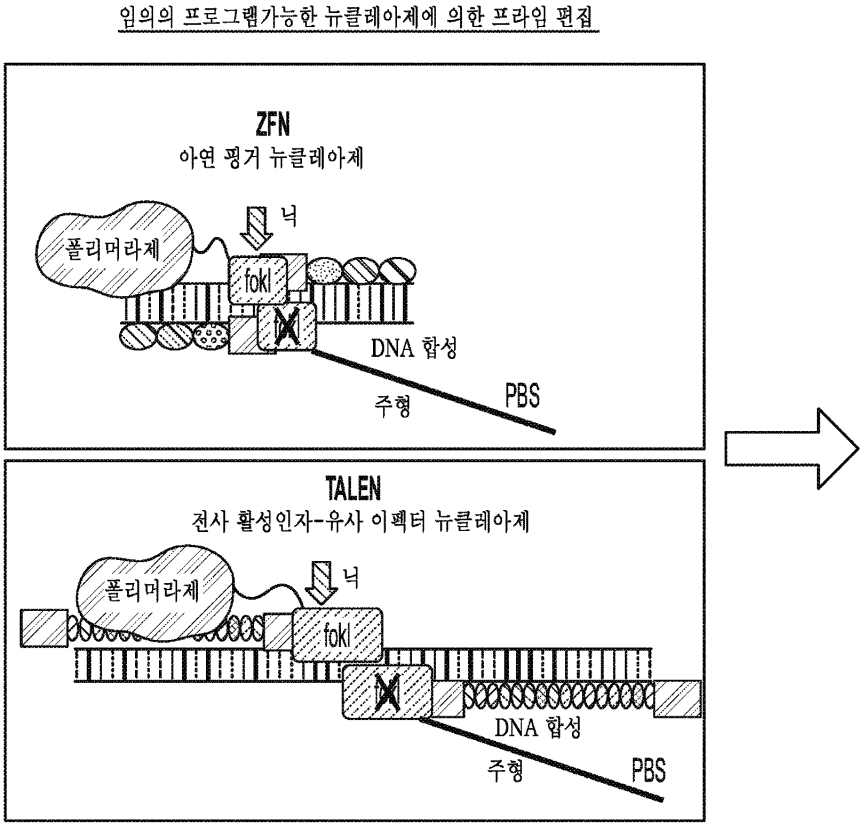
도면1i

PE3b 시간적 닉킹.

PE3b 편집 시스템: 상보적인 (비편집된) DNA 가닥의 편집-특이적 닉킹



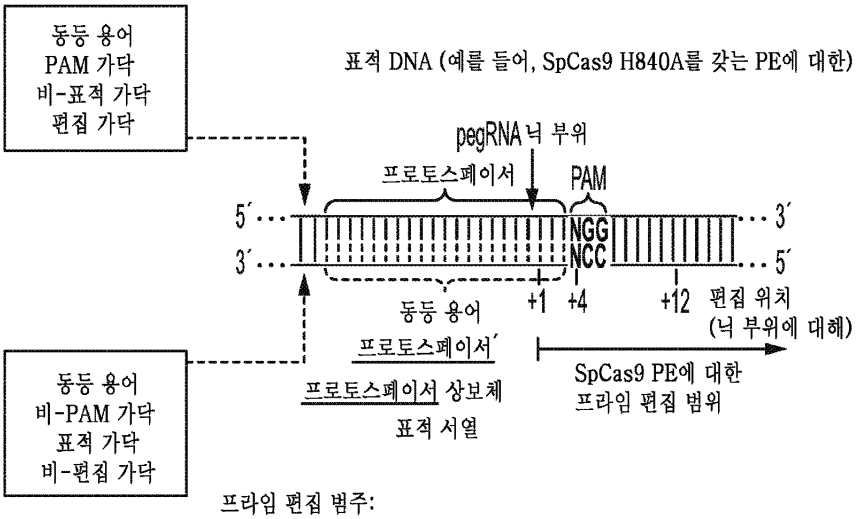
도면1j



도면1k

DNA 합성 주형에 대한 폴리머라제에 의해 코딩되고 목적하는 편집을 함유하는 대체 가닥의 폴리머라제-의존성 합성. 메카니즘은 프라임 편집과 유사하게 작동하도록 구상된다. 추가의 인자 또는 이펙터가 트랜스로 또는 융합체로서 부가되어 반응을 전체로서 용이하게 할 수 있다 (예를 들어, (a) 절단 부위에서 DNA를 풀어내어 3' 단부를 갖는 절단 가닥을 프라이머로서 이용가능하게 하는 헬리카제, (b) 절단 가닥 상의 내인성 가닥을 제거하도록 도와 합성된 가닥으로 내인성 가닥을 대체하는 쪽으로 반응을 구동하는 FEN1, 또는 (c) 대향 가닥 상에 제2 부위 닉을 생성하여, 비-편집된 가닥의 선호되는 세포 복구를 통해 합성 복구의 통합을 구동하는 것을 도울 수 있는 nCas9:gRNA 복합체)

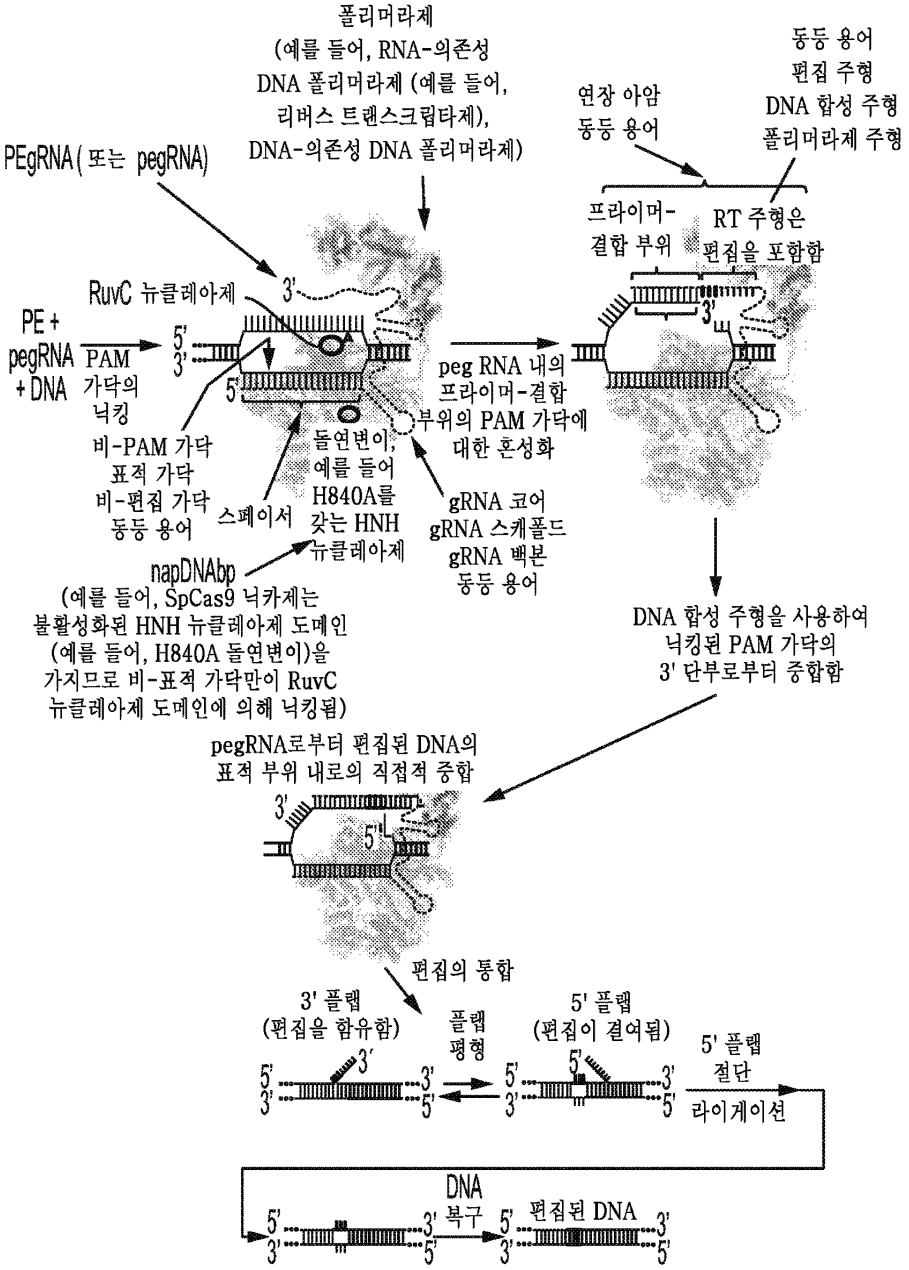
도면11



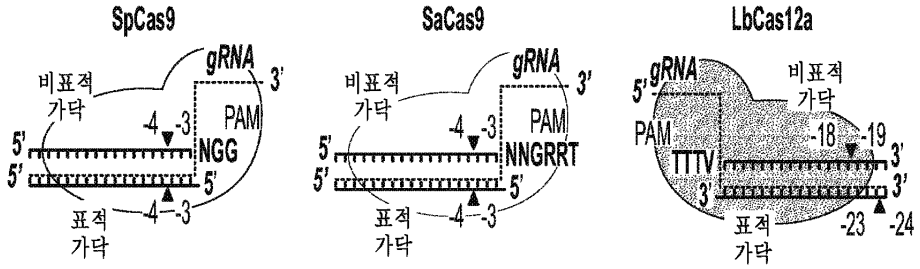
프라임 편집 범주:

모든 4개의 전이 점 돌연변이
모든 8개의 전환 점 돌연변이
삽입 (1 bp 내지 ≥ 44 bp)
결실 (1 bp 내지 ≥ 80 bp)
상기의 조합

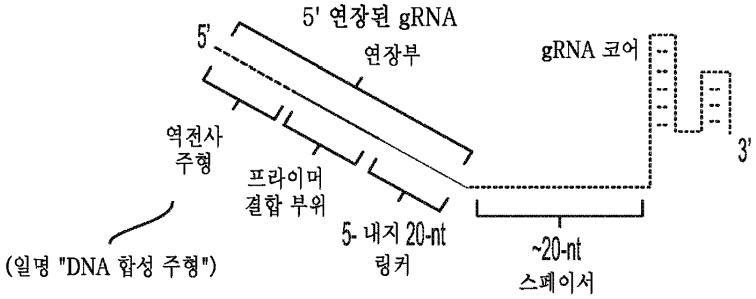
도면1m


폴리머라제
(예를 들어, RNA-의존성
DNA 폴리머라제 (예를 들어,
리버스 트랜스크립타제),
DNA-의존성 DNA 폴리머라제)
PEgRNA (또는 pegRNA)
연장 아암
동등 용어
동등 용어
편집 주형
DNA 합성 주형
폴리머라제 주형
프라이머-
결합 부위
RT 주형은
편집을 포함함
RuvC 뉴클레아제
PE +
pegRNA
+ DNA
PAM
가닥의
닉킹
peg RNA 내의
프라이머-결합
부위의 PAM 가닥에
대한 혼성화
비-PAM 가닥
표적 가닥
비-편집 가닥
동등 용어
스페이서
돌연변이,
예를 들어
H840A를
갖는 HNH
뉴클레아제
gRNA 코어
gRNA 스캐폴드
gRNA 백본
동등 용어
napDNAbp
(예를 들어, SpCas9 닉카제는
불활성화된 HNH 뉴클레아제 도메인
(예를 들어, H840A 돌연변이)을
가지므로 비-표적 가닥만이 RuvC
뉴클레아제 도메인에 의해 닉킹됨)
DNA 합성 주형을 사용하여
닉킹된 PAM 가닥의
3' 단부로부터 중합함
pegRNA로부터 편집된 DNA의
표적 부위 내로의 직접적 중합
편집의 통합
3' 플랩
(편집을 함유함)
플랩
평형
5' 플랩
(편집이 결여됨)
5' 플랩
절단
라이게이션
DNA
복구
편집된 DNA

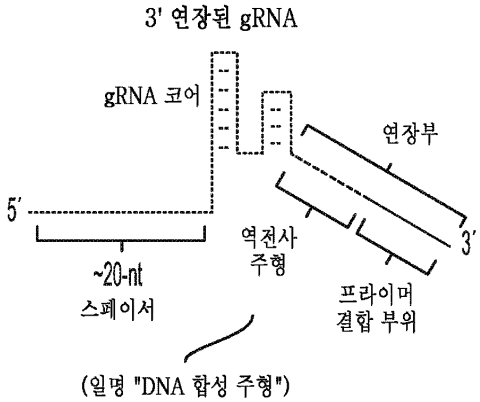
도면2


SpCas9
SaCas9
LbCas12a
gRNA
비표적
가닥
PAM
NGG
NNGRRT
TTTV
표적
가닥

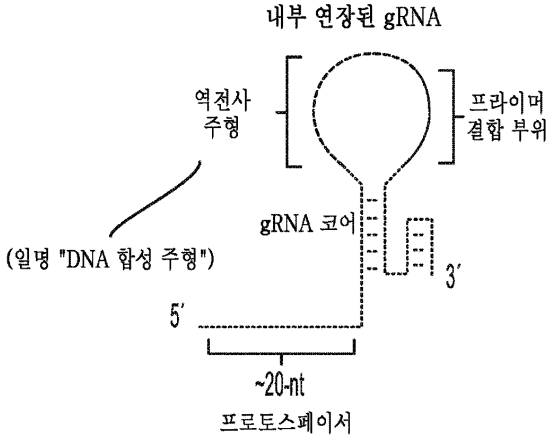
도면3a



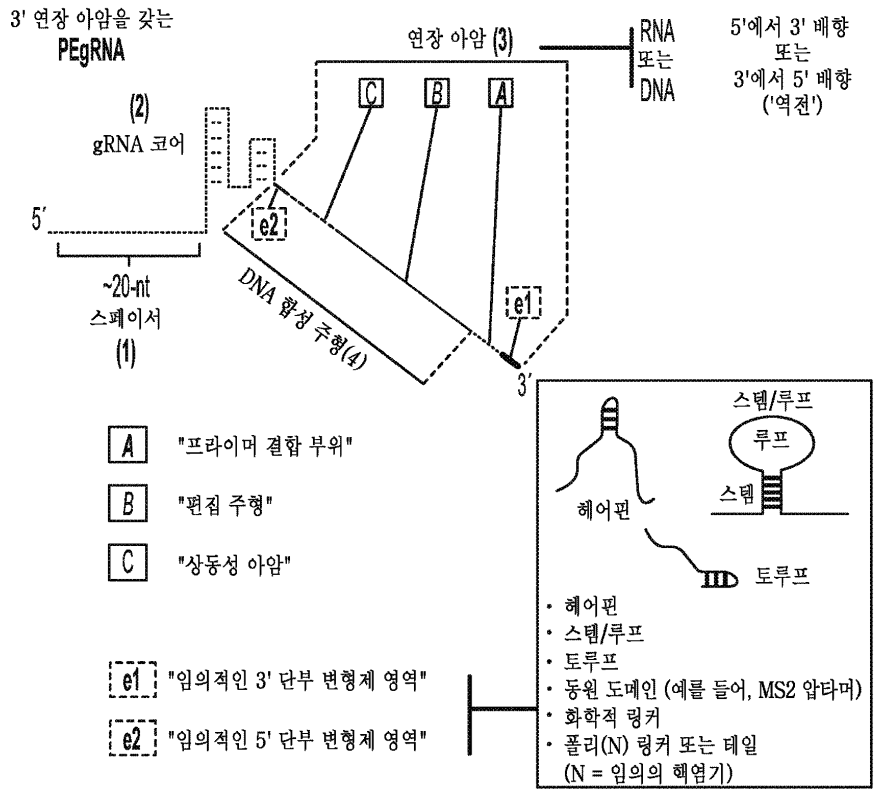
도면3b



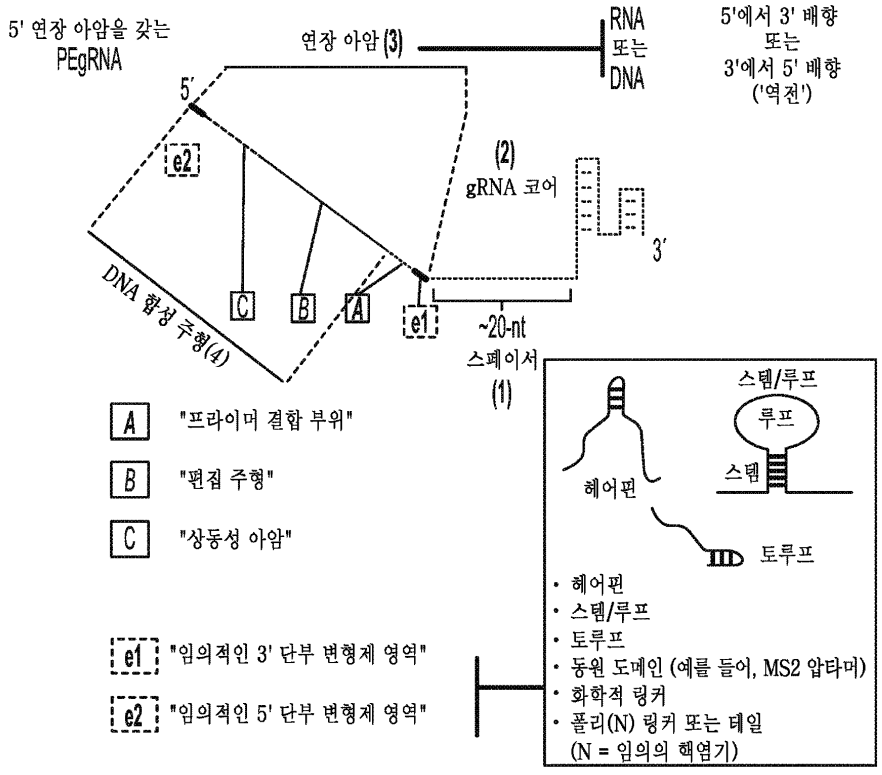
도면3c



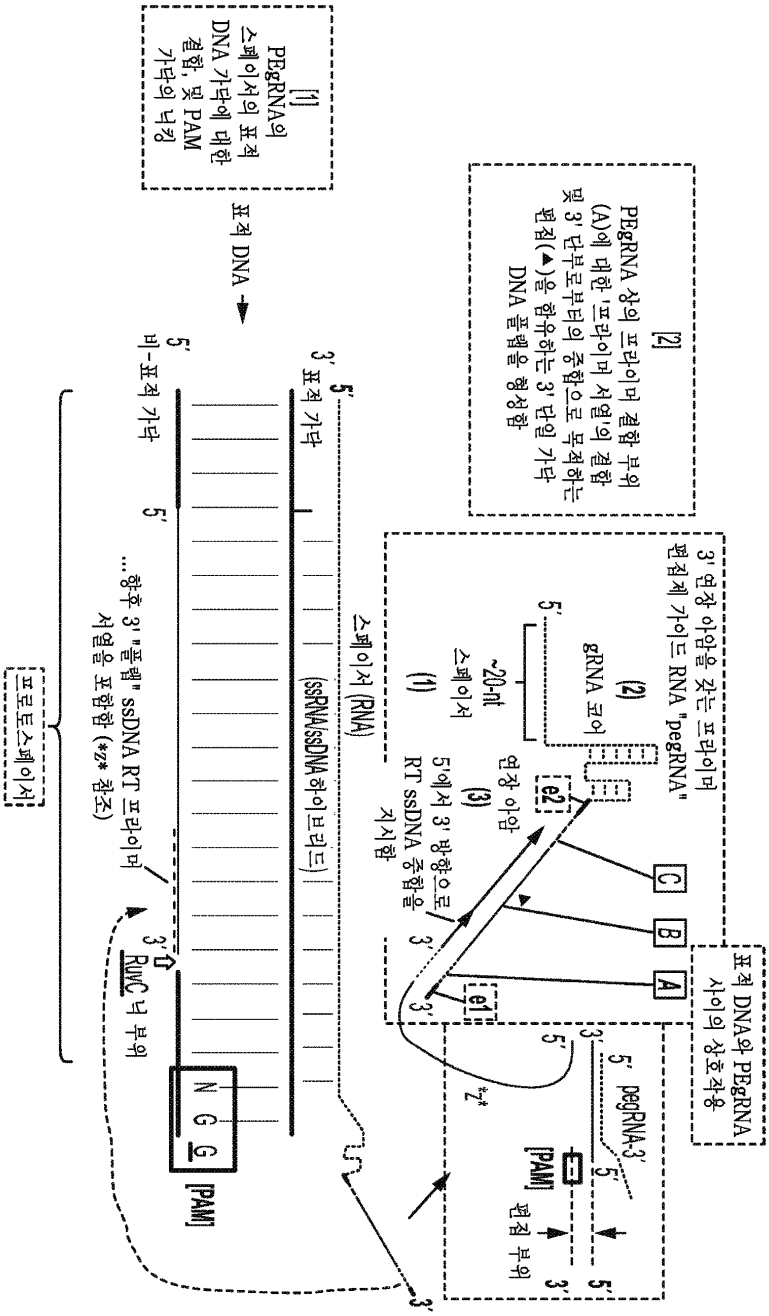
도면3d


3' 연장 아암을 갖는
PEgRNA
(2)
gRNA 코어
5'
~20-nt
스페이서
(1)
연장 아암 (3)
RNA
또는
DNA
5'에서 3' 배향
또는
3'에서 5' 배향
('역전')
C
B
A
e2
e1
DNA 합성 주형(4)
3'
A "프라이머 결합 부위"
B "편집 주형"
C "상동성 아암"
e1 "임의적인 3' 단부 변형제 영역"
e2 "임의적인 5' 단부 변형제 영역"
스템/루프
루프
스템
헤어핀
토루프
• 헤어핀
• 스템/루프
• 토루프
• 동원 도메인 (예를 들어, MS2 압타머)
• 화학적 링커
• 폴리(N) 링커 또는 테일
(N = 임의의 핵염기)

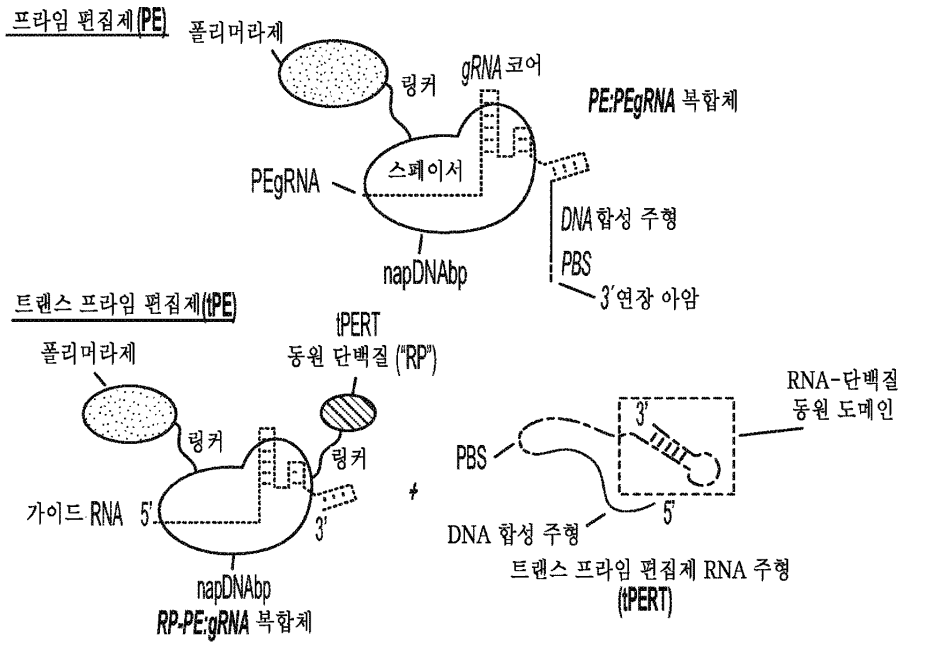
도면3e


5' 연장 아암을 갖는
PEgRNA
연장 아암 (3)
RNA
또는
DNA
5'에서 3' 배향
또는
3'에서 5' 배향
('역전')
5'
e2
(2)
gRNA 코어
3'
DNA 합성 주형(4)
C
B
A
e1
~20-nt
스페이서
(1)
A "프라이머 결합 부위"
B "편집 주형"
C "상동성 아암"
e1 "임의적인 3' 단부 변형제 영역"
e2 "임의적인 5' 단부 변형제 영역"
스템/루프
루프
스템
헤어핀
토루프
• 헤어핀
• 스템/루프
• 토루프
• 동원 도메인 (예를 들어, MS2 압타머)
• 화학적 링커
• 폴리(N) 링커 또는 테일
(N = 임의의 핵염기)

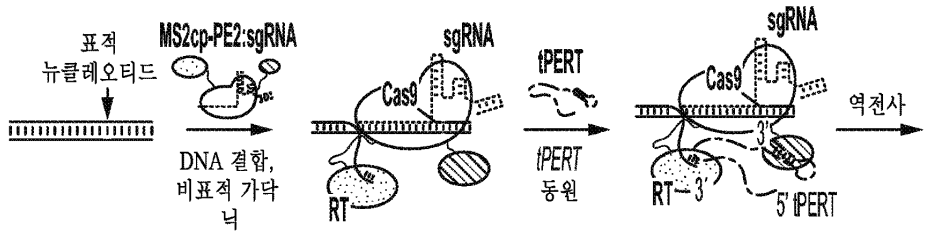
도면3f


[1]
PEgRNA의
스페이서의 표적
DNA 가닥에 대한
결합, 및 PAM
가닥의 닉킹
표적 DNA
5′
비-표적 가닥
5′
3′ 표적 가닥
5′
스페이서 (RNA)
(ssRNA:ssDNA 하이브리드)
...향후 3′ "플랩" ssDNA RT 프라이머
서열을 포함함 (*z* 참조)
3′
RuvC 닉 부위
N G G
[PAM]
프로토스페이서
[2]
PBgRNA 상의 프라이머 결합 부위
(A)에 대한 '프라이머 서열'의 결합
및 3′ 단부로부터의 중합으로 목적하는
편집(▲)을 함유하는 3′ 단일 가닥
DNA 플랩을 형성함
3′ 연장 아암을 갖는 프라이머
편집체 가이드 RNA "pegRNA"
(2)
gRNA 코어
5′
~20-nt
스페이서
(1)
e2
연장 아암
(3)
5′에서 3′ 방향으로
RT ssDNA 중합을
지시함
3′
C
B
A
e1
3′
5′
z
3
5′ pegRNA-3′
5′
[PAM]
편집 부위
3′
5′
표적 DNA와 PBgRNA
사이의 상호작용

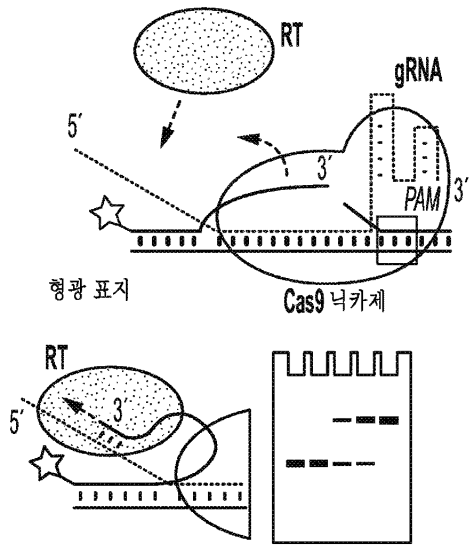
도면3g



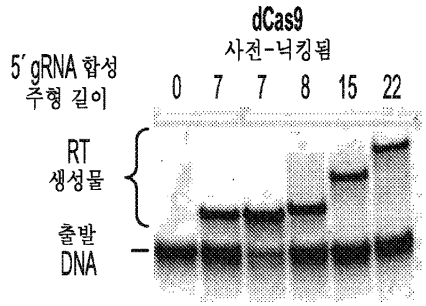
도면3h



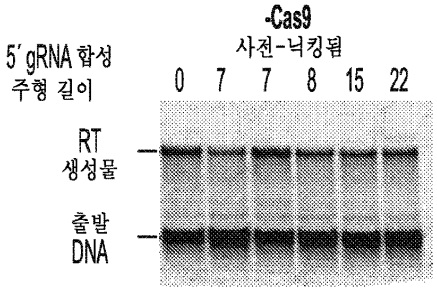
도면4a



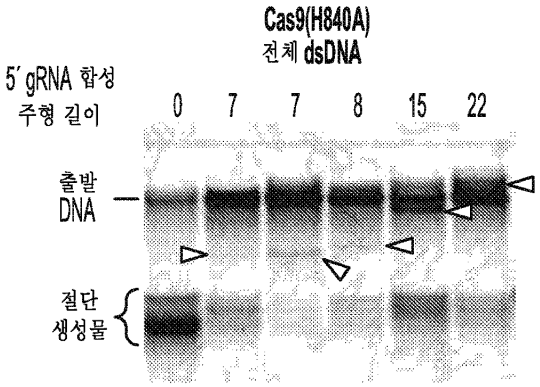
도면4b



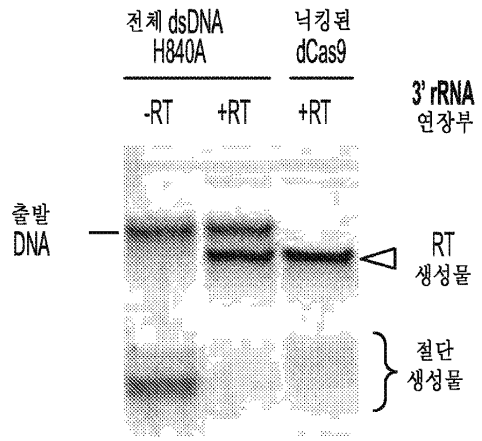
도면4c



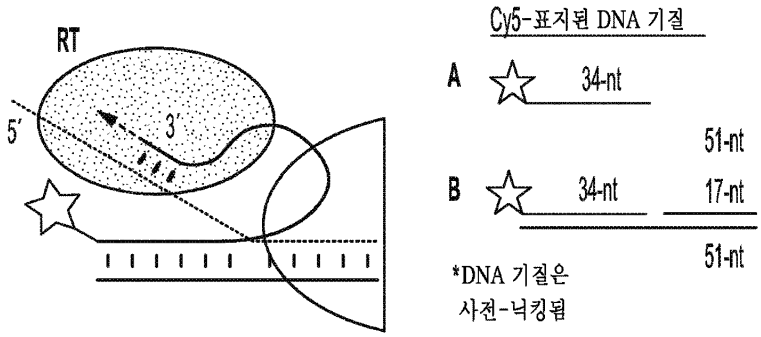
도면4d

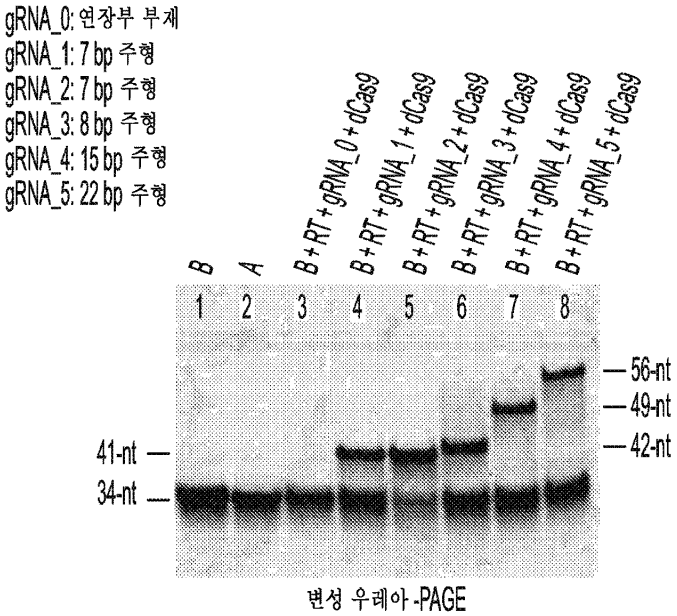


도면4e

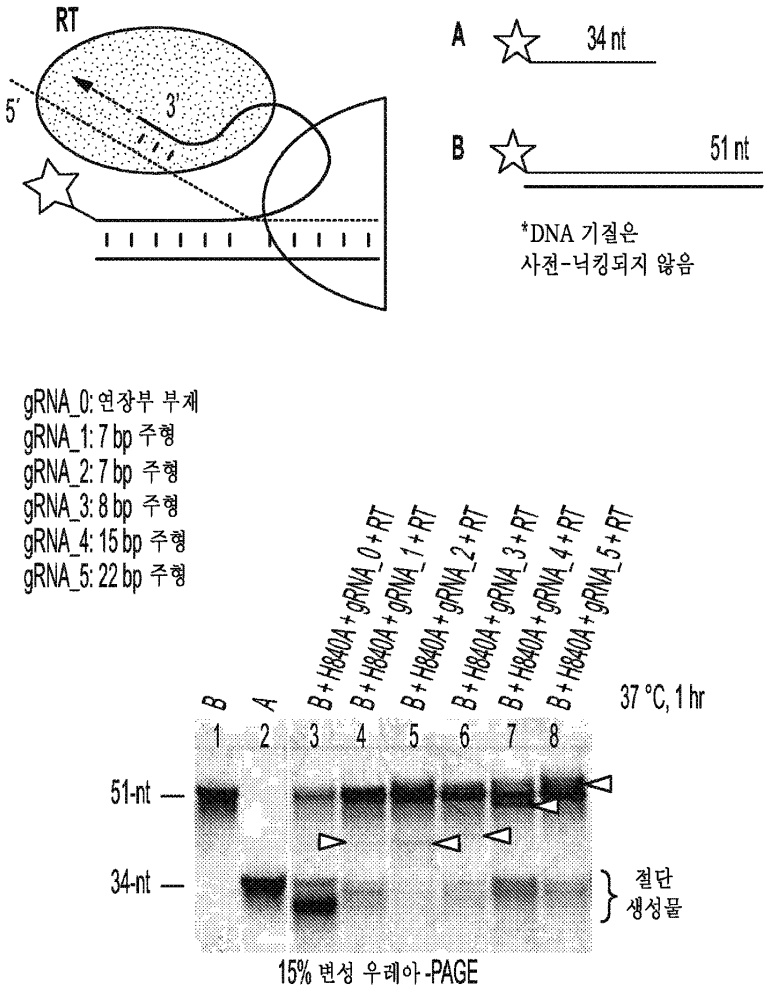

전체 dsDNA
H840A
닉킹된
dCas9
-RT
+RT
+RT
3' rRNA
연장부
출발
DNA
RT
생성물
절단
생성물

도면5

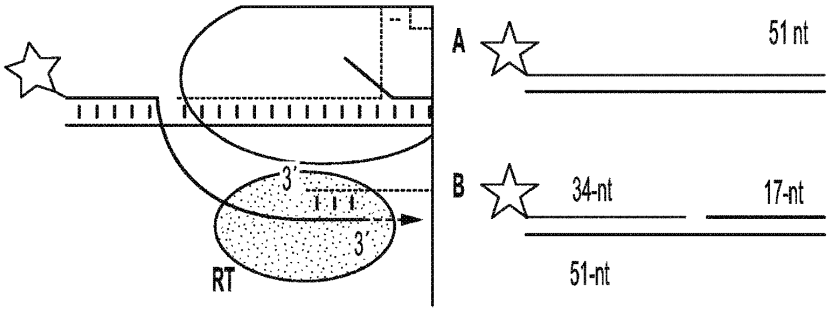

RT
5'
3'
Cy5-표지된 DNA 기질
A
34-nt
51-nt
B
34-nt
17-nt
51-nt
*DNA 기질은
사전-닉킹됨


gRNA_0: 연장부 부재
gRNA_1: 7 bp 주형
gRNA_2: 7 bp 주형
gRNA_3: 8 bp 주형
gRNA_4: 15 bp 주형
gRNA_5: 22 bp 주형
B
A
B+RT+gRNA_0+dCas9
B+RT+gRNA_1+dCas9
B+RT+gRNA_2+dCas9
B+RT+gRNA_3+dCas9
B+RT+gRNA_4+dCas9
B+RT+gRNA_5+dCas9
1 2 3 4 5 6 7 8
56-nt
49-nt
42-nt
41-nt
34-nt
변성 우레아-PAGE

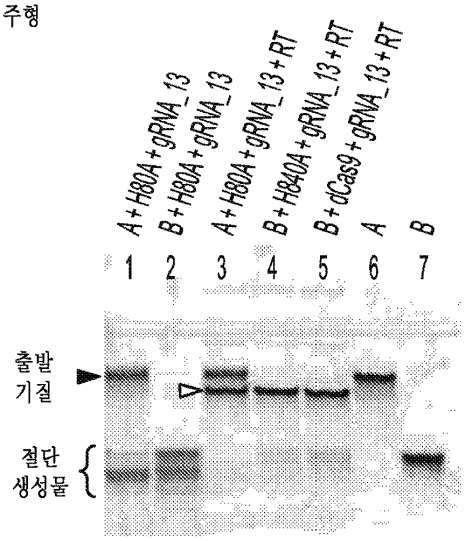
도면6


RT
5'
3'
A
34 nt
B
51 nt
*DNA 기질은
사전-닉킹되지 않음
gRNA_0: 연장부 부재
gRNA_1: 7 bp 주형
gRNA_2: 7 bp 주형
gRNA_3: 8 bp 주형
gRNA_4: 15 bp 주형
gRNA_5: 22 bp 주형
B
A
B + H840A + gRNA_0 + RT
B + H840A + gRNA_1 + RT
B + H840A + gRNA_2 + RT
B + H840A + gRNA_3 + RT
B + H840A + gRNA_4 + RT
B + H840A + gRNA_5 + RT
37 °C, 1 hr
1 2 3 4 5 6 7 8
51-nt
34-nt
절단
생성물
15% 변성 우레아-PAGE

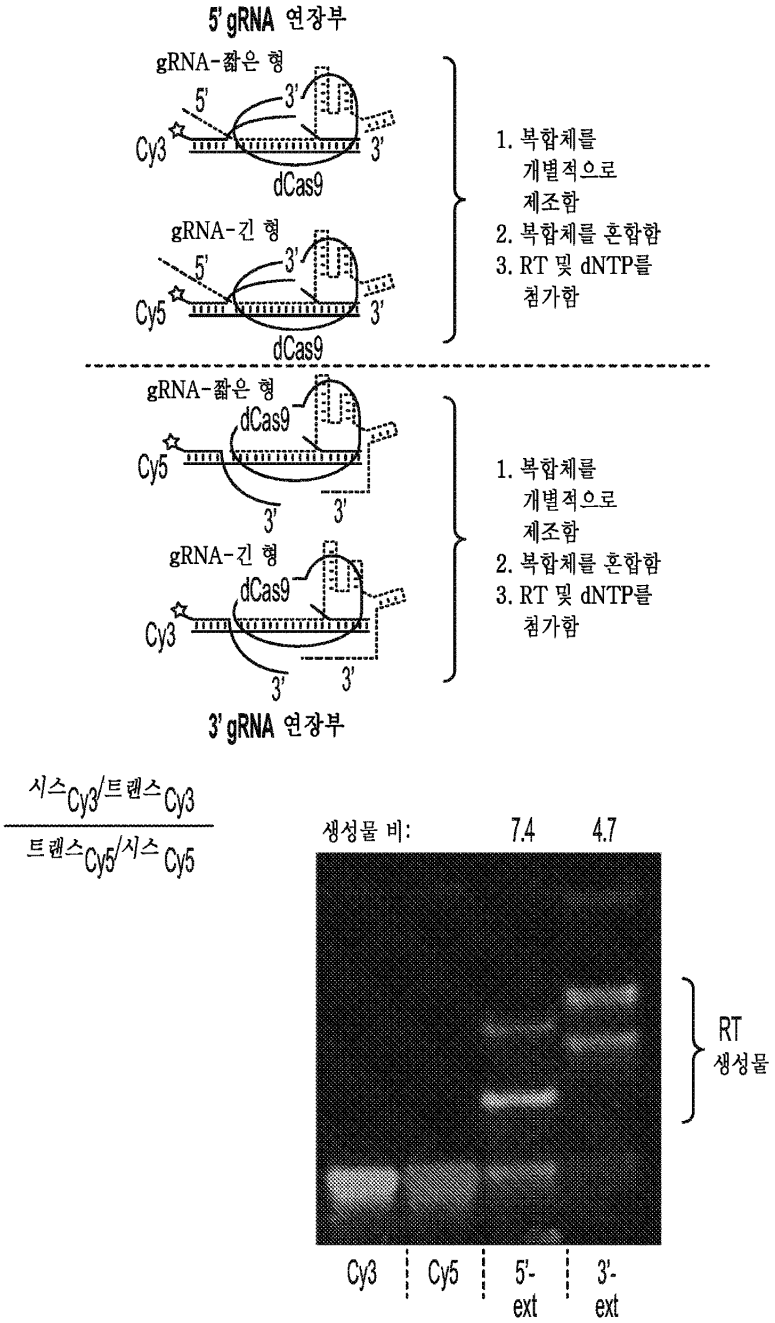
도면7



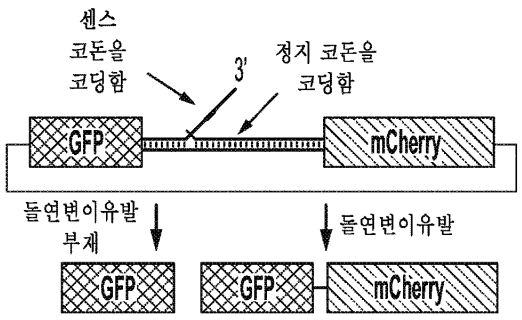
gRNA_0: 연장부 부재
gRNA_1: 15 bp 주형



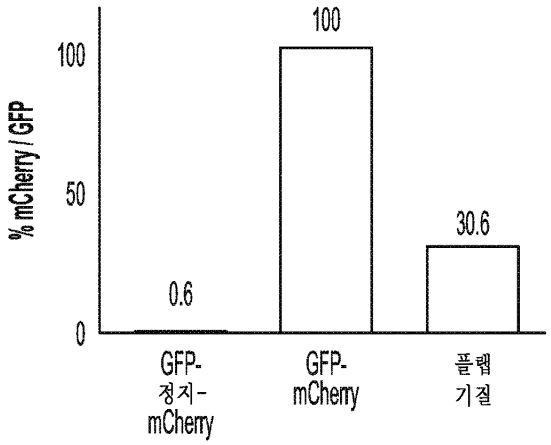
도면8


5' gRNA 연장부
gRNA-짧은 형
5'
3'
Cy3
3'
dCas9
gRNA-긴 형
5'
3'
Cy5
3'
dCas9
1. 복합체를 개별적으로 제조함
2. 복합체를 혼합함
3. RT 및 dNTP를 첨가함
gRNA-짧은 형
dCas9
Cy5
3'
3'
gRNA-긴 형
dCas9
Cy3
3'
3'
1. 복합체를 개별적으로 제조함
2. 복합체를 혼합함
3. RT 및 dNTP를 첨가함
3' gRNA 연장부
시스Cy3/트랜스Cy3
트랜스Cy5/시스Cy5
생성물 비:
7.4
4.7
RT 생성물
Cy3
Cy5
5'-ext
3'-ext

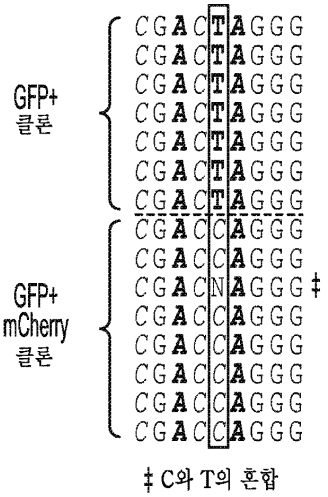
도면9a


센스 코돈을 코딩함
3'
정지 코돈을 코딩함
GFP
mCherry
돌연변이유발 부재
돌연변이유발
GFP
GFP
mCherry

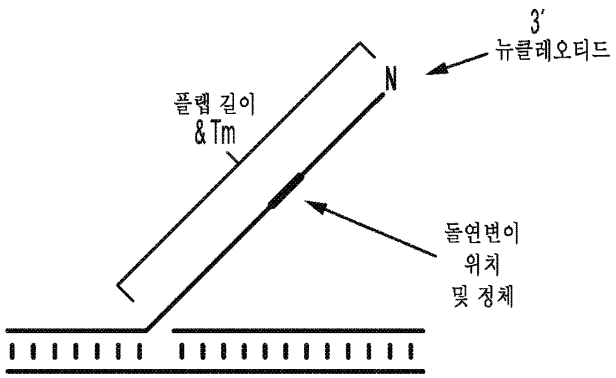
도면9b



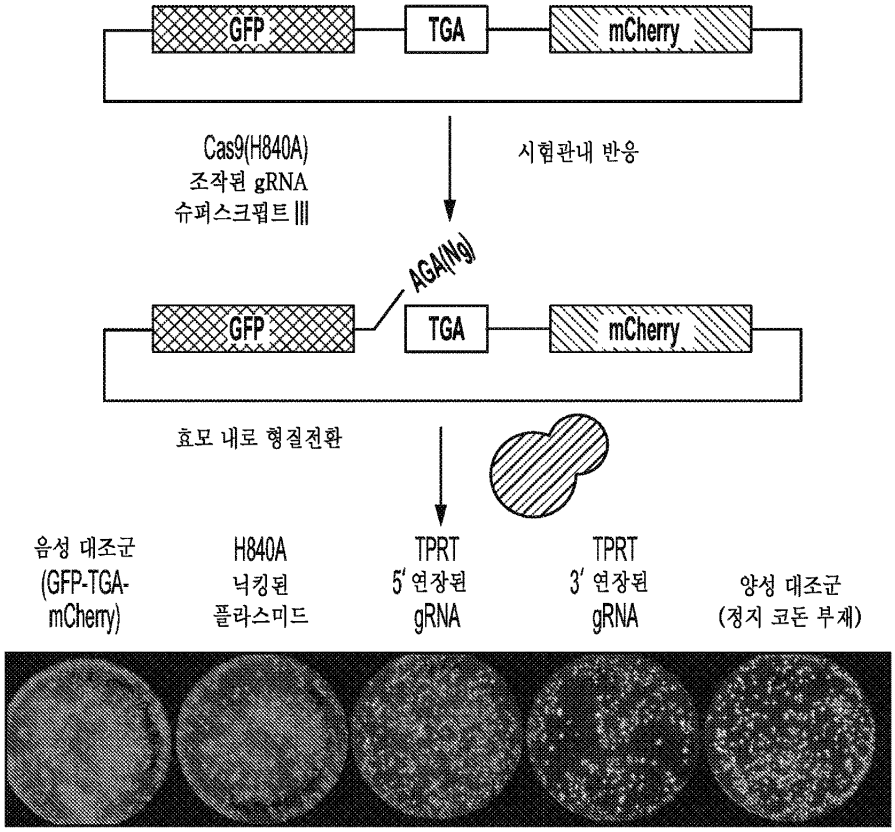
도면9c



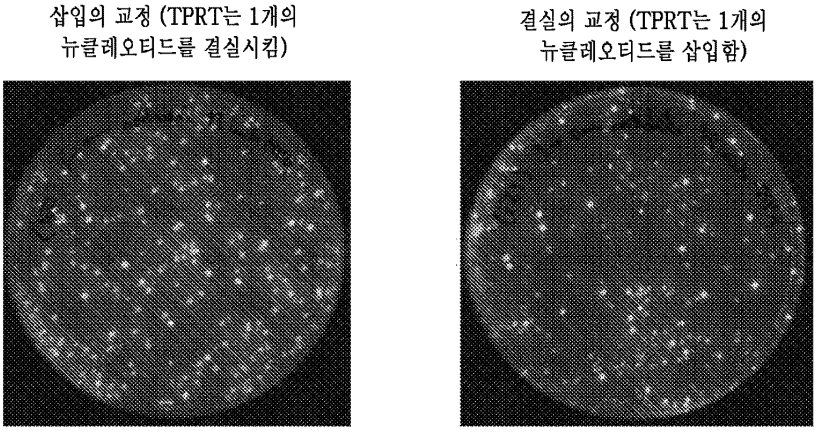
도면9d



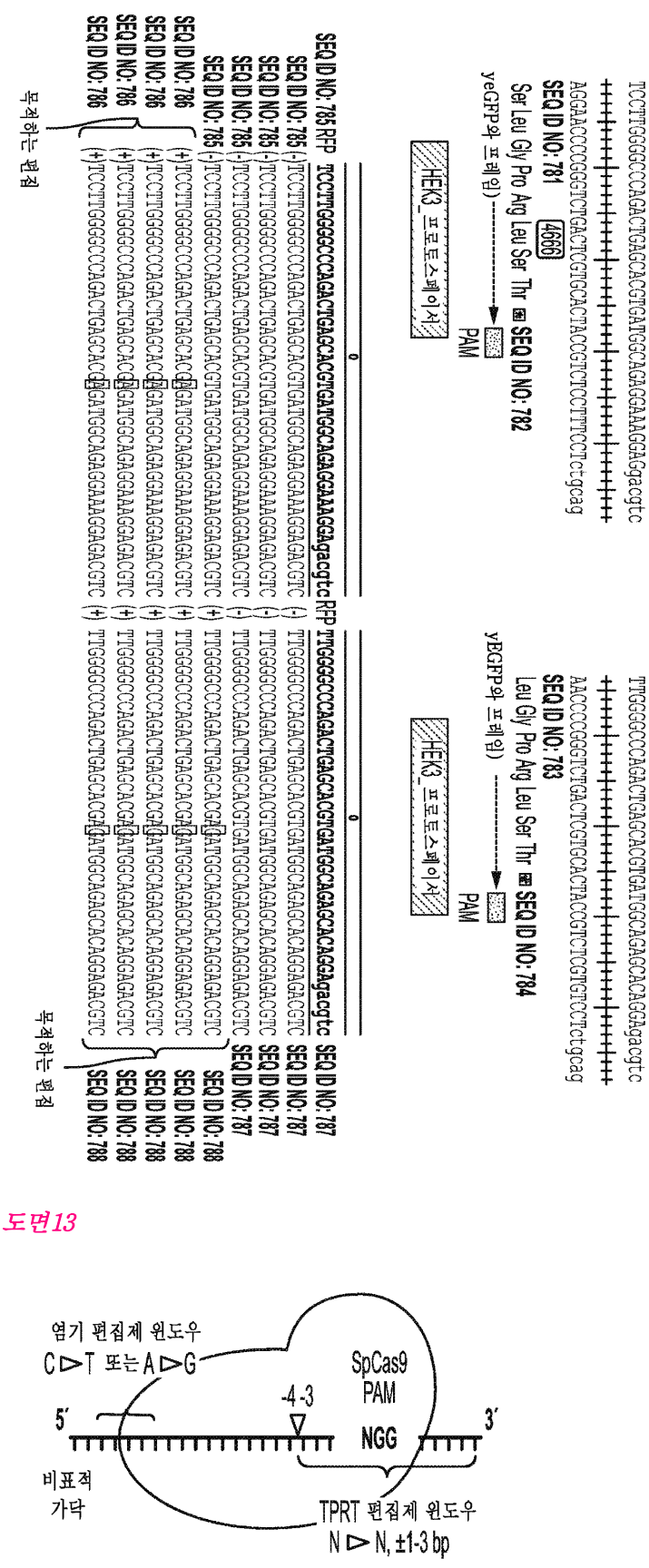
도면10


GFP
TGA
mCherry
Cas9(H840A)
조작된 gRNA
슈퍼스크립트 III
시험관내 반응
AGA(Ng)
GFP
TGA
mCherry
효모 내로 형질전환
음성 대조군
(GFP-TGA-
mCherry)
H840A
닉킹된
플라스미드
TPRT
5′ 연장된
gRNA
TPRT
3′ 연장된
gRNA
양성 대조군
(정지 코돈 부재)

도면11


삽입의 교정 (TPRT는 1개의
뉴클레오티드를 결실시킴)
결실의 교정 (TPRT는 1개의
뉴클레오티드를 삽입함)

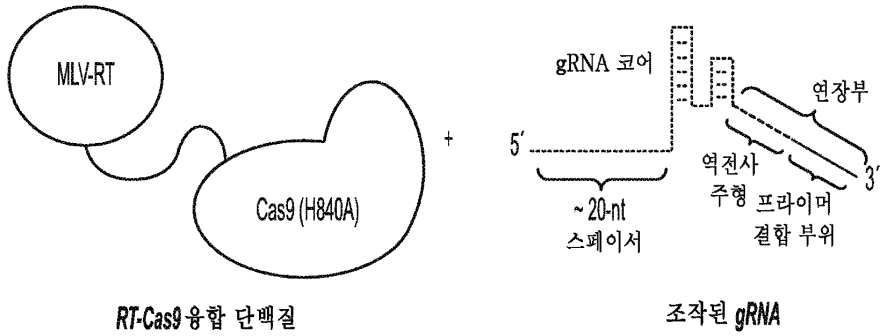
도면12


5′ -gRNA 연장부
SEQ ID NO: 785
TCCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAgacgtc
AGGAACCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCTctgcag
SEQ ID NO: 781
4666
Ser Leu Gly Pro Arg Leu Ser Thr
SEQ ID NO: 782
yeGFP와 프레임)
PAM
HEK3_ 프로토스페이서
3′ -gRNA 연장부
SEQ ID NO: 787
TTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAgacgtc
AACCCCGGGTCTGACTCGTGCACTACCGTCTCCTTTCCTctgcag
SEQ ID NO: 783
Leu Gly Pro Arg Leu Ser Thr
SEQ ID NO: 784
yEGFP와 프레임)
PAM
HEK3_ 프로토스페이서
SEQ ID NO: 785 RFP
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 785 (-)
SEQ ID NO: 786
SEQ ID NO: 786
SEQ ID NO: 786
SEQ ID NO: 786
SEQ ID NO: 786
목적하는 편집
SEQ ID NO: 787 RFP
SEQ ID NO: 787
SEQ ID NO: 787
SEQ ID NO: 787
SEQ ID NO: 787
SEQ ID NO: 787
SEQ ID NO: 788
SEQ ID NO: 788
SEQ ID NO: 788
SEQ ID NO: 788
SEQ ID NO: 788
목적하는 편집

도면13

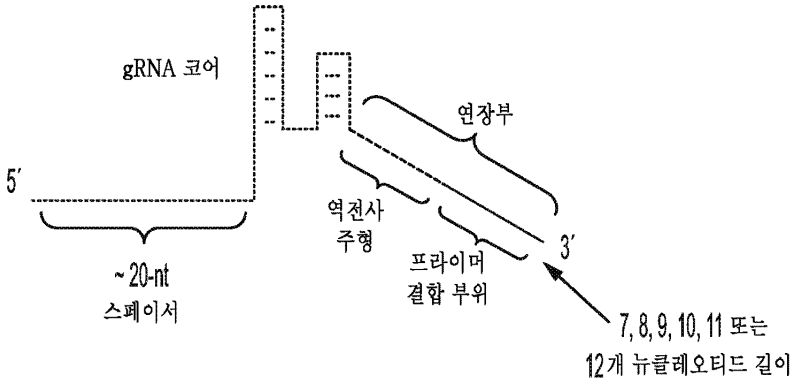
염기 편집제 윈도우
C▷T 또는 A▷G
SpCas9
PAM
-4 -3
5′
3′
NGG
비표적
가닥
TPRT 편집제 윈도우
N▷N, ±1-3 bp

도면14

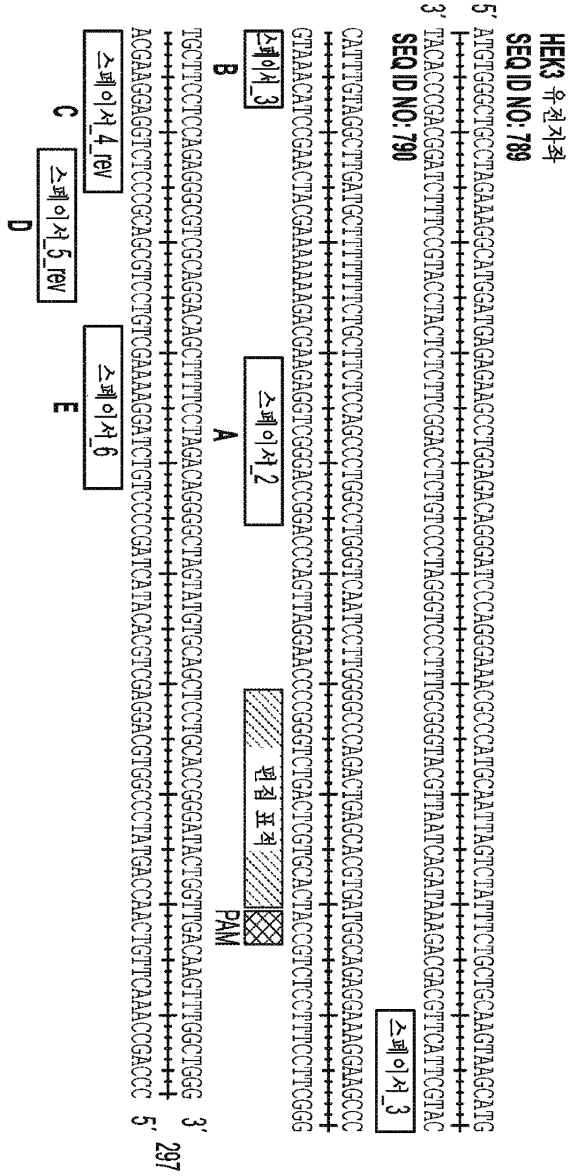


RT-Cas9 융합 단백질　　조작된 gRNA

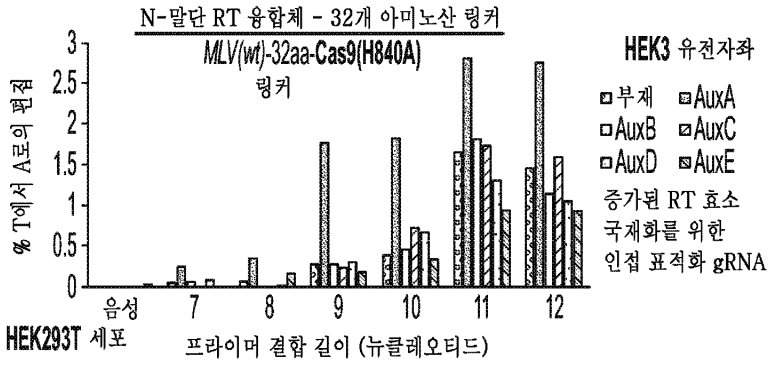
도면15



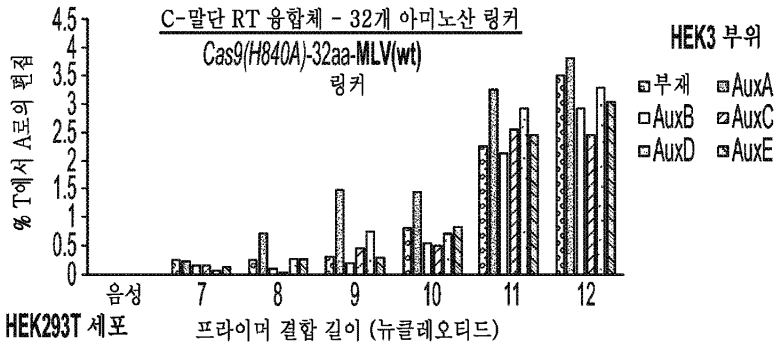
도면16


HEK3 유전자좌
SEQ ID NO: 789
SEQ ID NO: 790
스페이서_3
스페이서_3
스페이서_2
편집 표적
PAM
스페이서_4_rev
스페이서_5_rev
스페이서_6
A
B
C
D
E
297

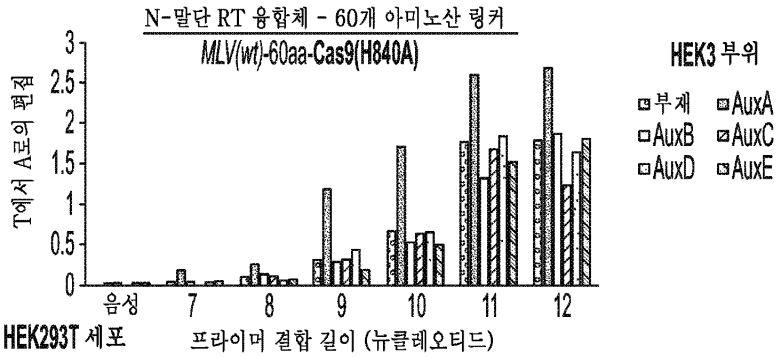
도면17a


N-말단 RT 융합체 - 32개 아미노산 링커
MLV(wt)-32aa-Cas9(H840A)
링커
HEK3 유전자좌
부재
AuxA
AuxB
AuxC
AuxD
AuxE
증가된 RT 효소
국재화를 위한
인접 표적화 gRNA
편집 스페이서 %
3
2.5
2
1.5
1
0.5
0
음성
7
8
9
10
11
12
HEK293T 세포
프라이머 결합 길이 (뉴클레오티드)

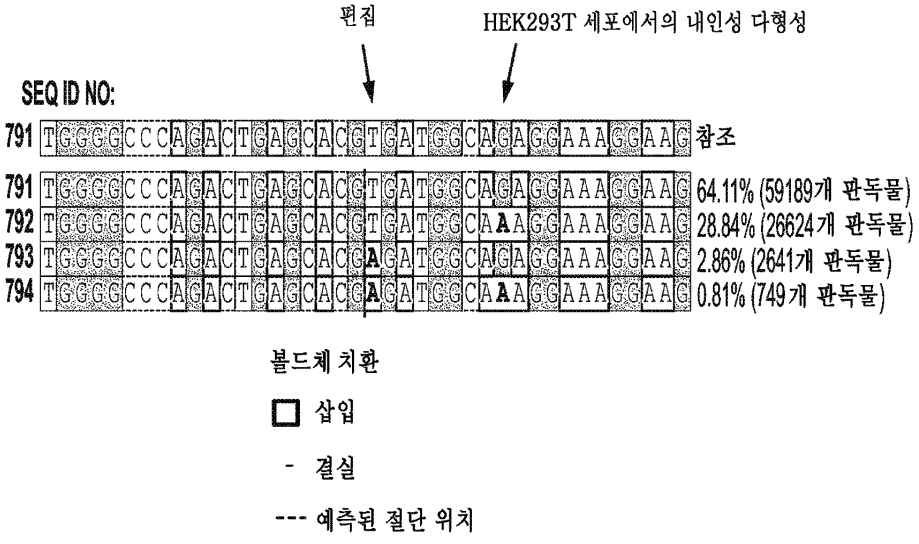
도면17b



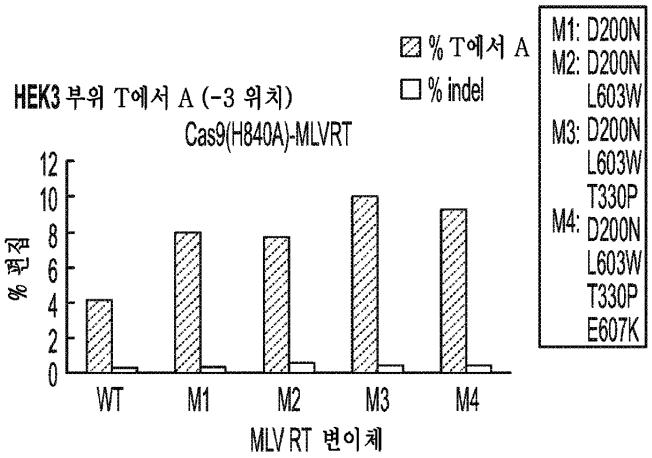
도면17c



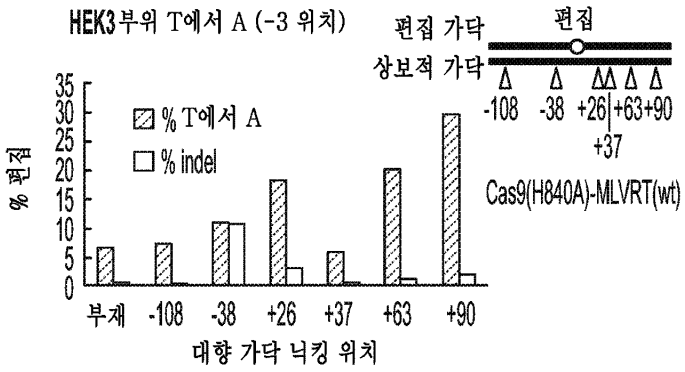
도면18



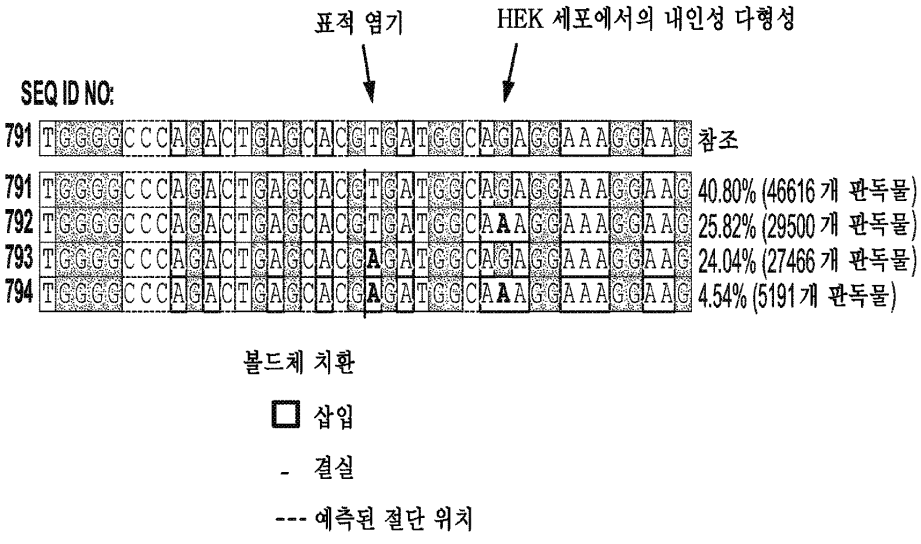
도면19



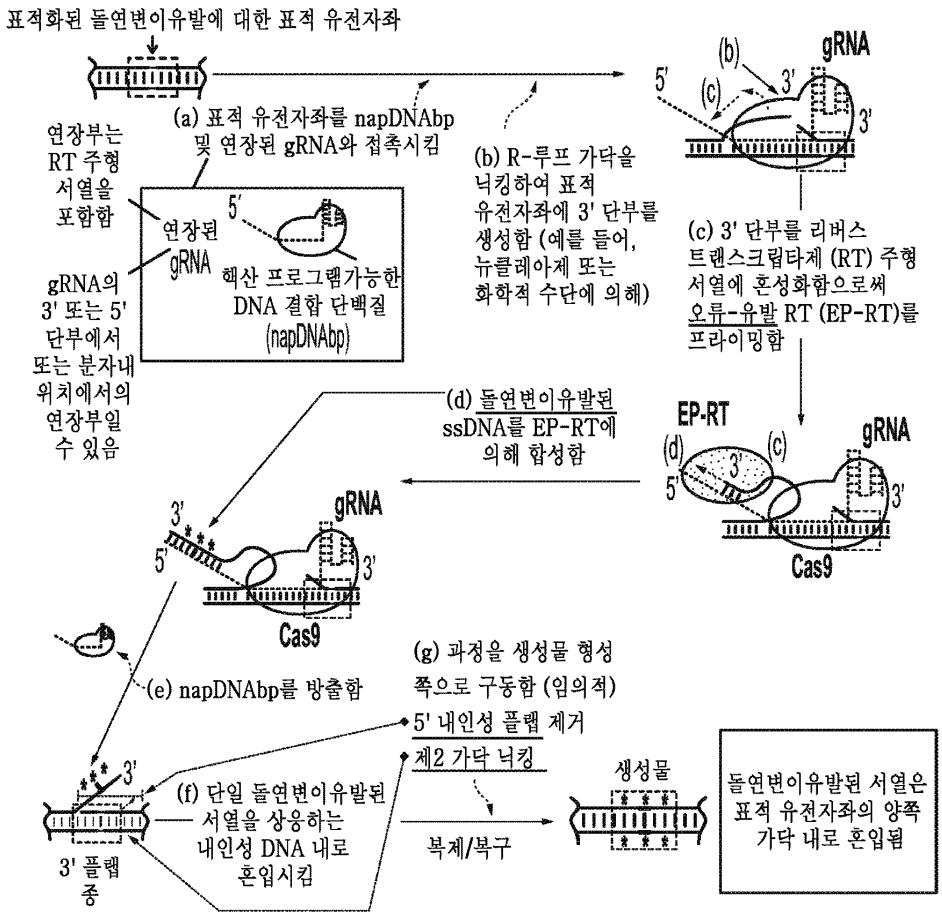
도면20



도면21

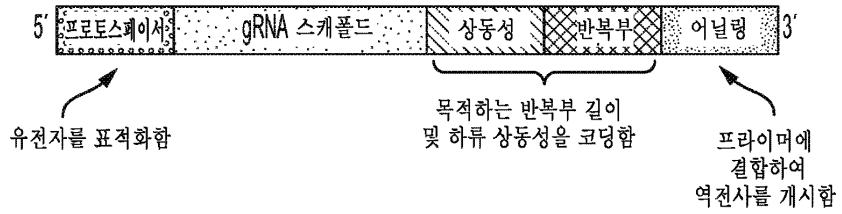


도면22

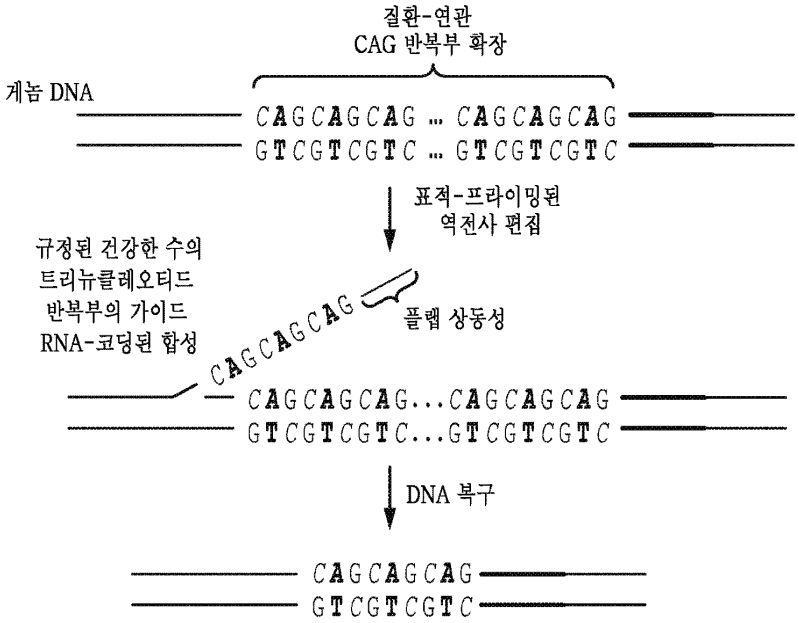

표적화된 돌연변이유발에 대한 표적 유전자좌
(a) 표적 유전자좌를 napDNAbp 및 연장된 gRNA와 접촉시킴
연장부는 RT 주형 서열을 포함함
gRNA의 3' 또는 5' 단부에서 또는 분자내 위치에서의 연장부일 수 있음
연장된 gRNA
핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)
(b) R-루프 가닥을 닉킹하여 표적 유전자좌에 3' 단부를 생성함 (예를 들어, 뉴클레아제 또는 화학적 수단에 의해)
(c) 3' 단부를 리버스 트랜스크립타제 (RT) 주형 서열에 혼성화함으로써 오류-유발 RT (EP-RT)를 프라이밍함
(d) 돌연변이유발된 ssDNA를 EP-RT에 의해 합성함
(e) napDNAbp를 방출함
(f) 단일 돌연변이유발된 서열을 상응하는 내인성 DNA 내로 혼입시킴
3' 플랩 종
(g) 과정을 생성물 형성 쪽으로 구동함 (임의적)
5' 내인성 플랩 제거
제2 가닥 닉킹
복제/복구
생성물
돌연변이유발된 서열은 표적 유전자좌의 양쪽 가닥 내로 혼입됨
gRNA
Cas9
EP-RT

도면23

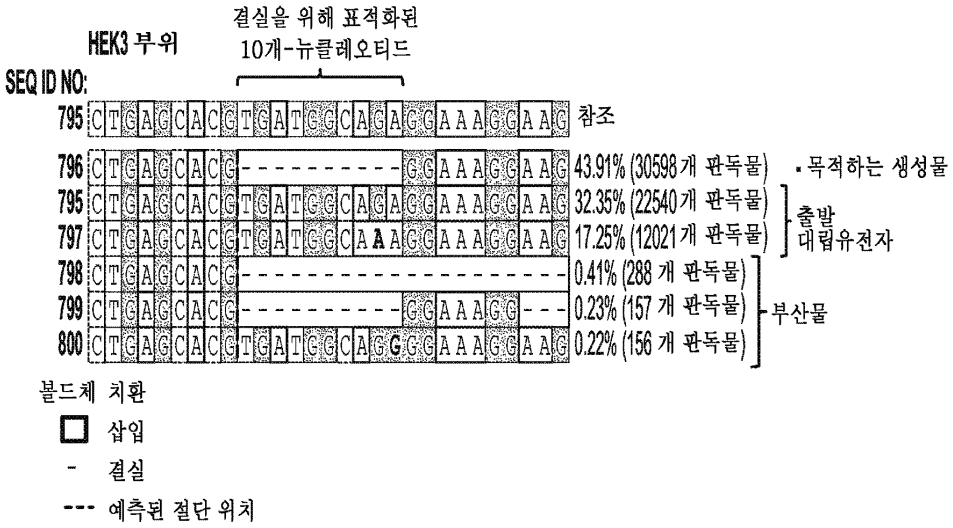
트리뉴클레오티드 반복 서열을 축소시키기 위한 gRNA 설계



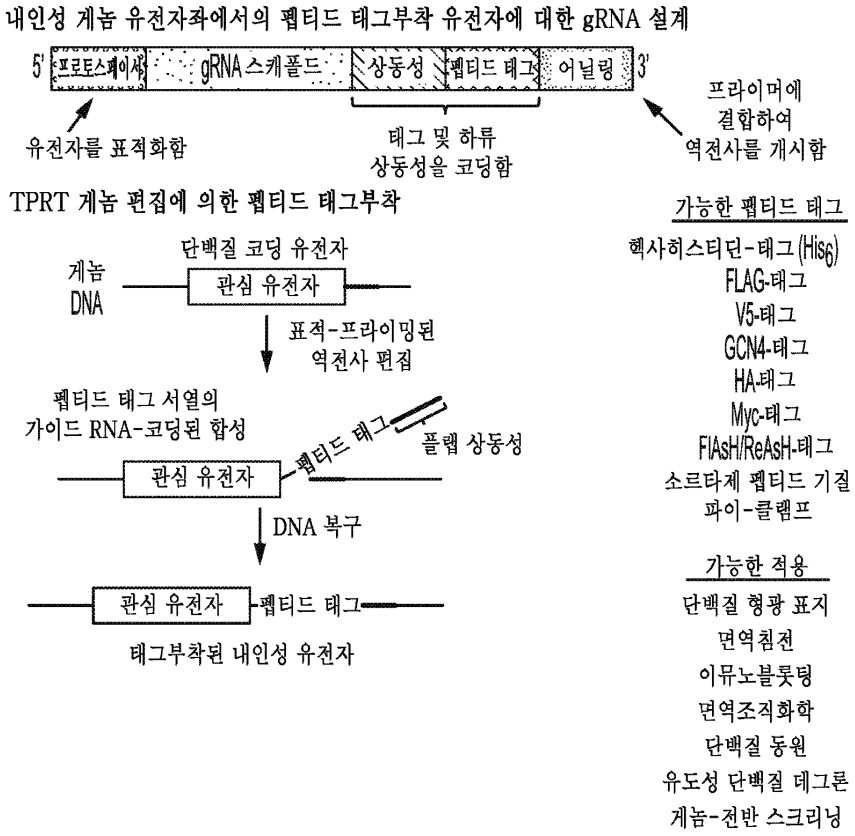
TPRT 게놈 편집에 의한 트리뉴클레오티드 반복부 축소



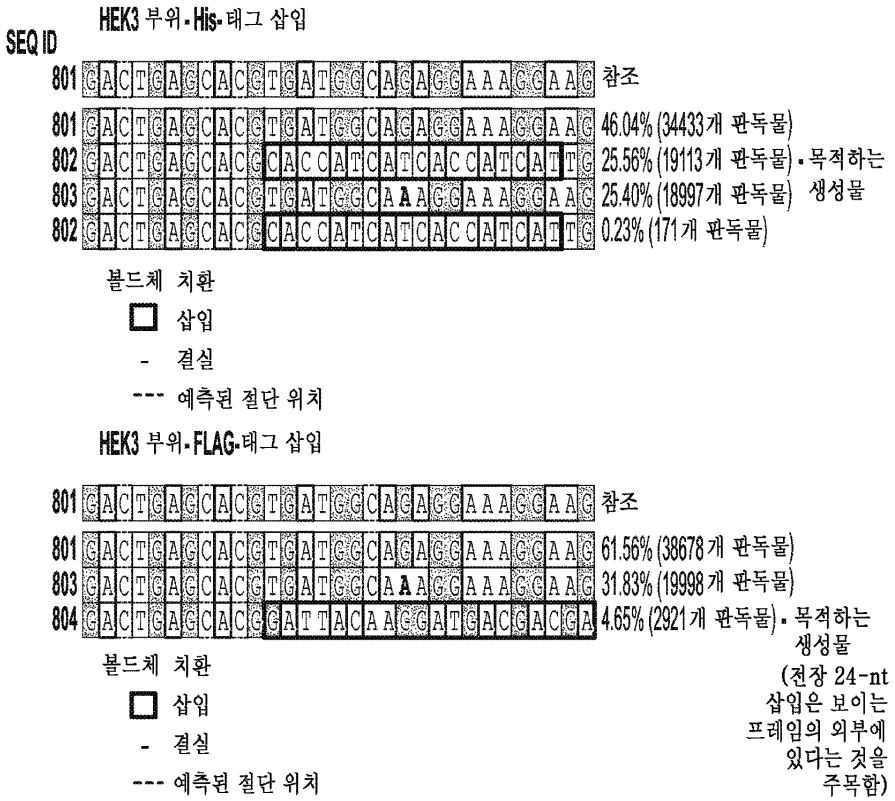
도면24



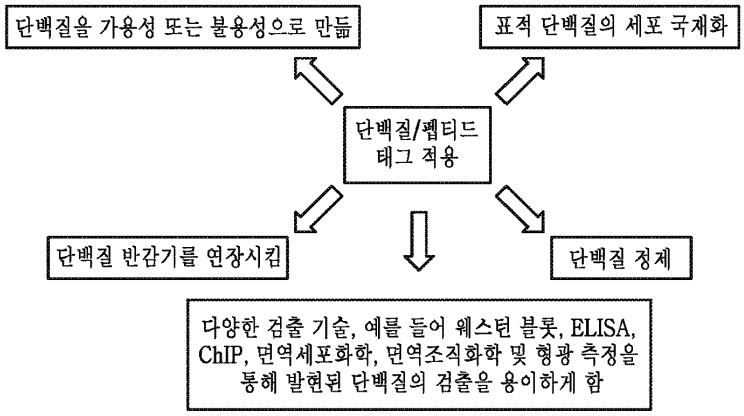
도면25


내인성 게놈 유전자좌에서의 펩티드 태그부착 유전자에 대한 gRNA 설계
5'
프로토스페이서
gRNA 스캐폴드
상동성
펩티드 태그
어닐링
3'
유전자를 표적화함
태그 및 하류 상동성을 코딩함
프라이머에 결합하여 역전사를 개시함
TPRT 게놈 편집에 의한 펩티드 태그부착
단백질 코딩 유전자
게놈 DNA
관심 유전자
표적-프라이밍된 역전사 편집
펩티드 태그 서열의 가이드 RNA-코딩된 합성
펩티드 태그
플랩 상동성
관심 유전자
DNA 복구
관심 유전자
펩티드 태그
태그부착된 내인성 유전자
가능한 펩티드 태그
헥사히스티딘-태그 (His6)
FLAG-태그
V5-태그
GCN4-태그
HA-태그
Myc-태그
FlAsH/ReAsH-태그
소르타제 펩티드 기질
파이-클램프
가능한 적용
단백질 형광 표지
면역침전
이뮤노블롯팅
면역조직화학
단백질 동원
유도성 단백질 데그론
게놈-전반 스크리닝

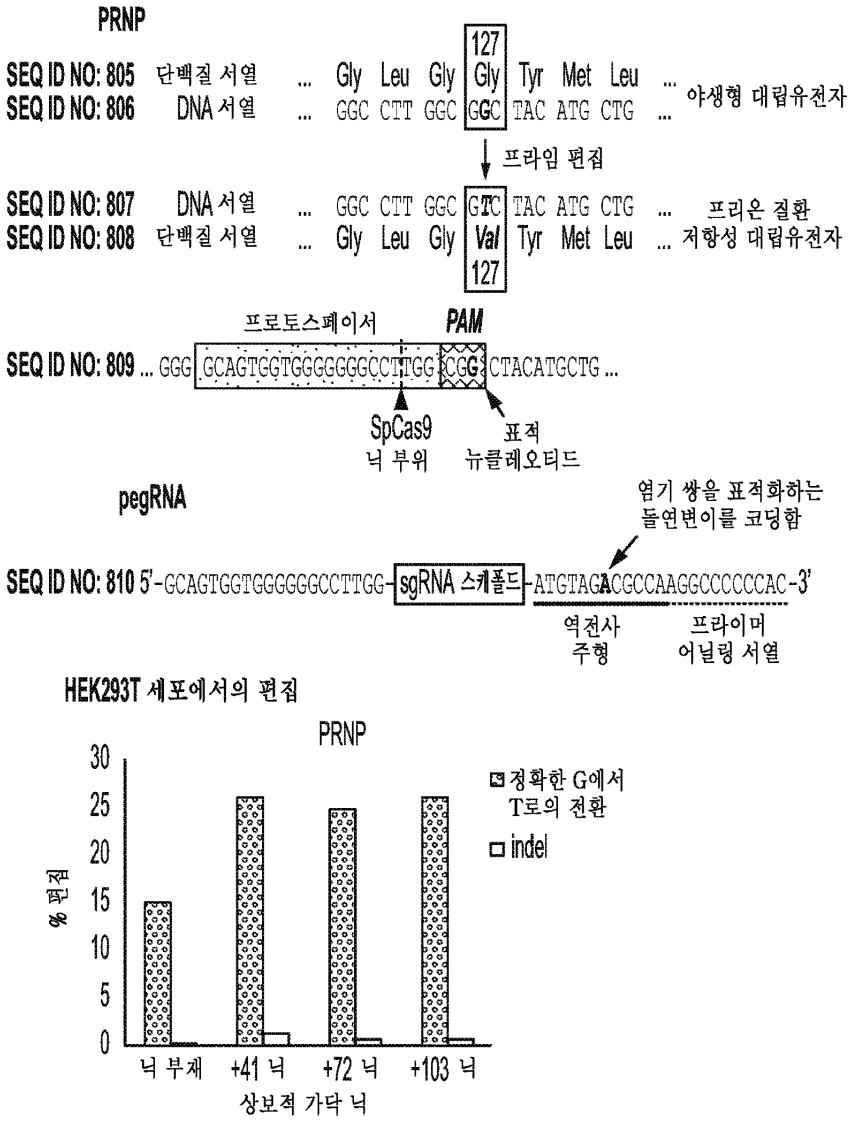
도면26a


HEK3 부위-His-태그 삽입
SEQ ID
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 참조
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 46.04% (34433개 판독물)
802 GACTGAGCACGCACCATCATCACCATCATTG 25.56% (19113개 판독물) - 목적하는
803 GACTGAGCACGTGATGGCAAAGGAAAGGAAG 25.40% (18997개 판독물) 생성물
802 GACTGAGCACGCACCATCATCACCATCATTG 0.23% (171개 판독물)
볼드체 치환
삽입
- 결실
--- 예측된 절단 위치
HEK3 부위-FLAG-태그 삽입
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 참조
801 GACTGAGCACGTGATGGCAGAGGAAAGGAAG 61.56% (38678개 판독물)
803 GACTGAGCACGTGATGGCAAAGGAAAGGAAG 31.83% (19998개 판독물)
804 GACTGAGCACGGATTACAAGGATGACGACGA 4.65% (2921개 판독물) - 목적하는 생성물
(전장 24-nt 삽입은 보이는 프레임의 외부에 있다는 것을 주목함)
볼드체 치환
삽입
- 결실
--- 예측된 절단 위치

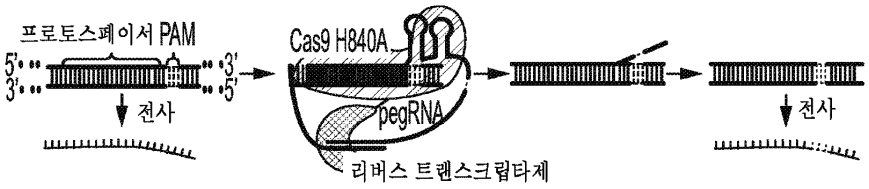
도면26b


단백질을 가용성 또는 불용성으로 만듦
표적 단백질의 세포 국제화
단백질/펩티드 태그 적용
단백질 반감기를 연장시킴
단백질 정제
다양한 검출 기술, 예를 들어 웨스턴 블롯, ELISA, ChIP, 면역세포화학, 면역조직화학 및 형광 측정을 통해 발현된 단백질의 검출을 용이하게 함

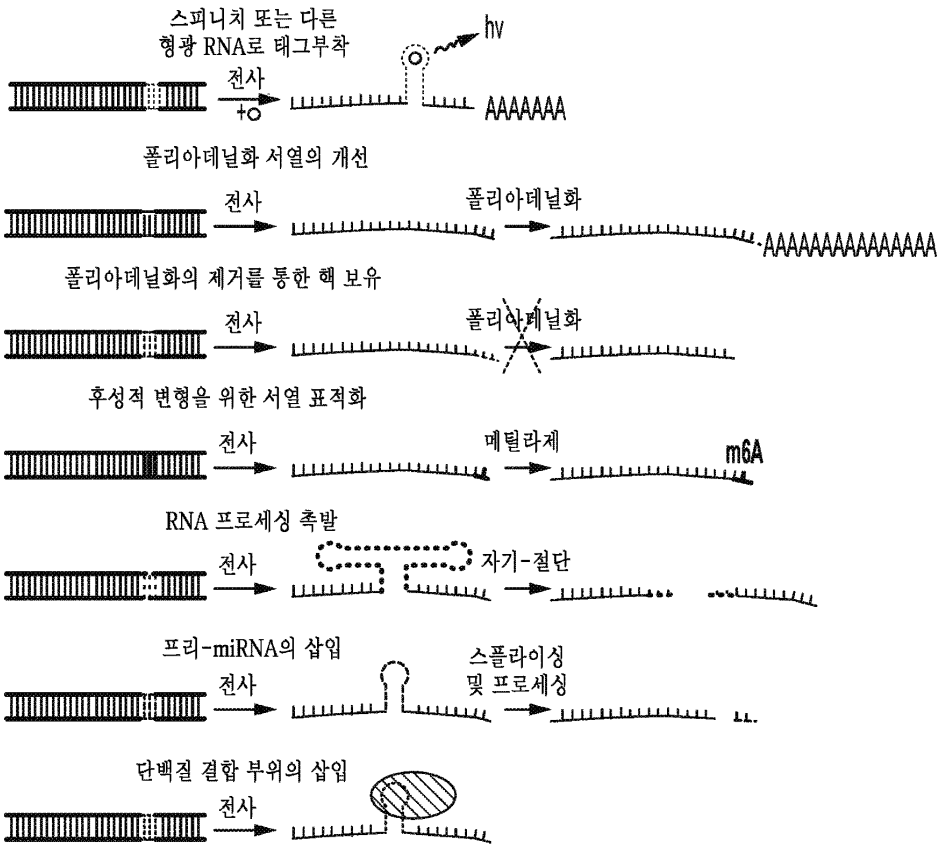
도면27


PRNP
SEQ ID NO: 805 단백질 서열 ... Gly Leu Gly Gly Tyr Met Leu ...
127
SEQ ID NO: 806 DNA 서열 ... GGC CTT GGC GGC TAC ATG CTG ... 야생형 대립유전자
프라임 편집
SEQ ID NO: 807 DNA 서열 ... GGC CTT GGC GTC TAC ATG CTG ... 프리온 질환
SEQ ID NO: 808 단백질 서열 ... Gly Leu Gly Val Tyr Met Leu ... 저항성 대립유전자
127
프로토스페이서
PAM
SEQ ID NO: 809 ... GGG GCAGTGGTGGGGGGGCCTTGG CGG CTACATGCTG ...
SpCas9 닉 부위
표적 뉴클레오티드
pegRNA
염기 쌍을 표적화하는 돌연변이를 코딩함
SEQ ID NO: 810 5'-GCAGTGGTGGGGGGCCTTGG-sgRNA 스케폴드-ATGTAGACGCCAAGGCCCCCCAC-3'
역전사 주형
프라이머 어닐링 서열
HEK293T 세포에서의 편집
PRNP
% 편집
30
25
20
15
10
5
0
정확한 G에서 T로의 전환
indel
닉 부재
+41 닉
+72 닉
+103 닉
상보적 가닥 닉

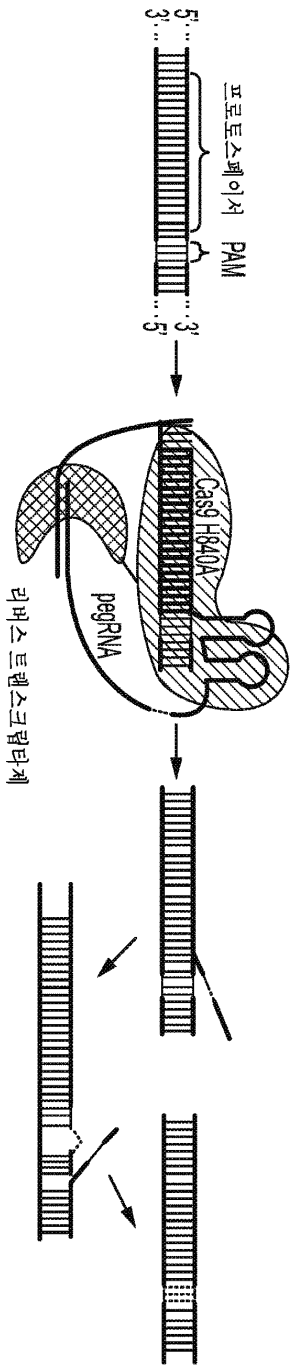
도면28a


프로토스페이서 PAM
5'
3'
3'
5'
전사
Cas9 H840A
pegRNA
리버스 트랜스크립타제
전사

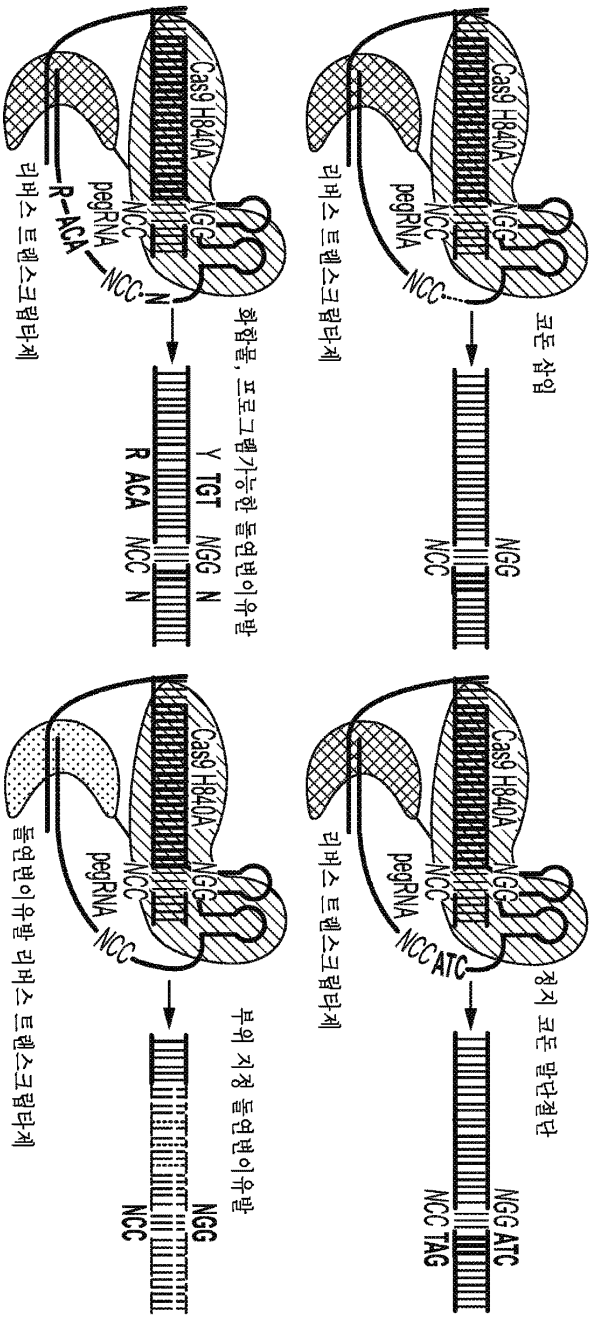
도면28b


스피니치 또는 다른
형광 RNA로 태그부착
전사
hv
AAAAAAA
폴리아데닐화 서열의 개선
전사
폴리아데닐화
AAAAAAAAAAAAAAA
폴리아데닐화의 제거를 통한 핵 보유
전사
폴리아데닐화
후성적 변형을 위한 서열 표적화
전사
메틸라제
m6A
RNA 프로세싱 촉발
전사
자기-절단
프리-miRNA의 삽입
전사
스플라이싱
및 프로세싱
단백질 결합 부위의 삽입
전사

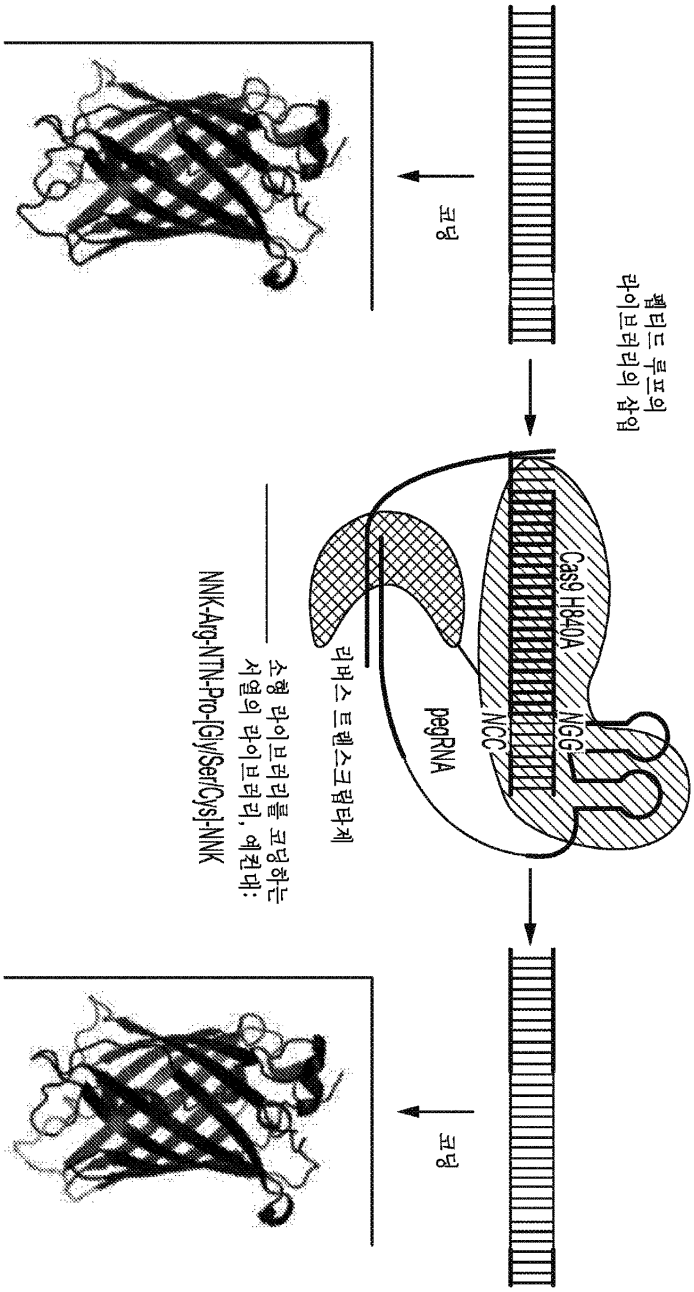
도면29a



도면29b


Cas9 H840A
pegRNA
NCC
NGG
리버스 트랜스크립타제
R-ACA
N
코돈 삽입
화합물, 프로그램가능한 돌연변이유발
Y TGT NGG N
R ACA NCC N
NGG
NCC
돌연변이유발 리버스 트랜스크립타제
부위 지정 돌연변이유발
NCC ATC
정지 코돈 말단절단
NCC TAG
NGG ATC

도면29c

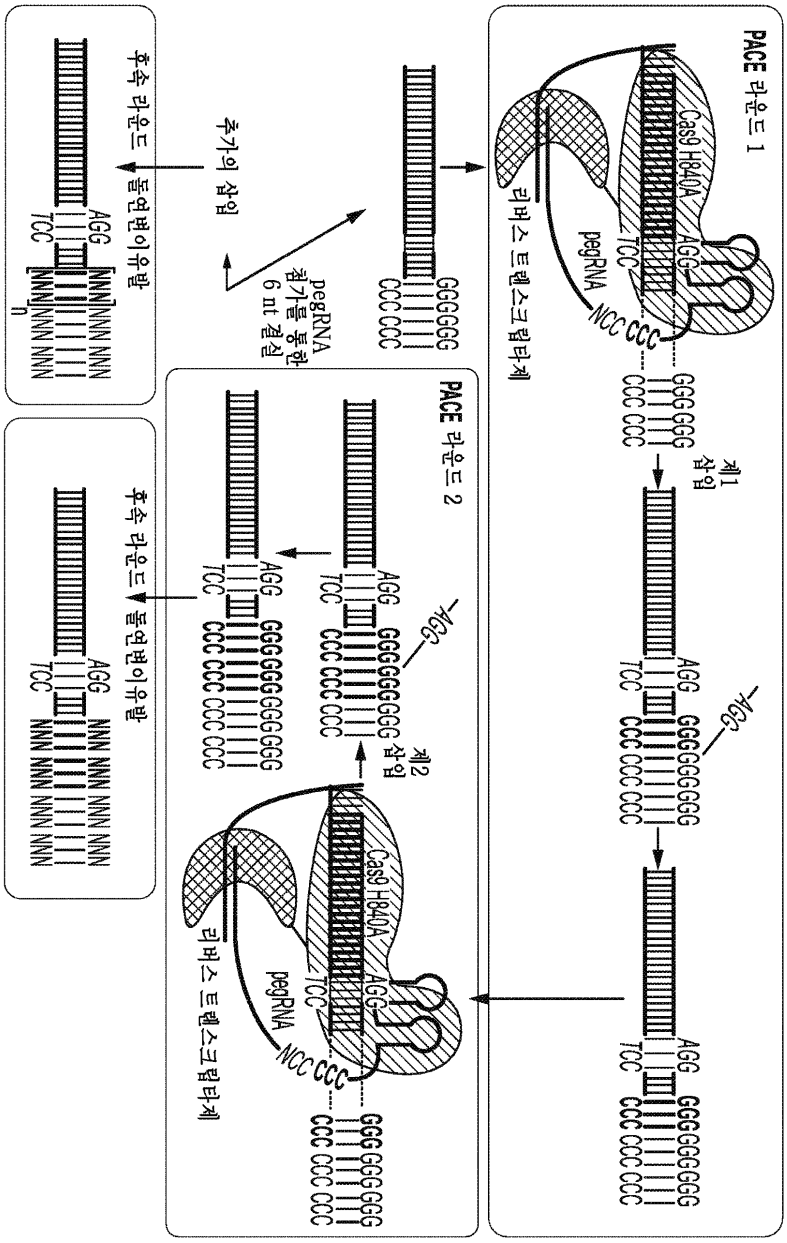


도면29d

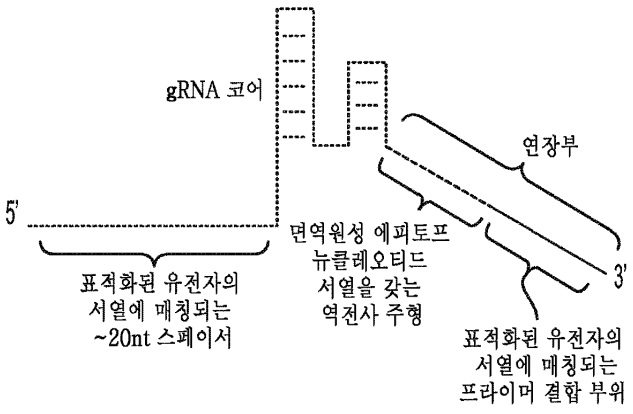
프로그램가능한 코돈 결실 및 말단절단

SEQ ID NO: 811	**GFP**	NFKIRHNIEDGSVQLADHYQQNTPI
SEQ ID NO: 812	del5	NFKI-HNIEDGSVQLADHYQQNTPI
SEQ ID NO: 813	del6	NFKIR-NIEDGSVQLADHYQQNTPI
SEQ ID NO: 814	del8	NFKIRHN-EDGSVQLADHYQQNTPI
SEQ ID NO: 815	del 8-13	NFKIRHN------QLADHYQQNTPI
SEQ ID NO: 816	del 1-5	-----HNIEDGSVQLADHYQQNTPI
SEQ ID NO: 817	del 15-25	NFKIRHNIEDGSVQ----------

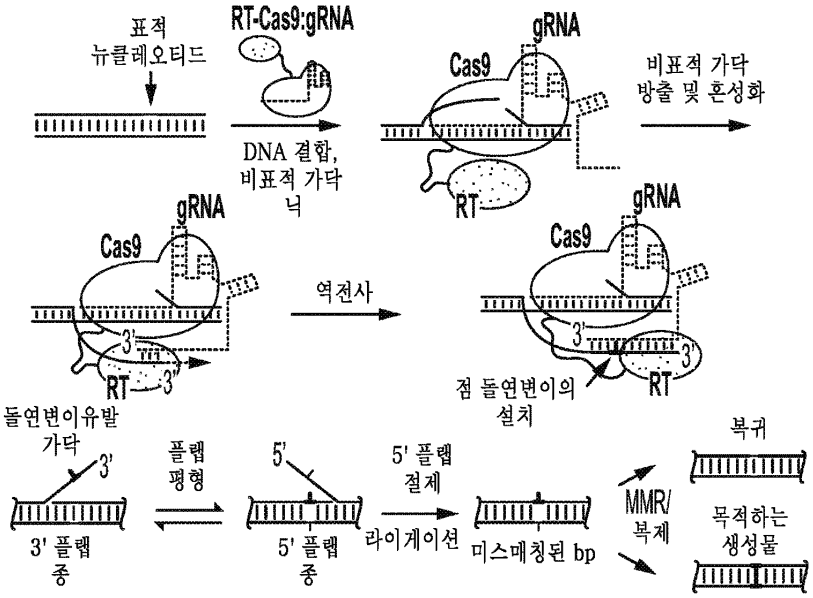
도면30


PACE 라운드 1
Cas9 H840A
pegRNA
리버스 트랜스크립타제
제1 삽입
PACE 라운드 2
제2 삽입
추가의 삽입
pegRNA 침가를 통한 6 nt 결실
후속 라운드 돌연변이유발

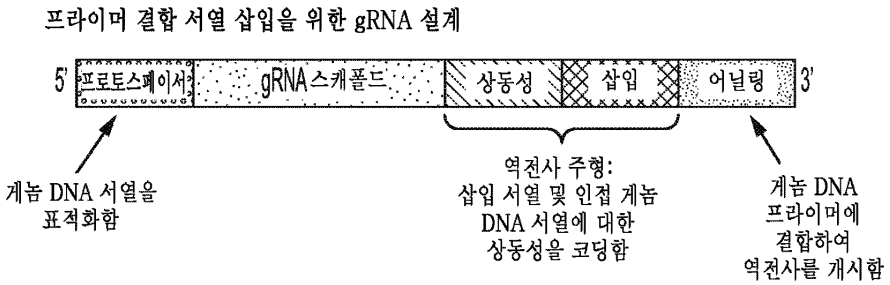
도면31

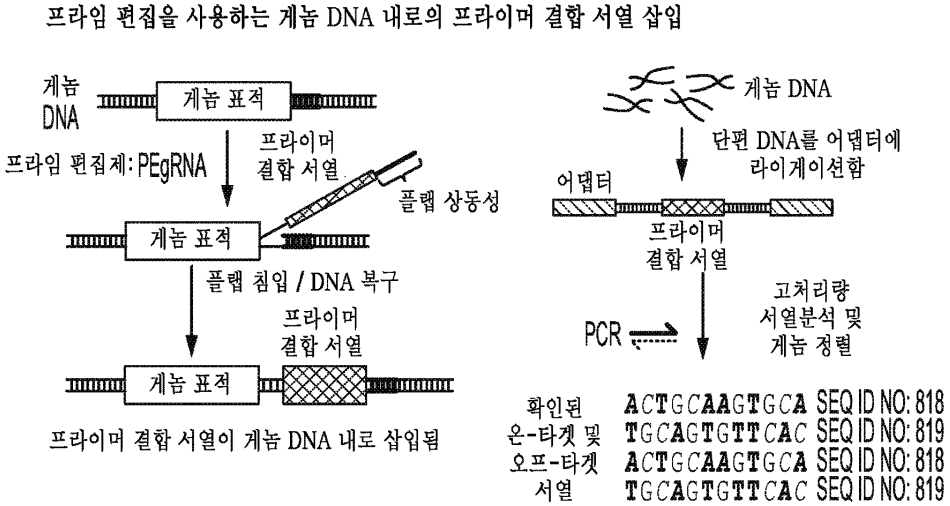

gRNA 코어
연장부
5'
3'
면역원성 에피토프 뉴클레오티드 서열을 갖는 역전사 주형
표적화된 유전자의 서열에 매칭되는 ~20nt 스페이서
표적화된 유전자의 서열에 매칭되는 프라이머 결합 부위

도면32

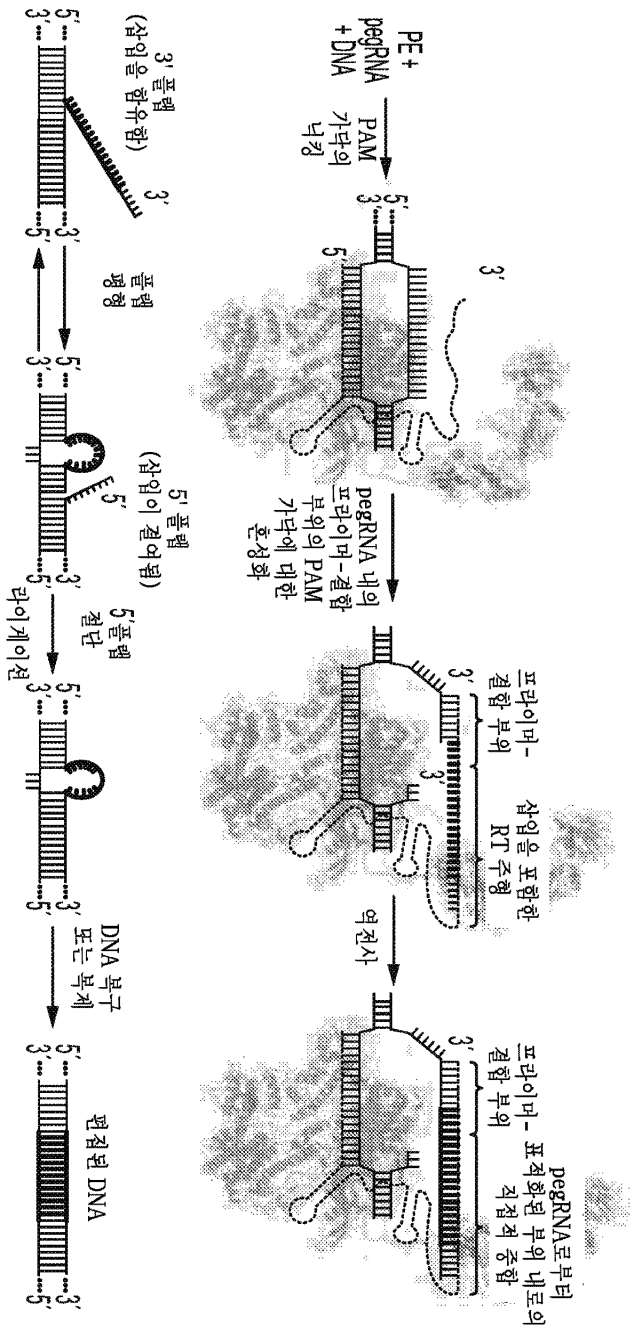

표적
뉴클레오티드
RT-Cas9:gRNA
DNA 결합,
비표적 가닥
닉
gRNA
Cas9
RT
비표적 가닥
방출 및 혼성화
역전사
3'
3''
돌연변이유발
가닥
점 돌연변이의
설치
플랩
평형
5'
5' 플랩
절제
라이게이션
미스매칭된 bp
MMR/
복제
복귀
목적하는
생성물
3' 플랩
종
5' 플랩
종

도면33

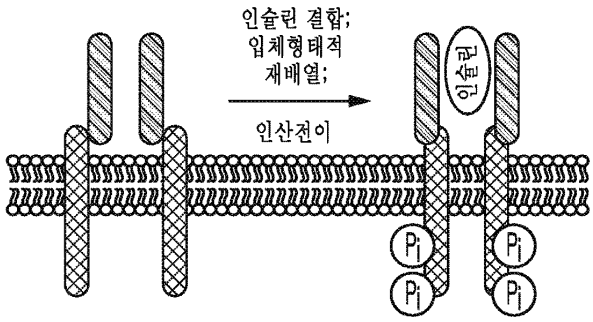

프라이머 결합 서열 삽입을 위한 gRNA 설계
5'
프로토스페이서
gRNA 스캐폴드
상동성
삽입
어닐링
3'
게놈 DNA 서열을
표적화함
역전사 주형:
삽입 서열 및 인접 게놈
DNA 서열에 대한
상동성을 코딩함
게놈 DNA
프라이머에
결합하여
역전사를 개시함


프라임 편집을 사용하는 게놈 DNA 내로의 프라이머 결합 서열 삽입
게놈
DNA
게놈 표적
프라임 편집제: PEgRNA
프라이머
결합 서열
플랩 상동성
플랩 침입 / DNA 복구
프라이머 결합 서열이 게놈 DNA 내로 삽입됨
게놈 DNA
단편 DNA를 어댑터에
라이게이션함
어댑터
PCR
고처리량
서열분석 및
게놈 정렬
확인된
온-타겟 및
오프-타겟
서열
ACTGCAAGTGCA SEQ ID NO: 818
TGCAGTGTTCAC SEQ ID NO: 819
ACTGCAAGTGCA SEQ ID NO: 818
TGCAGTGTTCAC SEQ ID NO: 819

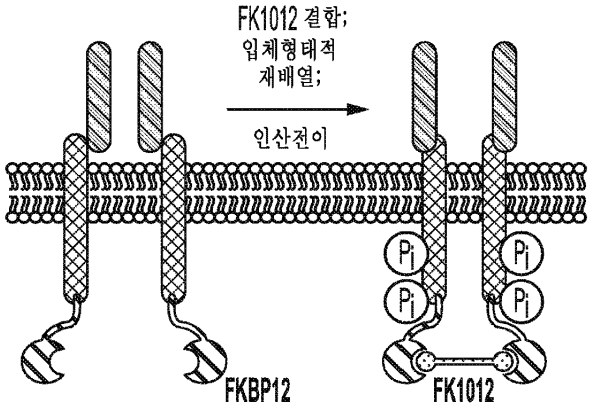
도면34


PE+
pegRNA
+DNA
PAM
가닥의
닉킹
pegRNA 내의
프라이머-결합
부위의 PAM
가닥에 대한
혼성화
프라이머-
결합 부위
삼입을 포함한
RT 주형
역전사
프라이머-
결합 부위
표적화된 부위 내로의
pegRNA로부터
직접적 중합
3' 플랩
(삼입을 함유함)
플랩
평형
5' 플랩
(삼입이 결여됨)
5'플랩
절단
라이게이션
DNA 복구
또는 복제
편집된 DNA

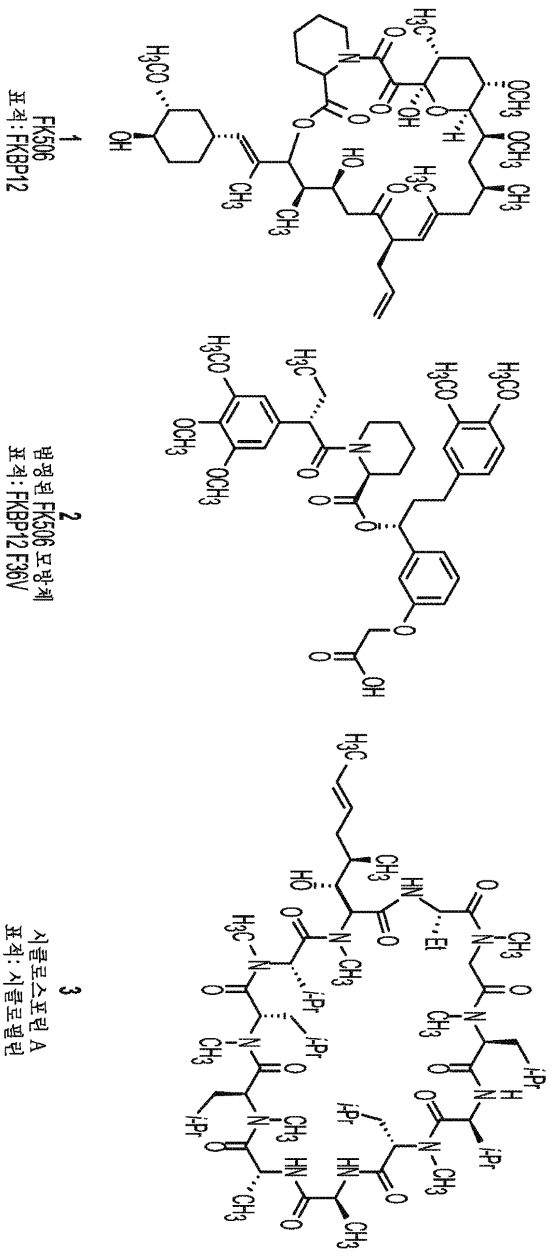
도면35a


인슐린 결합;
입체형태적
재배열;
인산전이
인슐린
Pi
Pi
Pi
Pi

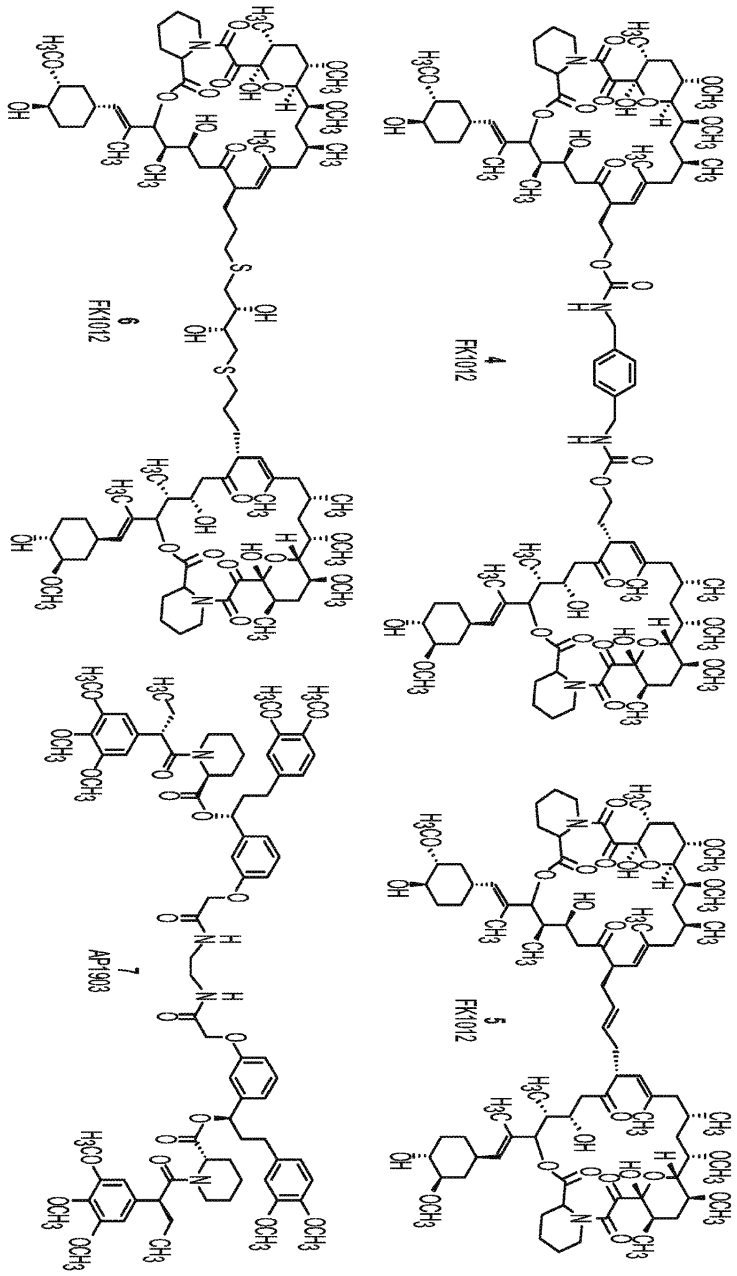
도면35b


FK1012 결합;
입체형태적
재배열;
인산전이
Pi
Pi
Pi
Pi
FKBP12
FK1012

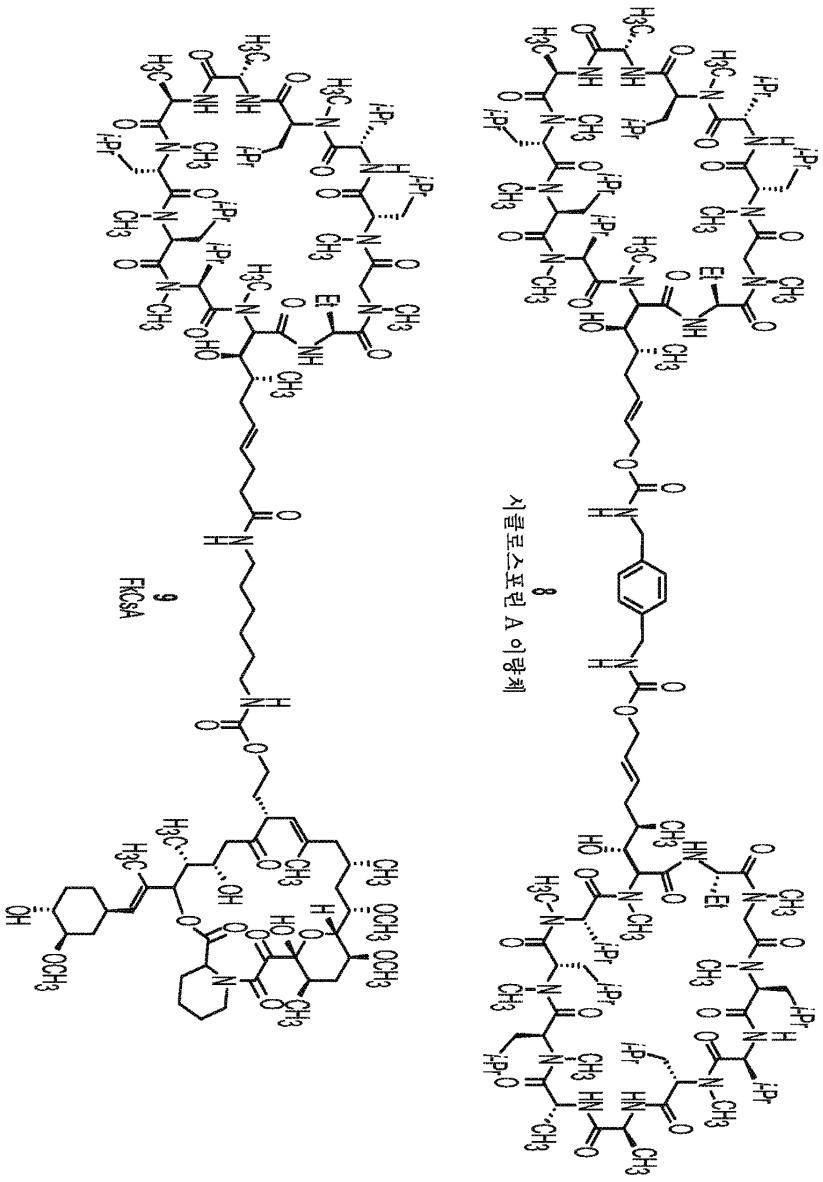
도면36


1
FK506
표적: FKBP12
2
합성된 FK506 모방체
표적: FKBP12 F36V
3
시클로스포린 A
표적: 시클로필린

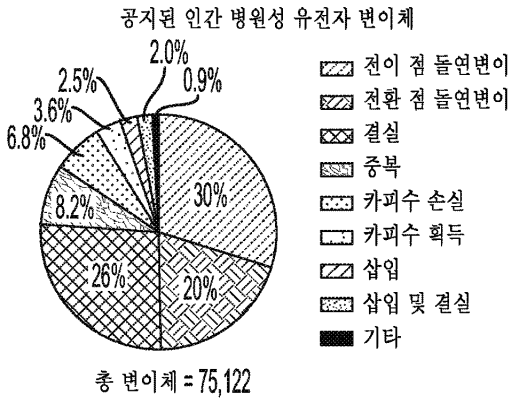
도면37a



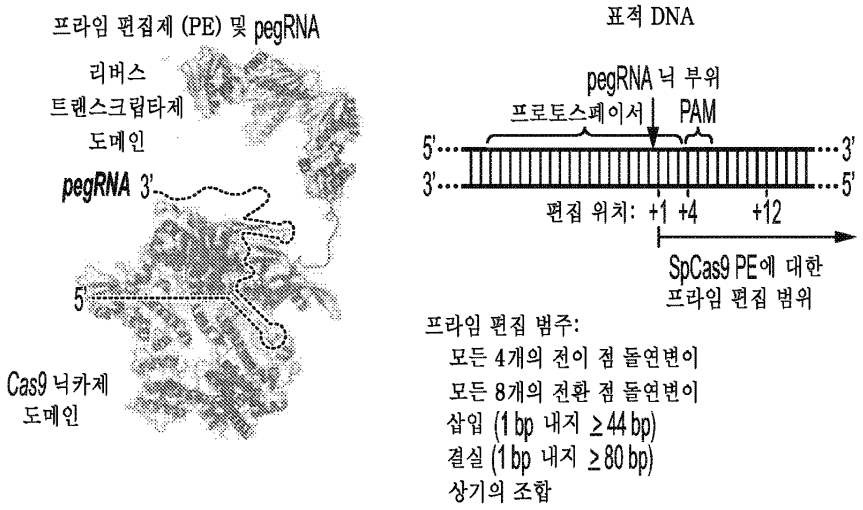
도면37b


9
FKCsA
8
시클로스포린 A 이량체

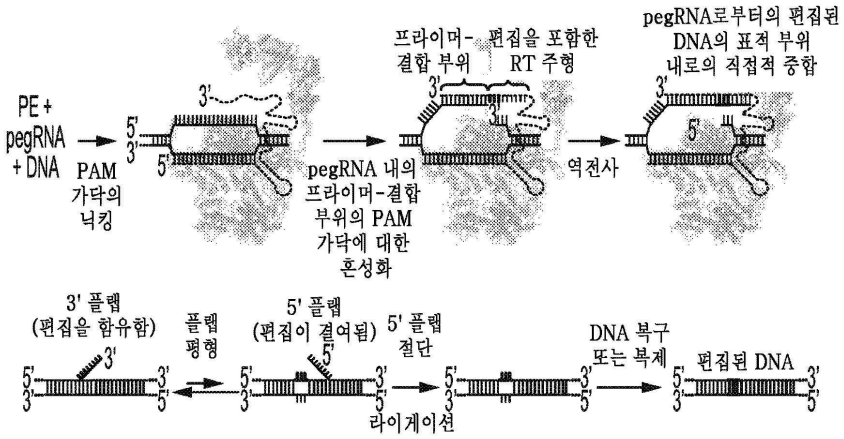
도면38a


공지된 인간 병원성 유전자 변이체
2.0%
0.9%
2.5%
3.6%
6.8%
8.2%
30%
26%
20%
전이 점 돌연변이
전환 점 돌연변이
결실
중복
카피수 손실
카피수 획득
삽입
삽입 및 결실
기타
총 변이체 = 75,122

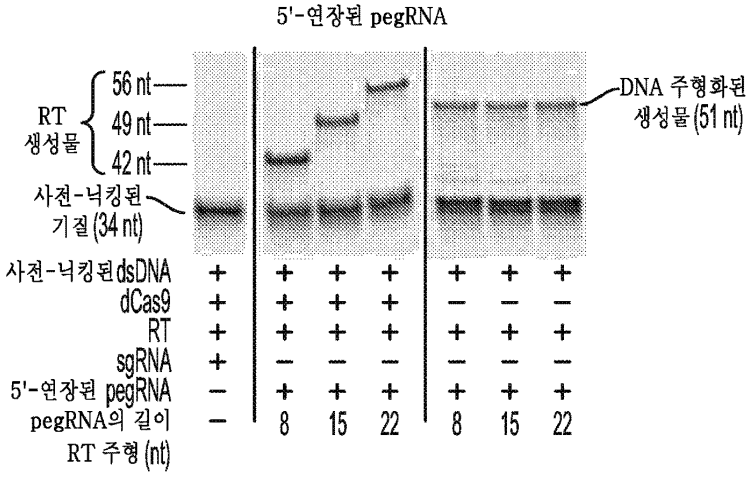
도면38b



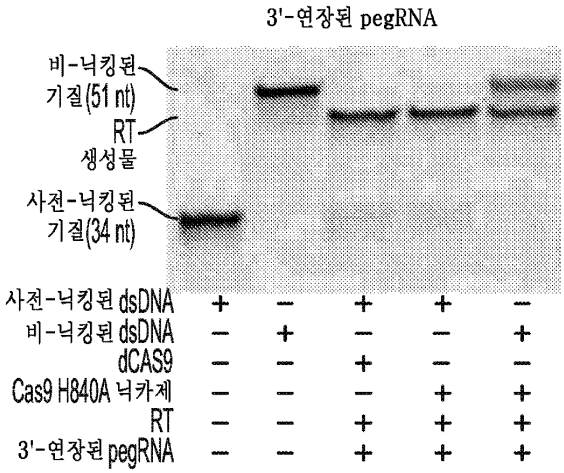
도면38c



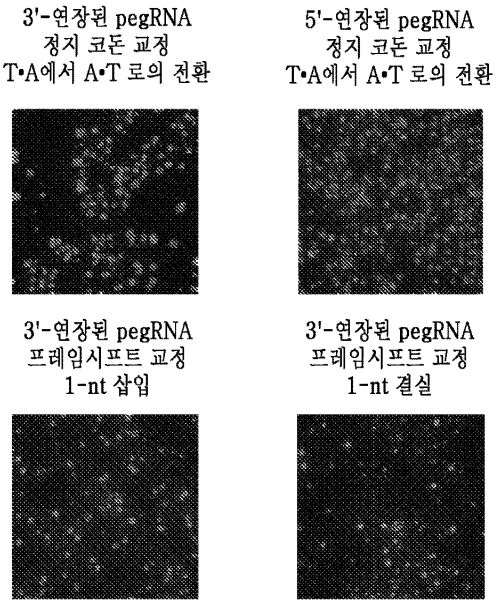
도면38d



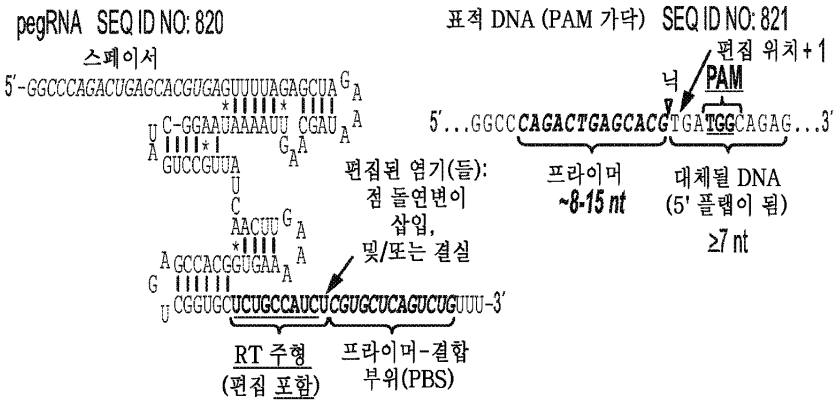
도면38e


3'-연장된 pegRNA
비-닉킹된 기질(51 nt)
RT 생성물
사전-닉킹된 기질(34 nt)
사전-닉킹된 dsDNA + − + + −
비-닉킹된 dsDNA − + − − +
dCAS9 − − + − −
Cas9 H840A 닉카제 − − − + +
RT − − + + +
3'-연장된 pegRNA − − + + +

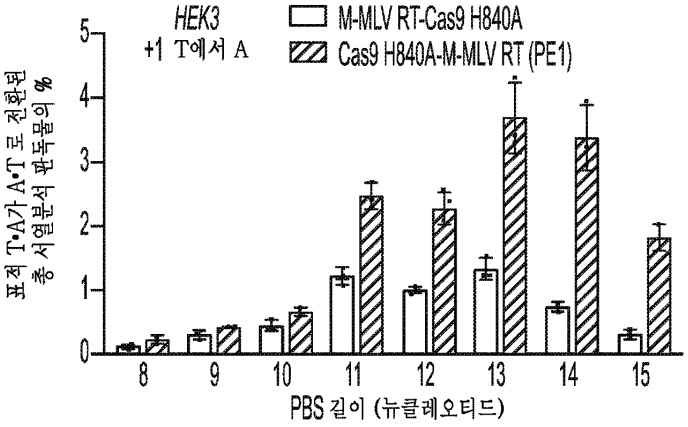
도면38f


3'-연장된 pegRNA
정지 코돈 교정
T•A에서 A•T 로의 전환
5'-연장된 pegRNA
정지 코돈 교정
T•A에서 A•T 로의 전환
3'-연장된 pegRNA
프레임시프트 교정
1-nt 삽입
3'-연장된 pegRNA
프레임시프트 교정
1-nt 결실

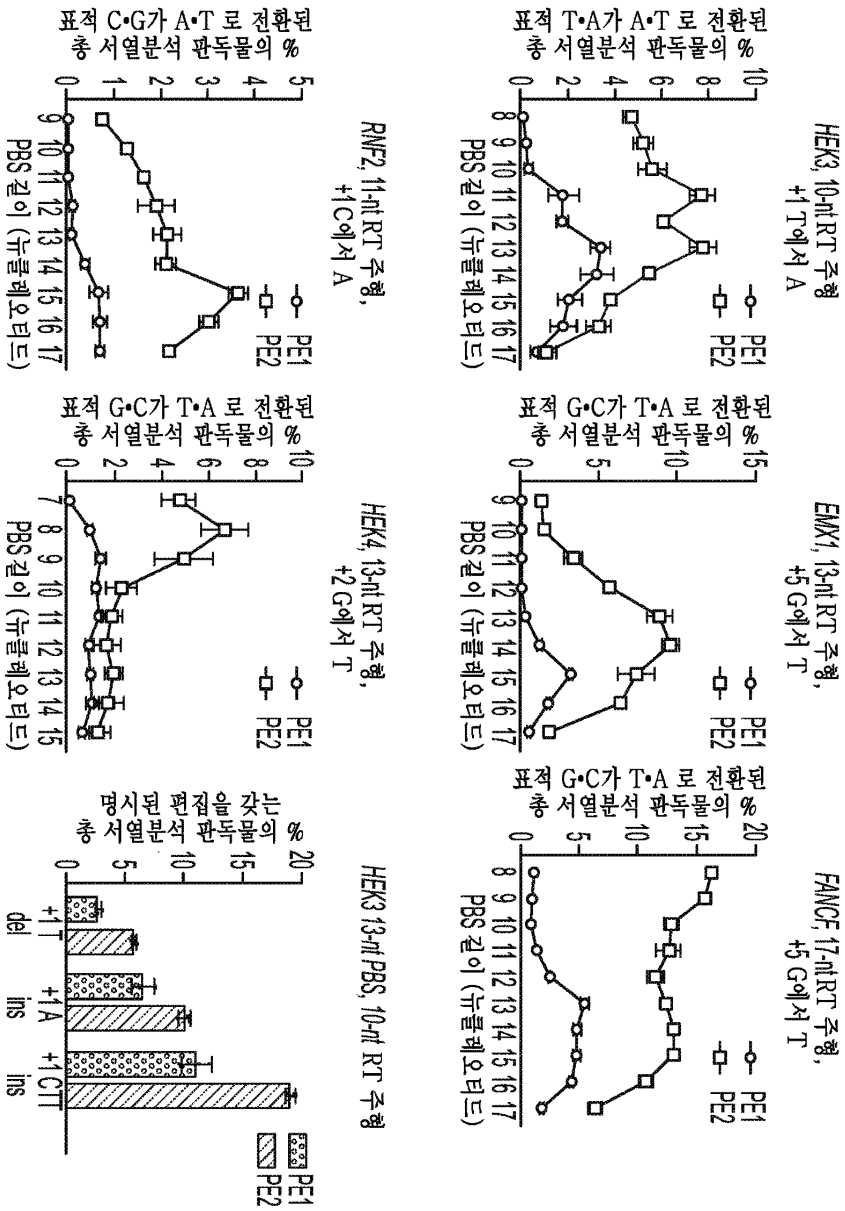
도면39a


pegRNA SEQ ID NO: 820
스페이서
5'-GGCCCAGACUGAGCACGUGAGUUUUAGAGCUA
편집된 염기(들):
점 돌연변이
삽입,
및/또는 결실
RT 주형
(편집 포함)
프라이머-결합
부위(PBS)
표적 DNA (PAM 가닥) SEQ ID NO: 821
편집 위치+1
닉
PAM
5'...GGCCCAGACTGAGCACGTGATGGCAGAG...3'
프라이머
~8-15 nt
대체될 DNA
(5' 플랩이 됨)
≥7 nt

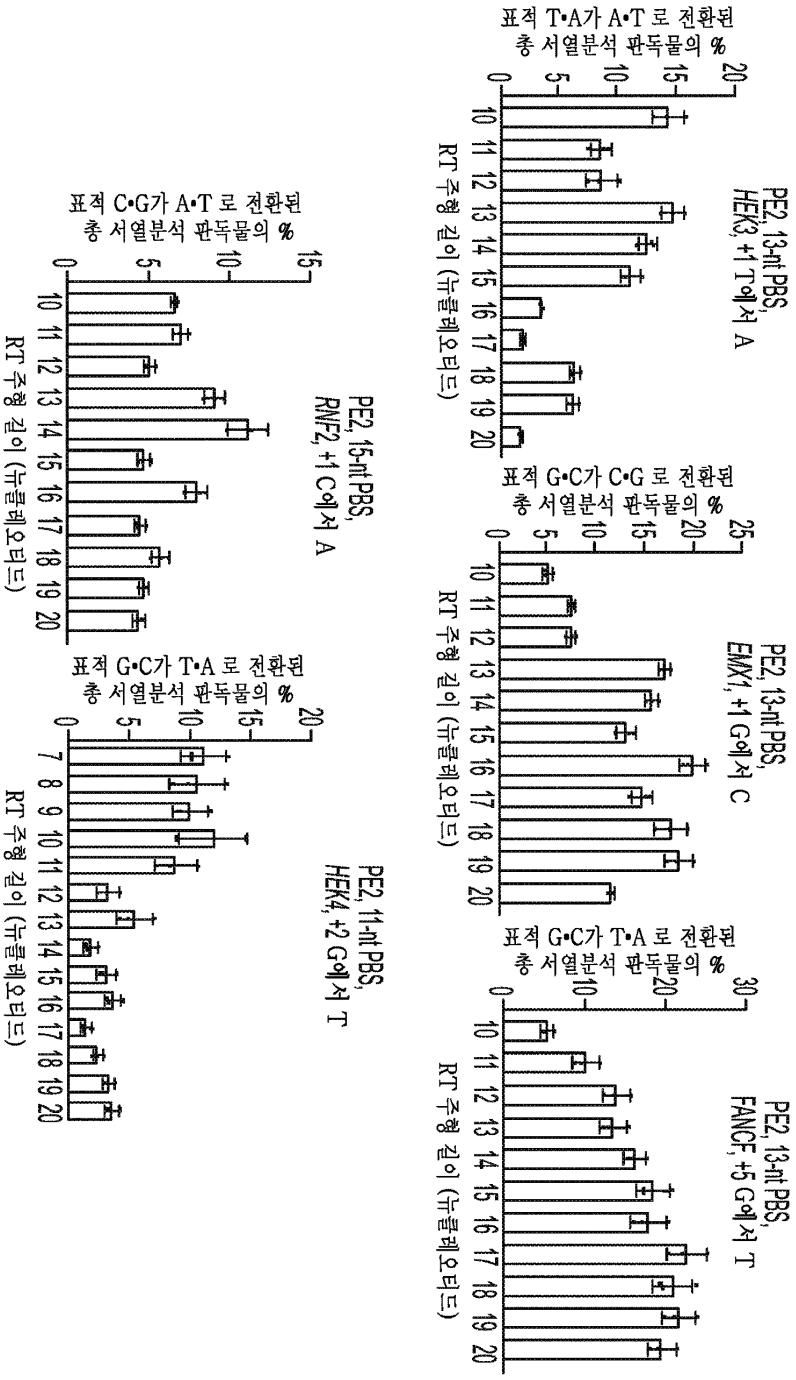
도면39b


HEK3
+1 T에서 A
M-MLV RT-Cas9 H840A
Cas9 H840A-M-MLV RT (PE1)
표적 T•A가 A•T로 전환된
총 서열분석 판독물의 %
0
1
2
3
4
5
8
9
10
11
12
13
14
15
PBS 길이 (뉴클레오티드)

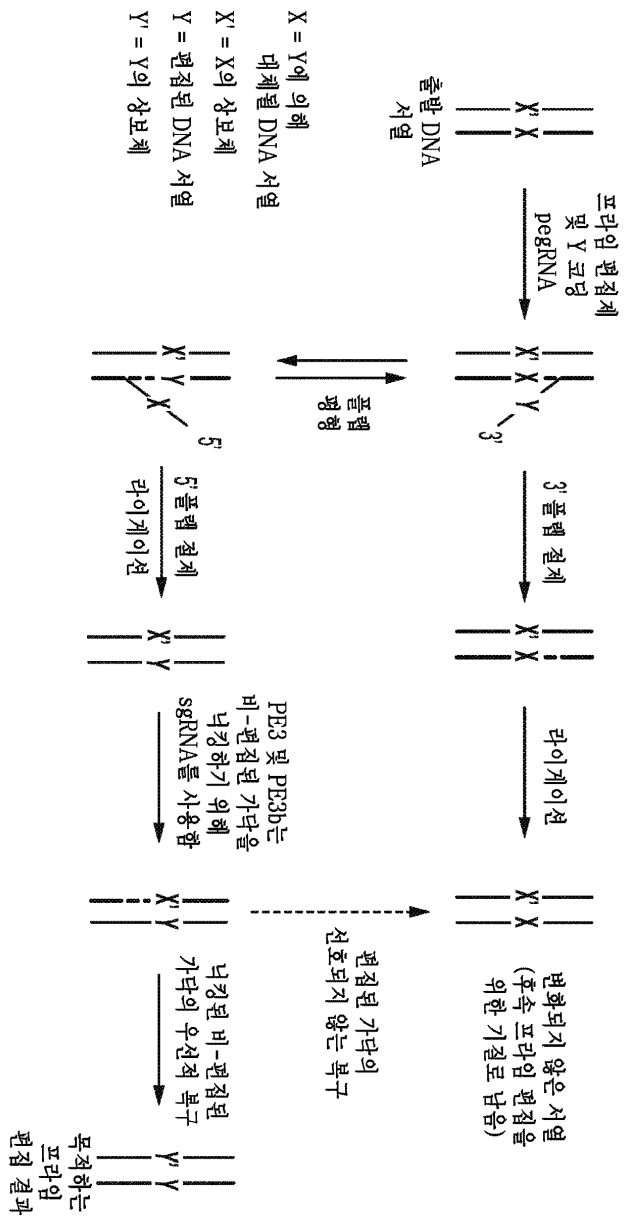
도면39c


HEK3, 10-nt RT 주형, +1 T에서 A
표적 T•A가 A•T 로 전환된 총 서열분석 판독물의 %
PBS길이 (뉴클레오티드)
PE1
PE2
EMX1, 13-nt RT 주형, +5 G에서 T
표적 G•C가 T•A 로 전환된 총 서열분석 판독물의 %
PBS길이 (뉴클레오티드)
PE1
PE2
FANCF, 17-nt RT 주형, +5 G에서 T
표적 G•C가 T•A 로 전환된 총 서열분석 판독물의 %
PBS길이 (뉴클레오티드)
PE1
PE2
RNF2, 11-nt RT 주형, +1 C에서 A
표적 C•G가 A•T 로 전환된 총 서열분석 판독물의 %
PBS길이 (뉴클레오티드)
PE1
PE2
HEK4, 13-nt RT 주형, +2 G에서 T
표적 G•C가 T•A 로 전환된 총 서열분석 판독물의 %
PBS길이 (뉴클레오티드)
PE1
PE2
HEK3 13-nt PBS, 10-nt RT 주형
명시된 편집을 갖는 총 서열분석 판독물의 %
+1 T del
+1 A ins
+1 CTT ins
PE1
PE2

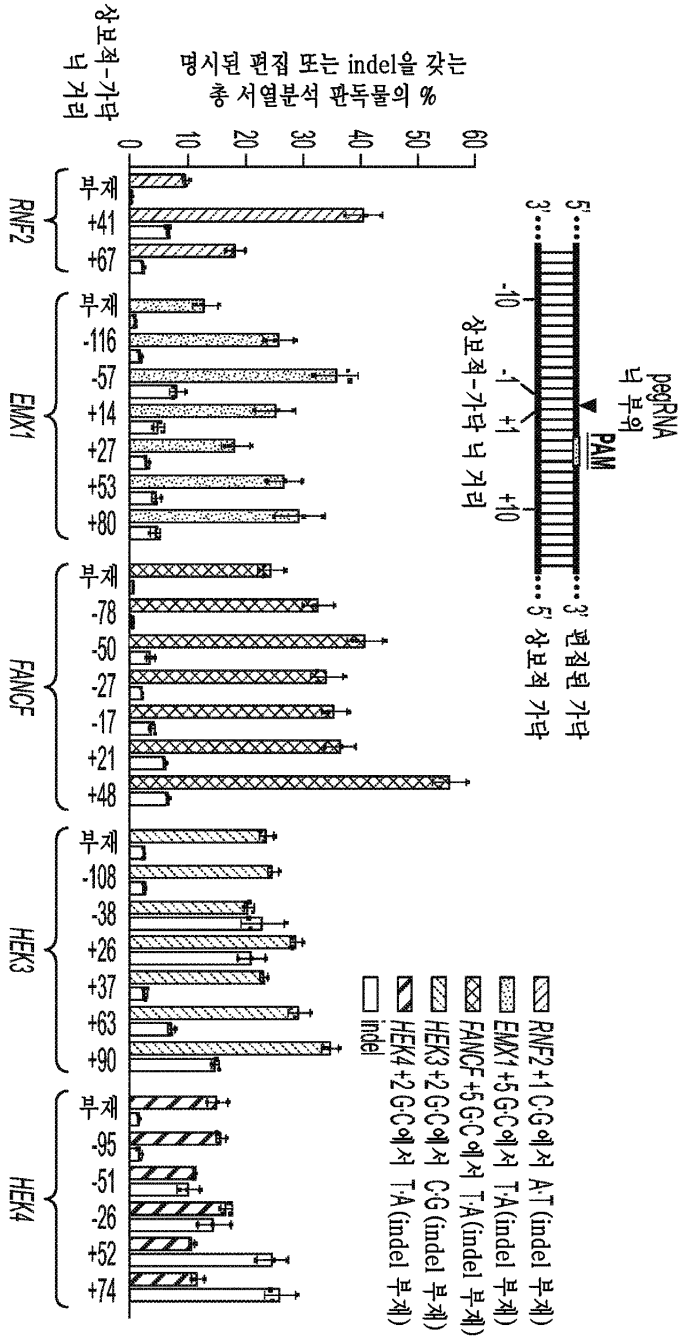
도면39d


표적 T•A가 A•T 로 전환된
총 서열분석 판독물의 %
PE2, 13-nt PBS,
HEK3, +1 T에서 A
RT 주형 길이 (뉴클레오티드)
표적 C•G가 A•T 로 전환된
총 서열분석 판독물의 %
PE2, 15-nt PBS,
RNF2, +1 C에서 A
RT 주형 길이 (뉴클레오티드)
표적 G•C가 C•G 로 전환된
총 서열분석 판독물의 %
PE2, 13-nt PBS,
EMX1, +1 G에서 C
RT 주형 길이 (뉴클레오티드)
표적 G•C가 T•A 로 전환된
총 서열분석 판독물의 %
PE2, 11-nt PBS,
HEK4, +2 G에서 T
RT 주형 길이 (뉴클레오티드)
표적 G•C가 T•A 로 전환된
총 서열분석 판독물의 %
PE2, 13-nt PBS,
FANCF, +5 G에서 T
RT 주형 길이 (뉴클레오티드)

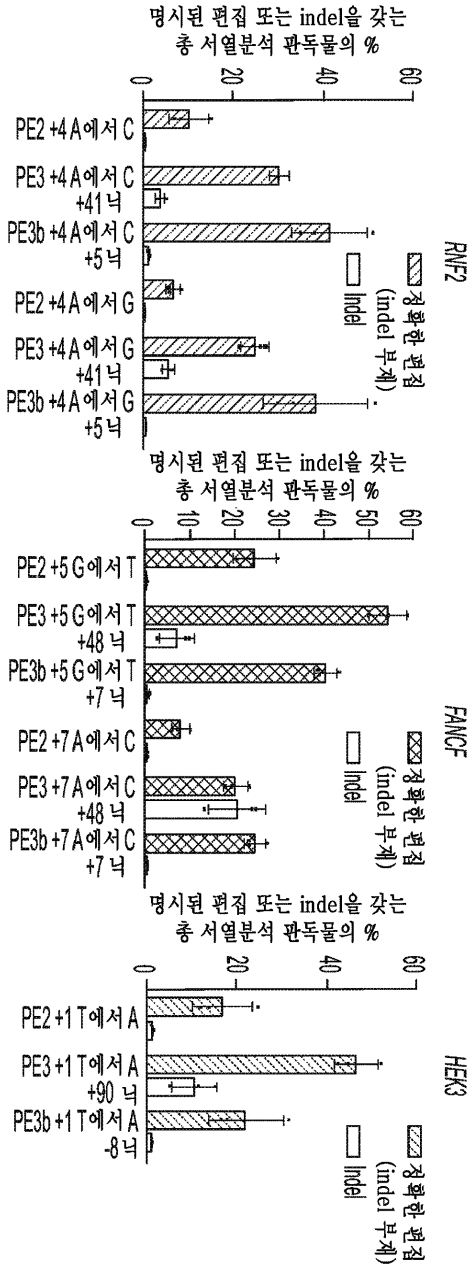
도면40a


프라임 편집제
및 Y 코딩
pegRNA
출발 DNA
서열
X = Y에 의해
대체될 DNA 서열
X' = X의 상보체
Y = 편집된 DNA 서열
Y' = Y의 상보체
3' 플랩 절제
플랩
평형
5'
5' 플랩 절제
라이게이션
라이게이션
변화되지 않은 서열
(후속 프라임 편집을
위한 기질로 남음)
PE3 및 PE3b는
비-편집된 가닥을
닉킹하기 위해
sgRNA를 사용함
편집된 가닥의
선호되지 않는 복구
닉킹된 비-편집된
가닥의 우선적 복구
목적하는
프라임
편집 결과

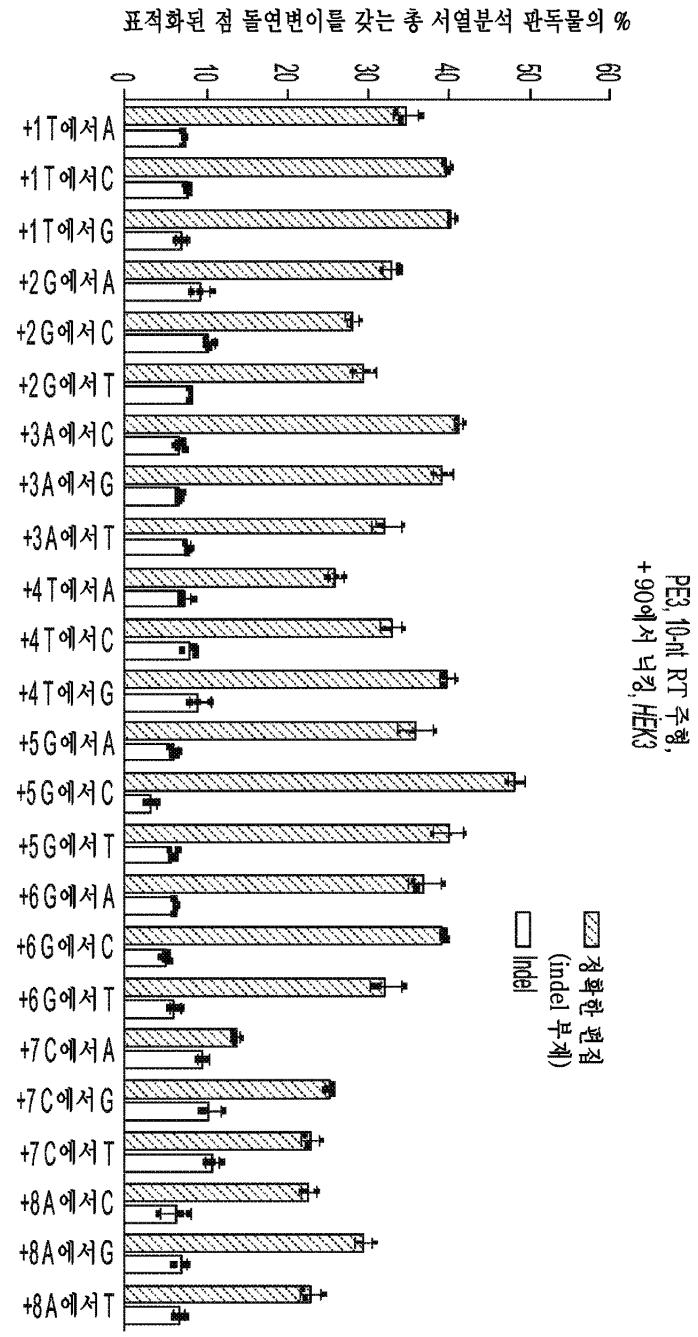
도면40b


명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0
10
20
30
40
50
60
상보좌-가닥
닉 거리
RNF2
EMX1
FANCF
HEK3
HEK4
부재
+41
+67
-116
-57
+14
+27
+53
+80
-78
-50
-27
-17
+21
+48
-108
-38
+26
+37
+63
+90
-95
-51
-26
+52
+74
pegRNA
닉 부위
PAM
5'
3'
-10
-1
+1
+10
상보좌-가닥 닉 거리
3' 편집된 가닥
5' 상보좌 가닥
RNF2 +1 C:G에서 A:T (indel 부재)
EMX1 +5 G:C에서 T:A (indel 부재)
FANCF +5 G:C에서 T:A (indel 부재)
HEK3 +2 G:C에서 C:G (indel 부재)
HEK4 +2 G:C에서 T:A (indel 부재)
indel

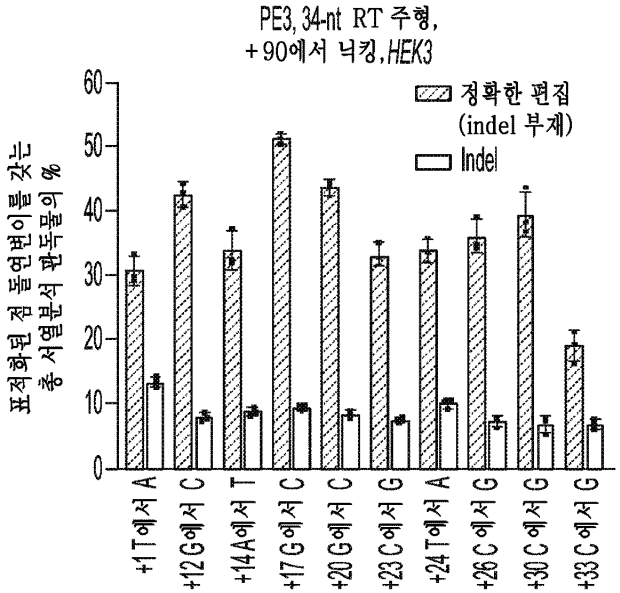
도면40c


명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0
20
40
60
PE2 +4 A에서 C
PE3 +4 A에서 C
+41 닉
PE3b +4 A에서 C
+5 닉
PE2 +4 A에서 G
PE3 +4 A에서 G
+41 닉
PE3b +4 A에서 G
+5 닉
정확한 편집
(indel 부재)
Indel
RNF2
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0
10
20
30
40
50
60
PE2 +5 G에서 T
PE3 +5 G에서 T
+48 닉
PE3b +5 G에서 T
+7 닉
PE2 +7 A에서 C
PE3 +7 A에서 C
+48 닉
PE3b +7 A에서 C
+7 닉
정확한 편집
(indel 부재)
Indel
FANCF
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0
20
40
60
PE2 +1 T에서 A
PE3 +1 T에서 A
+90 닉
PE3b +1 T에서 A
-8 닉
정확한 편집
(indel 부재)
Indel
HEK3

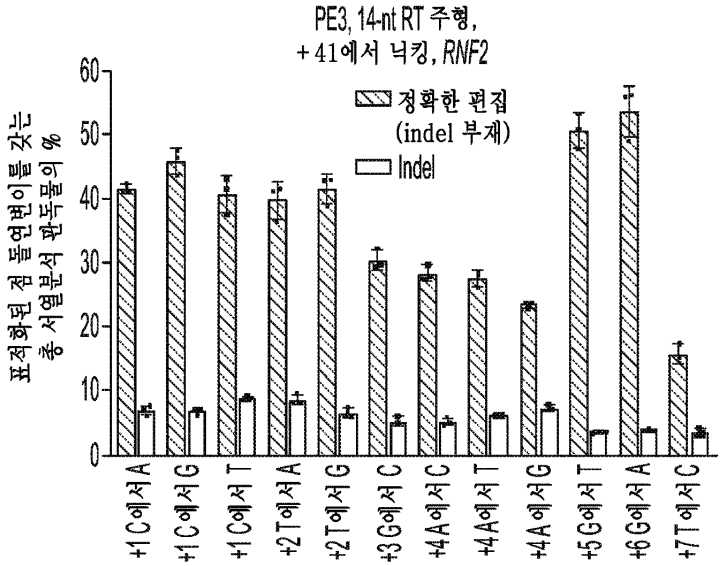
도면41a


표적화된 점 돌연변이를 갖는 총 서열분석 판독물의 %
0
10
20
30
40
50
60
+1 T에서 A
+1 T에서 C
+1 T에서 G
+2 G에서 A
+2 G에서 C
+2 G에서 T
+3 A에서 C
+3 A에서 G
+3 A에서 T
+4 T에서 A
+4 T에서 C
+4 T에서 G
+5 G에서 A
+5 G에서 C
+5 G에서 T
+6 G에서 A
+6 G에서 C
+6 G에서 T
+7 C에서 A
+7 C에서 G
+7 C에서 T
+8 A에서 C
+8 A에서 G
+8 A에서 T
PE3, 10-nt RT 주형, +90에서 닉킹, HEK3
정확한 편집 (indel 부재)
Indel

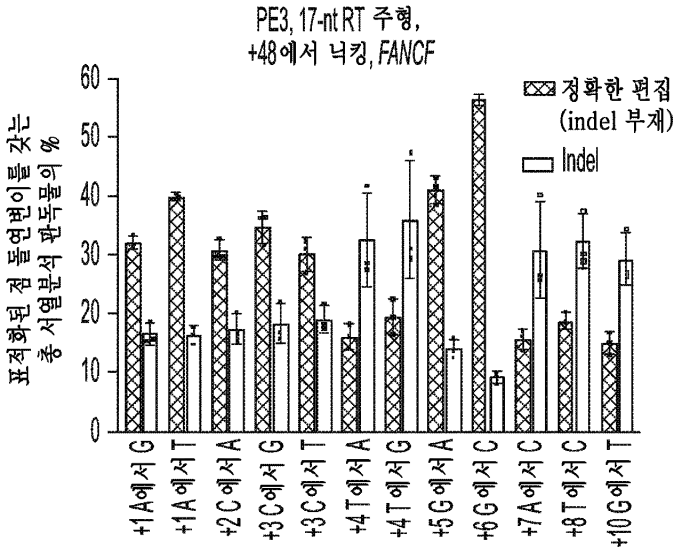
도면41b


PE3, 34-nt RT 주형,
+90에서 닉킹, HEK3
정확한 편집
(indel 부재)
Indel
표적화된 점 돌연변이를 갖는
총 서열분석 판독물의 %
60
50
40
30
20
10
0
+1 T에서 A
+12 G에서 C
+14 A에서 T
+17 G에서 C
+20 G에서 C
+23 C에서 G
+24 T에서 A
+26 C에서 G
+30 C에서 G
+33 C에서 G

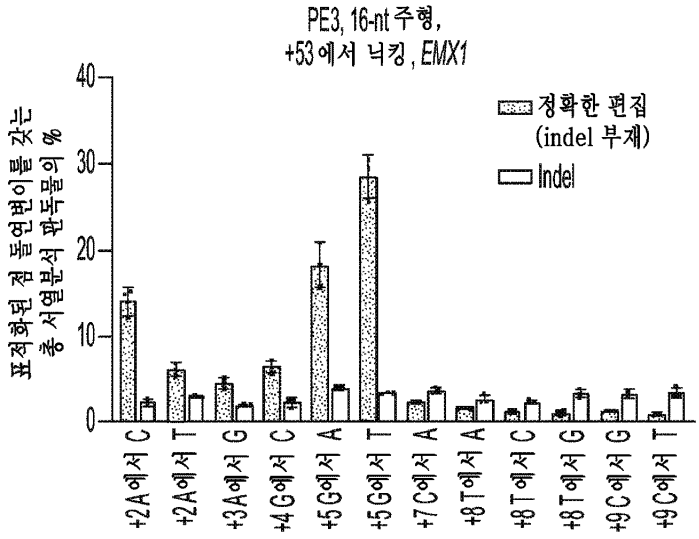
도면41c


PE3, 14-nt RT 주형,
+41에서 닉킹, RNF2
정확한 편집
(indel 부재)
Indel
표적화된 점 돌연변이를 갖는
총 서열분석 판독물의 %
60
50
40
30
20
10
0
+1 C에서 A
+1 C에서 G
+1 C에서 T
+2 T에서 A
+2 T에서 G
+3 G에서 C
+4 A에서 C
+4 A에서 T
+4 A에서 G
+5 G에서 T
+6 G에서 A
+7 T에서 C

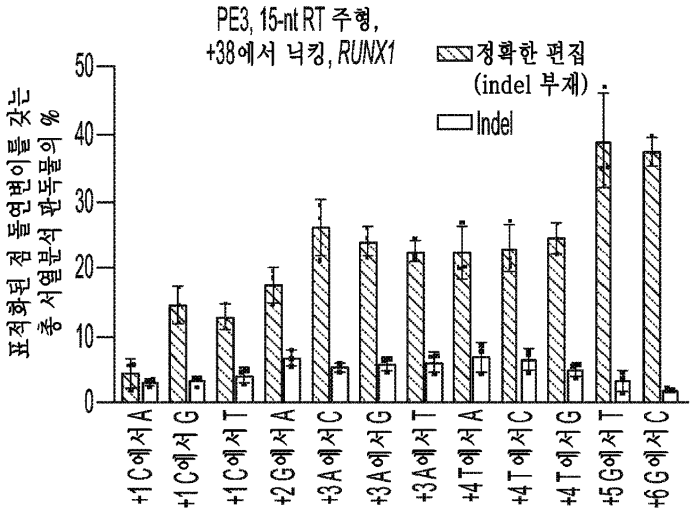
도면41d


PE3, 17-nt RT 주형,
+48에서 닉킹, FANCF
표적화된 점 돌연변이를 갖는
총 서열분석 판독물의 %
정확한 편집
(indel 부재)
Indel
0
10
20
30
40
50
60
+1A에서 G
+1A에서 T
+2C에서 A
+3C에서 G
+3C에서 T
+4T에서 A
+4T에서 G
+5G에서 A
+6G에서 C
+7A에서 C
+8T에서 C
+10G에서 T

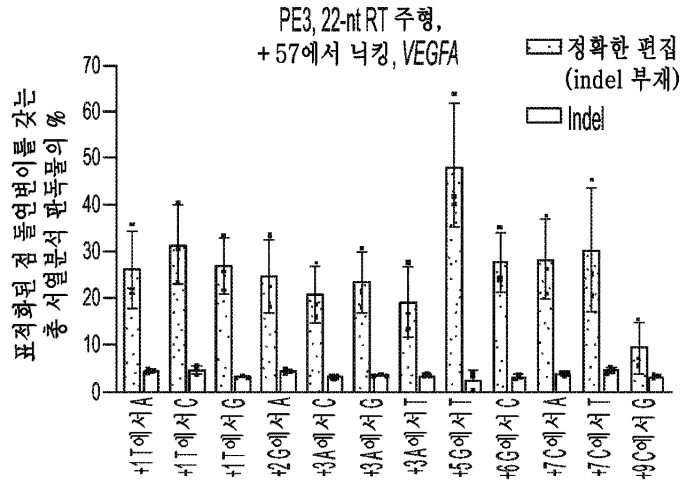
도면41e


PE3, 16-nt 주형,
+53에서 닉킹, EMX1
표적화된 점 돌연변이를 갖는
총 서열분석 판독물의 %
정확한 편집
(indel 부재)
Indel
0
10
20
30
40
+2A에서 C
+2A에서 T
+3A에서 G
+4G에서 C
+5G에서 A
+5G에서 T
+7C에서 A
+8T에서 A
+8T에서 C
+8T에서 G
+9C에서 G
+9C에서 T

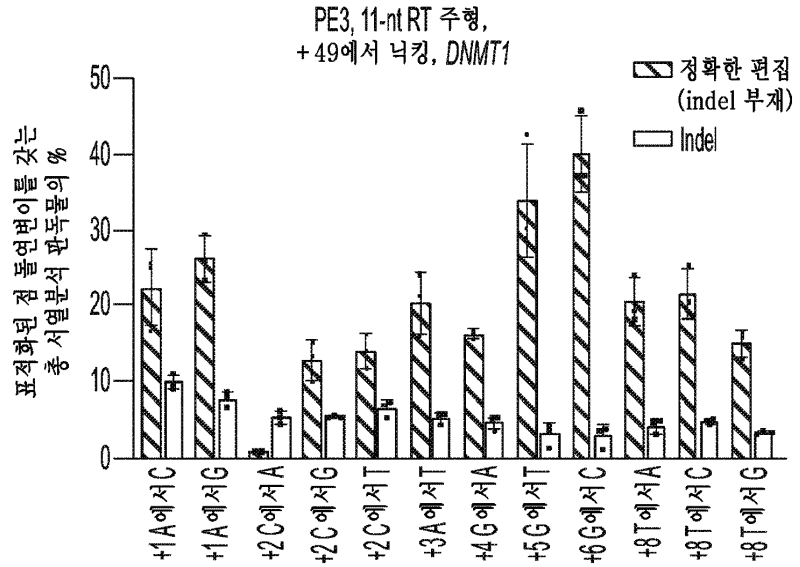
도면41f



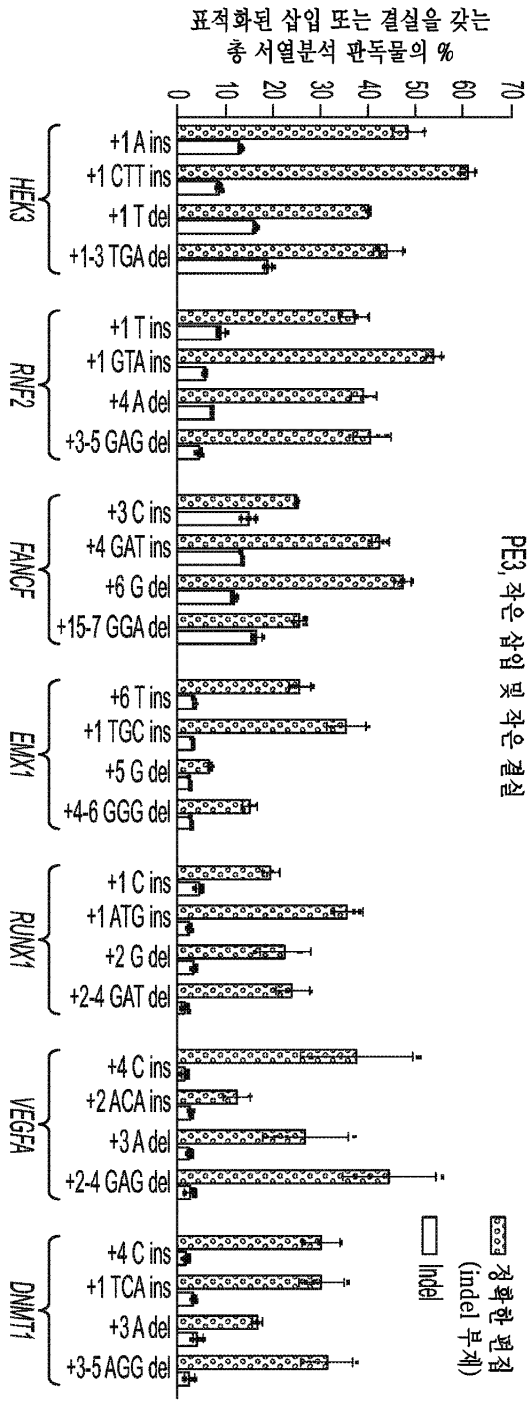
도면41g



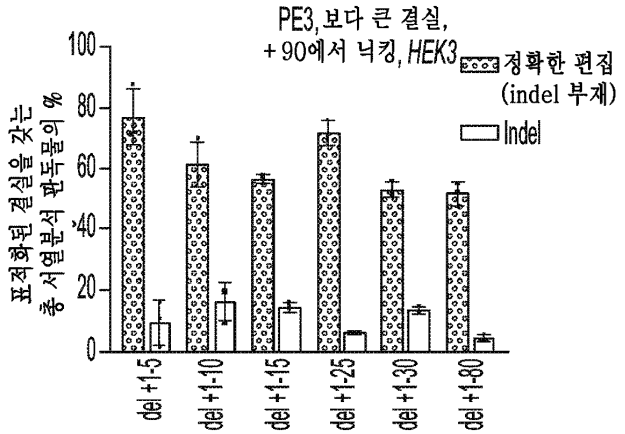
도면41h



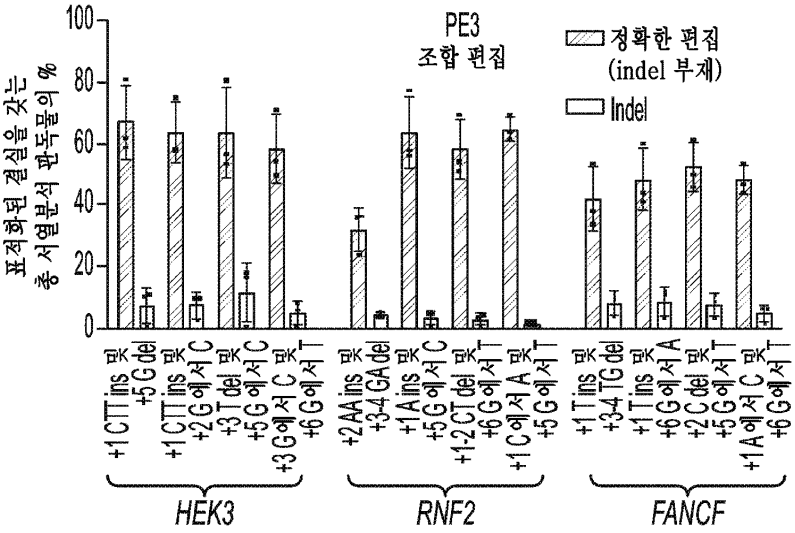
도면41i


표적화된 삽입 또는 결실을 갖는
총 서열분석 판독물의 %
0
10
20
30
40
50
60
70
HEK3
+1 A ins
+1 CTT ins
+1 T del
+1-3 TGA del
RNF2
+1 T ins
+1 GTA ins
+4 A del
+3-5 GAG del
FANCF
+3 C ins
+4 GAT ins
+6 G del
+15-7 GGA del
EMX1
+6 T ins
+1 TGC ins
+5 G del
+4-6 GGG del
RUNX1
+1 C ins
+1 ATG ins
+2 G del
+2-4 GAT del
VEGFA
+4 C ins
+2 ACA ins
+3 A del
+2-4 GAG del
DNMT1
+4 C ins
+1 TCA ins
+3 A del
+3-5 AGG del
PE3, 작은 삽입 및 작은 결실
정확한 편집
(indel 부재)
Indel

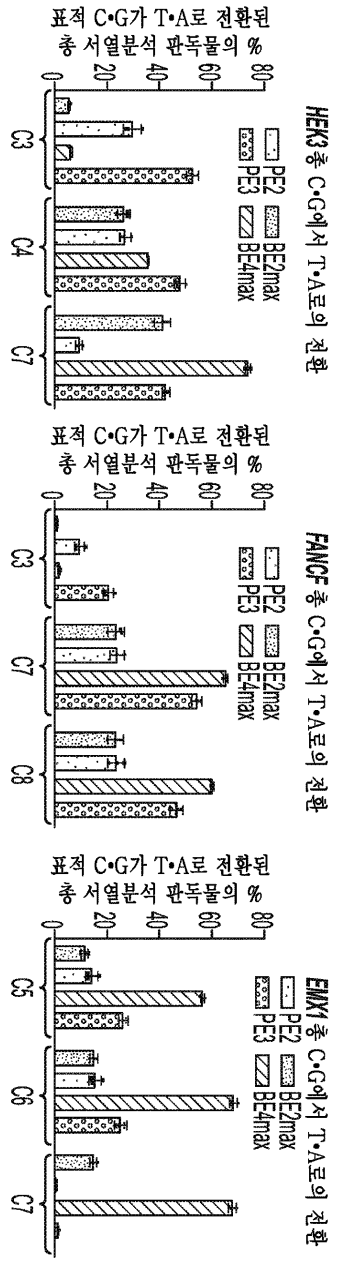
도면41j


PE3, 보다 큰 결실,
+90에서 닉킹, HEK3
정확한 편집
(indel 부재)
Indel
표적화된 결실을 갖는
총 서열분석 판독물의 %
100
80
60
40
20
0
del +1-5
del +1-10
del +1-15
del +1-25
del +1-30
del +1-80

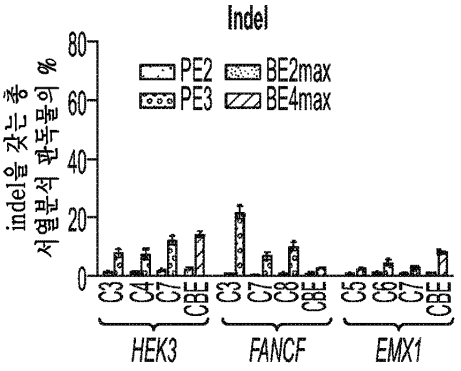
도면41k


PE3
조합 편집
정확한 편집
(indel 부재)
Indel
표적화된 결실을 갖는
총 서열분석 판독물의 %
100
80
60
40
20
0
+1 CTT ins 및 +5 G del
+1 CTT ins 및 +2 G에서 C
+3 T del 및 +5 G에서 C
+3 G에서 C 및 +6 G에서 T
+2 AA ins 및 +3-4 GA del
+1 A ins 및 +5 G에서 C
+1-2 CT del 및 +6 G에서 T
+1 C에서 A 및 +5 G에서 T
+1 T ins 및 +3-4 TG del
+1 T ins 및 +6 G에서 A
+2 C del 및 +5 G에서 T
+1 A에서 C 및 +6 G에서 T
HEK3
RNF2
FANCF

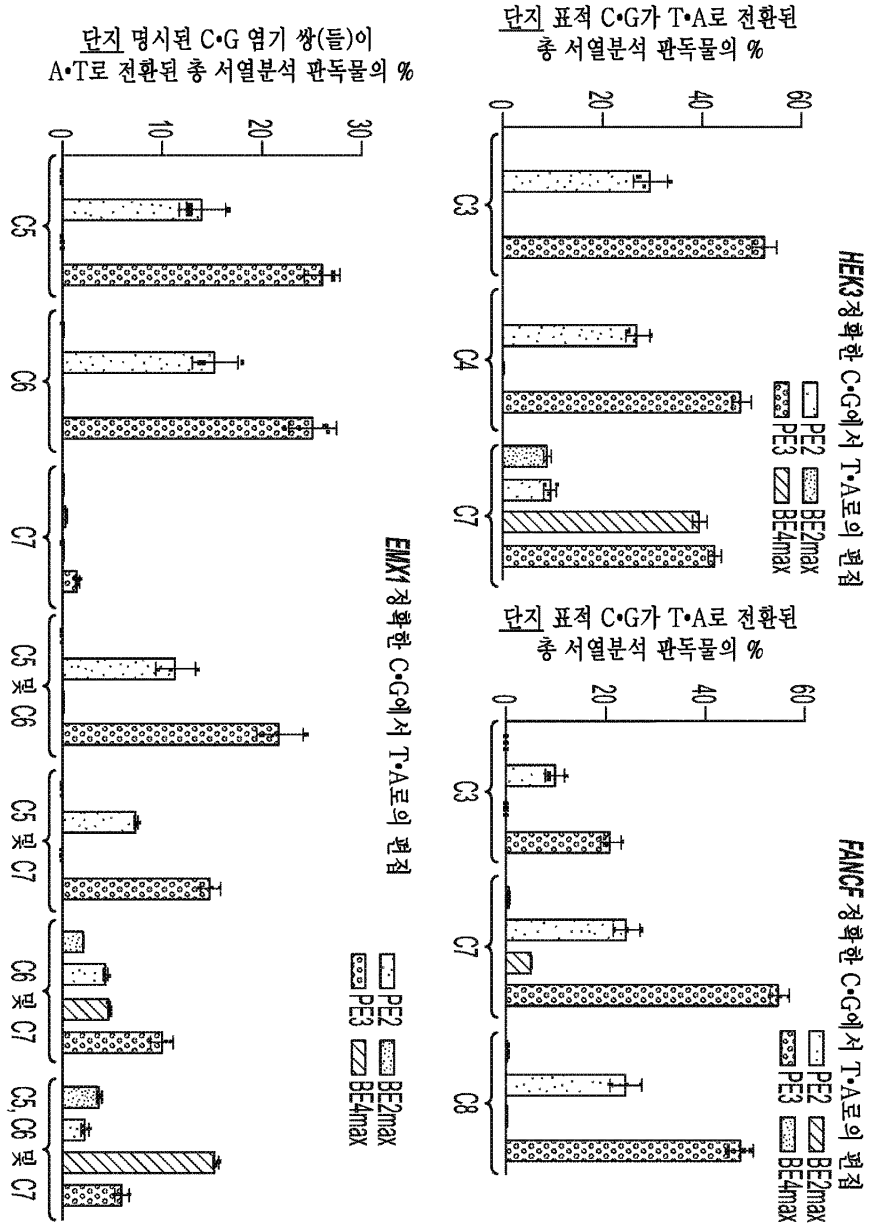
도면42a


표적 C•G가 T•A로 전환된
총 서열분석 판독물의 %
0
20
40
60
80
C3
C4
C7
PE2
PE3
BE2max
BE4max
HEK3 총 C•G에서 T•A로의 전환
표적 C•G가 T•A로 전환된
총 서열분석 판독물의 %
C3
C7
C8
FANCF 총 C•G에서 T•A로의 전환
표적 C•G가 T•A로 전환된
총 서열분석 판독물의 %
C5
C6
C7
EMX1 총 C•G에서 T•A로의 전환

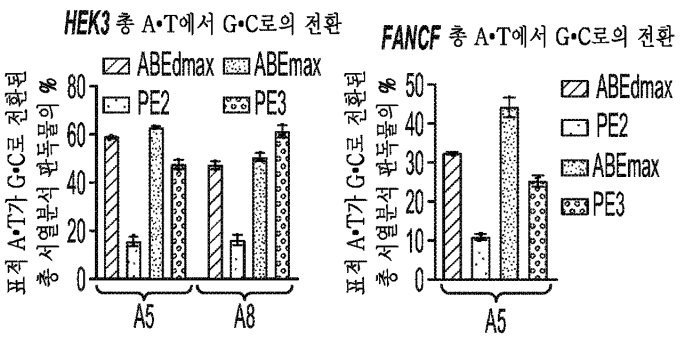
도면42b


Indel
indel을 갖는 총
서열분석 판독물의 %
0
20
40
60
80
PE2
PE3
BE2max
BE4max
C3
C4
C7
C8
C5
C6
HEK3
FANCF
EMX1

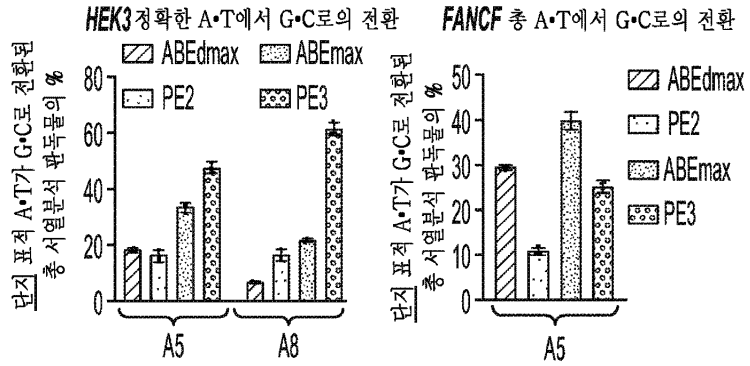
도면42c


단지 명시된 C•G 염기 쌍(들)이
A•T로 전환된 총 서열분석 판독물의 %
0
10
20
30
C5
C6
C7
C5 및 C6
C5 및 C7
C6 및 C7
C5, C6 및 C7
EMX1 정확한 C•G에서 T•A로의 편집
PE2
BE2max
PE3
BE4max
단지 표적 C•G가 T•A로 전환된
총 서열분석 판독물의 %
0
20
40
60
C3
C4
C7
HEK3 정확한 C•G에서 T•A로의 편집
PE2
BE2max
PE3
BE4max
단지 표적 C•G가 T•A로 전환된
총 서열분석 판독물의 %
0
20
40
60
C3
C7
C8
FANCF 정확한 C•G에서 T•A로의 편집
PE2
BE2max
PE3
BE4max

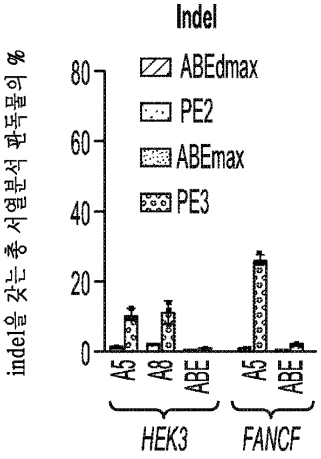
도면42d


HEK3 총 A•T에서 G•C로의 전환
표적 A•T가 G•C로 전환된
총 서열분석 판독물의 %
ABEdmax
ABEmax
PE2
PE3
0
20
40
60
80
A5
A8
FANCF 총 A•T에서 G•C로의 전환
표적 A•T가 G•C로 전환된
총 서열분석 판독물의 %
ABEdmax
PE2
ABEmax
PE3
0
10
20
30
40
50
A5

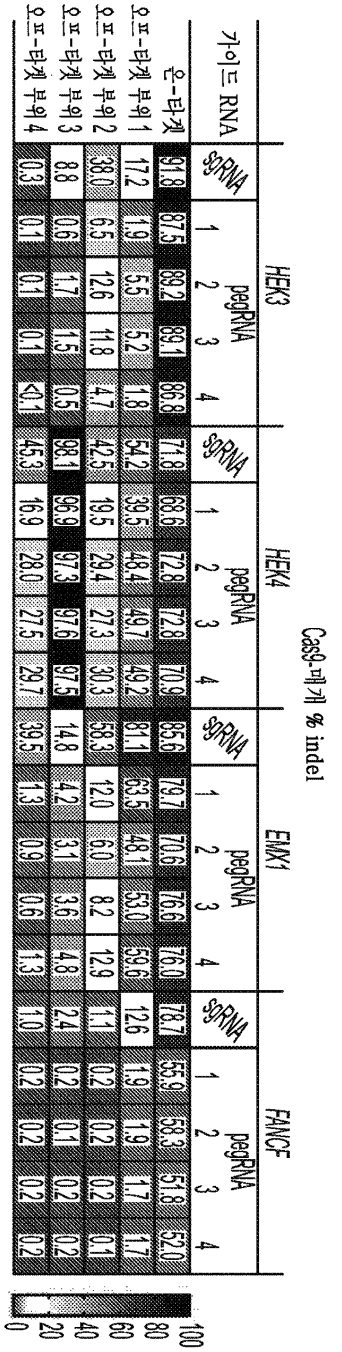
도면42e


HEK3 정확한 A•T에서 G•C로의 전환
FANCF 총 A•T에서 G•C로의 전환
단지 표적 A•T가 G•C로 전환된 총 서열분석 판독물의 %
ABEdmax
PE2
ABEmax
PE3
0
20
40
60
80
10
30
50
A5
A8

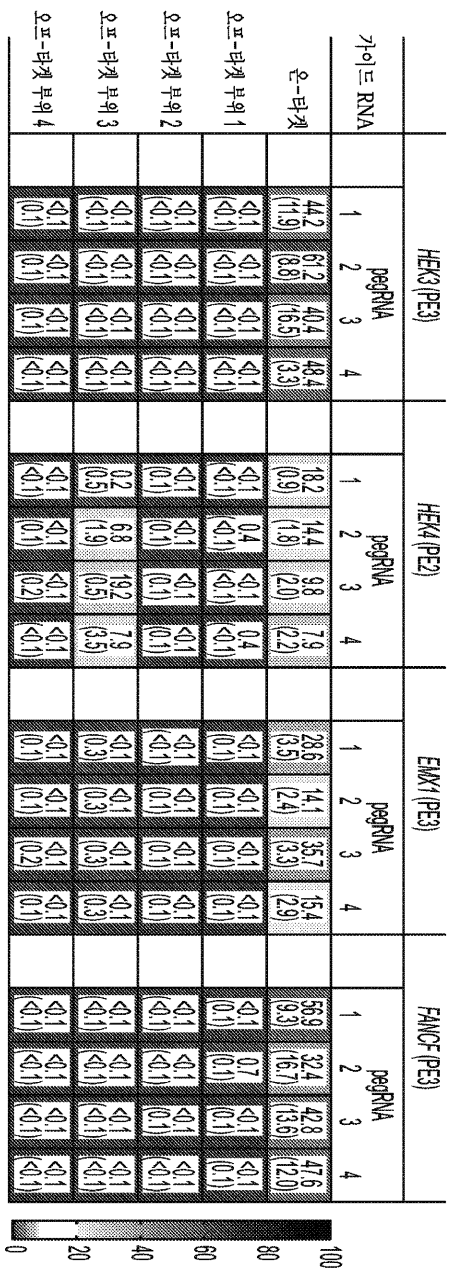
도면42f


Indel
indel을 갖는 총 서열분석 판독물의 %
ABEdmax
PE2
ABEmax
PE3
0
20
40
60
80
A5
A8
ABE
A5
ABE
HEK3
FANCF

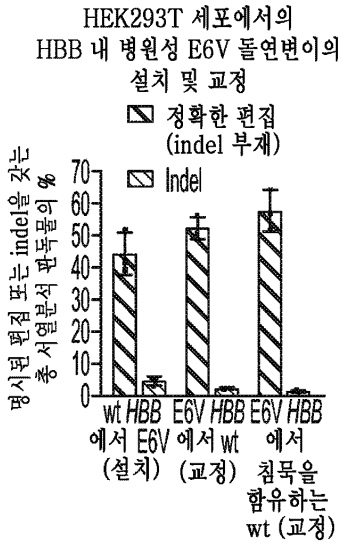
도면42g


Cas9-매개 % indel
HEK3
HEK4
EMX1
FANCF
가이드 RNA
sgRNA
pegRNA 1
pegRNA 2
pegRNA 3
pegRNA 4
온-타겟
오프-타겟 부위 1
오프-타겟 부위 2
오프-타겟 부위 3
오프-타겟 부위 4
91.8 87.5 89.2 89.1 86.8
17.2 1.9 5.5 5.2 1.8
38.0 6.5 12.6 11.8 4.7
8.8 0.6 1.7 1.5 0.5
0.3 0.1 0.1 0.1 <0.1
71.8 66.6 72.8 72.8 70.9
54.2 39.5 48.4 49.7 49.2
42.5 19.5 29.4 27.3 30.3
98.1 96.9 97.3 97.6 97.5
45.3 16.9 28.0 27.5 29.7
85.6 79.7 70.6 76.6 76.0
81.1 63.5 48.1 53.0 59.6
56.3 12.0 6.0 8.2 12.9
14.8 4.2 3.1 3.6 4.8
39.5 1.3 0.9 0.6 1.3
78.7 55.9 58.3 51.8 52.0
12.6 1.9 1.9 1.7 1.7
1.1 0.2 0.2 0.2 0.1
2.4 0.2 0.1 0.2 0.2
1.0 0.2 0.2 0.2 0.2
0 20 40 60 80 100

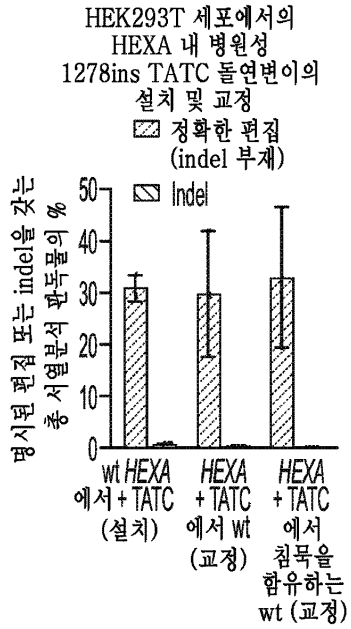
도면42h


PE-매개 % 편집 및 % indel (괄호 안)
HEK3 (PE3)
HEK4 (PE2)
EMX1 (PE3)
FANCF (PE3)
가이드 RNA
pegRNA
1
2
3
4
온-타겟
오프-타겟 부위 1
오프-타겟 부위 2
오프-타겟 부위 3
오프-타겟 부위 4
44.2 (11.9)
61.2 (8.6)
40.4 (16.5)
48.4 (3.3)
<0.1 (<0.1)
<0.1 (0.1)
18.2 (0.9)
14.4 (1.8)
9.8 (2.0)
7.9 (2.2)
0.4 (<0.1)
6.8 (1.9)
19.2 (0.5)
0.2 (0.5)
28.6 (3.5)
14.1 (2.4)
35.7 (3.3)
15.4 (2.9)
56.9 (9.3)
32.4 (16.7)
42.8 (13.6)
47.6 (12.0)
0.7 (0.1)
0
20
40
60
80
100

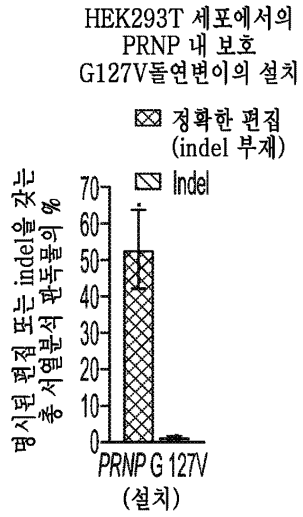
도면43a


HEK293T 세포에서의
HBB 내 병원성 E6V 돌연변이의
설치 및 교정
정확한 편집
(indel 부재)
Indel
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
70
60
50
40
30
20
10
0
wt HBB 에서 E6V (설치)
E6V HBB 에서 wt (교정)
E6V HBB 에서 침묵을 함유하는 wt (교정)

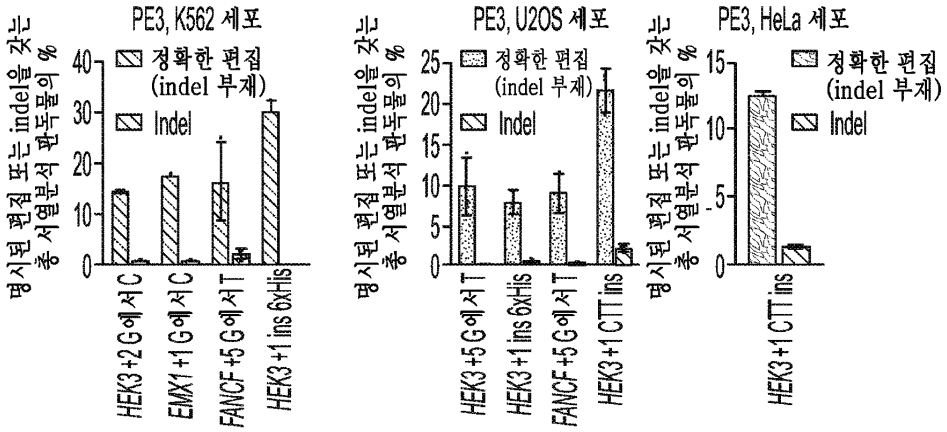
도면43b


HEK293T 세포에서의
HEXA 내 병원성
1278ins TATC 돌연변이의
설치 및 교정
정확한 편집
(indel 부재)
Indel
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
50
40
30
20
10
0
wt HEXA + TATC 에서 (설치)
HEXA + TATC 에서 wt (교정)
HEXA + TATC 에서 침묵을 함유하는 wt (교정)

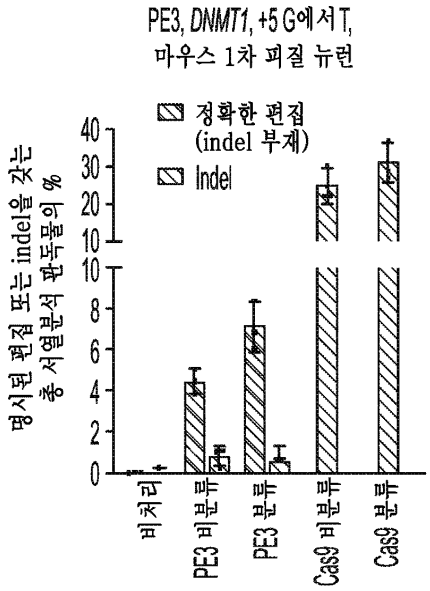
도면43c


HEK293T 세포에서의
PRNP 내 보호
G127V돌연변이의 설치
정확한 편집
(indel 부재)
Indel
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
70
60
50
40
30
20
10
0
PRNP G 127V
(설치)

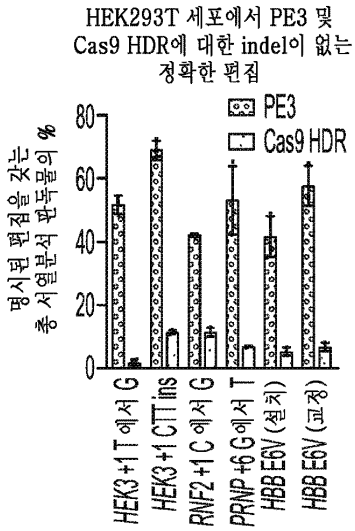
도면43d


PE3, K562 세포
PE3, U2OS 세포
PE3, HeLa 세포
정확한 편집
(indel 부재)
Indel
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
HEK3 +2 G에서 C
EMX1 +1 G에서 C
FANCF +5 G에서 T
HEK3 +1 ins 6xHis
HEK3 +5 G에서 T
HEK3 +1 6xHis
FANCF +5 G에서 T
HEK3 +1 CTT ins
HEK3 +1 CTT ins

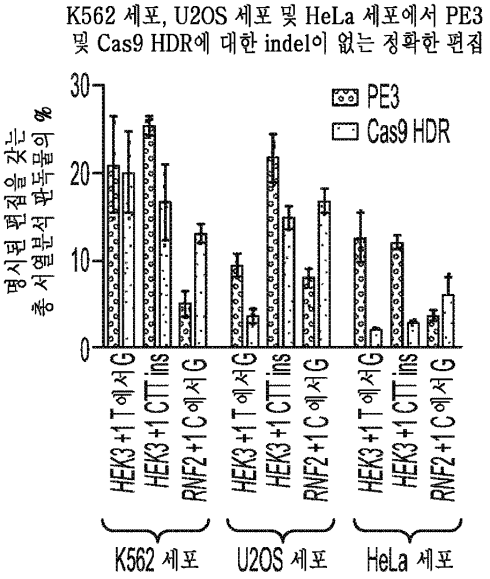
도면43e


PE3, DNMT1, +5 G에서 T,
마우스 1차 피질 뉴런
정확한 편집
(indel 부재)
Indel
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
40
30
20
10
10
8
6
4
2
0
비처리
PE3 비분류
PE3 분류
Cas9 비분류
Cas9 분류

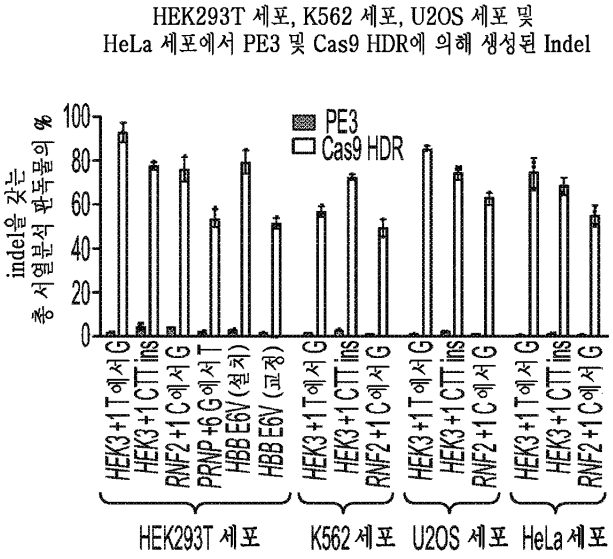
도면43f


HEK293T 세포에서 PE3 및
Cas9 HDR에 대한 indel이 없는
정확한 편집
PE3
Cas9 HDR
명시된 편집을 갖는
총 서열분석 판독물의 %
80
60
40
20
0
HEK3 +1 T 에서 G
HEK3 +1 CTT ins
RNF2 +1 C 에서 G
PRNP +6 G에서 T
HBB E6V (설치)
HBB E6V (교정)

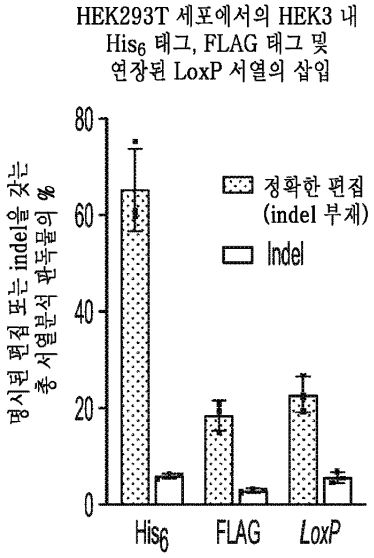
도면43g


K562 세포, U2OS 세포 및 HeLa 세포에서 PE3 및 Cas9 HDR에 대한 indel이 없는 정확한 편집
명시된 편집을 갖는 총 서열분석 판독물의 %
PE3
Cas9 HDR
HEK3 +1 T에서 G
HEK3 +1 CTT ins
RNF2 +1 C에서 G
K562 세포
U2OS 세포
HeLa 세포

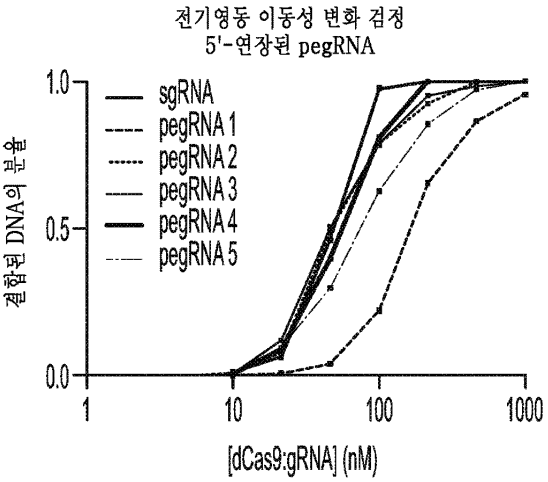
도면43h


HEK293T 세포, K562 세포, U2OS 세포 및 HeLa 세포에서 PE3 및 Cas9 HDR에 의해 생성된 Indel
indel을 갖는 총 서열분석 판독물의 %
PE3
Cas9 HDR
HEK3 +1 T에서 G
HEK3 +1 CTT ins
RNF2 +1 C에서 G
PRNP +6 G에서 T
HBB E6V (설치)
HBB E6V (교정)
HEK293T 세포
K562 세포
U2OS 세포
HeLa 세포

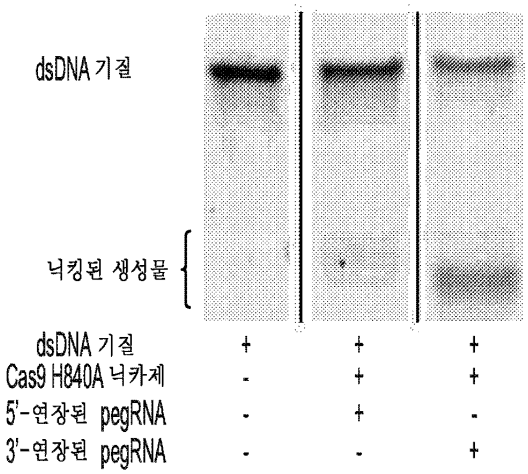
도면43i


HEK293T 세포에서의 HEK3 내
His6 태그, FLAG 태그 및
연장된 LoxP 서열의 삽입
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
80
60
40
20
0
정확한 편집
(indel 부재)
Indel
His6
FLAG
LoxP

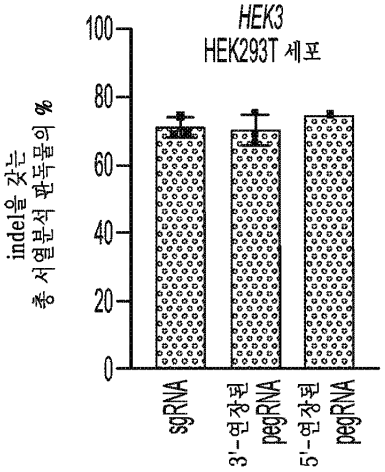
도면44a


전기영동 이동성 변화 검정
5'-연장된 pegRNA
결합된 DNA의 분율
1.0
0.5
0.0
sgRNA
pegRNA 1
pegRNA 2
pegRNA 3
pegRNA 4
pegRNA 5
1
10
100
1000
[dCas9:gRNA] (nM)

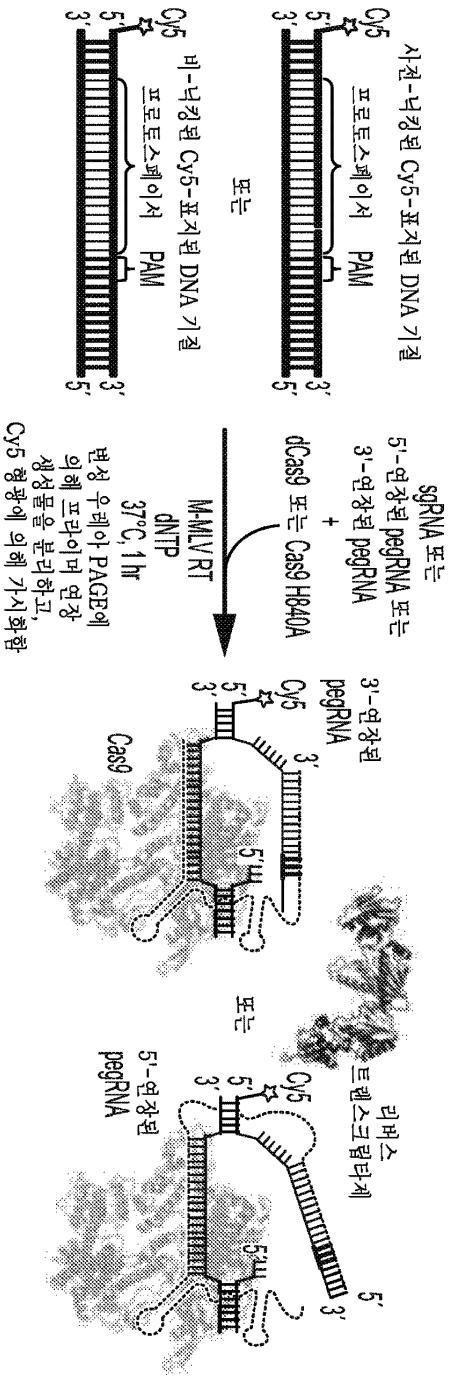
도면44b


dsDNA 기질
닉킹된 생성물
dsDNA 기질 + + +
Cas9 H840A 닉카제 - + +
5'-연장된 pegRNA - + -
3'-연장된 pegRNA - - +

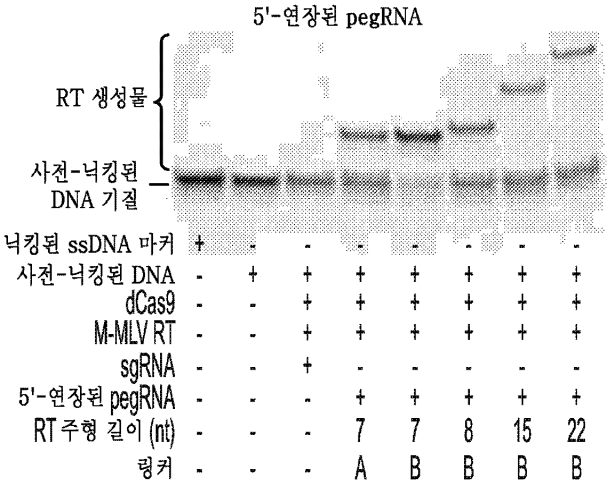
도면44c


HEK3
HEK293T 세포
indel을 갖는 총 서열분석 판독물의 %
100
80
60
40
20
0
sgRNA
3'-연장된 pegRNA
5'-연장된 pegRNA

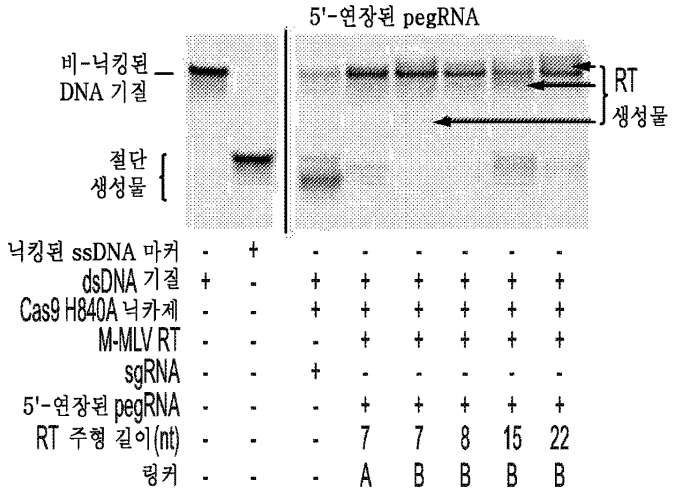
도면44d


사전-닉킹된 Cy5-표지된 DNA 기질
Cy5
5′
3′
프로토스페이서
PAM
3′
5′
또는
비-닉킹된 Cy5-표지된 DNA 기질
Cy5
5′
3′
프로토스페이서
PAM
3′
5′
sgRNA 또는
5′-연장된 pegRNA 또는
3′-연장된 pegRNA
+
dCas9 또는 Cas9 H840A
M-MLV RT
dNTP
37°C, 1 hr
변성 우레아 PAGE에
의해 프라이머 연장
생성물을 분리하고,
Cy5 형광에 의해 가시화함
3′-연장된
pegRNA
Cy5
5′
3′
3′
5′
Cas9
또는
리버스
트랜스크립타제
Cy5
5′
3′
5′
3′
5′-연장된
pegRNA

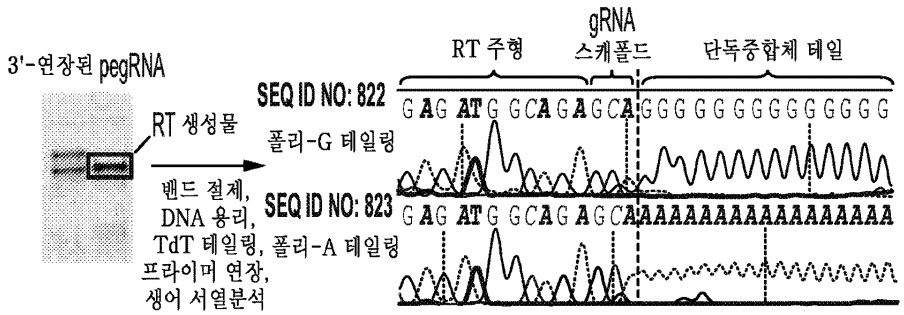
도면44e



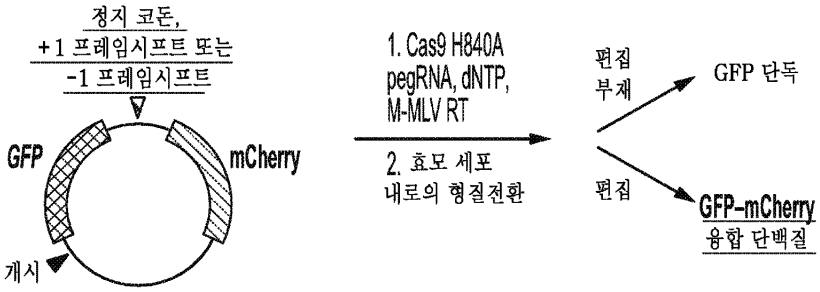
도면44f



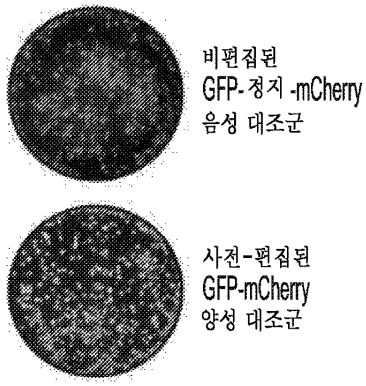
도면44g



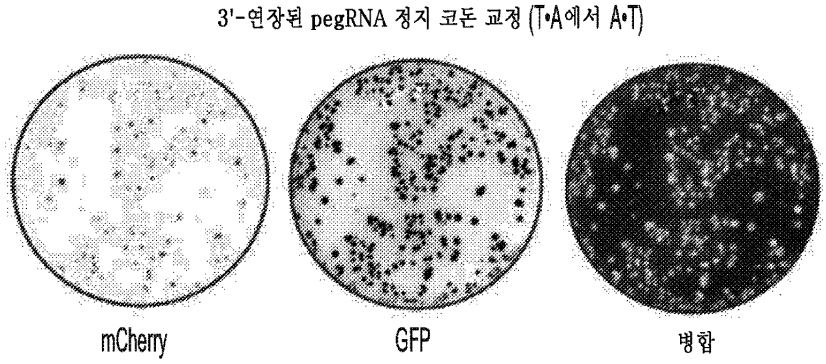
도면45a



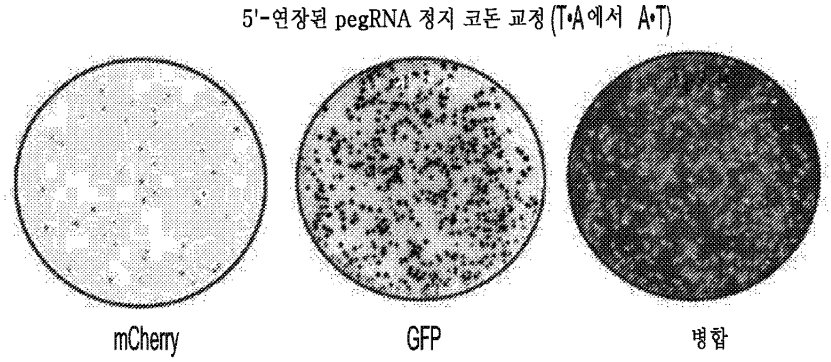
도면45b



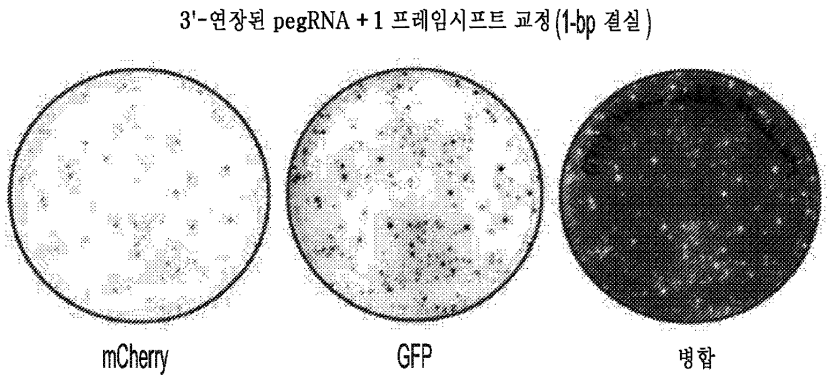
도면45c



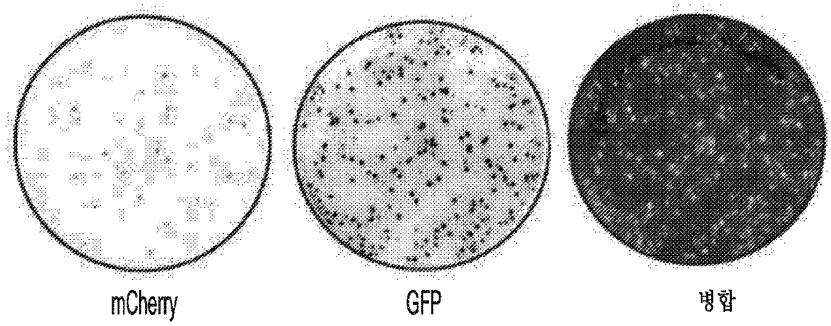
도면45d



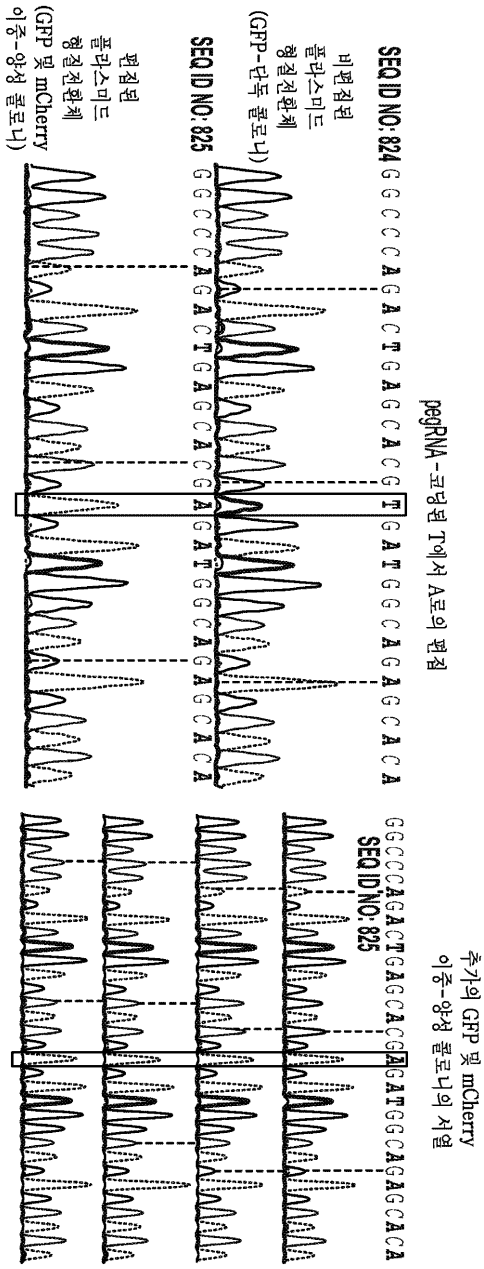
도면45e



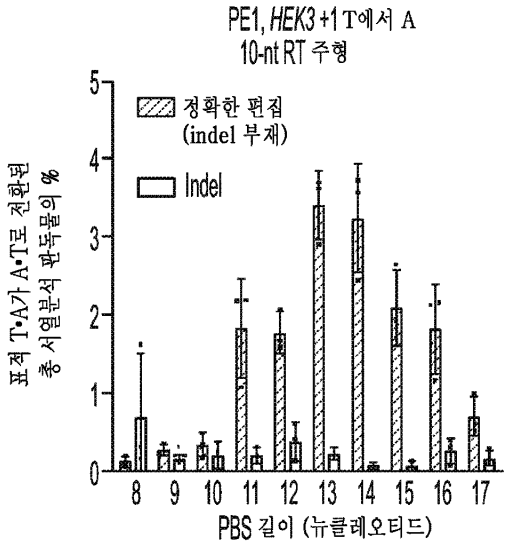
도면45f



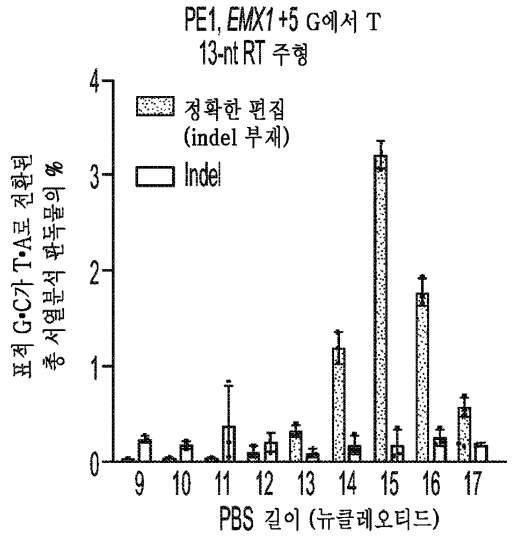
도면45g


SEQ ID NO: 824
G G C C C A G A C T G A G C A C G T G A T G G C A G A G C A C A
pegRNA-코딩된 T에서 A로의 편집
비편집된
플라스미드
형질전환체
(GFP-단독 콜로니)
SEQ ID NO: 825
G G C C C A G A C T G A G C A C G A G A T G G C A G A G C A C A
편집된
플라스미드
형질전환체
(GFP 및 mCherry
이중-양성 콜로니)
G G C C C A G A C T G A G C A C G A G A T G G C A G A G C A C A
SEQ ID NO: 825
추가의 GFP 및 mCherry
이중-양성 콜로니의 서열

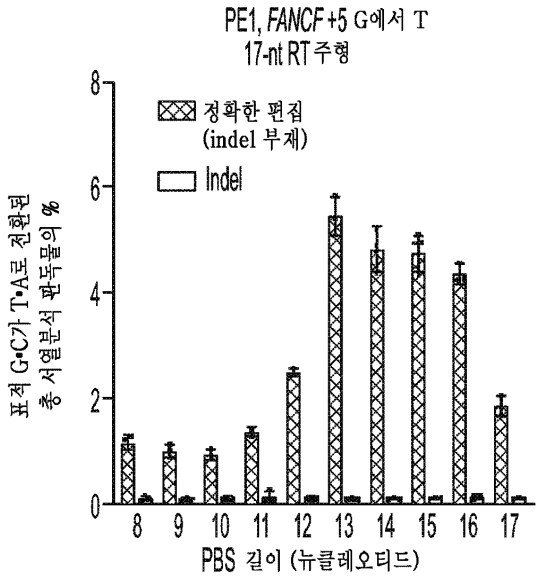
도면46a


PE1, HEK3 +1 T에서 A
10-nt RT 주형
정확한 편집
(indel 부재)
Indel
표적 T•A가 A•T로 전환된
총 서열분석 판독물의 %
0
1
2
3
4
5
8
9
10
11
12
13
14
15
16
17
PBS 길이 (뉴클레오티드)

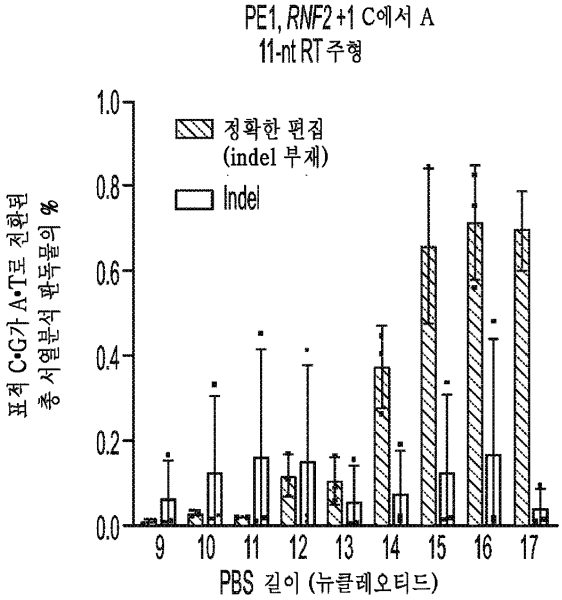
도면46b


PE1, EMX1 +5 G에서 T
13-nt RT 주형
정확한 편집
(indel 부재)
Indel
표적 G•C가 T•A로 전환된
총 서열분석 판독물의 %
0
1
2
3
4
9
10
11
12
13
14
15
16
17
PBS 길이 (뉴클레오티드)

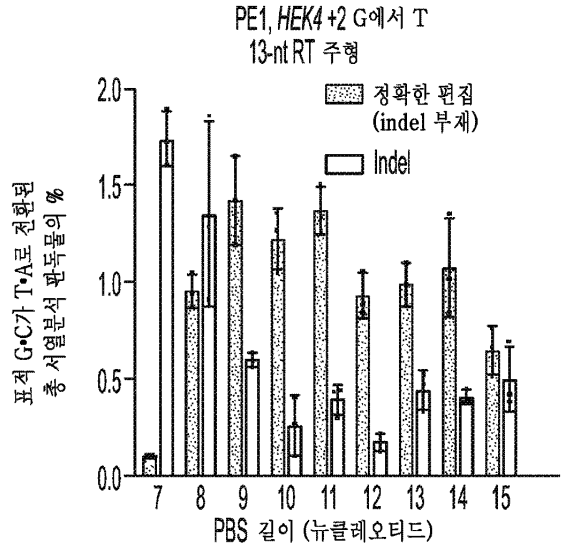
도면46c


PE1, FANCF +5 G에서 T
17-nt RT주형
정확한 편집
(indel 부재)
Indel
표적 G•C가 T•A로 전환된
총 서열분석 판독물의 %
0
2
4
6
8
8 9 10 11 12 13 14 15 16 17
PBS 길이 (뉴클레오티드)

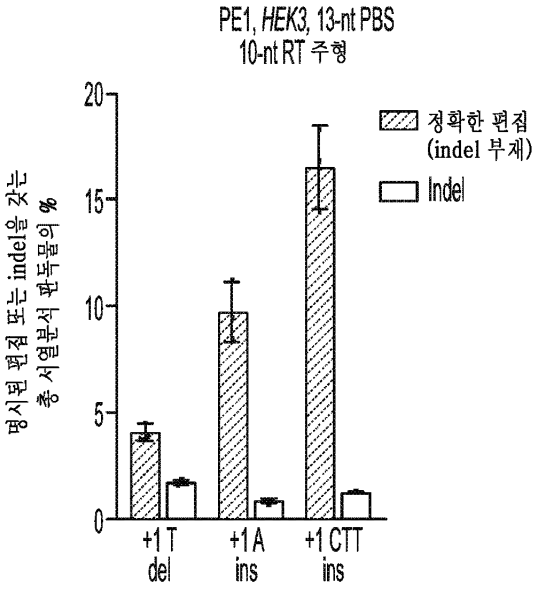
도면46d


PE1, RNF2 +1 C에서 A
11-nt RT주형
정확한 편집
(indel 부재)
Indel
표적 C•G가 A•T로 전환된
총 서열분석 판독물의 %
0.0
0.2
0.4
0.6
0.8
1.0
9 10 11 12 13 14 15 16 17
PBS 길이 (뉴클레오티드)

도면46e



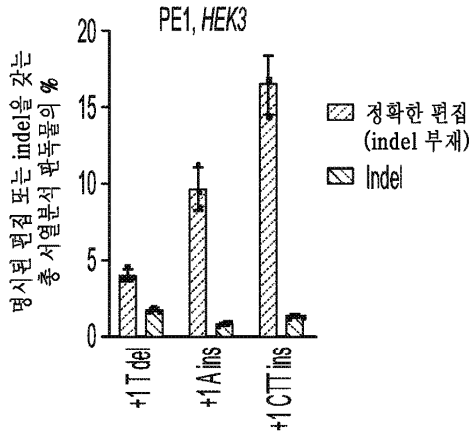
도면46f



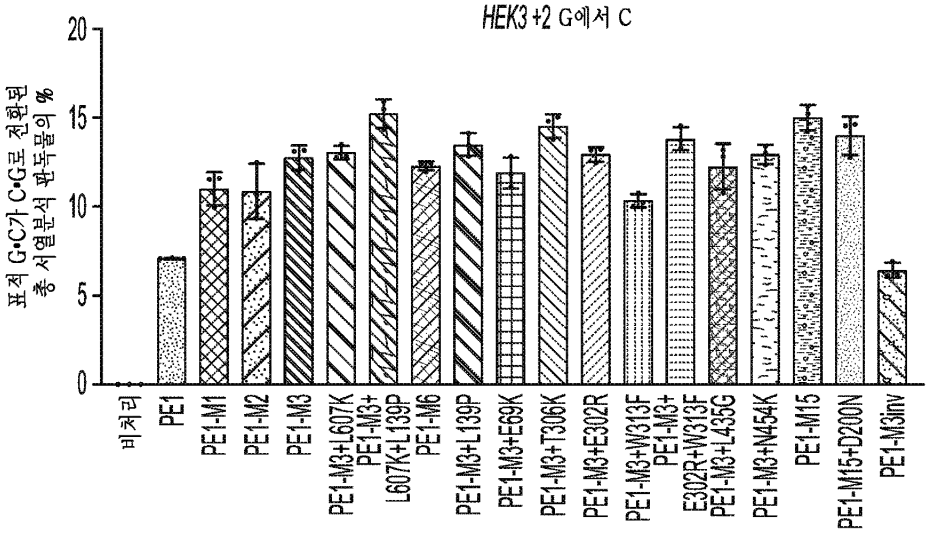
도면47a

PE1	Cas9 H840A-M-MLV RT (wt)
PE1-M1	Cas9 H840A-M-MLV RT (D200N)
PE1-M2	Cas9 H840A-M-MLV RT (D200N+L603W)
PE1-M3	Cas9 H840A-M-MLV RT (D200N+L603W+T330P)
PE1-M6	Cas9 H840A-M-MLV RT (D200N+L603W+T330P+D524G+E562Q+D583N)
PE1-M15	Cas9 H840A-M-MLV RT (P51L+S67R+E67K+T197A+H204R+E302K+F309N+W1331F+T330P+L345G+ N454K+D524G+D583N+H594Q+D653N)
PE1-M3inv	M-MLV RT (D200N+L603W+T330P)-Cas9 H840A
PE2	Cas9 H840A-M-MLV RT (D200N+L603W+T330P+T306K+W313F)

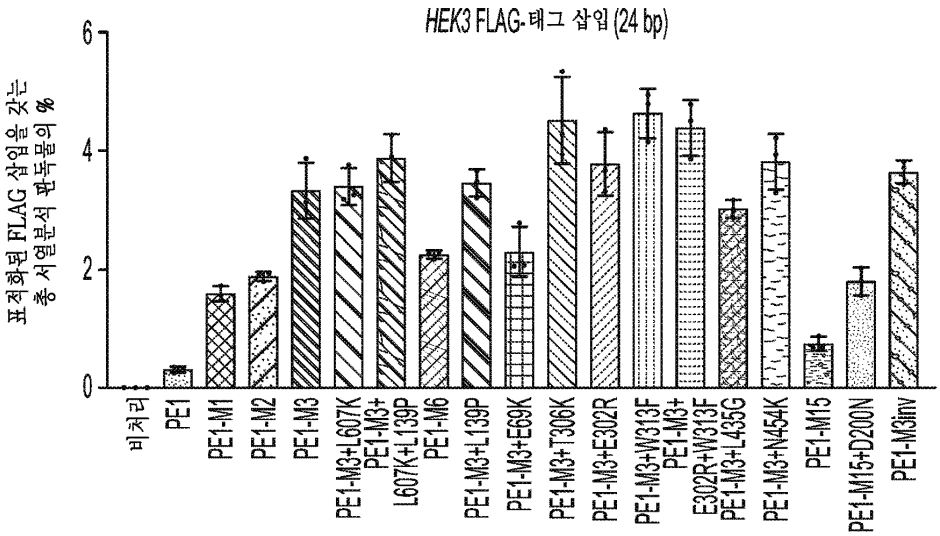
도면47b



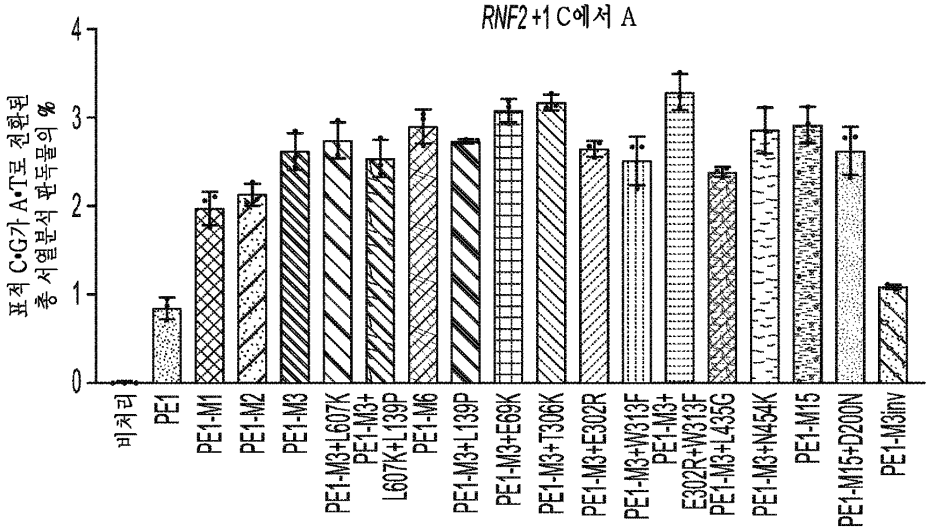
도면47c



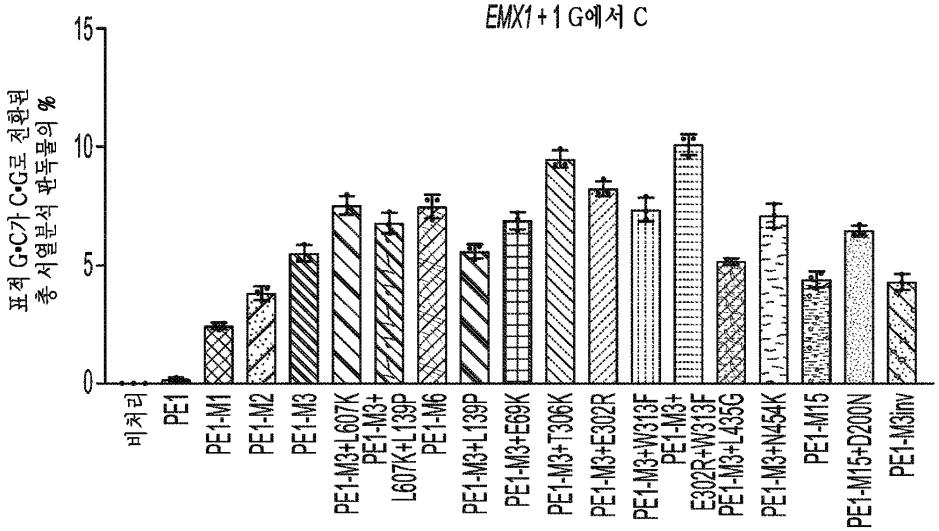
도면47d



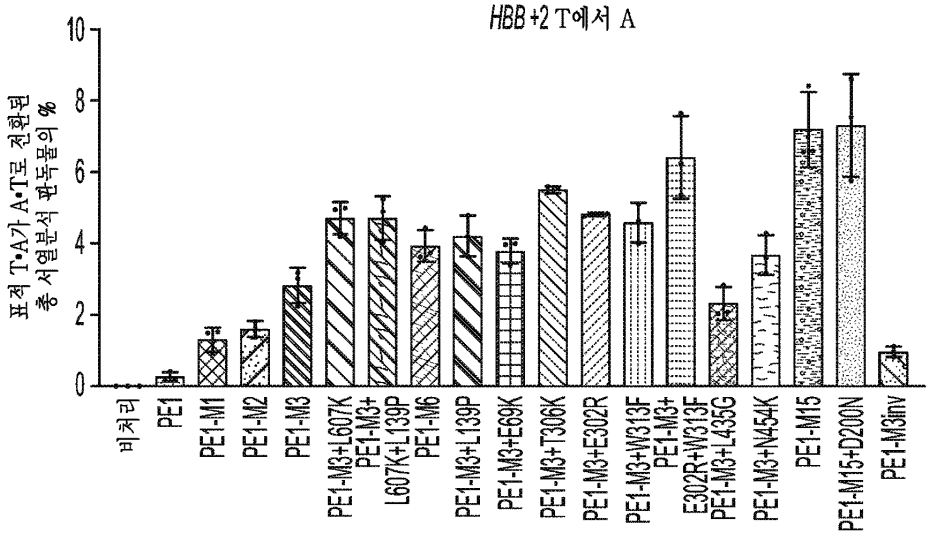
도면47e


RNF2 +1 C에서 A
표적 C•G가 A•T로 전환된
총 서열분석 판독물의 %
0
1
2
3
4
비처리
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+
L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+
E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

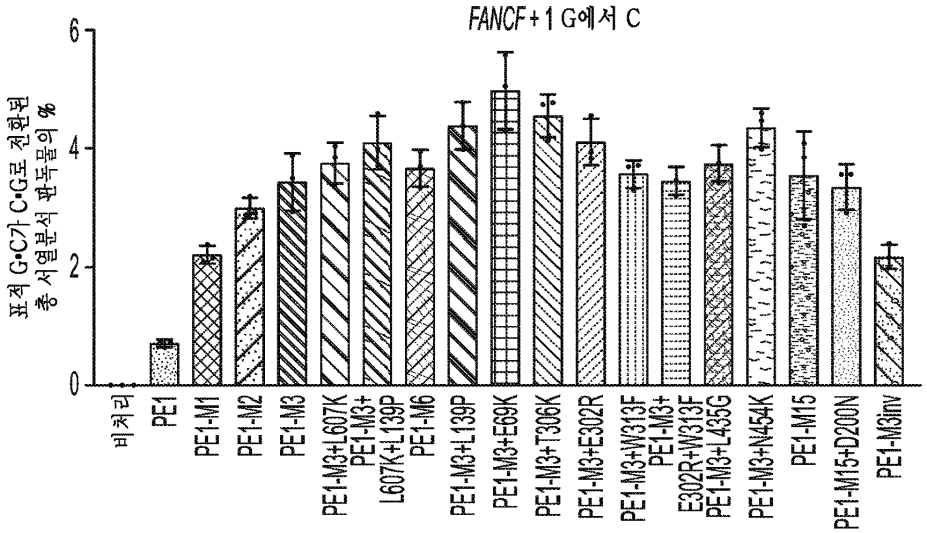
도면47f


EMX1 +1 G에서 C
표적 G•C가 C•G로 전환된
총 서열분석 판독물의 %
0
5
10
15
비처리
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+
L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+
E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

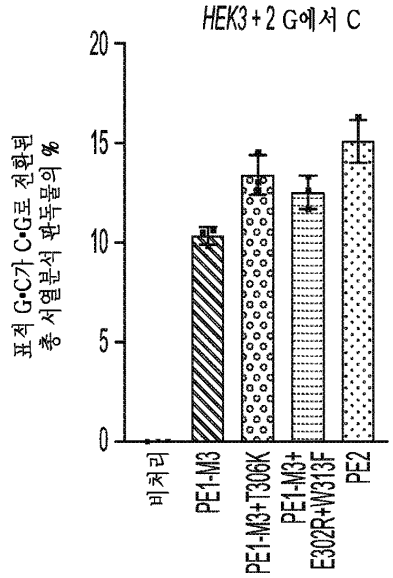
도면47g


HBB +2 T에서 A
표적 T•A가 A•T로 전환된 총 서열분석 판독물의 %
0
2
4
6
8
10
비처리
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+ L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+ E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

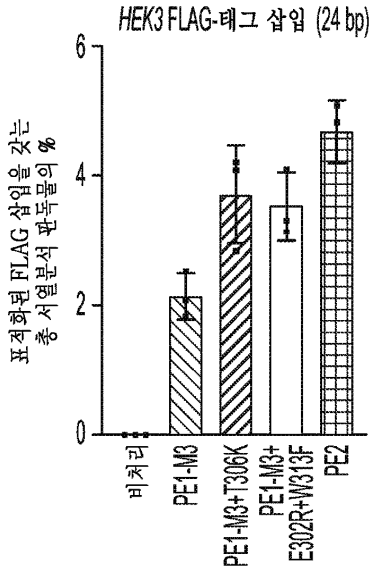
도면47h


FANCF +1 G에서 C
표적 G•C가 C•G로 전환된 총 서열분석 판독물의 %
0
2
4
6
비처리
PE1
PE1-M1
PE1-M2
PE1-M3
PE1-M3+L607K
PE1-M3+ L607K+L139P
PE1-M6
PE1-M3+L139P
PE1-M3+E69K
PE1-M3+T306K
PE1-M3+E302R
PE1-M3+W313F
PE1-M3+ E302R+W313F
PE1-M3+L435G
PE1-M3+N454K
PE1-M15
PE1-M15+D200N
PE1-M3inv

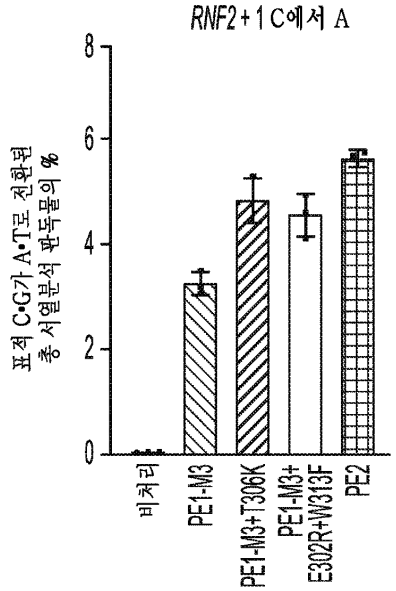
도면47i


HEK3+2 G에서 C
표적 G•C가 C•G로 전환된
총 서열분석 판독물의 %
20
15
10
5
0
비처리
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

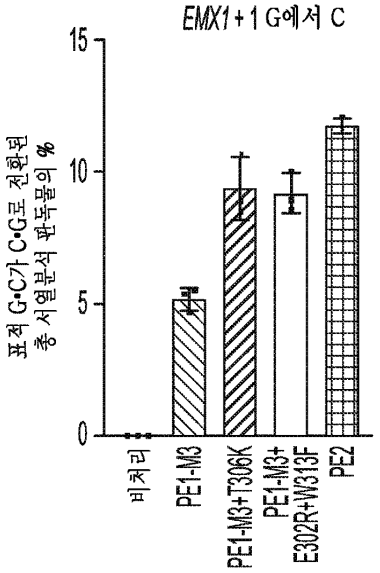
도면47j


HEK3 FLAG-태그 삽입 (24 bp)
표적화된 FLAG 삽입을 갖는
총 서열분석 판독물의 %
6
4
2
0
비처리
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

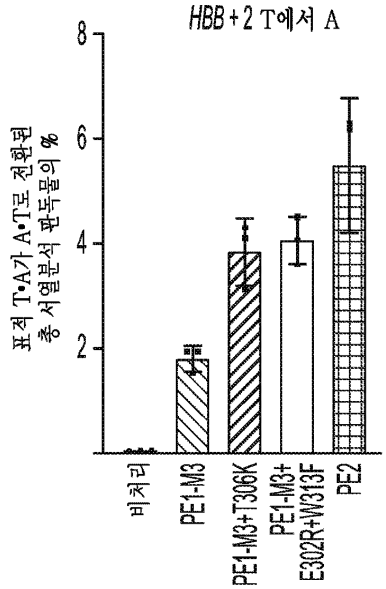
도면47k


RNF2+1C에서 A
표적 C•G가 A•T로 전환된
총 서열분석 판독물의 %
8
6
4
2
0
비처리
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

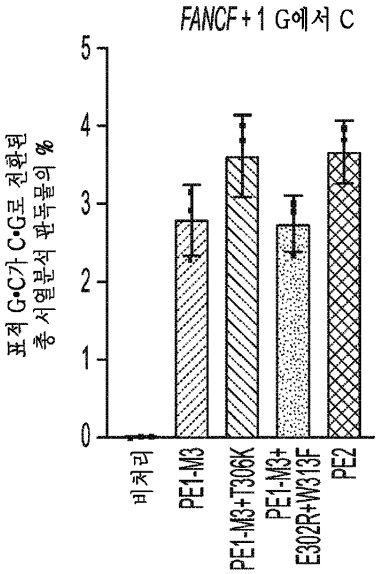
도면47l


EMX1+1G에서 C
표적 G•C가 C•G로 전환된
총 서열분석 판독물의 %
15
10
5
0
비처리
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

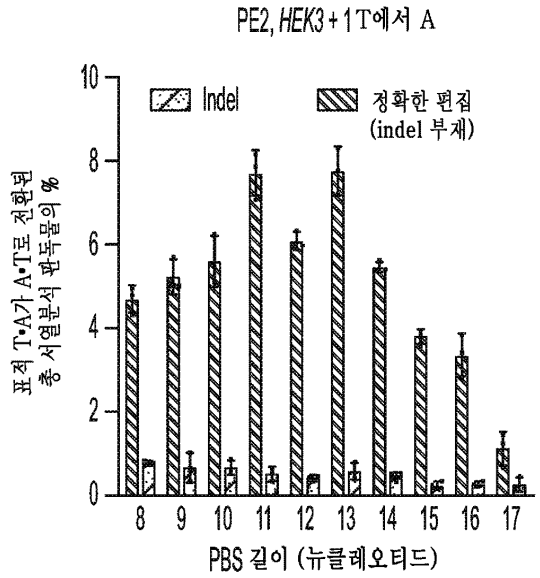
도면47m


HBB+2 T에서 A
표적 T•A가 A•T로 전환된
총 서열분석 판독물의 %
8
6
4
2
비처리
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

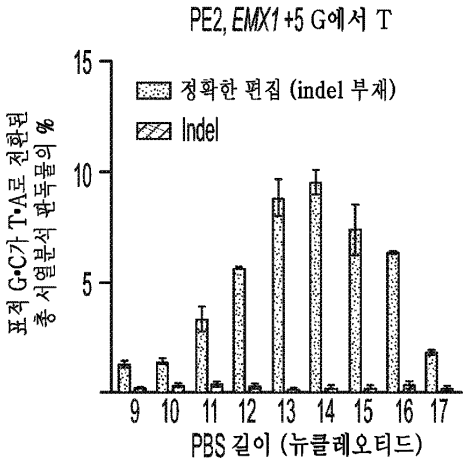
도면47n


FANCF+1 G에서 C
표적 G•C가 C•G로 전환된
총 서열분석 판독물의 %
5
4
3
2
1
0
비처리
PE1-M3
PE1-M3+T306K
PE1-M3+
E302R+W313F
PE2

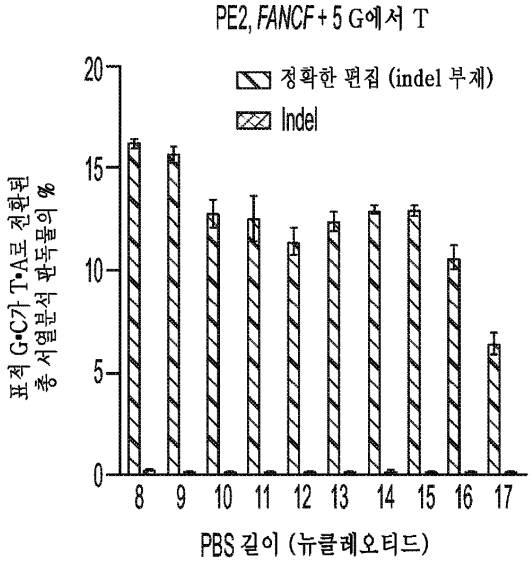
도면47o


PE2, HEK3+1T에서 A
Indel
정확한 편집
(indel 부재)
표적 T•A가 A•T로 전환된
총 서열분석 판독물의 %
0
2
4
6
8
10
8 9 10 11 12 13 14 15 16 17
PBS 길이 (뉴클레오티드)

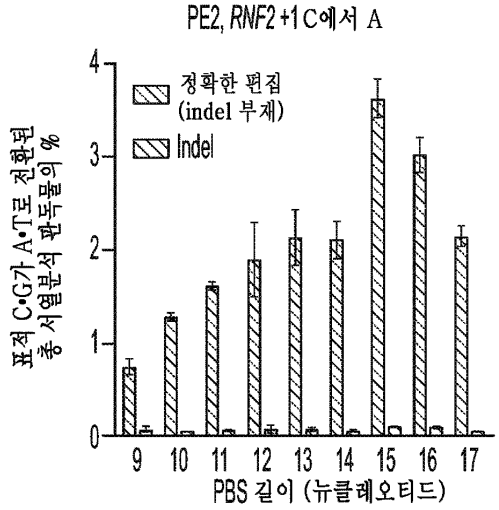
도면47p


PE2, EMX1+5 G에서 T
정확한 편집 (indel 부재)
Indel
표적 G•C가 T•A로 전환된
총 서열분석 판독물의 %
5
10
15
9 10 11 12 13 14 15 16 17
PBS 길이 (뉴클레오티드)

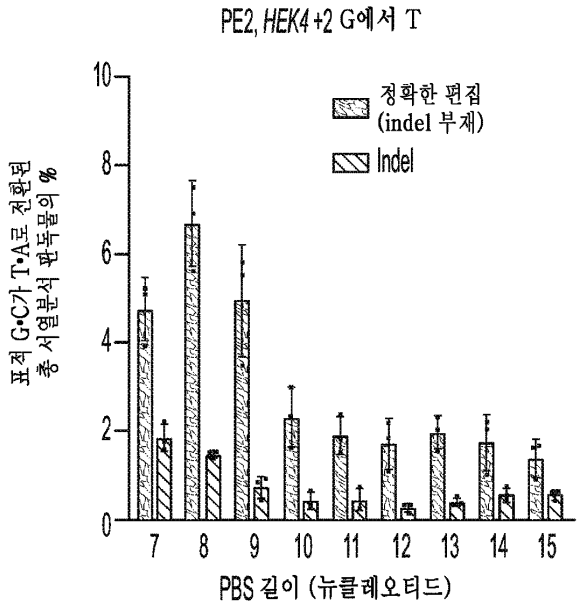
도면47q


PE2, FANCF + 5 G에서 T
정확한 편집 (indel 부재)
Indel
표적 G•C가 T•A로 전환된
총 서열분석 판독물의 %
20
15
10
5
0
8 9 10 11 12 13 14 15 16 17
PBS 길이 (뉴클레오티드)

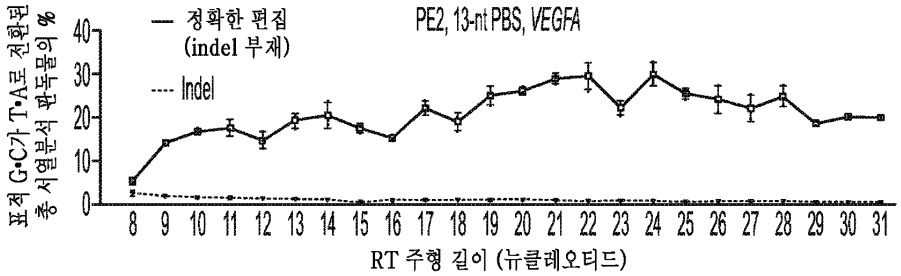
도면47r


PE2, RNF2 +1 C에서 A
정확한 편집
(indel 부재)
Indel
표적 C•G가 A•T로 전환된
총 서열분석 판독물의 %
4
3
2
1
0
9 10 11 12 13 14 15 16 17
PBS 길이 (뉴클레오티드)

도면47s



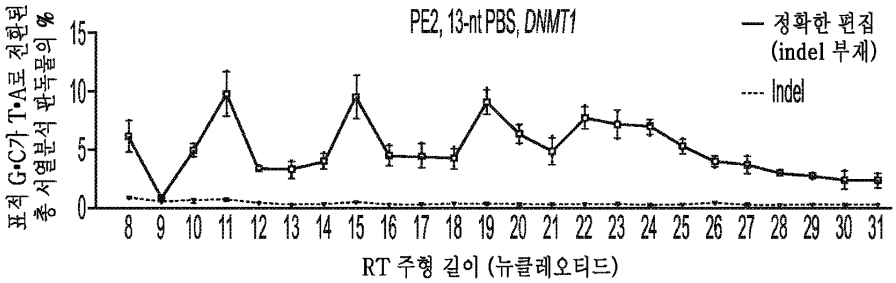
도면48a



마지막 주형화된 뉴클레오티드 T C C A G A T G G C A C A T T G T C A G A G G G

SEQ ID NO: 826

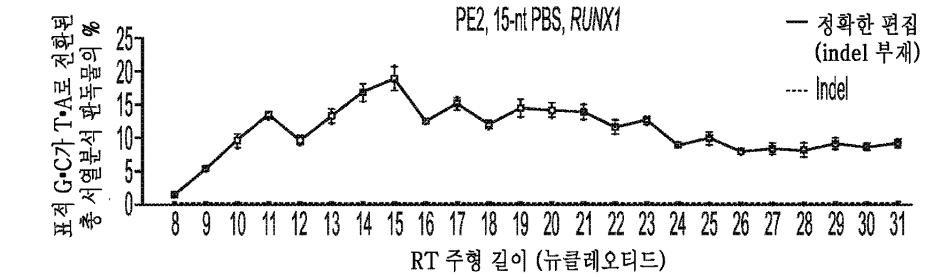
도면48b



마지막 주형화된 뉴클레오티드 T G A C G G G A G G G C A G A A C T A G T C C T

SEQ ID NO: 827

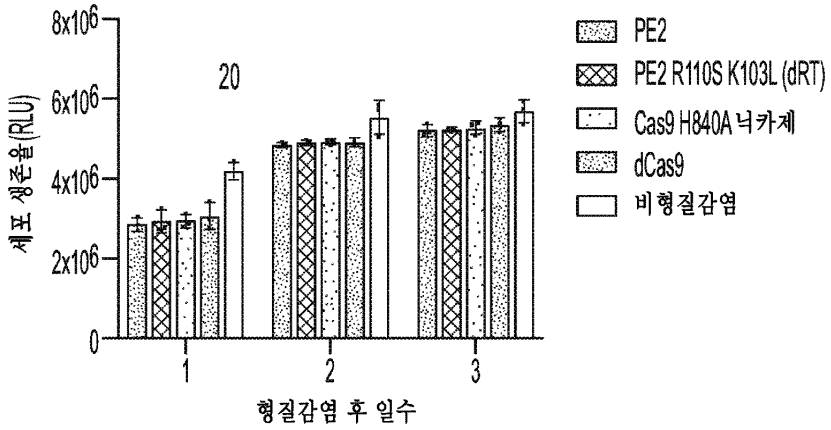
도면48c



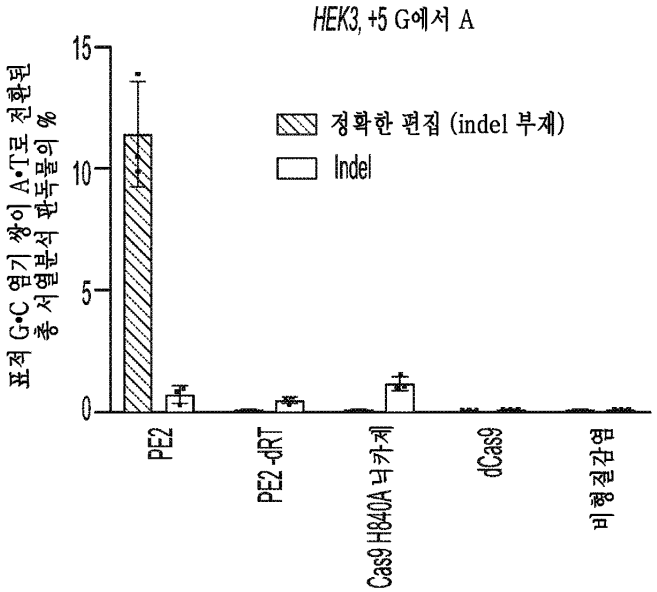
마지막 주형화된 뉴클레오티드 T T C A G A C A G C A T A T T T G A G T C A T T

SEQ ID NO: 828

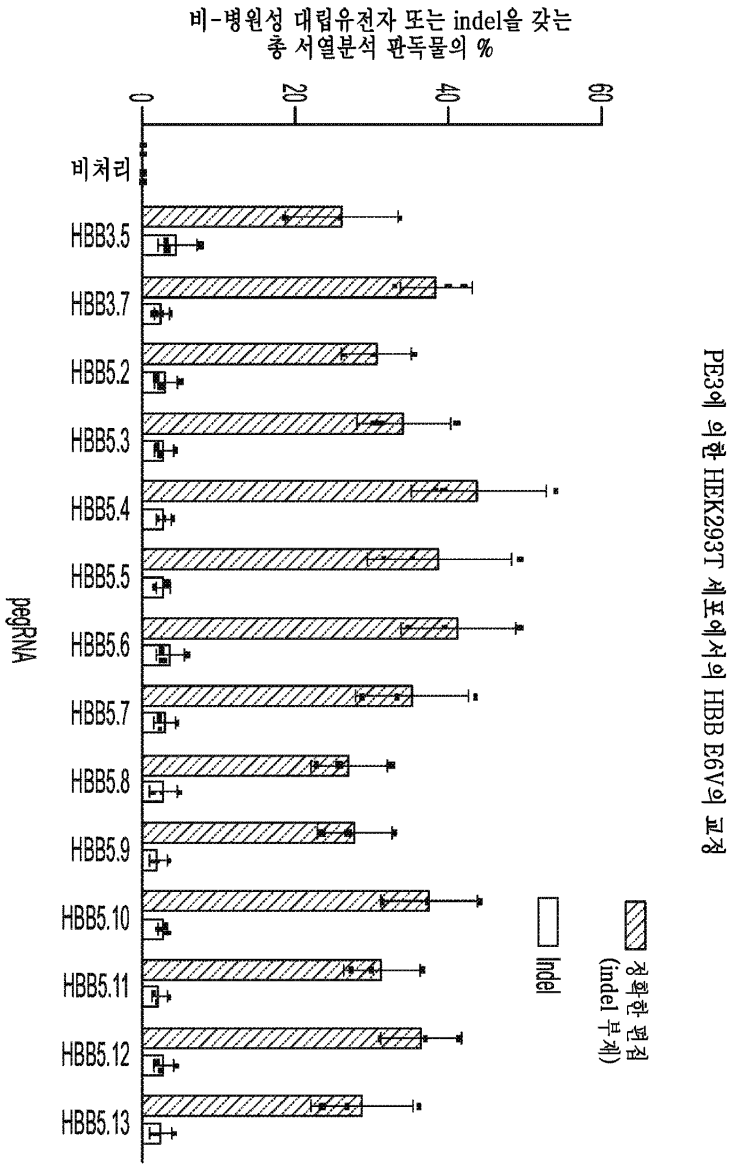
도면49a



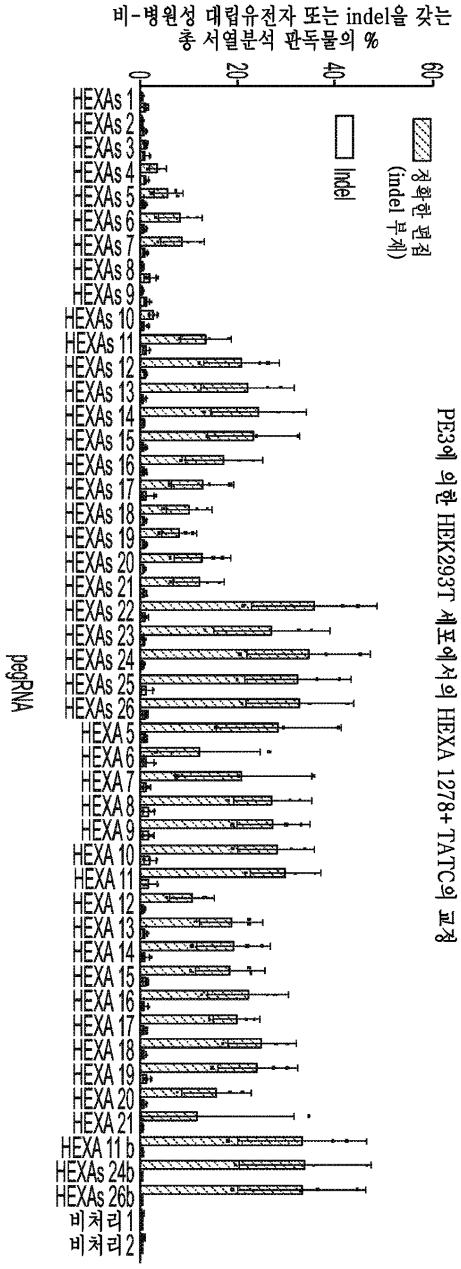
도면49b



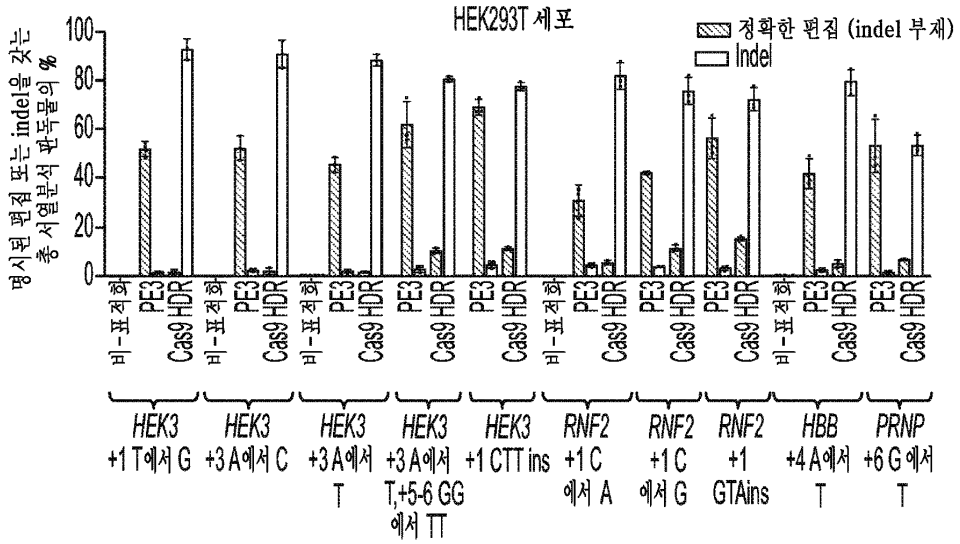
도면50a


PE3에 의한 HEK293T 세포에서의 HBB E6V의 교정
비-병원성 대립유전자 또는 indel을 갖는
총 서열분석 판독물의 %
0
20
40
60
pegRNA
비처리
HBB3.5
HBB3.7
HBB5.2
HBB5.3
HBB5.4
HBB5.5
HBB5.6
HBB5.7
HBB5.8
HBB5.9
HBB5.10
HBB5.11
HBB5.12
HBB5.13
정확한 편집
(indel 부재)
Indel

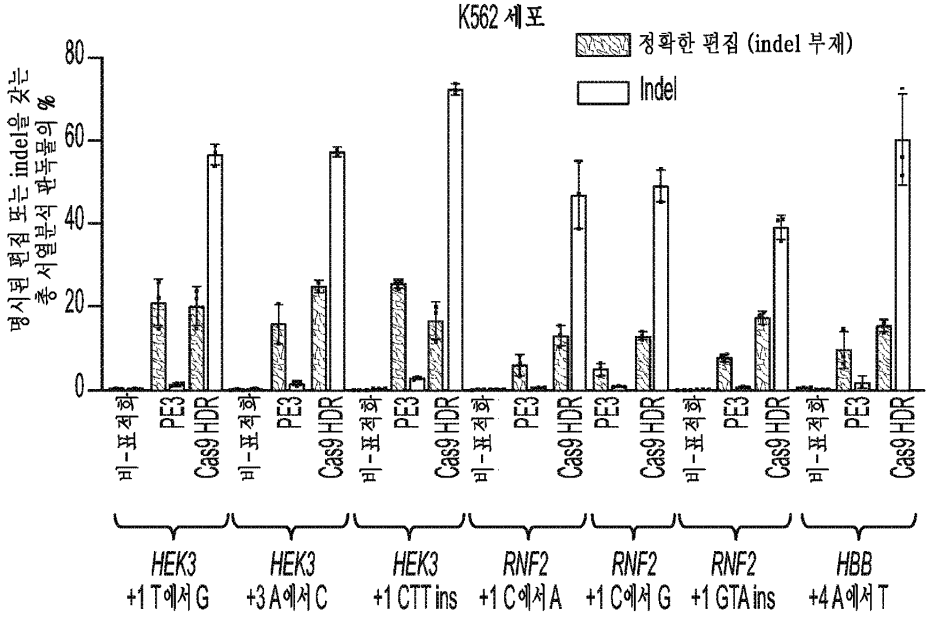
도면50b


PE3에 의한 HEK293T 세포에서의 HEXA 1278+TATC의 교정
비-병원성 대립유전자 또는 indel을 갖는
총 서열분석 판독물의 %
0
20
40
60
정확한 편집
(indel 부재)
indel
pegRNA
HEXAs 1
HEXAs 2
HEXAs 3
HEXAs 4
HEXAs 5
HEXAs 6
HEXAs 7
HEXAs 8
HEXAs 9
HEXAs 10
HEXAs 11
HEXAs 12
HEXAs 13
HEXAs 14
HEXAs 15
HEXAs 16
HEXAs 17
HEXAs 18
HEXAs 19
HEXAs 20
HEXAs 21
HEXAs 22
HEXAs 23
HEXAs 24
HEXAs 25
HEXAs 26
HEXA 5
HEXA 6
HEXA 7
HEXA 8
HEXA 9
HEXA 10
HEXA 11
HEXA 12
HEXA 13
HEXA 14
HEXA 15
HEXA 16
HEXA 17
HEXA 18
HEXA 19
HEXA 20
HEXA 21
HEXA 11 b
HEXAs 24b
HEXAs 26b
비처리 1
비처리 2

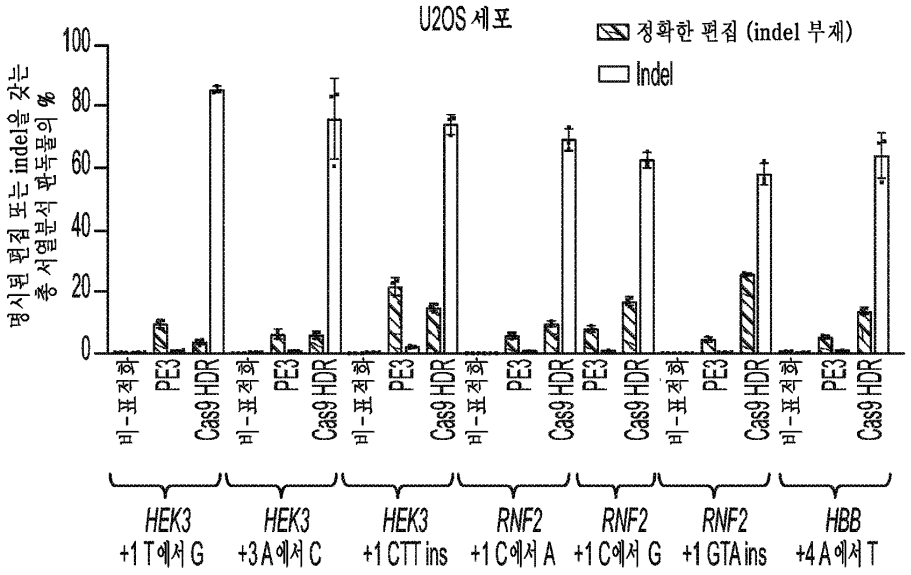
도면51a


HEK293T 세포
정확한 편집 (indel 부재)
Indel
명시된 편집 또는 indel을 갖는 총 서열분석 판독물의 %
100
80
60
40
20
0
비-표적화
PE3
Cas9 HDR
HEK3 +1 T에서 G
HEK3 +3 A에서 C
HEK3 +3 A에서 T
HEK3 +3 A에서 T,+5-6 GG에서 TT
HEK3 +1 CTT ins
RNF2 +1 C에서 A
RNF2 +1 C에서 G
RNF2 +1 GTAins
HBB +4 A에서 T
PRNP +6 G에서 T

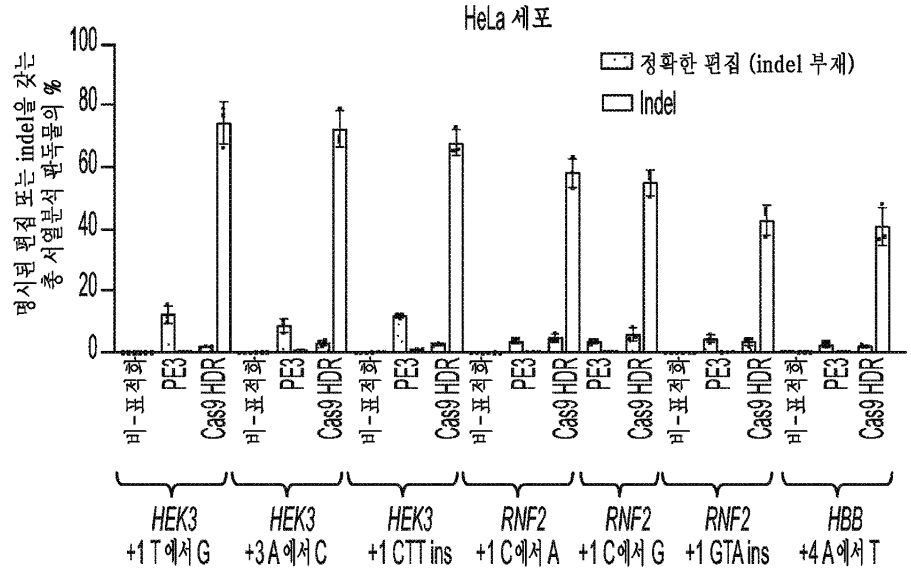
도면51b


K562 세포
정확한 편집 (indel 부재)
Indel
명시된 편집 또는 indel을 갖는 총 서열분석 판독물의 %
80
60
40
20
0
비-표적화
PE3
Cas9 HDR
HEK3 +1 T에서 G
HEK3 +3 A에서 C
HEK3 +1 CTT ins
RNF2 +1 C에서 A
RNF2 +1 C에서 G
RNF2 +1 GTA ins
HBB +4 A에서 T

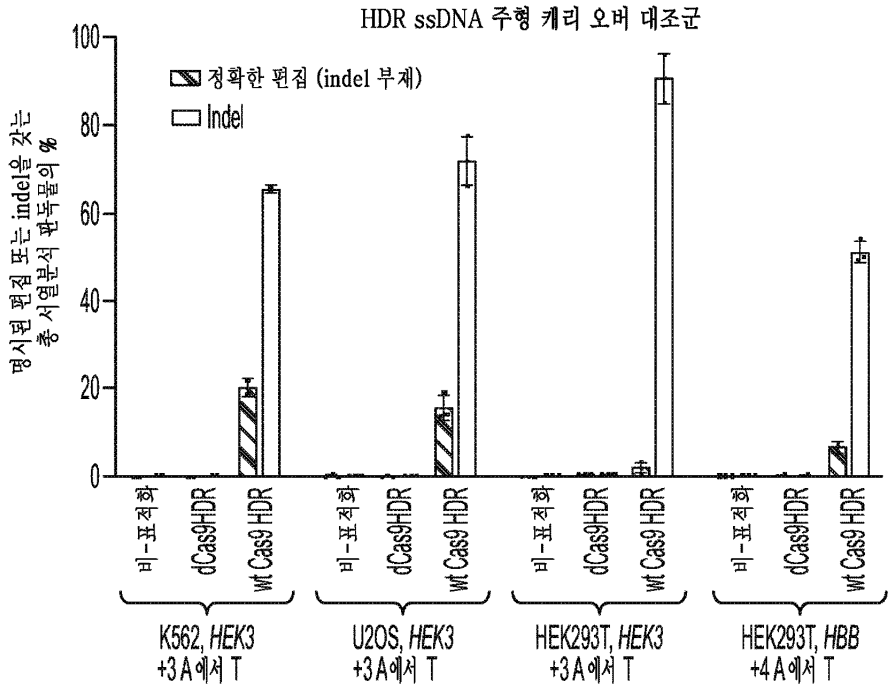
도면51c


U2OS 세포
정확한 편집 (indel 부재)
Indel
명시된 편집 또는 indel을 갖는 총 서열분석 판독물의 %
100
80
60
40
20
0
비-표적화
PE3
Cas9 HDR
HEK3 +1 T에서 G
HEK3 +3 A에서 C
HEK3 +1 CTT ins
RNF2 +1 C에서 A
RNF2 +1 C에서 G
RNF2 +1 GTA ins
HBB +4 A에서 T

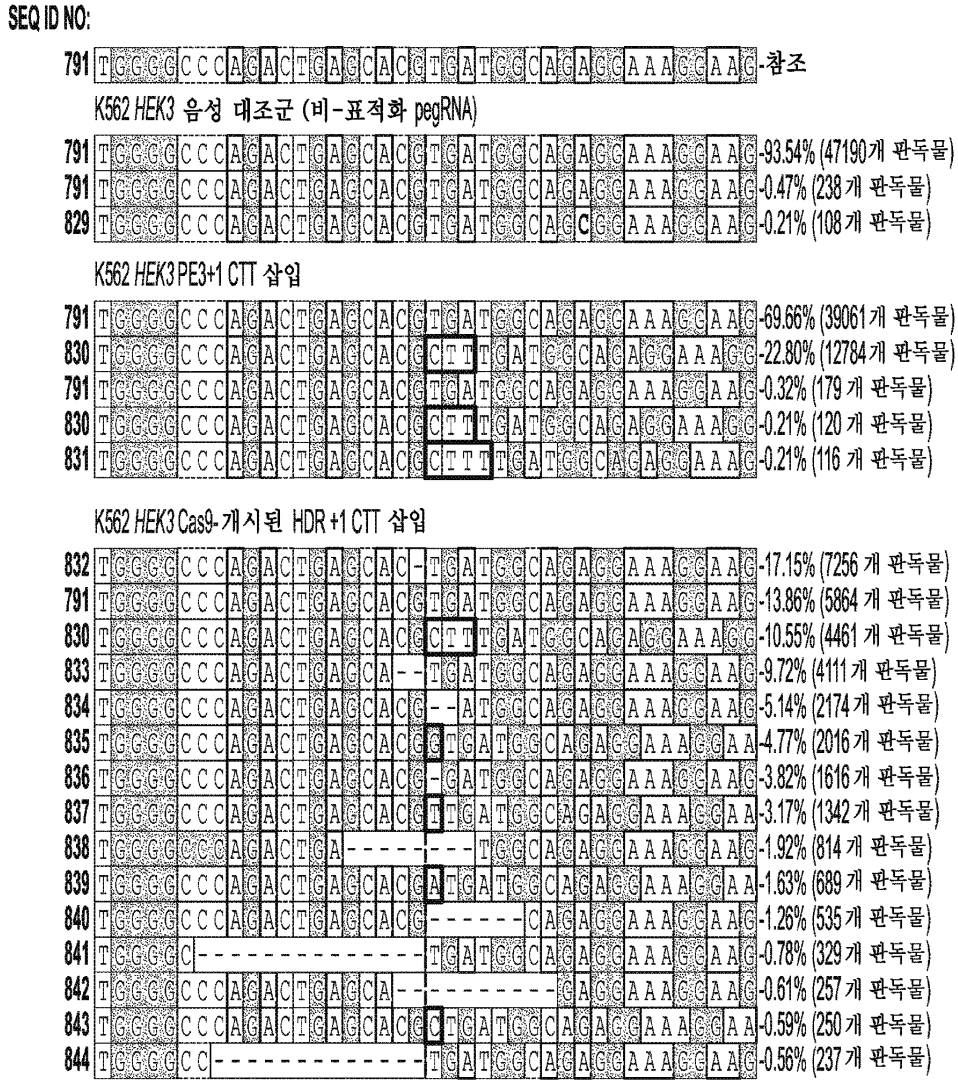
도면51d


HeLa 세포
정확한 편집 (indel 부재)
Indel
명시된 편집 또는 indel을 갖는 총 서열분석 판독물의 %
100
80
60
40
20
0
비-표적화
PE3
Cas9 HDR
HEK3 +1 T에서 G
HEK3 +3 A에서 C
HEK3 +1 CTT ins
RNF2 +1 C에서 A
RNF2 +1 C에서 G
RNF2 +1 GTA ins
HBB +4 A에서 T

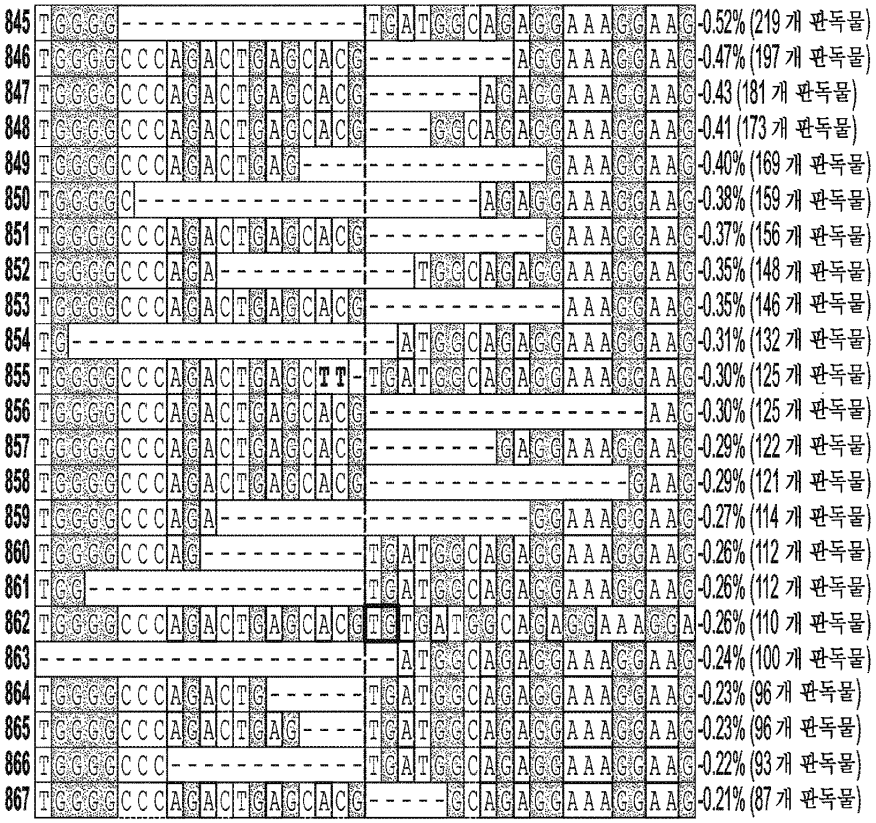
도면51e


HDR ssDNA 주형 캐리 오버 대조군
정확한 편집 (indel 부재)
Indel
명시된 편집 또는 indel을 갖는 총 서열분석 판독물의 %
100
80
60
40
20
0
비-표적화
dCas9HDR
wt Cas9 HDR
K562, HEK3 +3 A에서 T
U2OS, HEK3 +3 A에서 T
HEK293T, HEK3 +3 A에서 T
HEK293T, HBB +4 A에서 T

도면51f


SEQ ID NO:
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -참조
K562 HEK3 음성 대조군 (비-표적화 pegRNA)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -93.54% (47190개 판독물)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -0.47% (238 개 판독물)
829 TGGGGCCCAGACTGAGCACGTGATGGCAGCGGAAAGGAAG -0.21% (108 개 판독물)
K562 HEK3 PE3+1 CTT 삽입
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -69.66% (39061개 판독물)
830 TGGGGCCCAGACTGAGCACGCTTTGATGGCAGAGGAAAGGG -22.80% (12784개 판독물)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -0.32% (179 개 판독물)
830 TGGGGCCCAGACTGAGCACGCTTTGATGGCAGAGGAAAGGG -0.21% (120 개 판독물)
831 TGGGGCCCAGACTGAGCACGCTTTTGATGGCAGAGGAAAG -0.21% (116 개 판독물)
K562 HEK3 Cas9-개시된 HDR +1 CTT 삽입
832 TGGGGCCCAGACTGAGCAC-TGATGGCAGAGGAAAGGAAG -17.15% (7256 개 판독물)
791 TGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG -13.86% (5864 개 판독물)
830 TGGGGCCCAGACTGAGCACGCTTTGATGGCAGAGGAAAGGG -10.55% (4461 개 판독물)
833 TGGGGCCCAGACTGAGCA--TGATGGCAGAGGAAAGGAAG -9.72% (4111 개 판독물)
834 TGGGGCCCAGACTGAGCACG--ATGGCAGAGGAAAGGAAG -5.14% (2174 개 판독물)
835 TGGGGCCCAGACTGAGCACGGTGATGGCAGAGGAAAGGAA -4.77% (2016 개 판독물)
836 TGGGGCCCAGACTGAGCACG-GATGGCAGAGGAAAGGAAG -3.82% (1616 개 판독물)
837 TGGGGCCCAGACTGAGCACGTTGATGGCAGAGGAAAGGAA -3.17% (1342 개 판독물)
838 TGGGGCCCAGACTGA-----TGGCAGAGGAAAGGAAG -1.92% (814 개 판독물)
839 TGGGGCCCAGACTGAGCACGATGATGGCAGAGGAAAGGAA -1.63% (689 개 판독물)
840 TGGGGCCCAGACTGAGCACG------CAGAGGAAAGGAAG -1.26% (535 개 판독물)
841 TGGGGC----------TGATGGCAGAGGAAAGGAAG -0.78% (329 개 판독물)
842 TGGGGCCCAGACTGAGCA--------GAGGAAAGGAAG -0.61% (257 개 판독물)
843 TGGGGCCCAGACTGAGCACGCTGATGGCAGAGGAAAGGAA -0.59% (250 개 판독물)
844 TGGGGCC----------TGATGGCAGAGGAAAGGAAG -0.56% (237 개 판독물)

도면51g



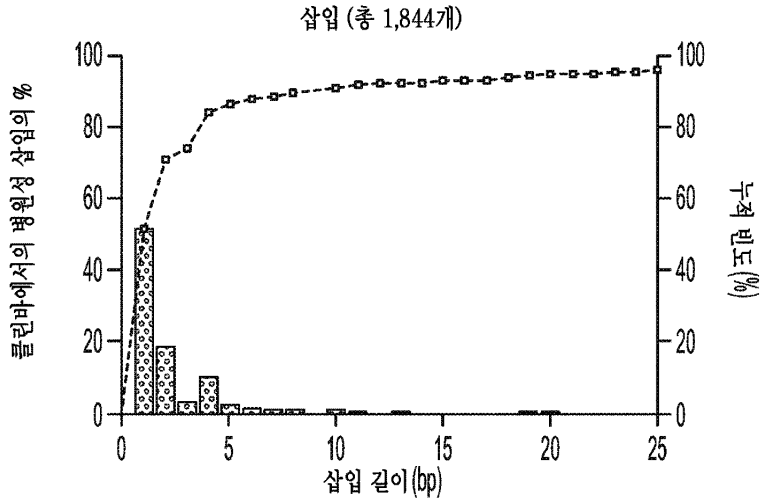
볼드체 치환

[ ] 삽입

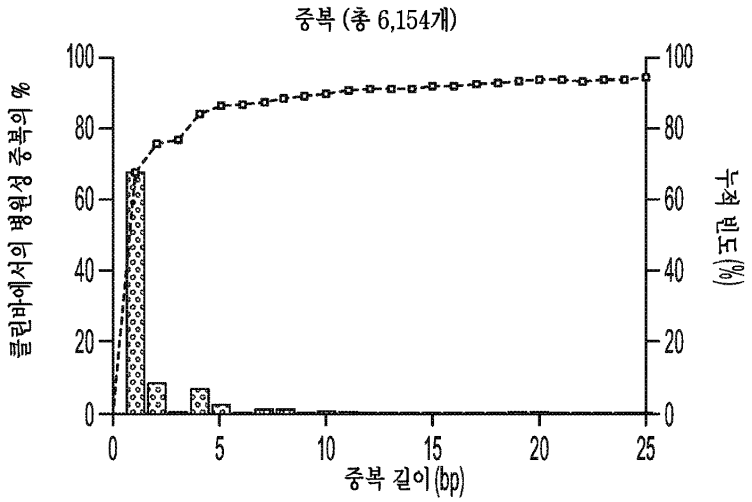
- 결실

···· 예측된 절단 위치

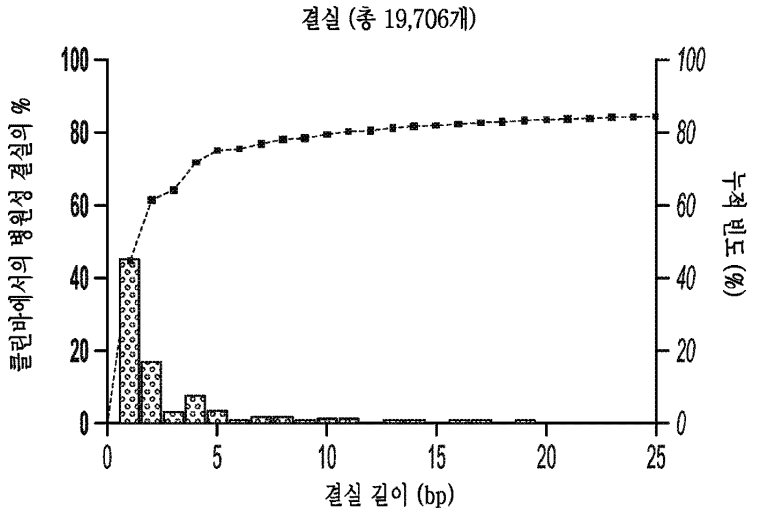
도면52a



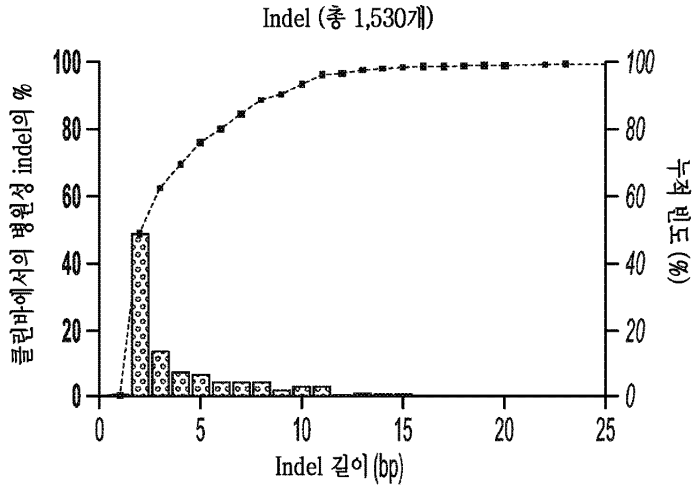
도면52b



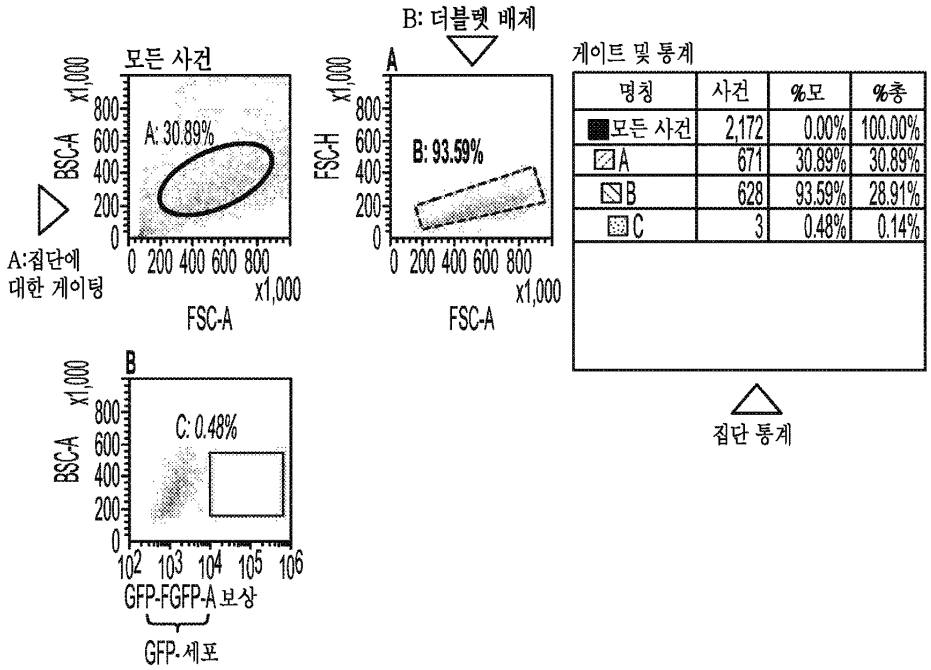
도면52c



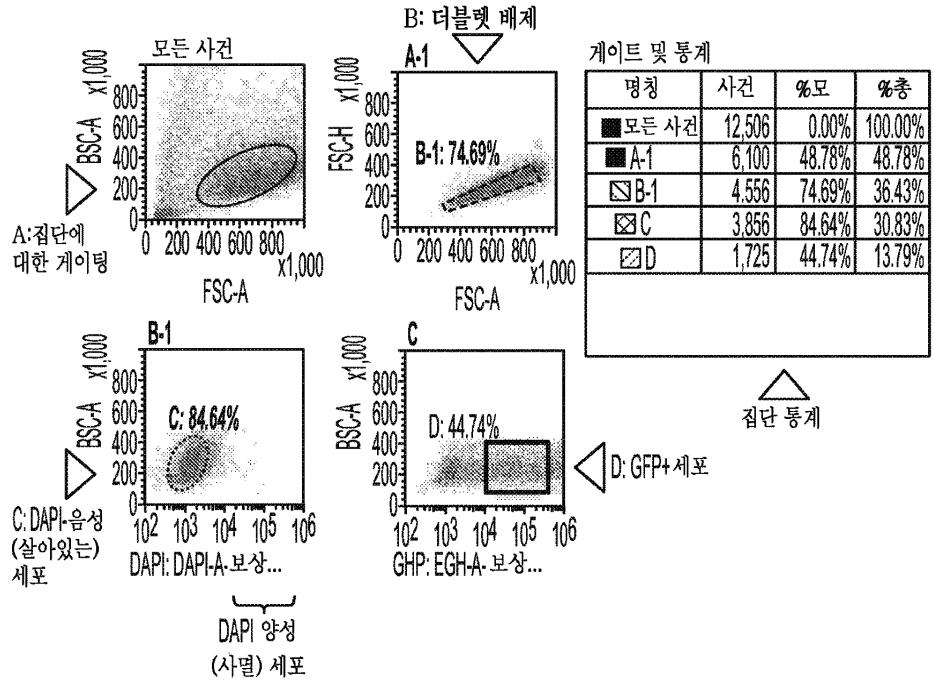
도면52d



도면53a

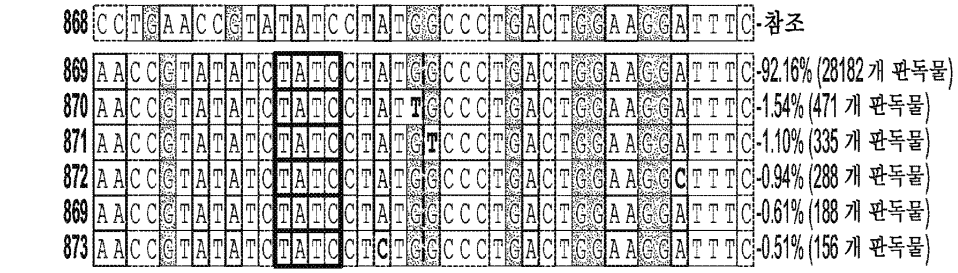

모든 사건
BSC-A x1,000
A: 30.89%
FSC-A
A:집단에 대한 게이팅
B: 더블렛 배제
A
FSC-H x1,000
B: 93.59%
FSC-A
게이트 및 통계
명칭 사건 %모 %총
모든 사건 2,172 0.00% 100.00%
A 671 30.89% 30.89%
B 628 93.59% 28.91%
C 3 0.48% 0.14%
집단 통계
B
BSC-A x1,000
C: 0.48%
GFP-FGFP-A 보상
GFP-세포

도면53b

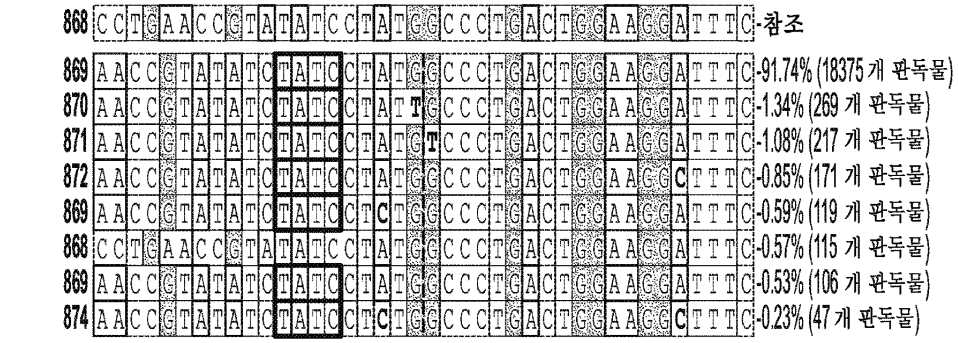

모든 사건
BSC-A x1,000
A:집단에 대한 게이팅
FSC-A
B: 더블렛 배제
A-1
FSC-H x1,000
B-1: 74.69%
FSC-A
게이트 및 통계
명칭 사건 %모 %총
모든 사건 12,506 0.00% 100.00%
A-1 6,100 48.78% 48.78%
B-1 4,556 74.69% 36.43%
C 3,856 84.64% 30.83%
D 1,725 44.74% 13.79%
집단 통계
B-1
BSC-A x1,000
C: 84.64%
C: DAPI-음성 (살아있는) 세포
DAPI: DAPI-A-보상...
DAPI 양성 (사멸) 세포
C
BSC-A x1,000
D: 44.74%
D: GFP+세포
GHP: EGH-A-보상...

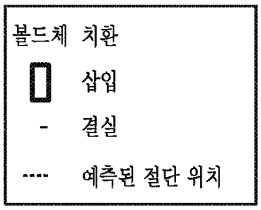
도면53c

세포주 1

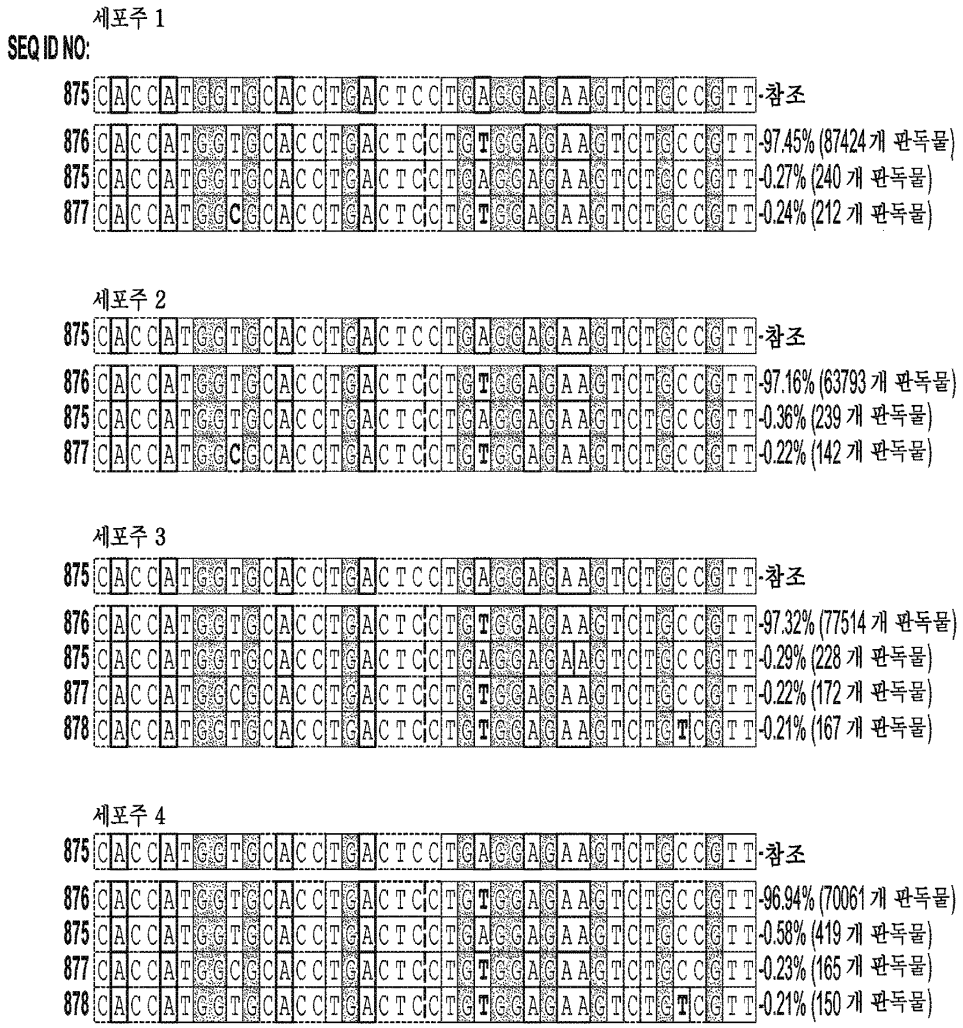


세포주 2

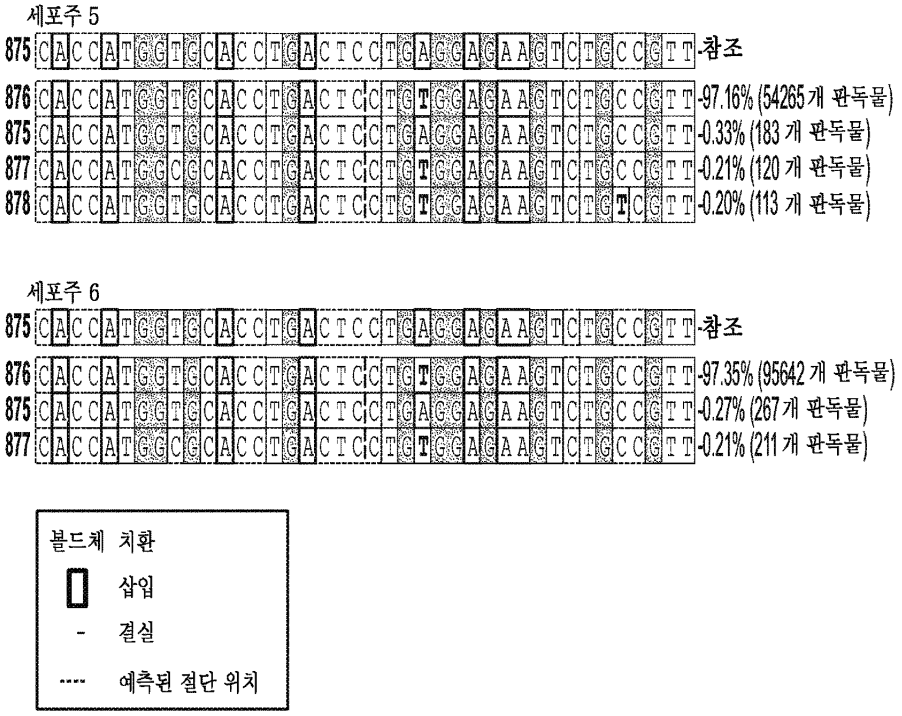




도면53d


세포주 1
SEQ ID NO:
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-참조
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-97.45% (87424 개 판독물)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-0.27% (240 개 판독물)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-0.24% (212 개 판독물)
세포주 2
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-참조
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-97.16% (63793 개 판독물)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-0.36% (239 개 판독물)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-0.22% (142 개 판독물)
세포주 3
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-참조
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-97.32% (77514 개 판독물)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-0.29% (228 개 판독물)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-0.22% (172 개 판독물)
878 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGTCGTT-0.21% (167 개 판독물)
세포주 4
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-참조
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-96.94% (70061 개 판독물)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-0.58% (419 개 판독물)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT-0.23% (165 개 판독물)
878 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGTCGTT-0.21% (150 개 판독물)

도면53e


세포주 5
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -참조
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -97.16% (54265 개 판독물)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.33% (183 개 판독물)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.21% (120 개 판독물)
878 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGTCGTT -0.20% (113 개 판독물)
세포주 6
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -참조
876 CACCATGGTGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -97.35% (95642 개 판독물)
875 CACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT -0.27% (267 개 판독물)
877 CACCATGGCGCACCTGACTCCTGTGGAGAAGTCTGCCGTT -0.21% (211 개 판독물)
볼드체 치환
삽입
- 결실
---- 예측된 절단 위치

도면54

스페이서 서열 (표적 부위) 및 PBS 및 RT 주형을 함유하는 pegRNA 3' 연장부를 인간 U6 프로모터 pegRNA 발현 벡터 내로 동시에 클로닝하기 위한 프로토콜:

[성분 1] 비소화된 pU6-pegRNA-GG-벡터 개략도:



플라스미드	설명	특징부	하기로 절단
pU6-pegRNA-GG-벡터	인간 pegRNA U6발현 벡터	U6_BsaI-mRFP-카세트_U6-Term	BsaI-HFV2

[성분 2] 인간 pegRNA 벡터에 대한 스페이서 (표적) 서열을 생성하기 위해 올리고뉴클레오티드를 정렬함:

상단 및 하단 올리고를 어닐링한 다음 소화된 pU6-pegRNA-GG-벡터 내로 조립함:

SEQ ID NO: 879 5'-CACCNNNNNNNNNNNNNNNNNNNGTTTT -3'

SEQ ID NO: 880 3'- NNNNNNNNNNNNNNNNNNNCAAAATCTC-5'

(제시된 N20 모의 프로토스페이서 서열은 필요에 따라 가변적 길이를 가질 수 있음)

[성분 3] 인간 pegRNA 벡터에 대한 SpCas9 sgRNA 스캐폴드를 생성하기 위해 올리고뉴클레오티드를 정렬함:

상단 및 하단 올리고를 어닐링한 다음 소화된 pU6-pegRNA-GG-벡터 내로 조립함

{Phos}5'-AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG -3' SEQ ID NO: 881

SEQ ID NO: 882 3'- GATCTTTATCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACG-5'{Phos}

(이들 올리고는 pegRNA 벡터 조립을 허용하기 위해 5' 인산화되어야 한다는 것을 주목함)

[성분 4] 인간 pegRNA 벡터에 대한 PBS 및 RT 주형 서열을 함유하는 pegRNA 3' 연장부를 생성하기 위해 올리고뉴클레오티드를 정렬함:

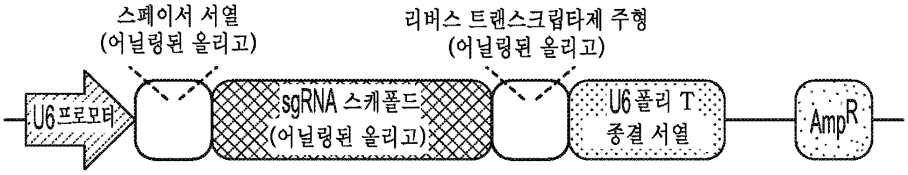
상단 및 하단 올리고를 어닐링한 다음 소화된 pU6-pegRNA-GG-벡터 내로 조립함:

SEQ ID NO: 883 5'-GTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNN -3'

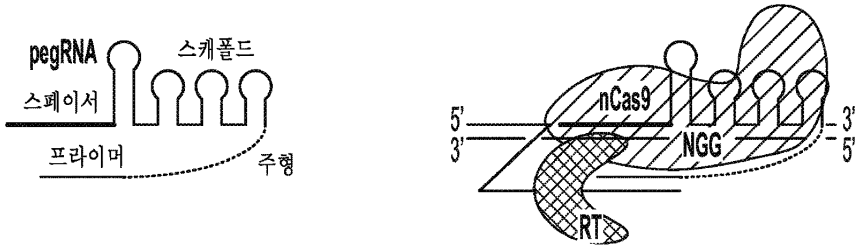
SEQ ID NO: 884 3'- NNNNNNNNNNNNNNNNNNNNNNNNNNNNAAA-5'

(제시된 N30 모의 리버스 트랜스크립타제 주형 서열은 필요에 따라 가변적 길이를 가질 수 있음)

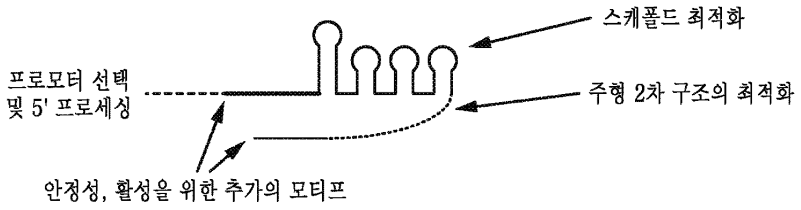
성분 1, 2, 3 및 4로부터의 최종 조립된 U6 pegRNA 발현 벡터의 개략도:



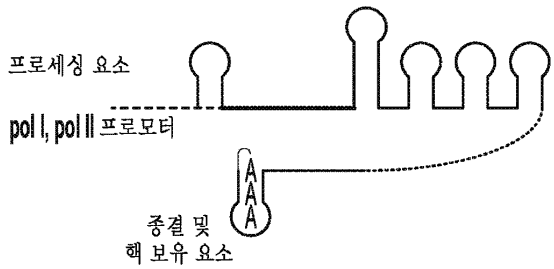
도면55a



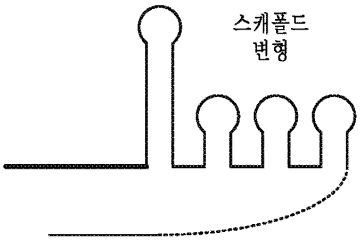
도면55b



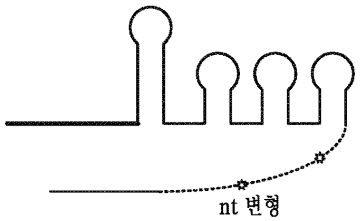
도면55c



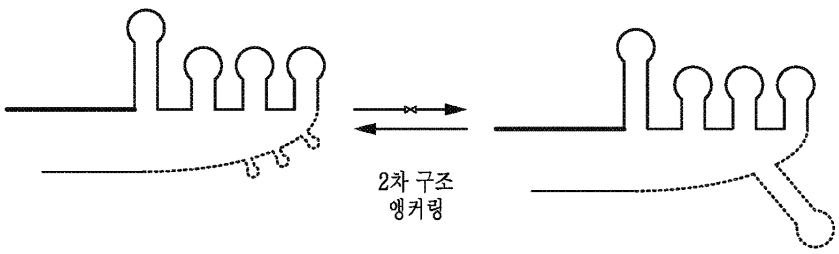
도면55d



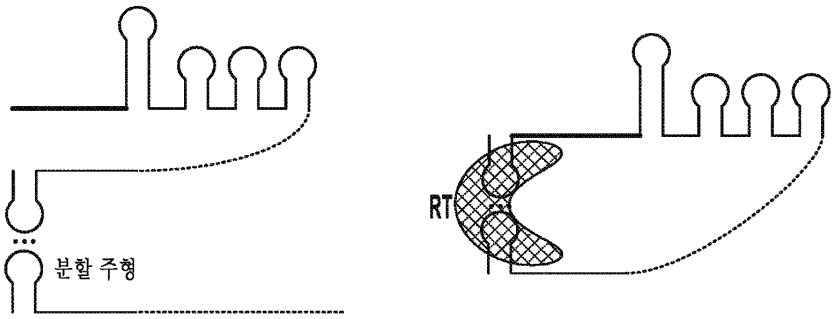
도면55e



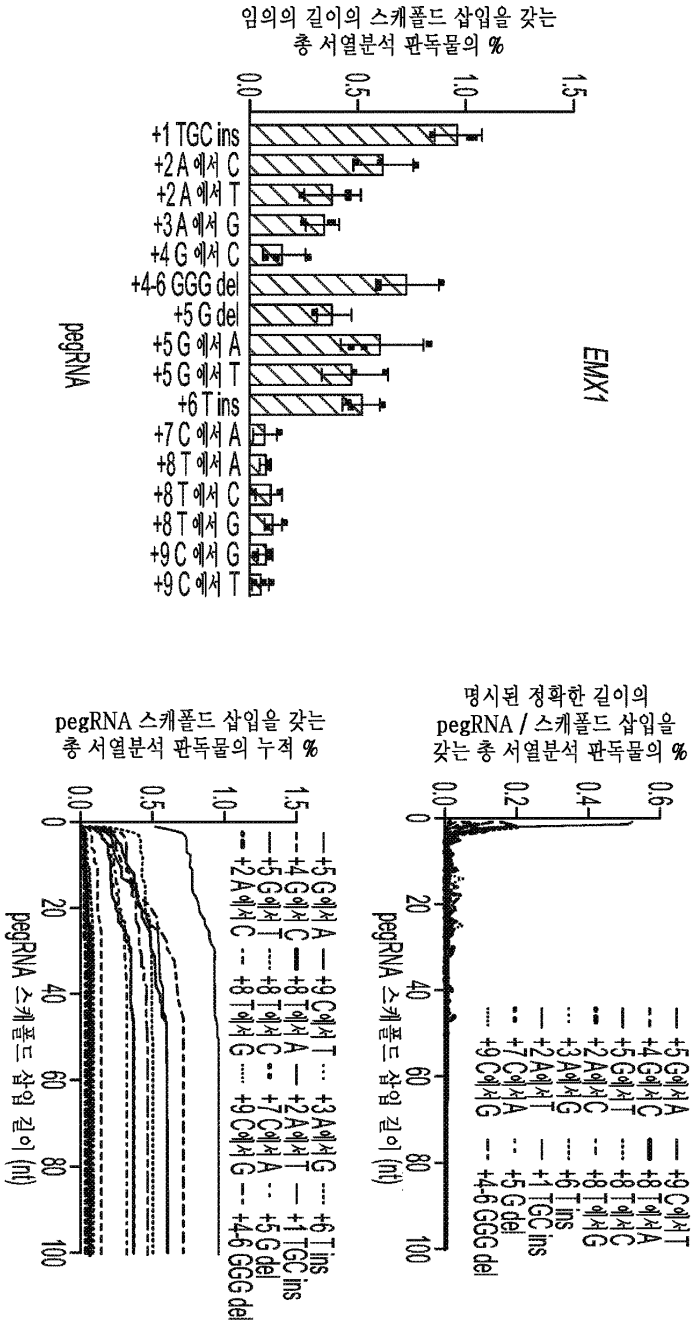
도면55f



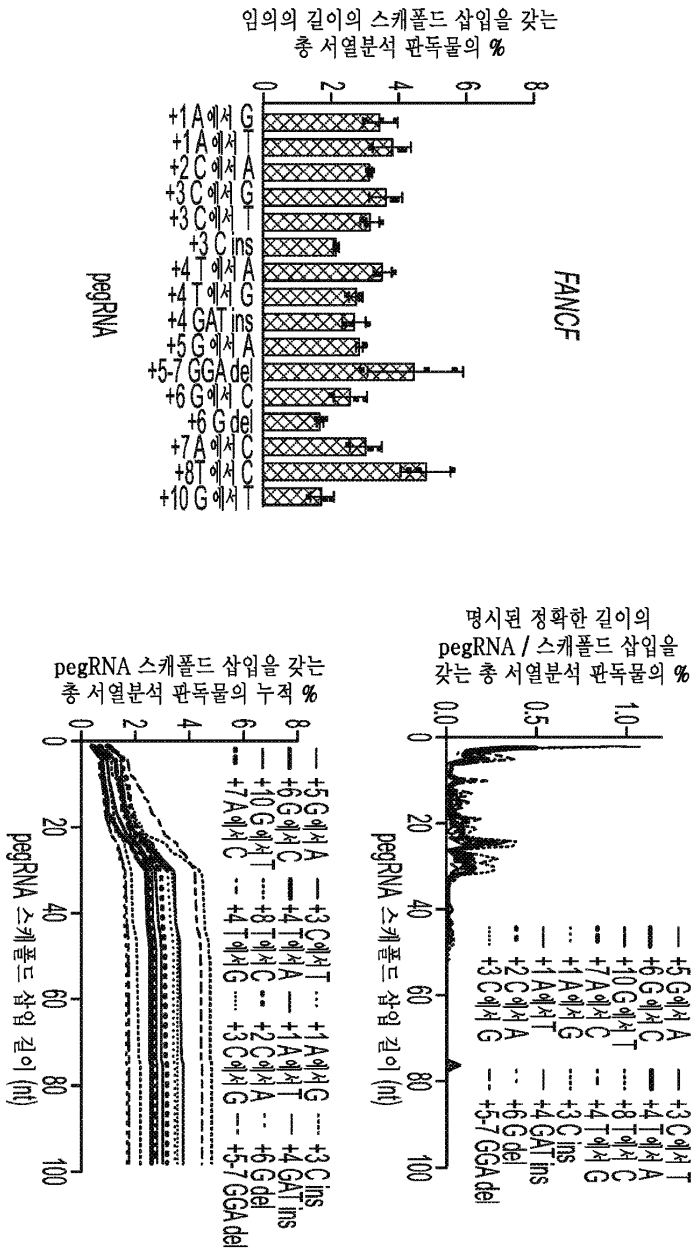
도면55g


분할 주형
RT

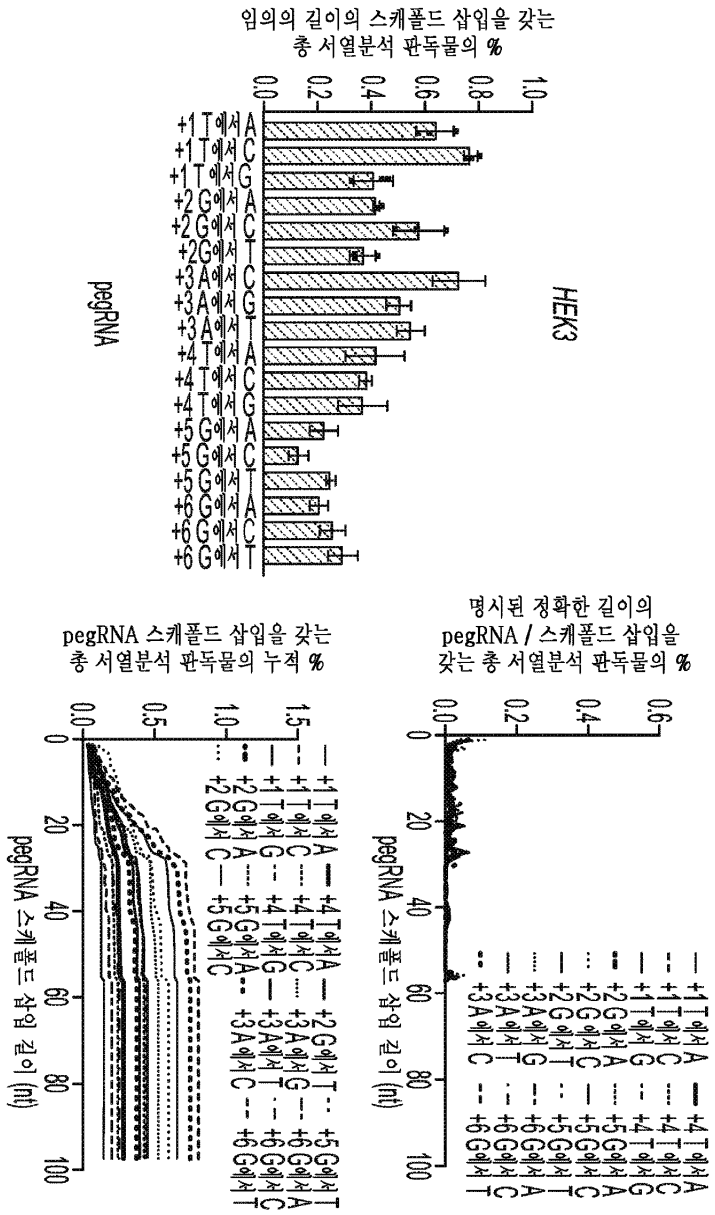
도면56a


임의의 길이의 스케폴드 삽입을 갖는
총 서열분석 판독물의 %
0.0
0.5
1.0
1.5
+1 TGC ins
+2 A 에서 C
+2 A 에서 T
+3 A 에서 G
+4 G 에서 C
+4-6 GGG del
+5 G del
+5 G 에서 A
+5 G 에서 T
+6 T ins
+7 C 에서 A
+8 T 에서 A
+8 T 에서 C
+8 T 에서 G
+9 C 에서 G
+9 C 에서 T
pegRNA
EMX1
명시된 정확한 길이의
pegRNA / 스케폴드 삽입을
갖는 총 서열분석 판독물의 %
0.0
0.2
0.4
0.6
0
20
40
60
80
100
pegRNA 스케폴드 삽입 길이 (nt)
pegRNA 스케폴드 삽입을 갖는
총 서열분석 판독물의 누적 %
0.0
0.5
1.0
1.5
0
20
40
60
80
100
pegRNA 스케폴드 삽입 길이 (nt)

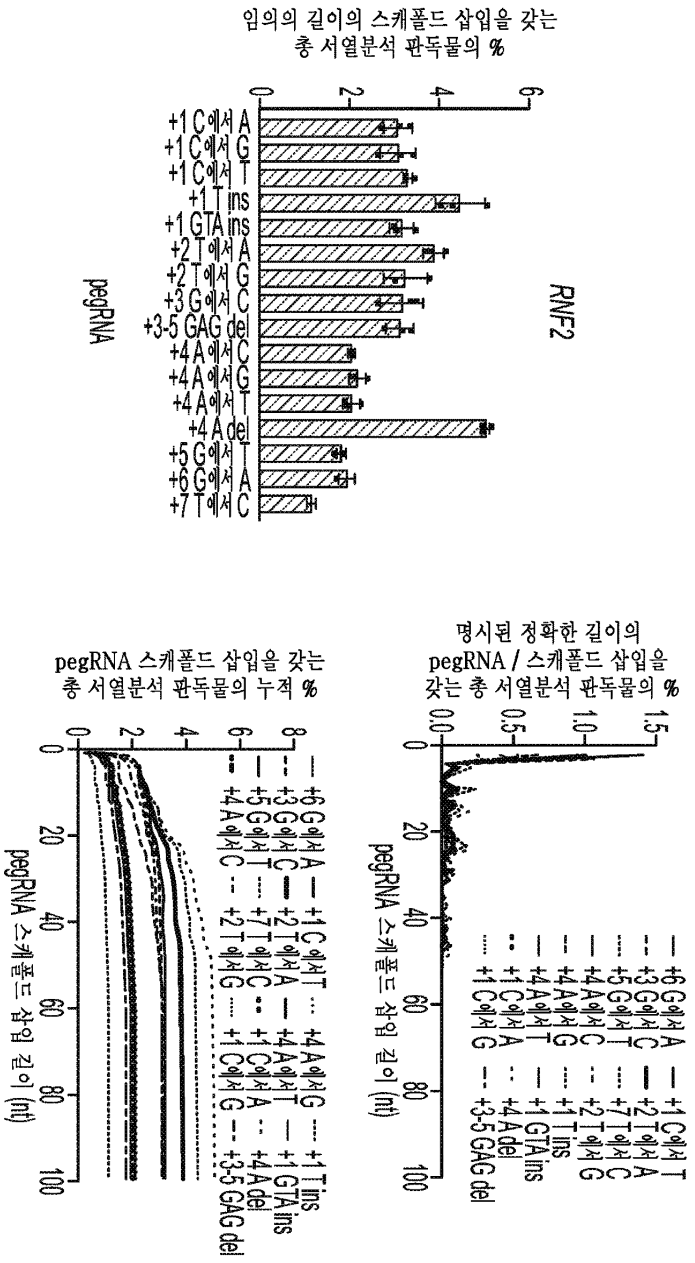
도면56b


임의의 길이의 스캐폴드 삽입을 갖는
총 서열분석 판독물의 %
pegRNA
FANCF
+1 A에서 G
+1 A에서 T
+2 C에서 A
+3 C에서 G
+3 C에서 T
+3 C ins
+4 T에서 A
+4 T에서 G
+4 GAT ins
+5 G에서 A
+5-7 GGA del
+6 G에서 C
+6 G del
+7 A에서 C
+8 T에서 C
+10 G에서 T
명시된 정확한 길이의
pegRNA / 스캐폴드 삽입을
갖는 총 서열분석 판독물의 %
pegRNA 스캐폴드 삽입 길이 (nt)
pegRNA 스캐폴드 삽입을 갖는
총 서열분석 판독물의 누적 %
pegRNA 스캐폴드 삽입 길이 (nt)

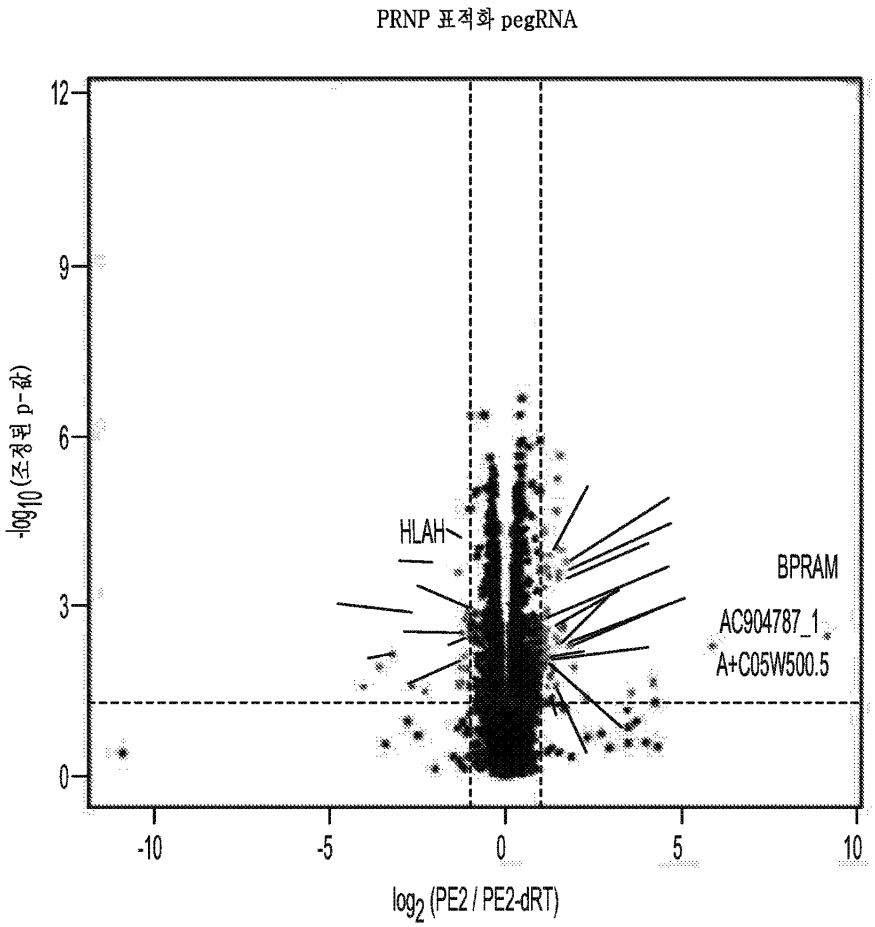
도면56c


임의의 길이의 스케폴드 삽입을 갖는
총 서열분석 판독물의 %
pegRNA
HEK3
pegRNA 스케폴드 삽입을 갖는
총 서열분석 판독물의 누적 %
명시된 정확한 길이의
pegRNA / 스케폴드 삽입을
갖는 총 서열분석 판독물의 %
pegRNA 스케폴드 삽입 길이 (nt)

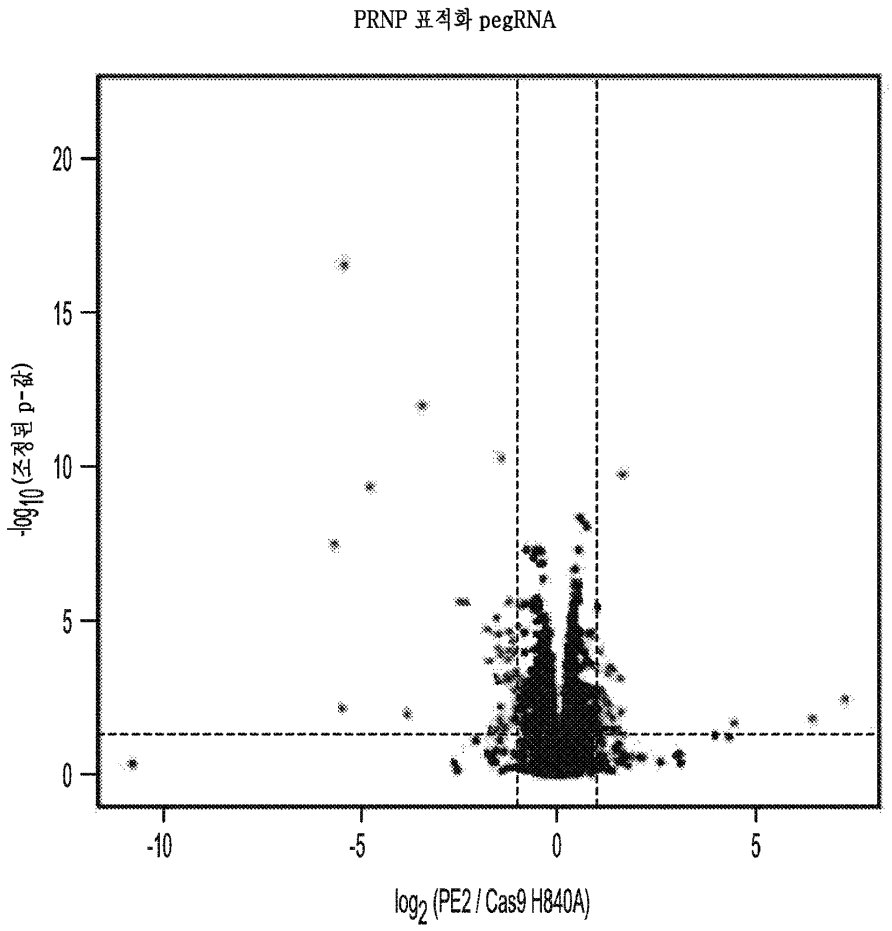
도면56d


임의의 길이의 스케폴드 삽입을 갖는
총 서열분석 판독물의 %
pegRNA
RNF2
+1 C에서 A
+1 C에서 G
+1 C에서 T
+1_T ins
+1 GTA ins
+2 T에서 A
+2 T에서 G
+3 G에서 C
+3-5 GAG del
+4 A에서 C
+4 A에서 G
+4 A에서 T
+4 A del
+5 G에서 T
+6 G에서 A
+7 T에서 C
pegRNA 스케폴드 삽입을 갖는
총 서열분석 판독물의 누적 %
pegRNA 스케폴드 삽입 길이 (nt)
명시된 정확한 길이의
pegRNA / 스케폴드 삽입을
갖는 총 서열분석 판독물의 %
pegRNA 스케폴드 삽입 길이 (nt)

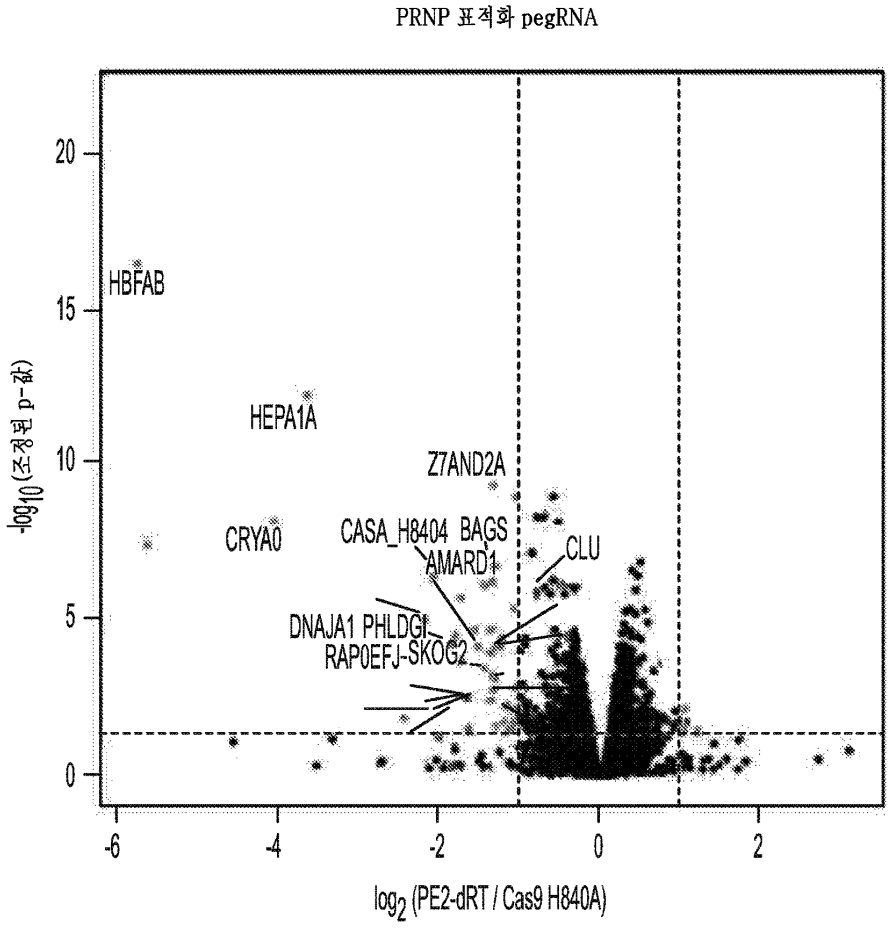
도면57a


PRNP 표적화 pegRNA
12
9
6
3
0
-log10 (조정된 p-값)
HLAH
BPRAM
AC904787_1
A+C05W500.5
-10
-5
0
5
10
log2 (PE2 / PE2-dRT)

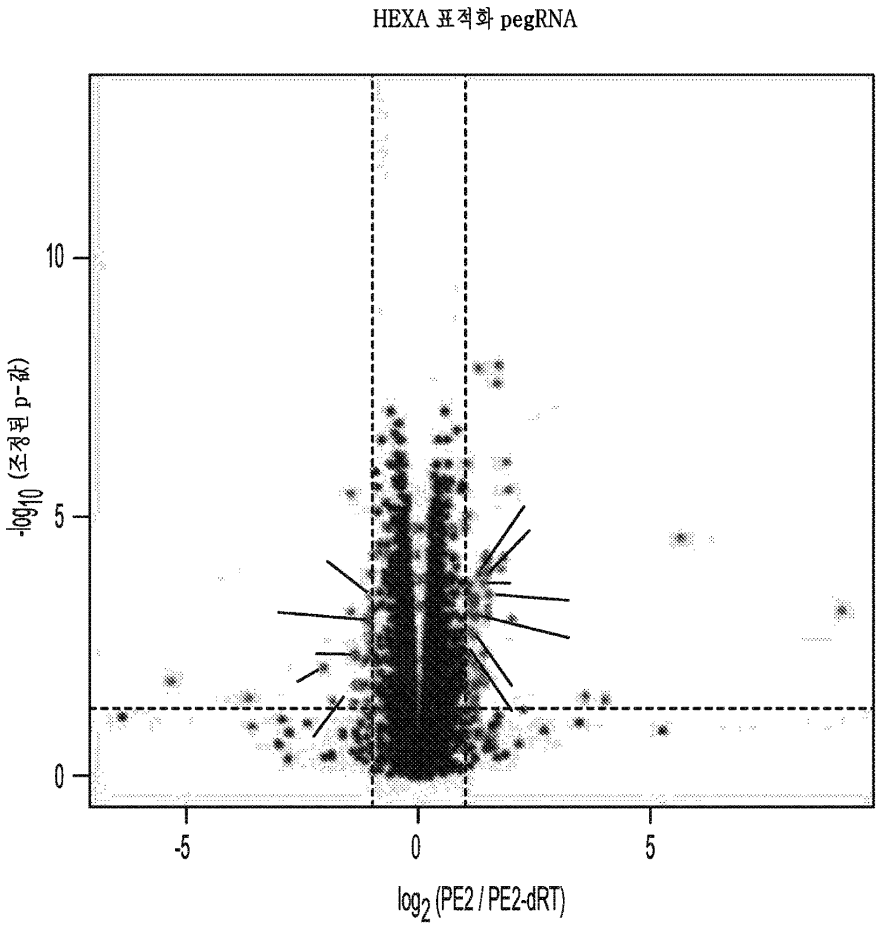
도면57b


PRNP 표적화 pegRNA
20
15
10
5
0
-10
-5
0
5
-log10 (조정된 p-값)
log2 (PE2 / Cas9 H840A)

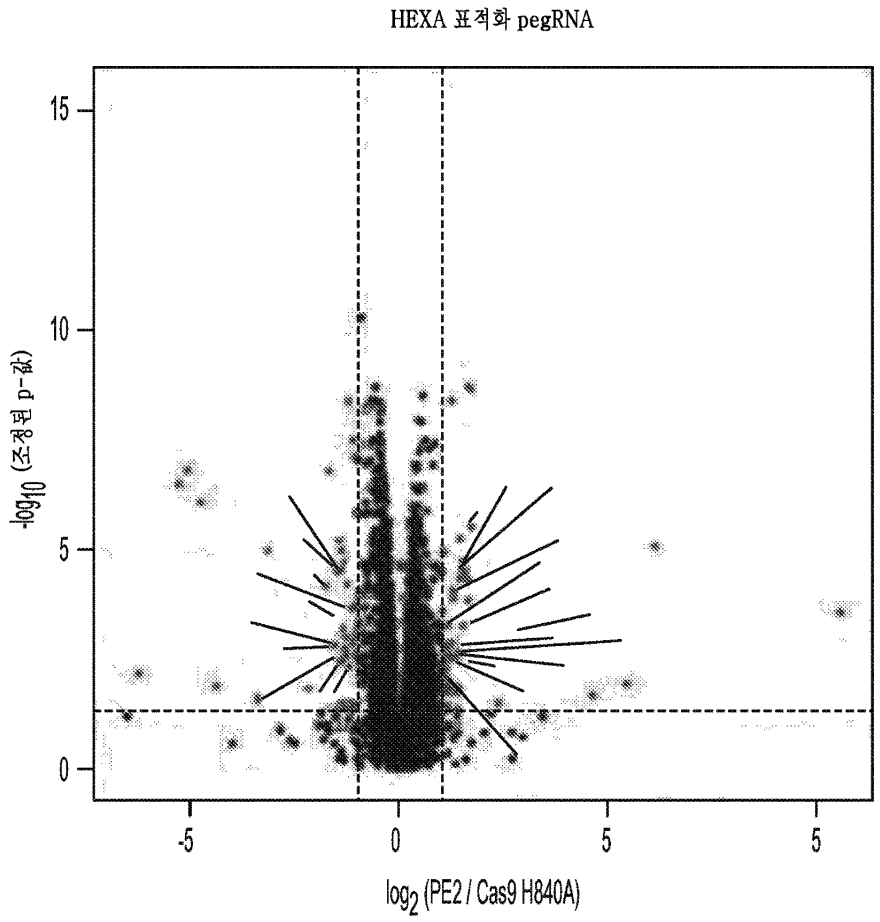
도면57c


PRNP 표적화 pegRNA
20
15
10
5
0
HBFAB
HEPA1A
Z7AND2A
CRYAB
CASA_H8404
BAGS
AMARD1
CLU
DNAJA1
PHLDGI
RAP0EFJ-SKOG2
-6
-4
-2
0
2
-log10(조정된 p-값)
log2 (PE2-dRT / Cas9 H840A)

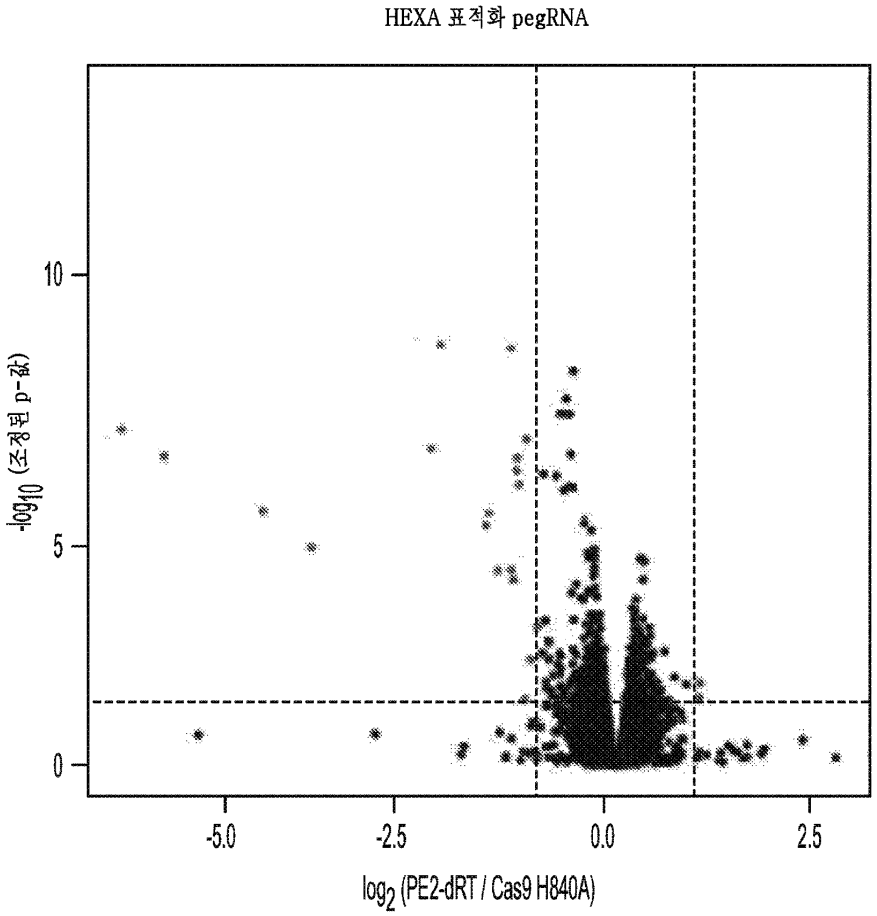
도면57d



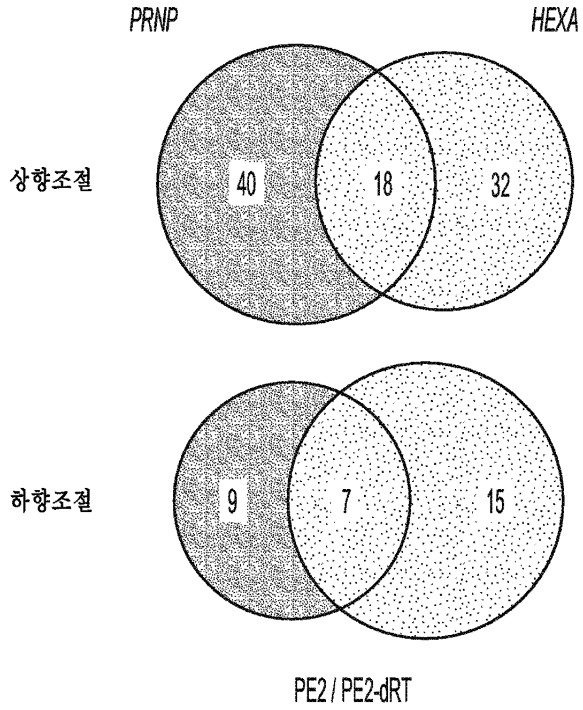
도면57e



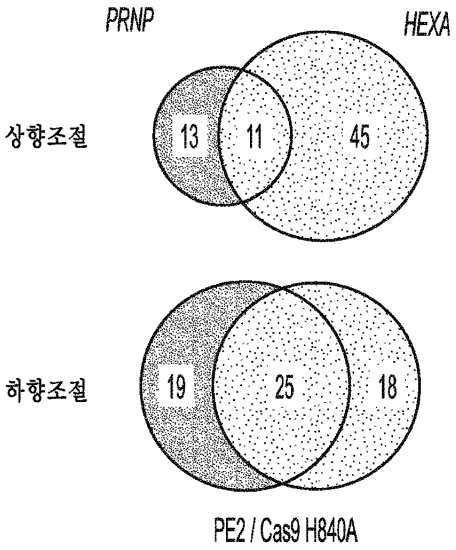
도면57f


HEXA 표적화 pegRNA
10
5
0
-log10 (조정된 p-값)
-5.0
-2.5
0.0
2.5
log2 (PE2-dRT / Cas9 H840A)

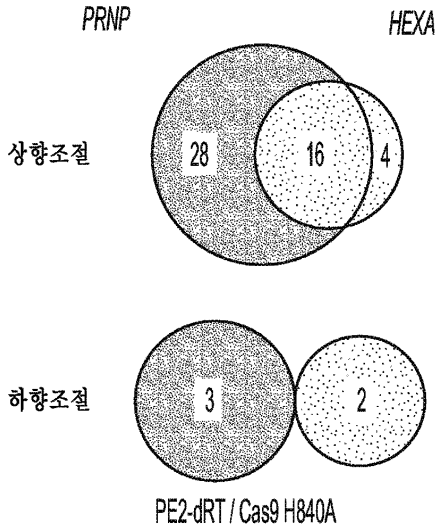
도면57g


PRNP
HEXA
상향조절
40
18
32
하향조절
9
7
15
PE2 / PE2-dRT

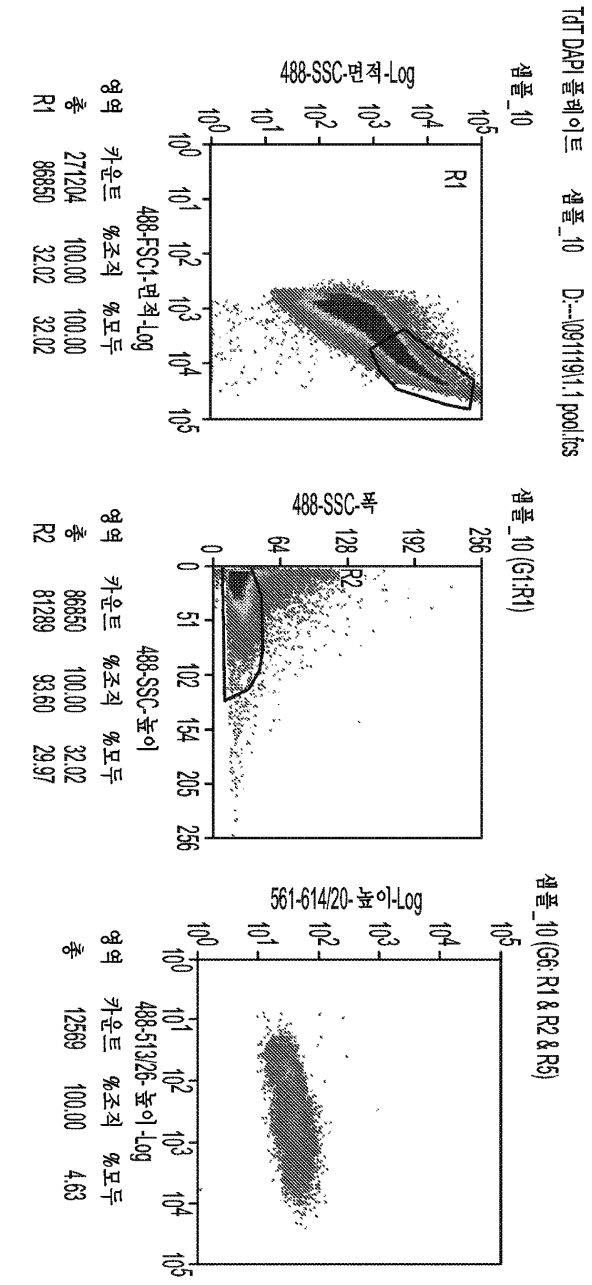
도면57h


PRNP
HEXA
상향조절
13
11
45
하향조절
19
25
18
PE2 / Cas9 H840A

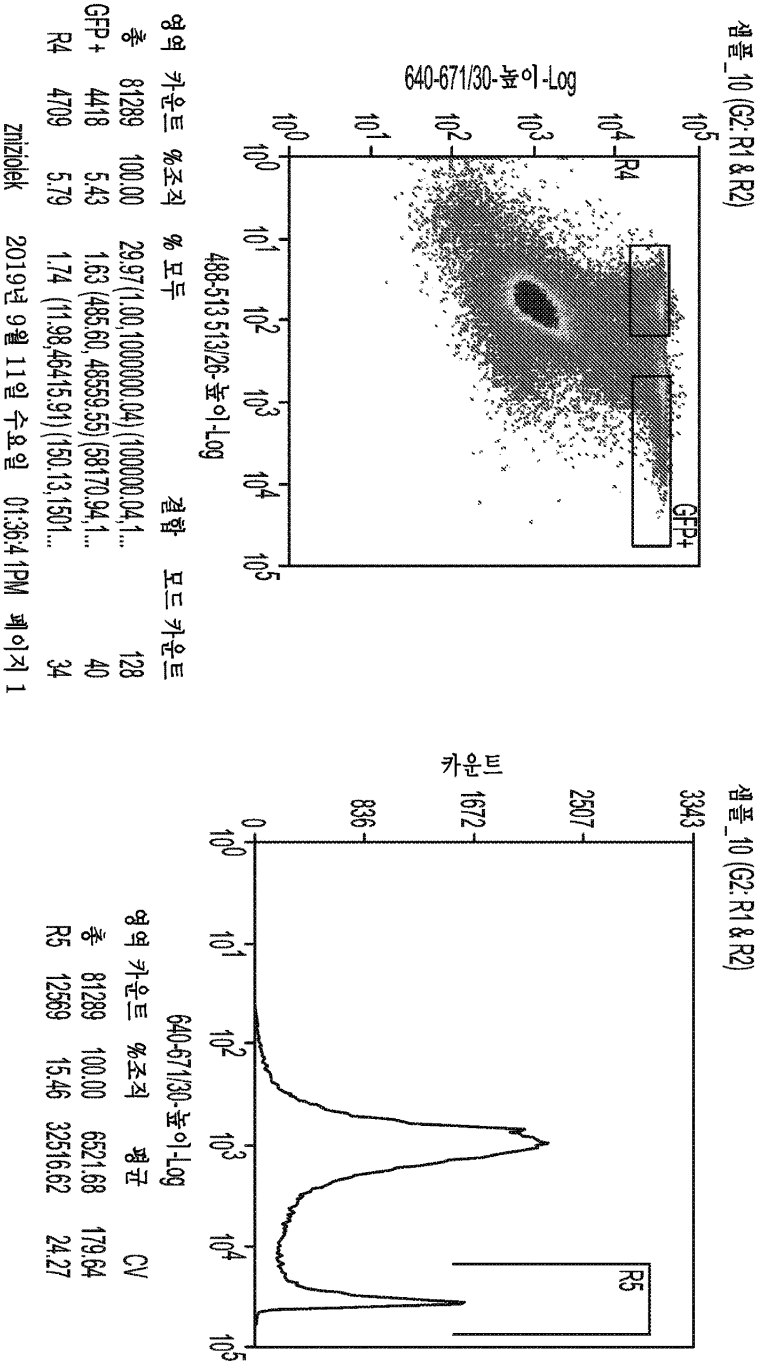
도면57i


PRNP
HEXA
상향조절
28
16
4
하향조절
3
2
PE2-dRT / Cas9 H840A

도면58a


TdT DAPI 플레이트 샘플_10 D:---\091119\1.1 pool.fcs
샘플_10
488-SSC-면적-Log
488-FSC1-면적-Log
R1
영역 카운트 %조직 %모두
총 271204 100.00 100.00
R1 86850 32.02 32.02
샘플_10 (G1:R1)
488-SSC-폭
488-SSC-높이
R2
영역 카운트 %조직 %모두
총 86650 100.00 32.02
R2 81289 93.60 29.97
샘플_10 (G6: R1 & R2 & R5)
561-614/20-폭이-Log
488-513/26-높이-Log
영역 카운트 %조직 %모두
총 12569 100.00 4.63

도면58b


샘플_10 (G2: R1 & R2)
640-671/30-높이-Log
488-513 513/26-높이-Log
R4
GFP+
영역 카운트 %조직 % 모두 결합 모드 카운트
총 81289 100.00 29.97(1.00,1000000.04) (100000.04,1... 128
GFP+ 4418 5.43 1.63 (485.60, 48559.55) (58170.94,1... 40
R4 4709 5.79 1.74 (11.98,46415.91) (150.13,1501... 34
znizlolek 2019년 9월 11일 수요일 01:36:4 1PM 페이지 1
샘플_10 (G2: R1 & R2)
카운트
640-671/30-높이-Log
R5
영역 카운트 %조직 평균 CV
총 81289 100.00 6521.68 179.64
R5 12569 15.46 32516.62 24.27

도면59a

클로닝 개관

1. BsaI로 pU6-pegRNA-GG-벡터 플라스미드 (성분 1)를 소화시키고, 복제 기점, U6 프로모터, U6 폴리-T 종결 서열 및 Amp^R 유전자를 함유하는 플라스미드 단편 (~2.2kb)을 단리한다.
2. 하기를 위해올리고뉴클레오티드를 정렬한다:
   a. 표시된 오버행이 플랭킹되어 있는 목적하는 스페이서 (표적) 서열 (성분 2)
      i. 상단 가닥 올리고뉴클레오티드 (5' CACC 및 3' GTTTT 오버행을 포함함)를 위해 목적하는 표적의 5'-3' 서열을 사용하고, 하단 가닥 올리고뉴클레오티드 (5' CTCTAAAAC 오버행을 포함함)를 위해 표적 서열의 역 상보체를 사용한다. 스페이서 서열은 효율적인 전사 개시를 위해 G 뉴클레오티드로 시작하여야 한다.
   b. 표시된 오버행이 플랭킹되어 있는 목적하는 pegRNA 3' 연장 주형 (성분 3)
      i. 상단 가닥 올리고뉴클레오티드 (5' GTGC 오버행을 특색으로 함)로서 RNA 센스 서열을 사용하고, 하단 가닥 올리고뉴클레오티드 (5' AAAA 오버행을 특색으로 함)로서 이 서열의 역 상보체를 사용한다.
   c. 상용성 골든 게이트 오버행을 특색으로 하는 SpCas9 sgRNA 스캐폴드 서열 (성분 4)
      i. 나머지 성분으로부터의 오버행이 여러 누락 뉴클레오티드에 기여하므로, 이들 올리고뉴클레오티드는 완전한 스캐폴드 서열이 아니다.
      ii. 주: 이들 올리고뉴클레오티드는 5' 인산화되어야 한다. 올리고뉴클레오티드는 제조업체에 의해 5' 인산화될 수 있거나 또는 T4 PNK를 사용하여 효소적으로 5' 인산화될 수 있다 (하기 프로토콜 참조).
3. 하기 프로토콜에 따라 별개의 어닐링 반응으로 성분 2, 3 및 4에 대한 상단 및 하단 올리고뉴클레오티드를 어닐링한다. SpCas9 sgRNA 스캐폴드 서열 (성분 4)이 인산화되지 않은 경우에, T4 PNK로 인산화한다.
4. 성분 1로부터의 단리된 2.2-kb 단편을 성분 2, 3 및 4와 골든 게이트 조립한다.
5. 라이게이션 생성물을 이. 콜라이 내로 형질전환시킨다. pU6-pegRNA-GG-벡터 플라스미드로부터의 성분 1에 의해 부여된 항생제 내성은 암피실린 내성 및 카르베니실린 내성이다.
6. 생성된 클론 형질전환체로부터 플라스미드를 단리하고 서열분석한다.

도면59b

pegRNA 클로닝 프로토콜

단계 1: **pU6-pegRNA-GG-**벡터 플라스미드 (성분 1)를 소화시킨다

하기를 PCR 튜브 내에서 합한다:

2000 ng pU6-pegRNA-GG-벡터 (성분 1)	X μL
Bsa1-HFv2 (NEB)	1.0 μL
10x 컷스마트 완충제	3.0 μL
H_2O	30.0 μL까지
총 반응 부피	30.0 μL

37℃에서 4-16시간 동안 인큐베이션한다
절단 플라스미드로부터 ~2.2-kb 단편을 단리한다.

도면59c

단계 2 및 3: 올리고뉴클레오티드 부분 (성분 2, 3 및 4)을 정렬하고 어닐링한다

물질

어닐링 완충제: 10 mM 트리스-Cl pH 8.5 및 50 mM NaCl로 보충된 H_2O
상보적 올리고뉴클레오티드 쌍

프로토콜

하기를 PCR 튜브 내에서 합한다:

상단 올리고뉴클레오티드, 100 μM	1.0 μL
하단 올리고뉴클레오티드, 100 μM	1.0 μL
어닐링 완충제 (성분 2, 3 및 4)	23.0 μL
총 반응 부피	25.0 μL

써모사이클러에서, 95℃에서 3분 동안 가열한 다음, 점차 (0.1 ℃/s) 22℃로 냉각시킨다
어닐링된 올리고뉴클레오티드를 75 μL H_2O 첨가에 의해 1:4로 희석한다. 이러한 희석 후 각각의 올리고뉴클레오티드의 최종 농도는 1 μM일 것이다. 단계 2.5에서 PNK에 의해 인산화하는 경우 sgRNA 스캐폴드 (성분 4)를 희석하지 않는다.

도면59d

단계 2.b.ii.: sgRNA 스캐폴드 인산화 (인산화된 올리고뉴클레오티드를 구입한 경우 불필요)

프로토콜

하기를 PCR 튜브 내에서 합한다:

단계 1로부터의 4 μM 올리고뉴클레오티드 듀플렉스	6.25 μL
10x T4 DNA 리가제 완충제 (NEB)	2.50 μL
T4 PNK (NEB)	0.50 μL
H_2O	15.75 μL
총 반응 부피	20.0 μL

써모사이클러에서, 37℃에서 60분 동안 인큐베이션한다
이러한 인산화 후, 어닐링된 스캐폴드 올리고뉴클레오티드는 이제 1 μM의 농도이다.
단계 3으로 진행한다.

도면59e

단계 4: pegRNA 조립
pegRNA, 골든 게이트 조립 반응

소화된 pU6-pegRNA-GG 플라스미드-벡터 - 사전-절단된, 단리된 2.2-kb 단편	1.00 µL @ 30 ng/µL
어닐링된 프로토스페이서 올리고뉴클레오티드 (성분 2)	1.00 µL @ 1 µM
어닐링된 pegRNA 3'-연장 올리고뉴클레오티드 (성분 4)	1.00 µL @ 1 µM
sgRNA 스캐폴드 어닐링된 올리고뉴클레오티드 (성분 3) - 올리고뉴클레오티드는 인산화되어야 한다	1.00 µL @ 1 µM
BsaI-HFv2 (NEB)	0.25 µL
T4 DNA 리가제 (NEB)	0.50 µL
10x T4 DNA 리가제 완충제(NEB)	1.00 µL
H_2O	4.25 µL
총 반응 부피:	10.0 µL

실온에서 10분 동안 인큐베이션한다
인큐베이션 후, 써모사이클러에서 하기 프로그램을 수행한다:
37℃에서 5분, 이어서 80℃에서 5분, 이어서 12℃에서 유지

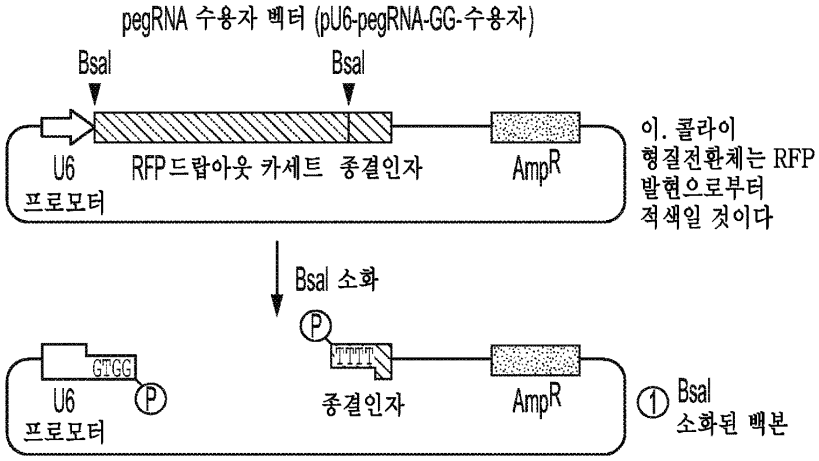
도면59f

단계 5 및 6: 조립된 플라스미드의 형질전환
10 µL 조립 반응물 중 1 µL를 10 µL의 적격 세포 내로 형질전환시킨다. 목적하는 형질전환체는 암피실린 및 카르베니실린에 대한 내성을 가질 것이다.

하기 다이어그램은 pegRNA 클로닝 프로토콜을 요약한 것이다.

도면59g

골든 게이트 조립에 의한 pegRNA 클로닝



② pegRNA 스페이서 어닐링된 올리고

SEQ ID NO: 879

5′ CACC*NNNNNNNNNNNNNNNNNNNN**GTTTT*** 3′
3′ *NNNNNNNNNNNNNNNNNNNN**CAAAATCTC*** 5′

SEQ ID NO: 880

③ pegRNA 3′ 연장부 어닐링된 올리고

SEQ ID NO: 883

5′ *GTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN* 3′
3′ *NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAA* 5′

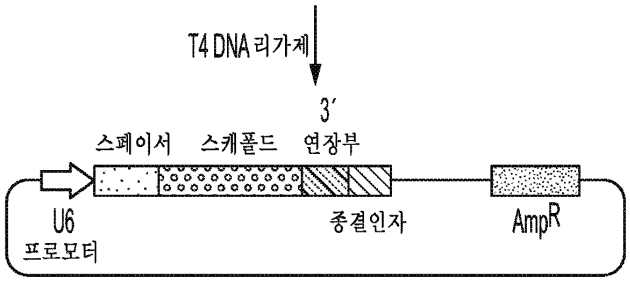
SEQ ID NO: 884

④ pegRNA 스캐폴드 어닐링된 올리고 (5' 인산화됨)

SEQ ID NO: 881

5′ [Phos]-***AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG*** 3′
3′ ***GATCTTTATCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACG***-[Phos] 5′

SEQ ID NO: 882 (이들 올리고는 pegRNA 벡터 조립을 허용하기 위해 5' 인산화되어야 한다는 것을 주목함)



조립된 pegRNA 벡터

도면60a

```
## sgRNA scaffold sequence search ##

import pandas as pd
import Bio as bio
from Bio import SeqIO
import glob

#generates list of fastq files to analyze
sources = glob.glob('*.fastq')

#reads the fastq files into a dictionary with the file names as keys
fastqdict = {}
for i in range(len(sources)):
        temp = list(SeqIO.parse(source[i],"fastq"))
        fastqdict[sources[i]]= [str(temp[k].seq) for k in range(len(temp))]

#the referenced sequence to be searched for is entered into the following dictionary with
#an appropriate key
scaffdict =
{'HEK3':'CAGAGGACCGACTCGGTCCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT
GCTATTTCTAGCTCTAAAACTCACGTGCTCAGTCTGGGCCGGTG', (SEQ ID NO: 885)
'EMX1':'ATCACGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT
TGCTATTTCTAGCTCTAAAACTTCTTCTTCTGCTCGGACTCGGTG', (SEQ ID NO: 886)
'FANCF':'TTTCCGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC
TTGCTATTTCTAGCTCTAAAACGGTGCTGCAGAAGGGATTCCGGTG', (SEQ ID NO: 887)
'RNF2':'TCGTTGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTT
GCTATTTCTAGCTCTAAAACCAGGTAATGACTAAGATGACGGTG'} (SEQ ID NO: 888)

#matches and counts iterative slices of the reference string to the appropriate fastq files
#reference key must be contained in the name of the fastq file
#generated values represent cumulative counts for a minimum degree of sgRNA integration
#i.e. a given value x means x reads contain y or more bases of the scaffold
resultdict = dict.fromkeys(sources)
for key in fastqdict:
   for scaffold in scaffdict:
      if scaffold in str(key):
         resultlist = []
         for j in range(len(scaffdict[scaffold])):
```

도면60b

```
                extent = scaffdict[scaffold][0:(j+1)]
                counter = 0
                for i in range(len(fastqdict[key])):
                    if extent in fastqdict[key][i]:
                        counter = counter + 1
                resultlist.append(counter)
            resultdict[key]=resultlist

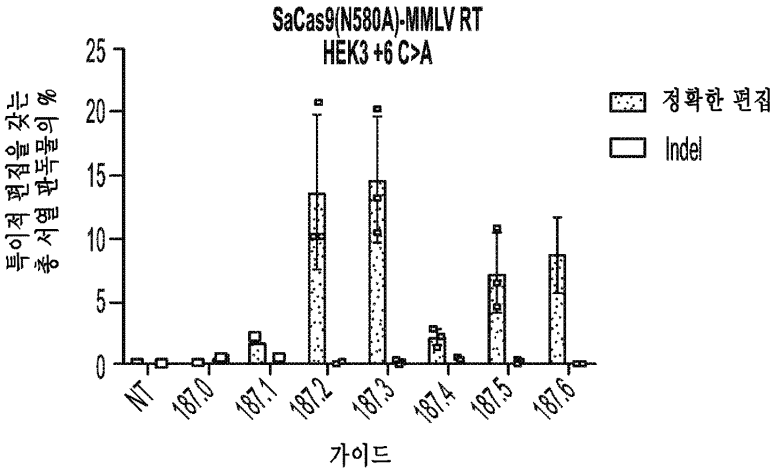
#writes the results into a dataframe indexed from 1
resultdf = pd.DataFrame.from_dict(resultdict)
resultdf = resultdf.reindex(sorted(resultdf.columns), axis=1)
resultdf.index = range(1,len(resultdf)+1)

#converts the cumulative count values into specific counts
#i.e. a given value x means x reads contain exactly y bases of the scaffold
resultdf2=resultdf.copy()
for entry in resultdf:
    for i in range(1,len(resultdf[entry])+1):
        try:
            resultdf2[entry][i] = resultdf[entry][i]-resultdf[entry][i+1]
        except:
            resultdf2[entry][i] = resultdf[entry][i]

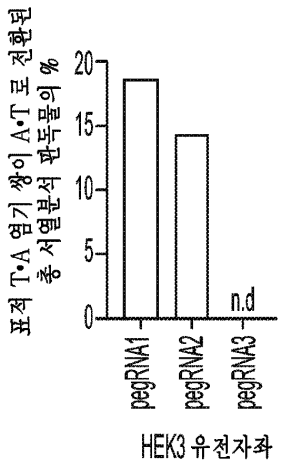
#converts the specific counts values into frequencies
resultdf3=resultdf2.copy()
for entry in resultdf3:
    resultdf3[entry]=resultdf2[entry].div(resultdf[entry][1])*100

#reads the results into excel files
resultdf.to_excel('cumulativecounts.xlsx')
resultdf2.to_excel('specificcounts.xlsx')
resultdf3.to_excel('specificfrequencies.xlsx')
```

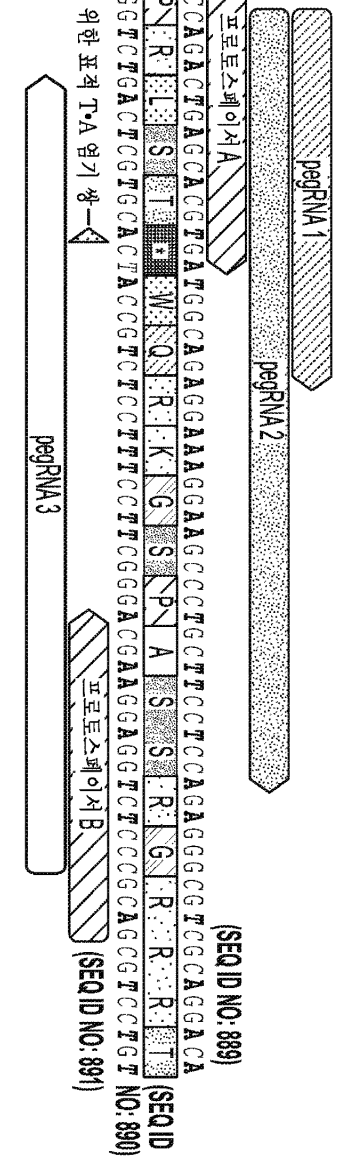
도면61



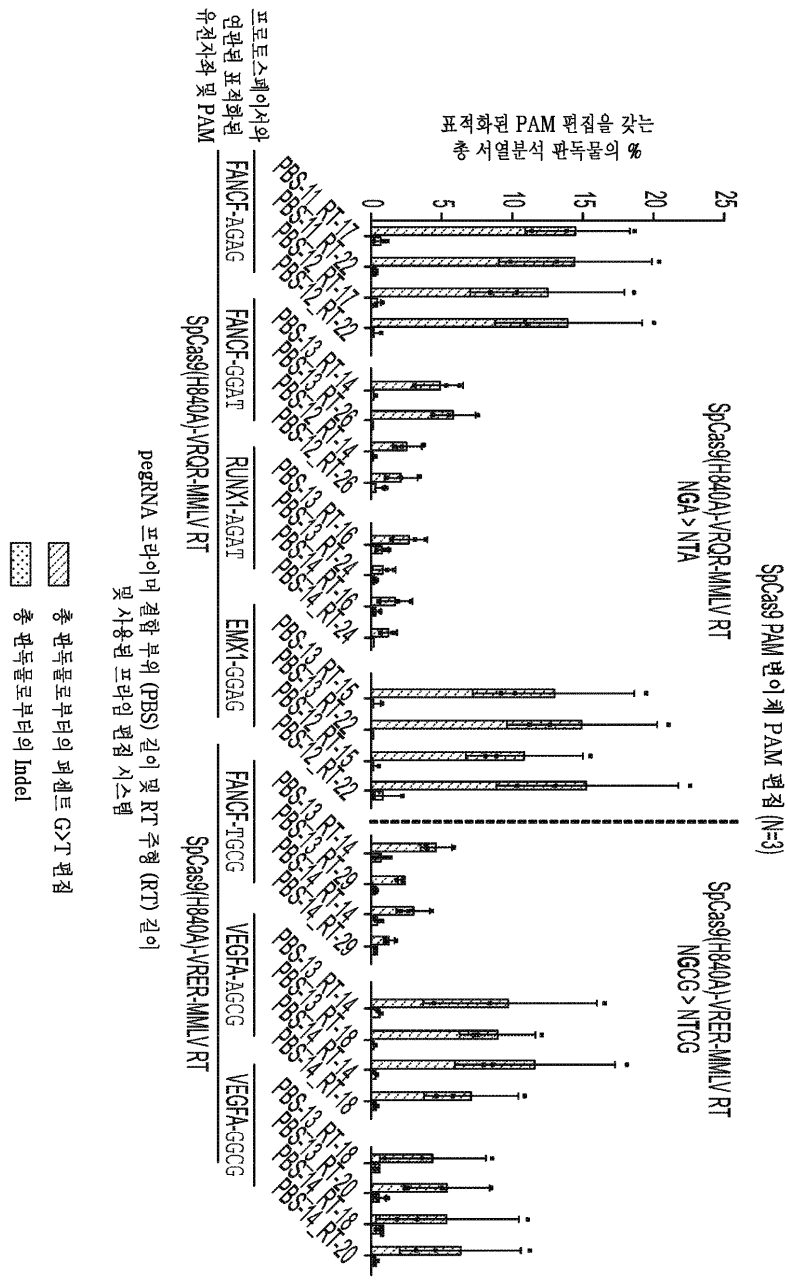
도면62a


표적 T•A 염기 쌍이 A•T 로 전환된
총 서열분석 판독물의 %
20
15
10
5
0
n.d
pegRNA1
pegRNA2
pegRNA3
HEK3 유전자좌

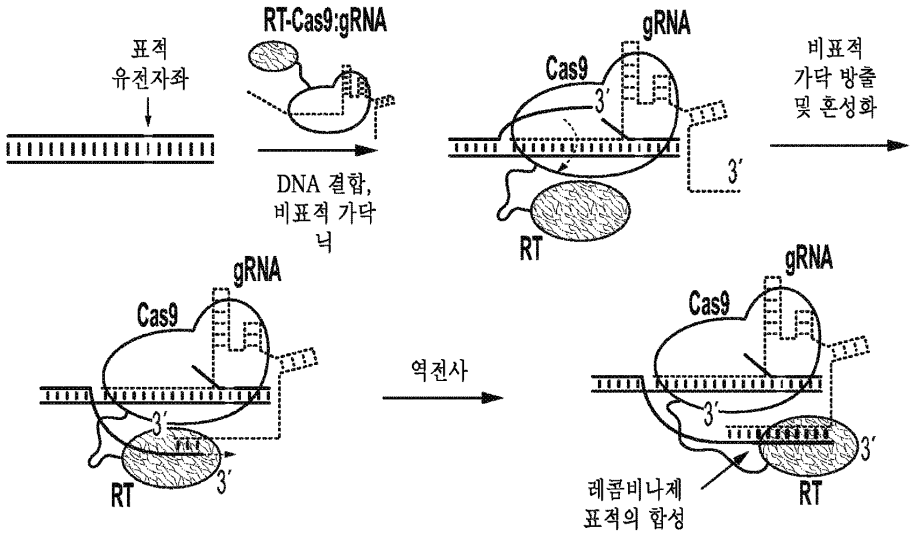
도면62b


pegRNA 1
pegRNA 2
프로토스페이서 A
(SEQ ID NO: 889)
편집을 위한 표적 T•A 염기 쌍
pegRNA 3
프로토스페이서 B
(SEQ ID NO: 891)
(SEQ ID NO: 890)

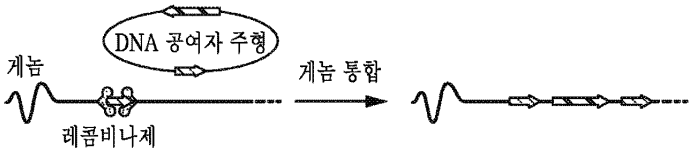
도면63


SpCas9 PAM 변이체 PAM 편집 (N=3)
표적화된 PAM 편집을 갖는
총 서열분석 판독물의 %
0
5
10
15
20
25
SpCas9(H840A)-VRQR-MMLV RT
NGA > NTA
SpCas9(H840A)-VRER-MMLV RT
NGCG > NTCG
프로토스페이서와
연관된 표적화된
유전자좌 및 PAM
FANCF-AGAG
FANCF-GGAT
RUNX1-AGAT
EMX1-GGAG
FANCF-TGCG
VEGFA-AGCG
VEGFA-GGCG
SpCas9(H840A)-VRQR-MMLV RT
SpCas9(H840A)-VRER-MMLV RT
pegRNA 프라이머 결합 부위 (PBS) 길이 및 RT 주형 (RT) 길이
및 사용된 프라임 편집 시스템
총 판독물로부터의 퍼센트 G>T 편집
총 판독물로부터의 Indel

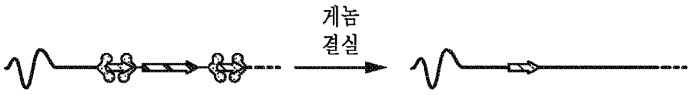
도면64a



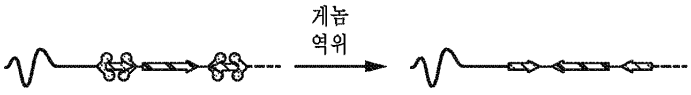
도면64b



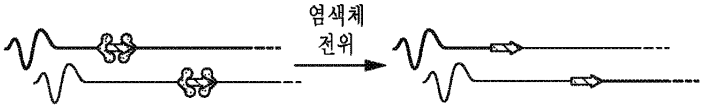
도면64c



도면64d



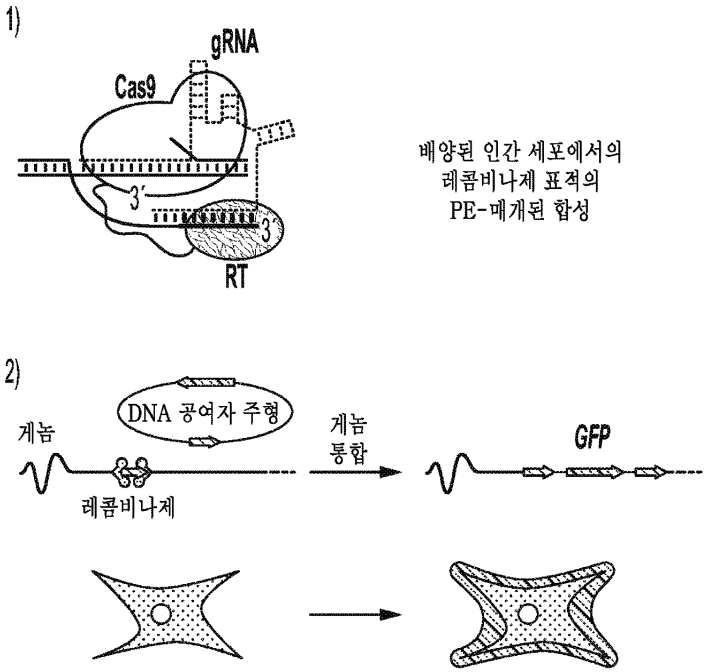
도면64e



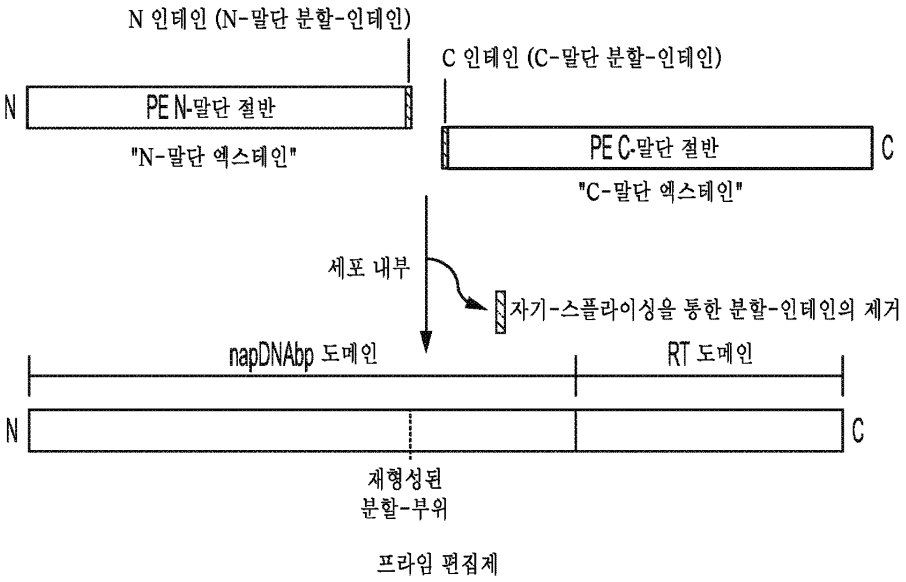
도면64f



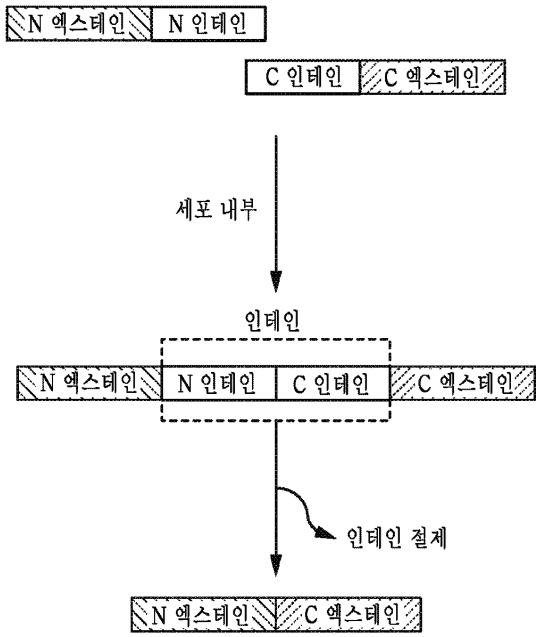
도면65


1)
gRNA
Cas9
3'
3'
RT
배양된 인간 세포에서의
레콤비나제 표적의
PE-매개된 합성
2)
DNA 공여자 주형
게놈
레콤비나제
게놈
통합
GFP

도면66

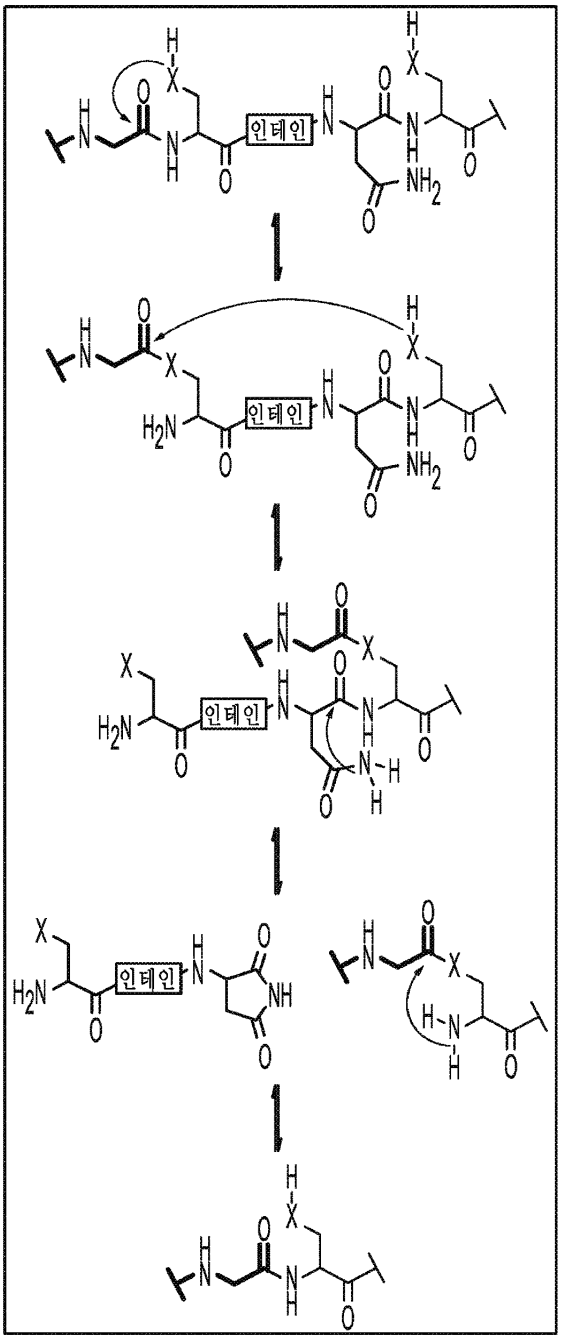

N 인테인 (N-말단 분할-인테인)
C 인테인 (C-말단 분할-인테인)
N
PE N-말단 절반
"N-말단 엑스테인"
PE C-말단 절반
C
"C-말단 엑스테인"
세포 내부
자기-스플라이싱을 통한 분할-인테인의 제거
napDNAbp 도메인
RT 도메인
N
C
재형성된
분할-부위
프라임 편집제

도면67a

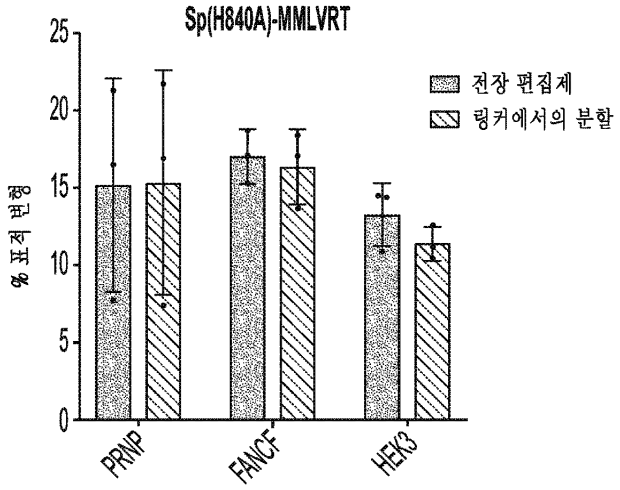

N 엑스테인
N 인테인
C 인테인
C 엑스테인
세포 내부
인테인
N 엑스테인
N 인테인
C 인테인
C 엑스테인
인테인 절제
N 엑스테인
C 엑스테인

도면67b

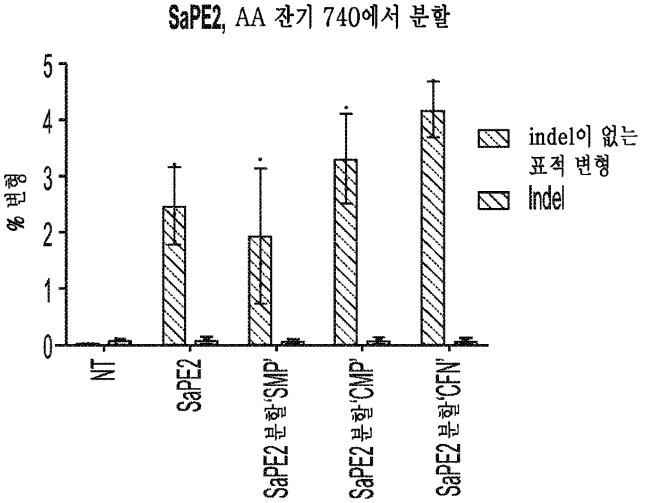
인테인 절제 메카니즘



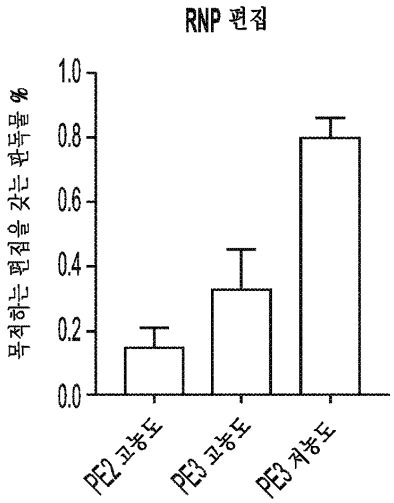
도면68a



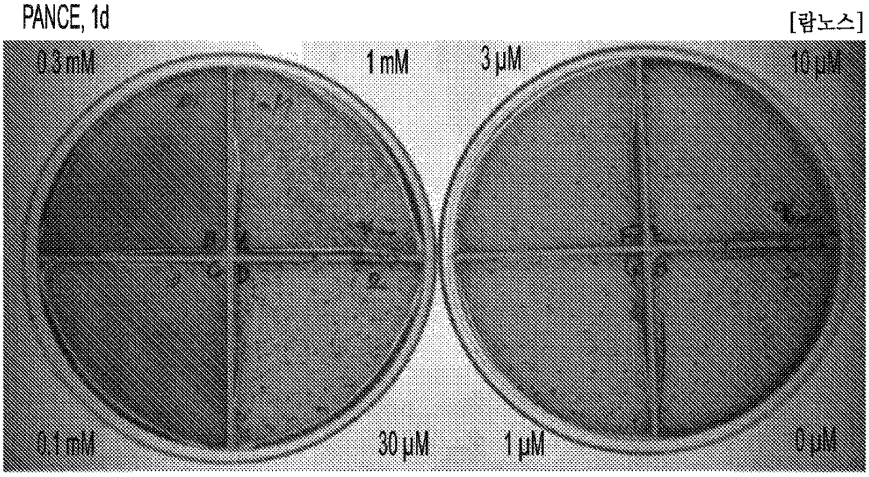
도면68b



도면68c



도면69



도면70a

출발 대립유전자

(SEQ ID NO: 892)

5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA

(SEQ ID NO: 893)

편집 표적 (A•T 에서 C•G)

ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG

CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC

도면70b

출발 대립유전자

(SEQ ID NO: 892)

5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA

(SEQ ID NO: 893)

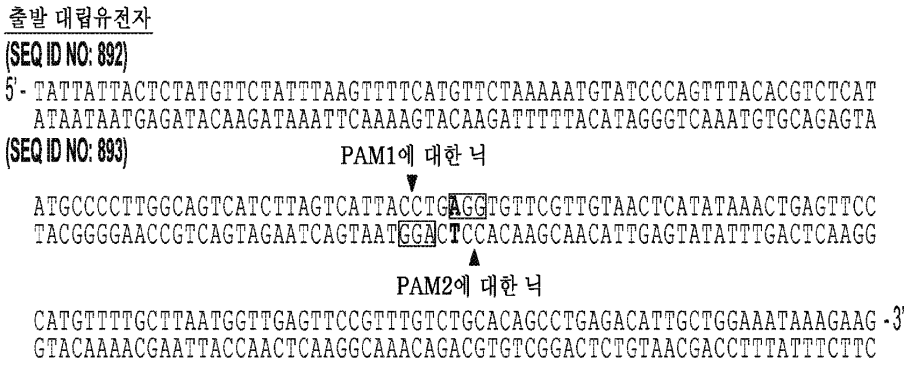
NGG PAM 1

ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACTCCACAAGCAACATTGAGTATATTTGACTCAAGG

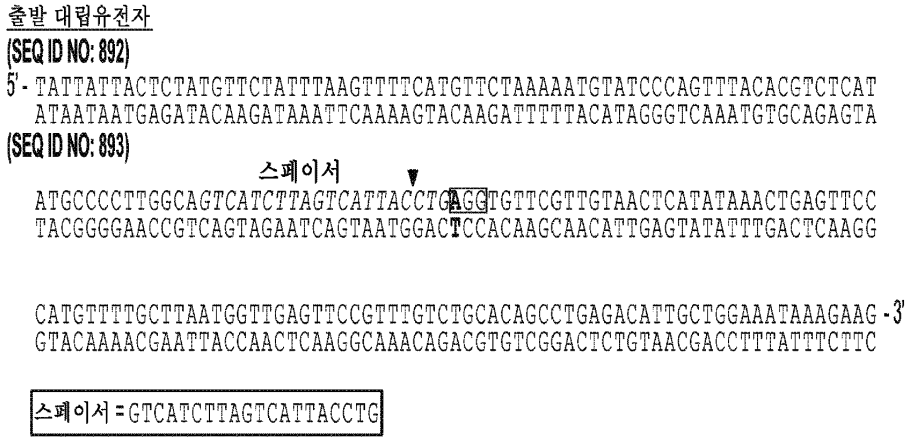
NGG PAM 2

CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG -3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC

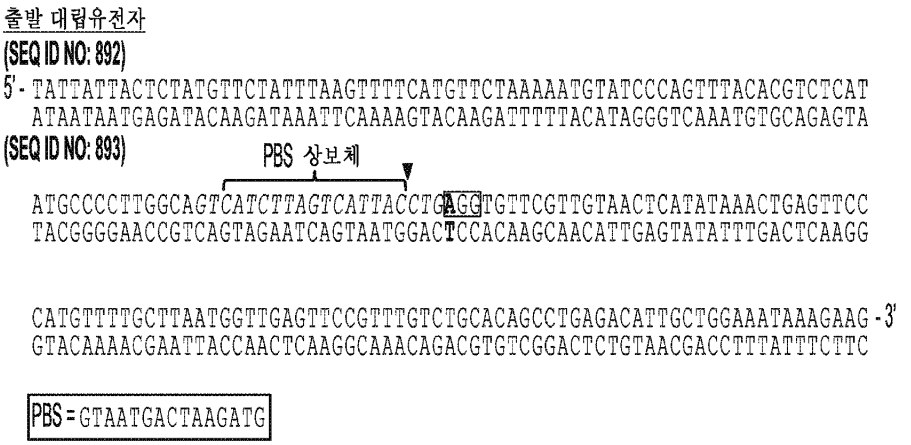
도면70c



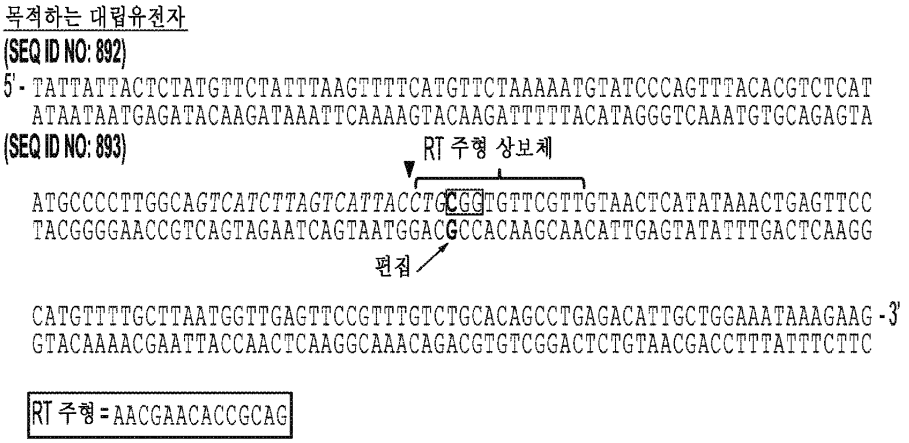
도면70d



도면70e



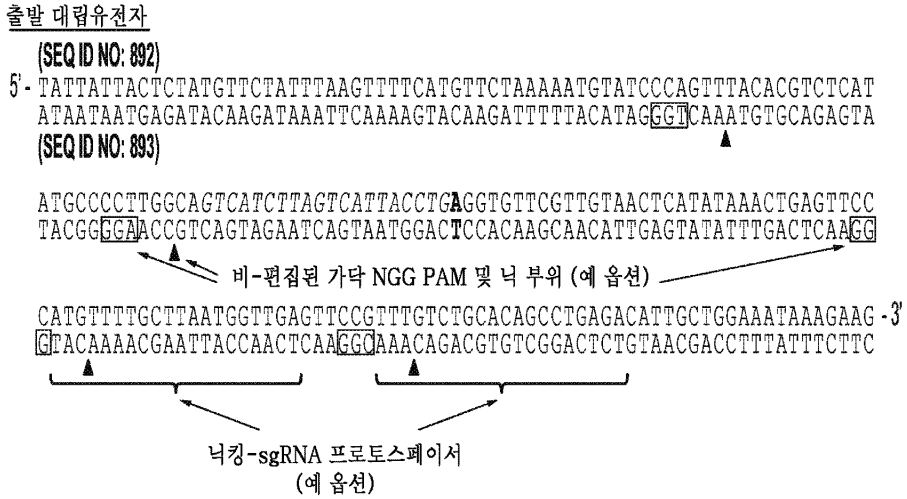
도면70f



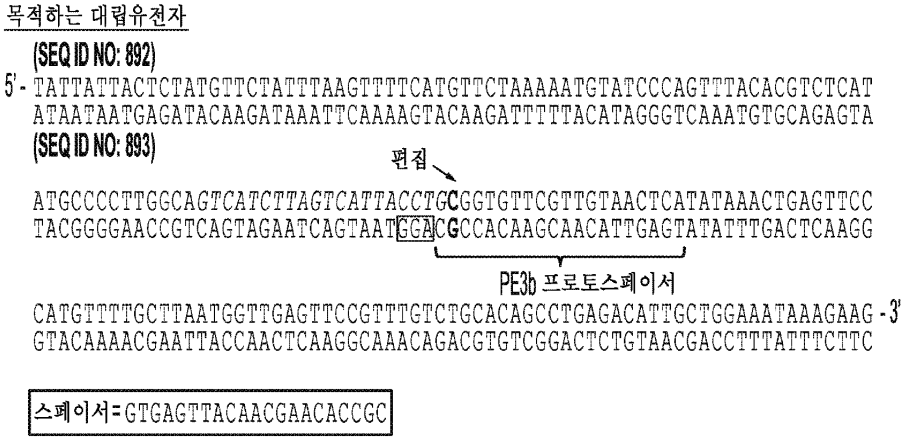
도면70g



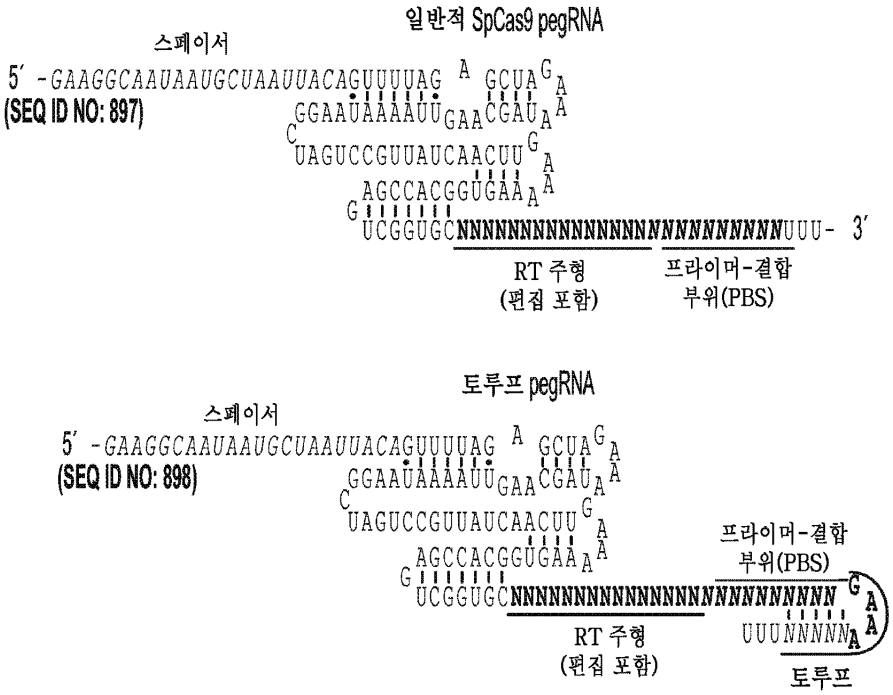
도면70h



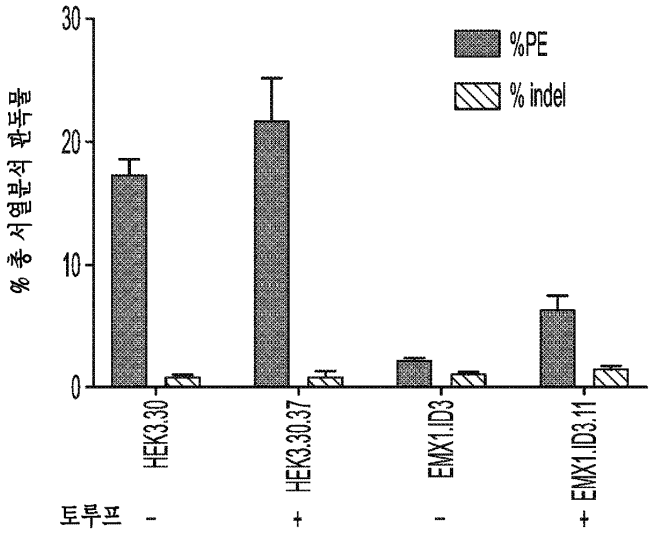
도면70i


목적하는 대립유전자
(SEQ ID NO: 892)
5'- TATTATTACTCTATGTTCTATTTAAGTTTTCATGTTCTAAAAATGTATCCCAGTTTACACGTCTCAT
ATAATAATGAGATACAAGATAAATTCAAAAGTACAAGATTTTTACATAGGGTCAAATGTGCAGAGTA
(SEQ ID NO: 893)
편집
ATGCCCCTTGGCAGTCATCTTAGTCATTACCTGCGGTGTTCGTTGTAACTCATATAAACTGAGTTCC
TACGGGGAACCGTCAGTAGAATCAGTAATGGACGCCACAAGCAACATTGAGTATATTTGACTCAAGG
PE3b 프로토스페이서
CATGTTTTGCTTAATGGTTGAGTTCCGTTTGTCTGCACAGCCTGAGACATTGCTGGAAATAAAGAAG - 3'
GTACAAAACGAATTACCAACTCAAGGCAAACAGACGTGTCGGACTCTGTAACGACCTTTATTTCTTC
스페이서 = GTGAGTTACAACGAACACCGC

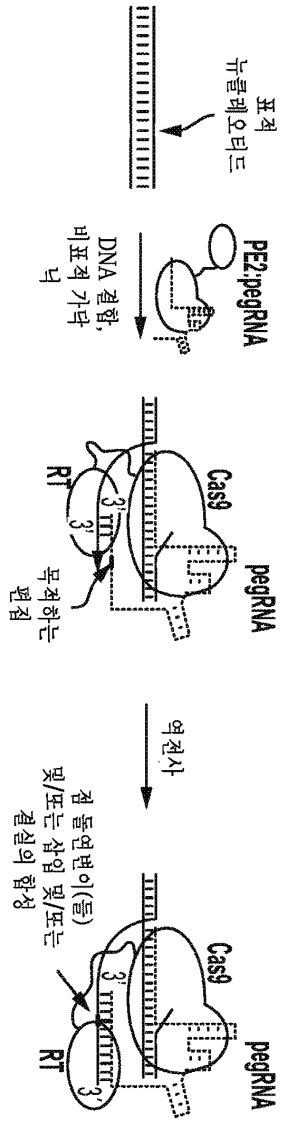
도면71a


일반적 SpCas9 pegRNA
스페이서
5' - GAAGGCAAUAAUGCUAAUUACAGUUUUAG A GCUA G A
(SEQ ID NO: 897)
GGAAUAAAAUU GAA CGAU A A
C
UAGUCCGUUAUCAACUU G A
AGCCACGGUGAA A A
G
UCGGUGCNNNNNNNNNNNNNNNNNNNNNNUUU - 3'
RT 주형
(편집 포함)
프라이머-결합
부위(PBS)
토루프 pegRNA
스페이서
5' - GAAGGCAAUAAUGCUAAUUACAGUUUUAG A GCUA G A
(SEQ ID NO: 898)
GGAAUAAAAUU GAA CGAU A A
C
UAGUCCGUUAUCAACUU G A
AGCCACGGUGAA A A
G
UCGGUGCNNNNNNNNNNNNNNNNNNNNNNNNNNNN G A
UUUNNNNN A A
프라이머-결합
부위(PBS)
RT 주형
(편집 포함)
토루프

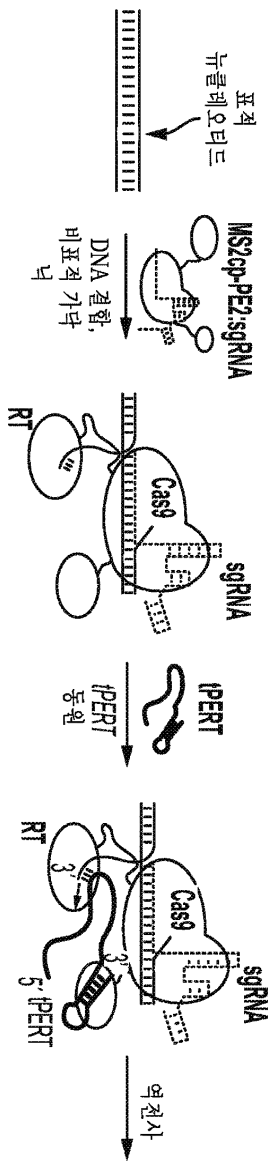
도면71b


30
20
10
0
% 총 서열분석 판독물
%PE
% indel
HEK3.30
HEK3.30.37
EMX1.ID3
EMX1.ID3.11
토루프
-
+
-
+

도면72a


표적
뉴클레오티드
PE2:pegRNA
DNA 결합,
비표적 가닥
닉
Cas9
pegRNA
RT
3'
3'
목적하는
편집
역전사
원 돌연변이(들)
및/또는 삽입 및/또는
결실의 합성
Cas9
pegRNA
RT
3'
3'

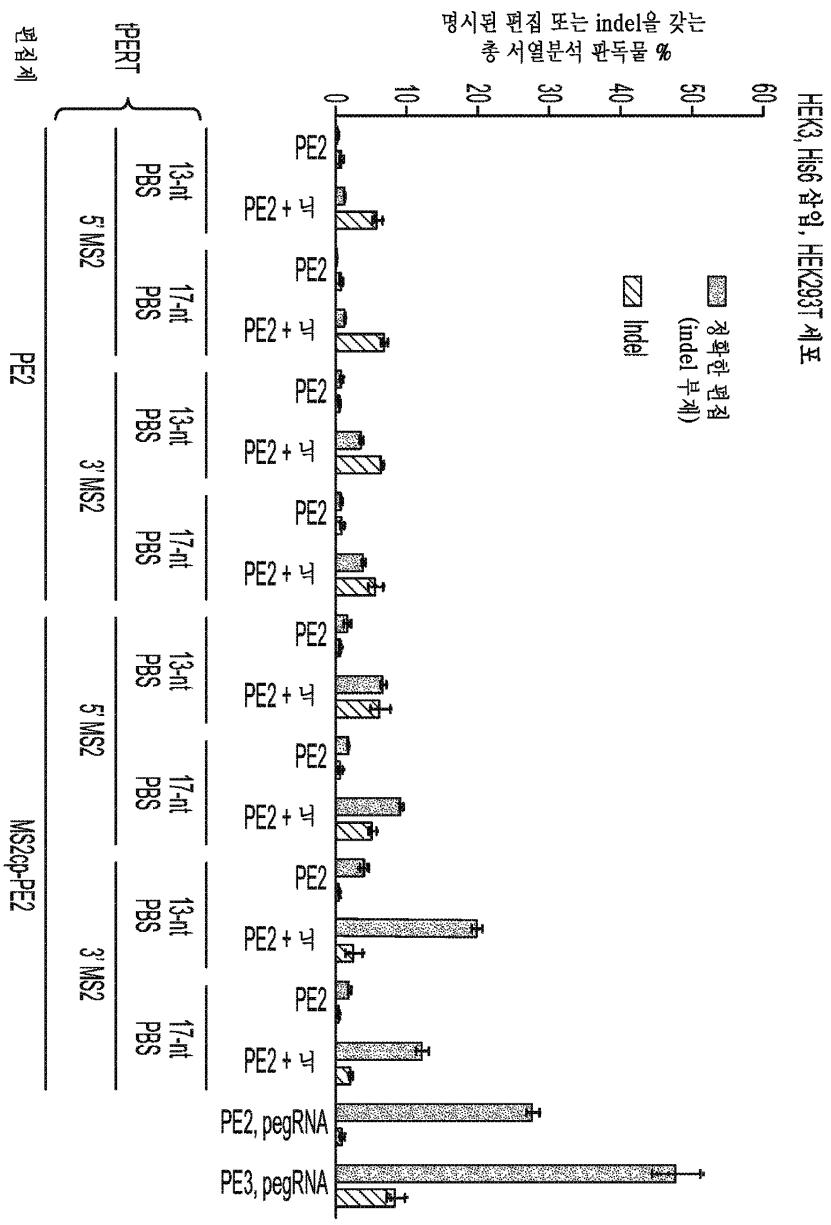
도면72b


표적
뉴클레오티드
MS2cp-PE2:sgRNA
DNA 결합,
비표적 가닥
닉
RT
Cas9
sgRNA
tPERT
tPERT
동원
RT
3'
Cas9
3'
sgRNA
5'
tPERT
역전사

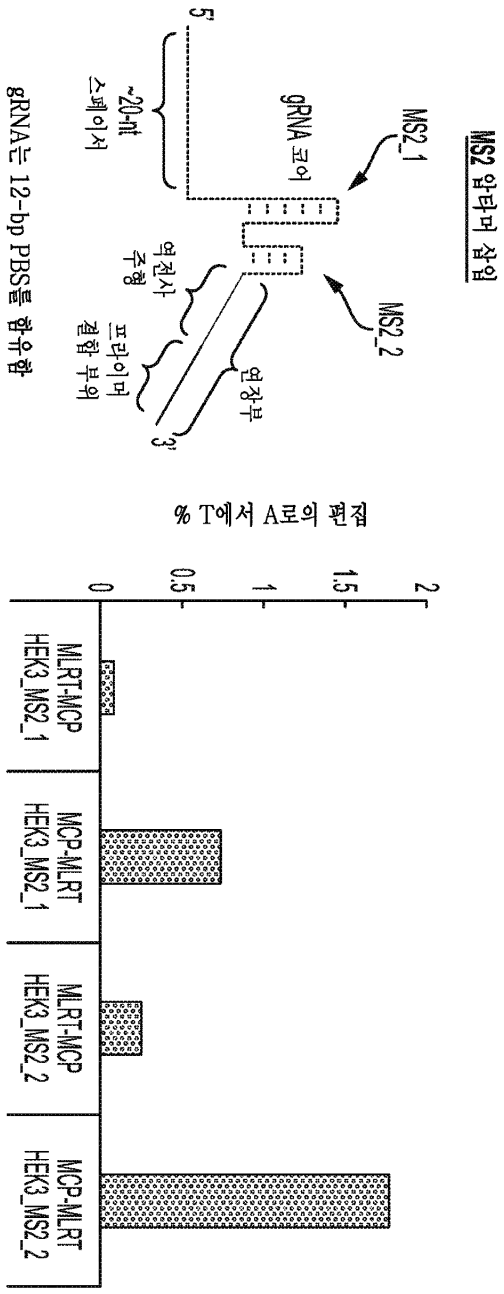
도면72c

하이브리드 RNA/DNA "pegRNA"
5'
RNA
DNA
3'
DNA 연장부는 프라이머 결합 부위 및 DNA 합성 주형을 함유함
DNA 합성 주형과 sgRNA 사이의 화학적 링커
5'
RNA
RNA또는DNA
3'
화학적 링커는 sgRNA 스캐폴드의 역전사를 방지함
역전된 DNA 합성 주형
5'
RNA
3'
RNA또는DNA
5'
5'에서 3'로의 배향은 DNA 합성 주형에서 스위칭됨
DNA-의존성 DNA 폴리머라제가 리버스 트랜스크립타제 대신에 사용될 수 있음

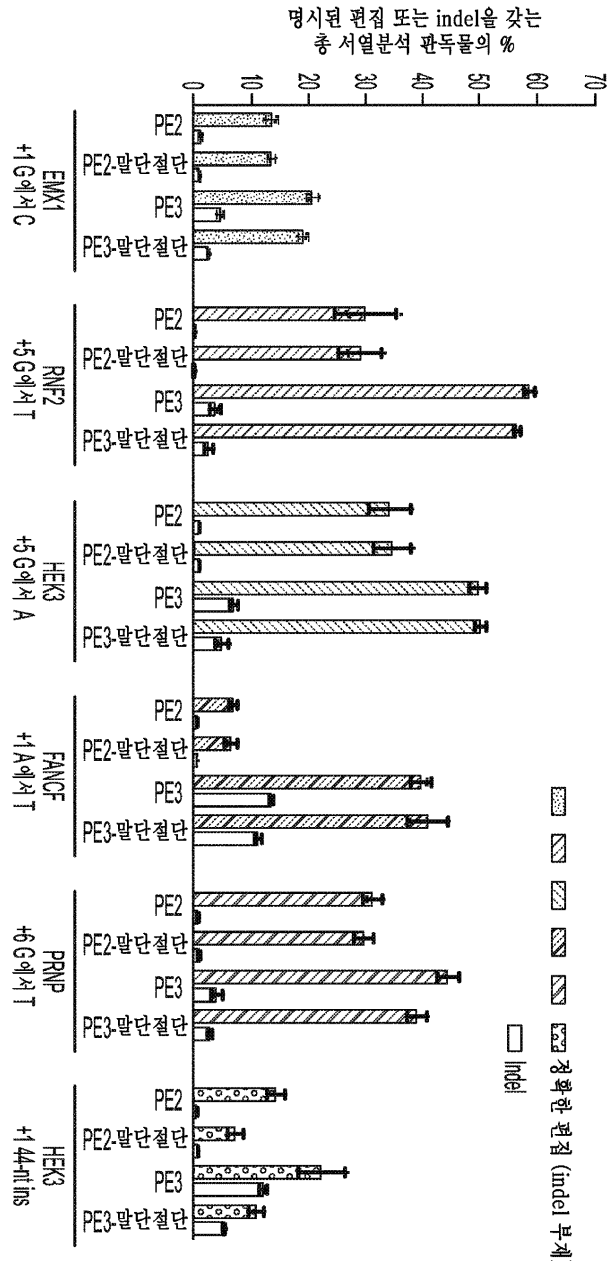
도면73


HEK3, His6 삽입, HEK293T 세포
명시된 편집 또는 indel을 갖는
총 서열분석 판독물 %
0
10
20
30
40
50
60
정확한 편집
(indel 부재)
Indel
PE2
PE2 + 닉
PE2, pegRNA
PE3, pegRNA
편집제
PERT
PE2
MS2cp-PE2
5' MS2
3' MS2
13-nt
17-nt
PBS

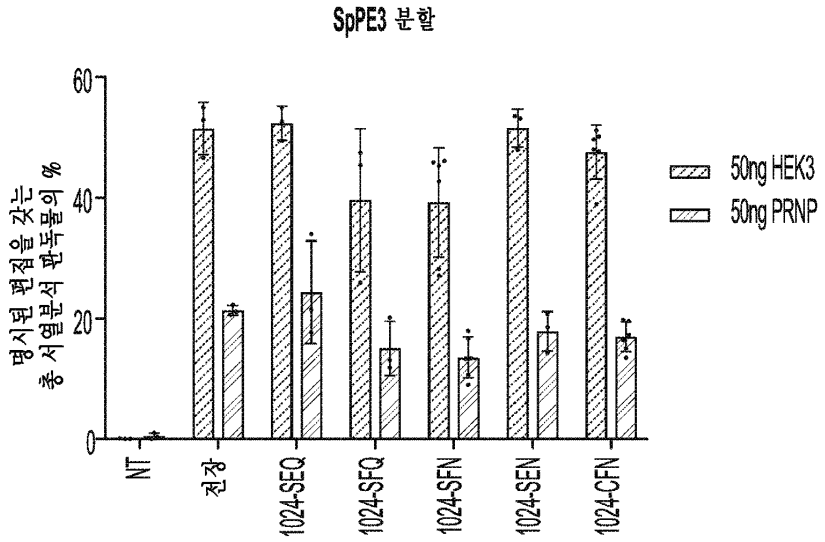
도면74


MS2 압타머 삽입
MS2_1
MS2_2
gRNA 코어
~20-nt
스페이서
5'
3'
역전사
주형
프라이머
결합 부위
연장부
gRNA는 12-bp PBS를 함유함
% T에서 A로의 편집
0
0.5
1
1.5
2
MLRT-MCP
HEK3_MS2_1
MCP-MLRT
HEK3_MS2_1
MLRT-MCP
HEK3_MS2_2
MCP-MLRT
HEK3_MS2_2

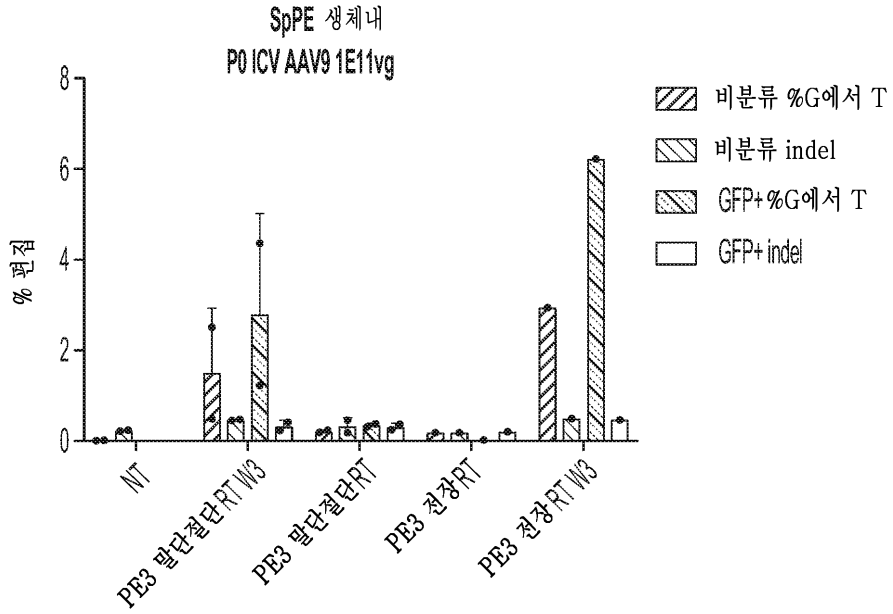
도면75


명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0
10
20
30
40
50
60
70
EMX1
+1 G에서 C
RNF2
+5 G에서 T
HEK3
+5 G에서 A
FANCF
+1 A에서 T
PRNP
+6 G에서 T
HEK3
+1 44-nt ins
PE2
PE2-말단절단
PE3
PE3-말단절단
Indel
정확한 편집 (indel 부재)

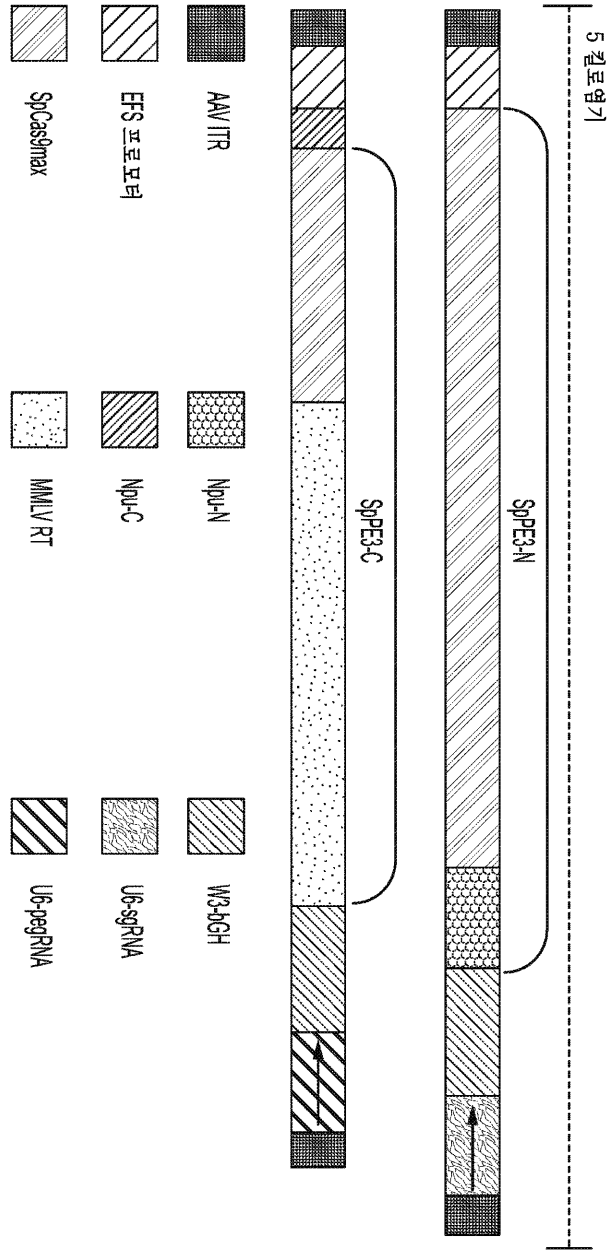
도면76


SpPE3 분할
명시된 편집을 갖는
총 서열분석 판독물의 %
60
40
20
0
NT
천창
1024-SEQ
1024-SFQ
1024-SFN
1024-SEN
1024-CFN
50ng HEK3
50ng PRNP

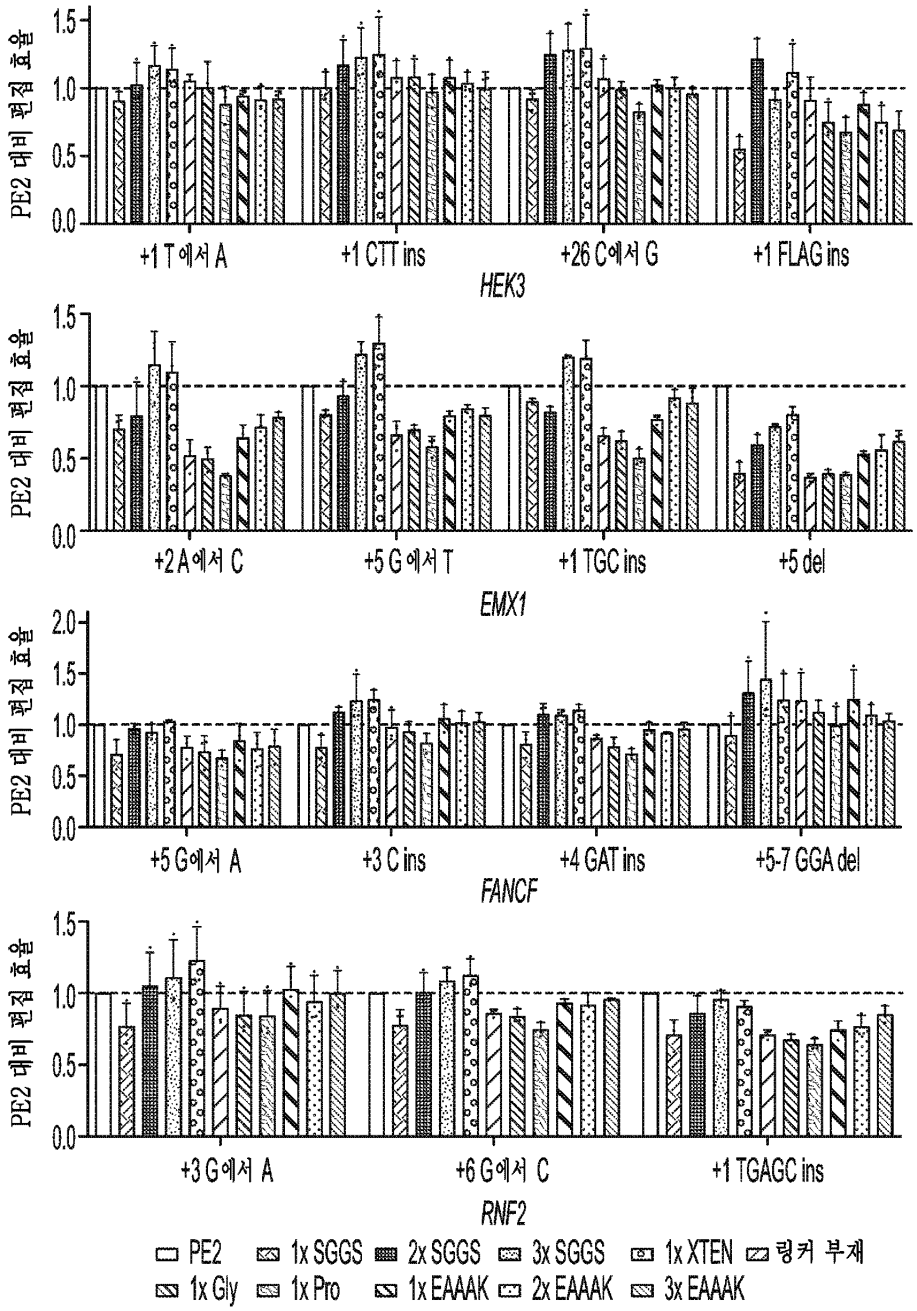
도면77


SpPE 생체내
P0 ICV AAV9 1E11vg
% 편집
8
6
4
2
0
NT
PE3 말단절단RT W3
PE3 말단절단RT
PE3 천장RT
PE3 천장RT W3
비분류 %G에서 T
비분류 indel
GFP+%G에서 T
GFP+ indel

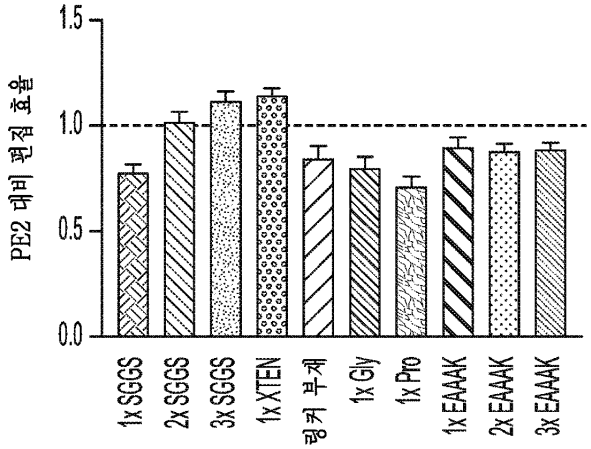
도면78


5 킬로염기
SpPE3-N
SpPE3-C
AAV ITR
EFS 프로모터
SpCas9max
Npu-N
Npu-C
MMLV RT
W3-bGH
U6-sgRNA
U6-pegRNA

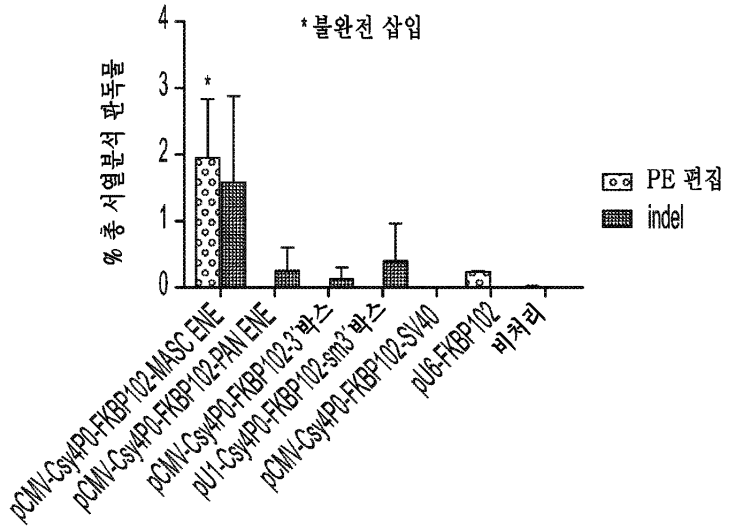
도면79


PE2 대비 편집 효율
1.5
1.0
0.5
0.0
+1 T 에서 A
+1 CTT ins
+26 C에서 G
+1 FLAG ins
HEK3
+2 A 에서 C
+5 G 에서 T
+1 TGC ins
+5 del
EMX1
2.0
+5 G에서 A
+3 C ins
+4 GAT ins
+5-7 GGA del
FANCF
+3 G에서 A
+6 G에서 C
+1 TGAGC ins
RNF2
PE2
1x SGGS
2x SGGS
3x SGGS
1x XTEN
링커 부재
1x Gly
1x Pro
1x EAAAK
2x EAAAK
3x EAAAK

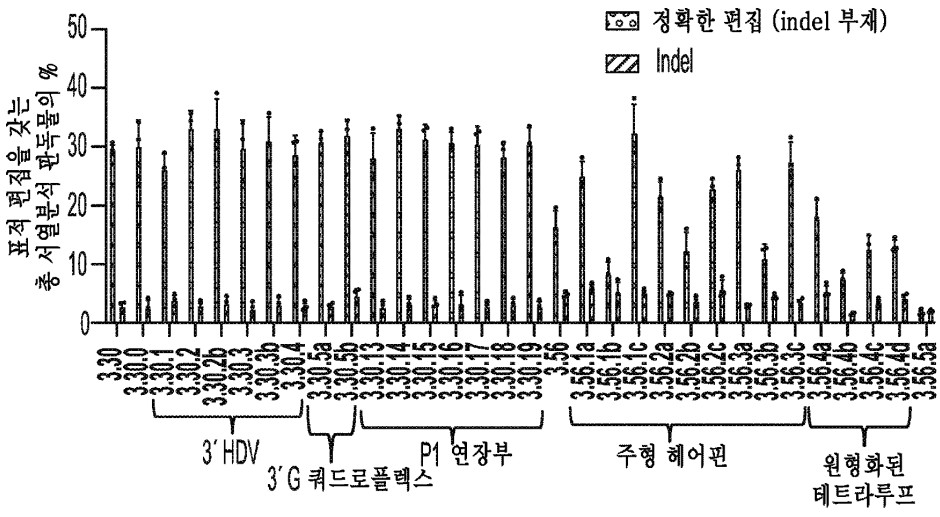
도면80



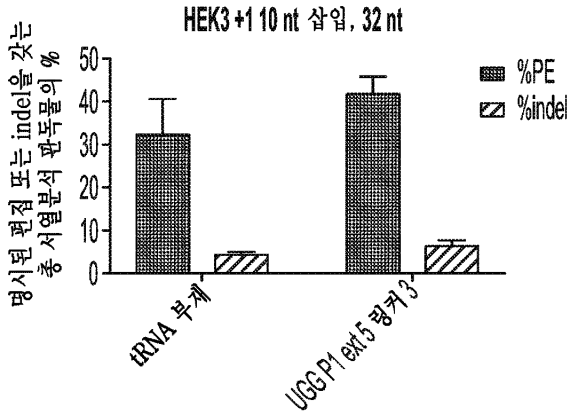
도면81



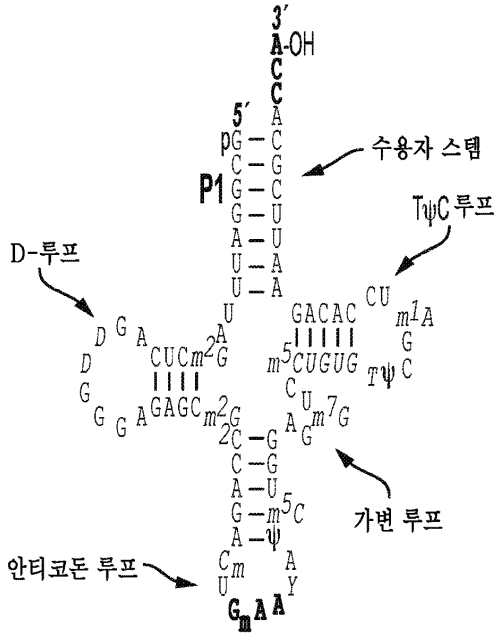
도면82



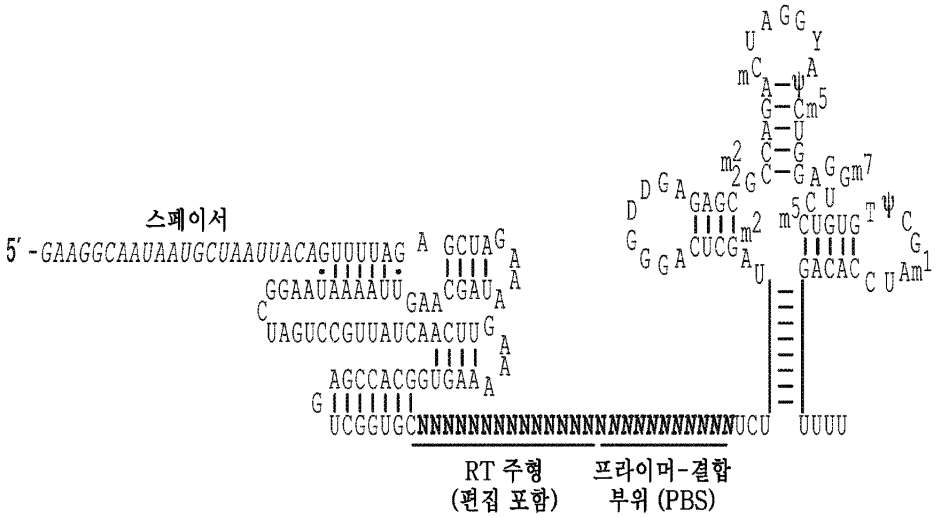
도면83


HEK3 +1 10 nt 삽입, 32 nt
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
50
40
30
20
10
0
%PE
%indel
tRNA 루케
UGGP1 ext5 링커3

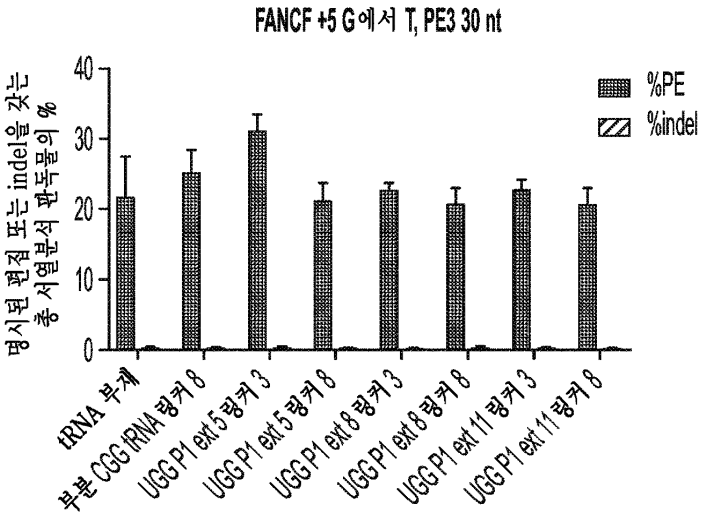
도면84a


3′
A-OH
5′
pG
P1
수용자 스템
TψC 루프
D-루프
가변 루프
안티코돈 루프

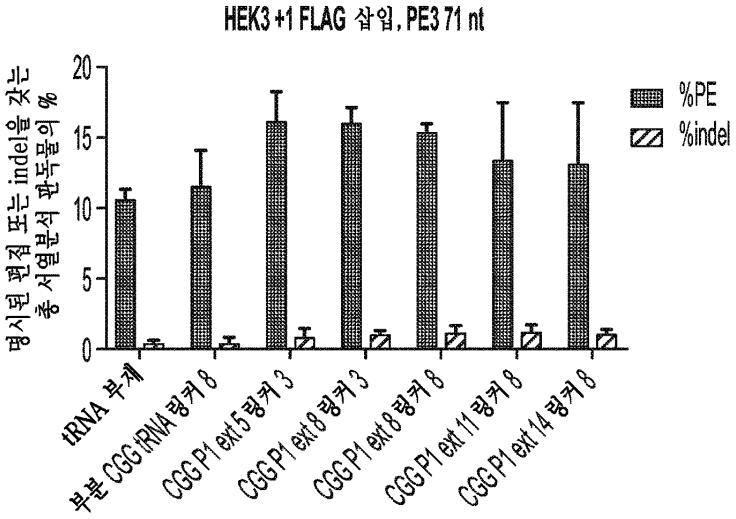
도면84b


스페이서
5′-GAAGGCAAUAAUGCUAAUUACAGUUUUAG
GCUAG
GGAAUAAAAUU
GAA
CGAU
UAGUCCGUUAUCAACUU
AGCCACGGUGAA
UCGGUGC
NNNNNNNNNNNNNNNNNNNNNNNNNN
UCU
UUUU
RT 주형
(편집 포함)
프라이머-결합
부위 (PBS)

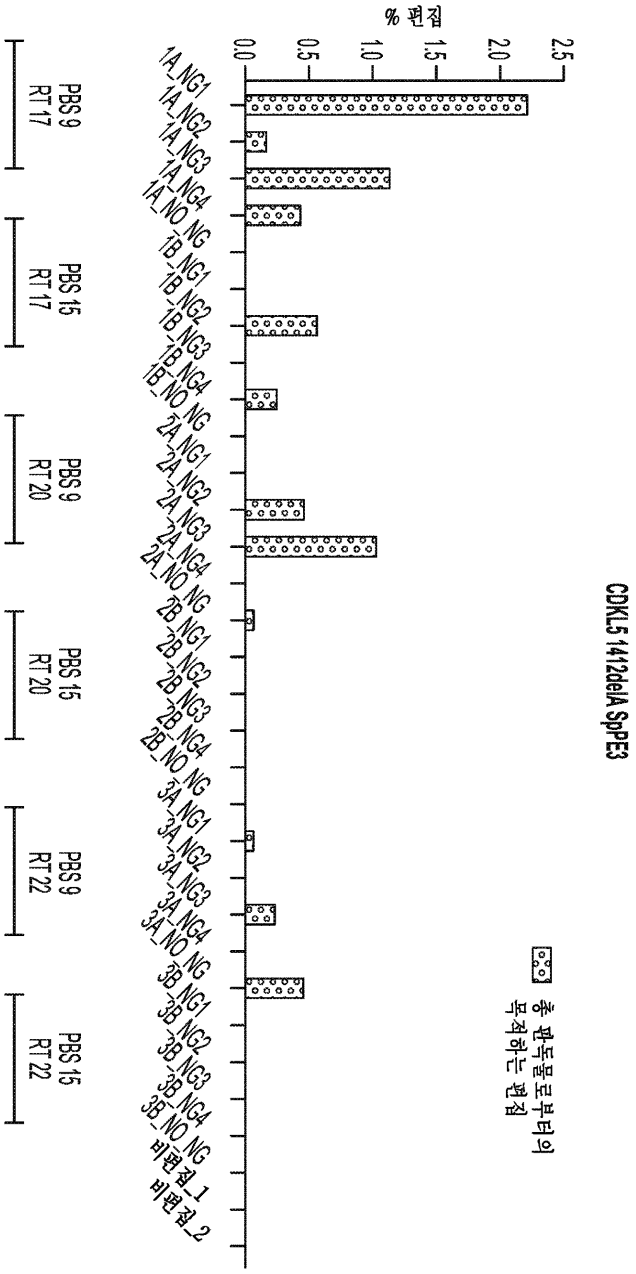
도면85


FANCF +5 G에서 T, PE3 30 nt
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
40
30
20
10
0
%PE
%indel
tRNA 부재
부분 CGG tRNA 링커 8
UGG P1 ext 5 링커 3
UGG P1 ext 5 링커 8
UGG P1 ext 8 링커 3
UGG P1 ext 8 링커 8
UGG P1 ext 11 링커 3
UGG P1 ext 11 링커 8

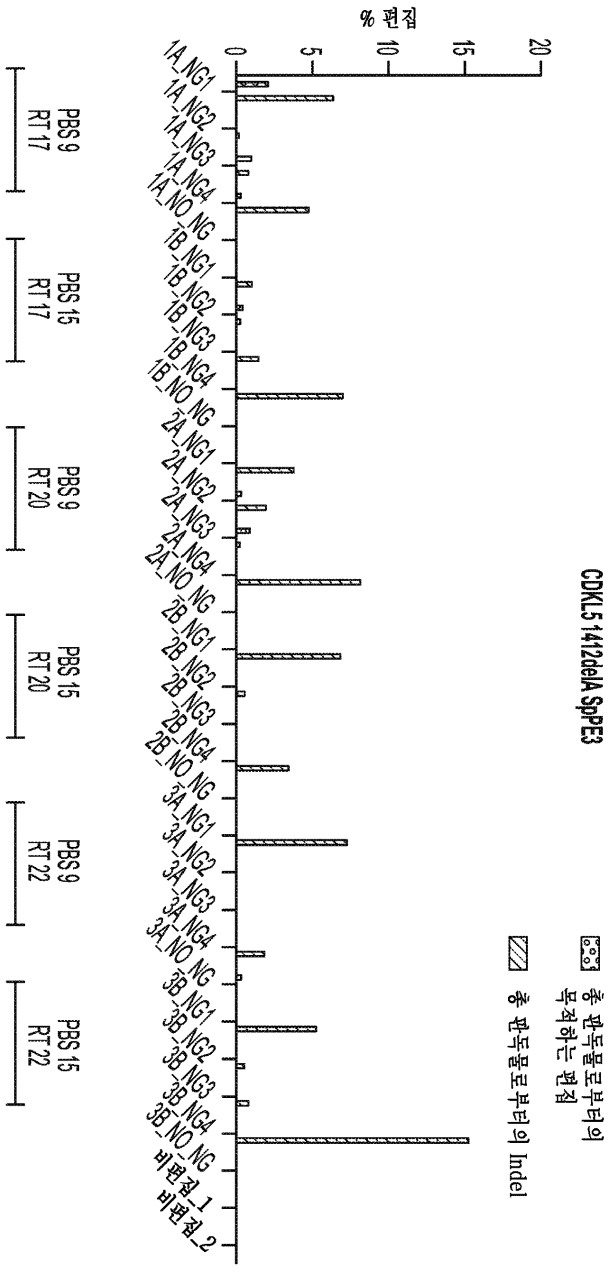
도면86


HEK3 +1 FLAG 삽입, PE3 71 nt
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
20
15
10
5
0
%PE
%indel
tRNA 부재
부분 CGG tRNA 링커 8
CGG P1 ext 5 링커 3
CGG P1 ext 8 링커 3
CGG P1 ext 8 링커 8
CGG P1 ext 11 링커 8
CGG P1 ext 14 링커 8

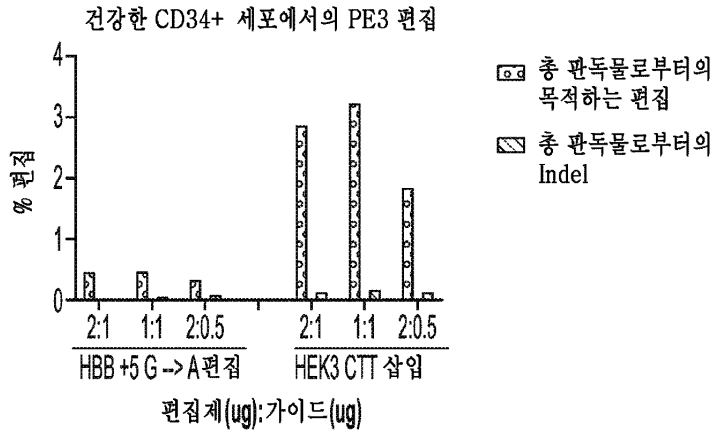
도면87


% 편집
0.0
0.5
1.0
1.5
2.0
2.5
1A_NG1
1A_NG2
1A_NG3
1A_NG4
1A_NO_NG
1B_NG1
1B_NG2
1B_NG3
1B_NG4
1B_NO_NG
2A_NG1
2A_NG2
2A_NG3
2A_NG4
2A_NO_NG
2B_NG1
2B_NG2
2B_NG3
2B_NG4
2B_NO_NG
3A_NG1
3A_NG2
3A_NG3
3A_NG4
3A_NO_NG
3B_NG1
3B_NG2
3B_NG3
3B_NG4
3B_NO_NG
비편집_1
비편집_2
PBS 9
RT 17
PBS 15
RT 17
PBS 9
RT 20
PBS 15
RT 20
PBS 9
RT 22
PBS 15
RT 22
CDKL5 1412delA SpPE3
총 편집물로부터의
목적하는 편집

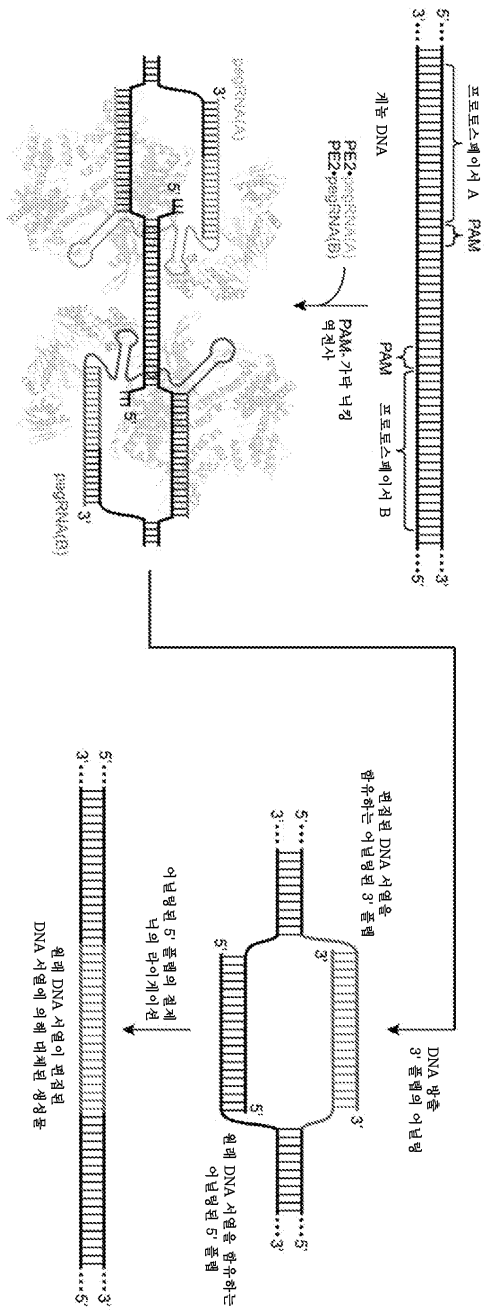
도면88


CDKL5 1412delA SpPE3
총 편독물로부터의 목적하는 편집
총 편독물로부터의 Indel
% 편집
0
5
10
15
20
1A_NG1
1A_NG2
1A_NG3
1A_NG4
1A_NO_NG
1B_NG1
1B_NG2
1B_NG3
1B_NG4
1B_NO_NG
2A_NG1
2A_NG2
2A_NG3
2A_NG4
2A_NO_NG
2B_NG1
2B_NG2
2B_NG3
2B_NG4
2B_NO_NG
3A_NG1
3A_NG2
3A_NG3
3A_NG4
3A_NO_NG
3B_NG1
3B_NG2
3B_NG3
3B_NG4
3B_NO_NG
비편집_1
비편집_2
PBS 9 RT 17
PBS 15 RT 17
PBS 9 RT 20
PBS 15 RT 20
PBS 9 RT 22
PBS 15 RT 22

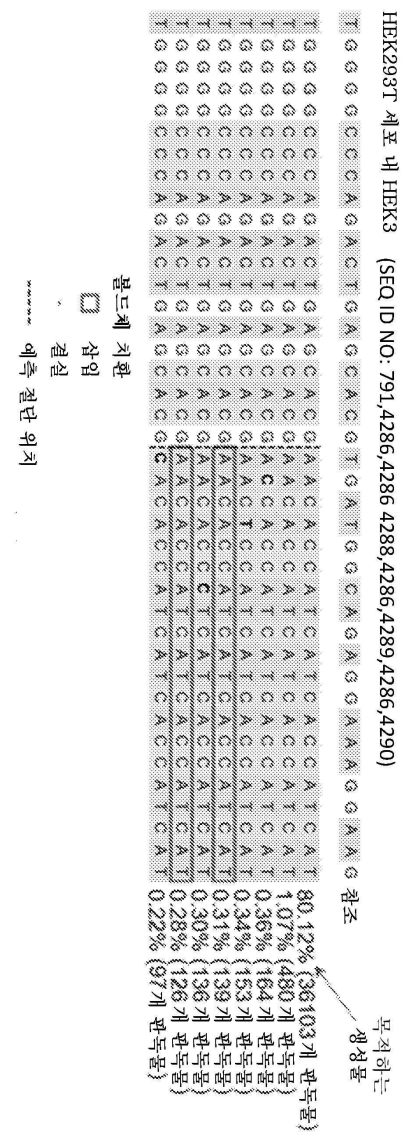
도면89


건강한 CD34+ 세포에서의 PE3 편집
% 편집
4
3
2
1
0
2:1
1:1
2:0.5
2:1
1:1
2:0.5
HBB +5 G -> A편집
HEK3 CTT 삽입
편집제(ug):가이드(ug)
총 판독물로부터의 목적하는 편집
총 판독물로부터의 Indel

도면90


프로토스페이서 A PAM
게놈 DNA
PAM 프로토스페이서 B
PAM-가닌 닉킹 엑젠사
pegRNA(B)
pegRNA(A)
DNA 방출
3' 플랩의 어닐링
변경된 DNA 서열을 함유하는 어닐링된 3' 플랩
어닐링된 5' 플랩의 제거
닉의 라이게이션
원래 DNA 서열을 함유하는 어닐링된 5' 플랩
원래 DNA 서열이 편집된 DNA 서열에 의해 대체된 생성물

도면91a


HEK293T 세포 내 HEK3 (SEQ ID NO: 791,4286,4286 4288,4286,4289,4286,4290)
목적하는 생성물
참조
80.12% {361037개 편독물}
1.07% {4807개 편독물}
0.36% {1647개 편독물}
0.34% {1537개 편독물}
0.31% {1397개 편독물}
0.30% {1367개 편독물}
0.28% {1267개 편독물}
0.22% {977개 편독물}
블드체 치환
삽입
결실
예측 절단 위치

도면91b

pegRNA1: (SEQ ID NO: 4291)

5' – GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC

CGTTATCAACTTGAAAAGTGGCACCGAGTCGGTGCAAATGATGGTGATGATGGTGTTCGTGCTCAGTCTG – 3'

pegRNA2: (SEQ ID NO: 4292)

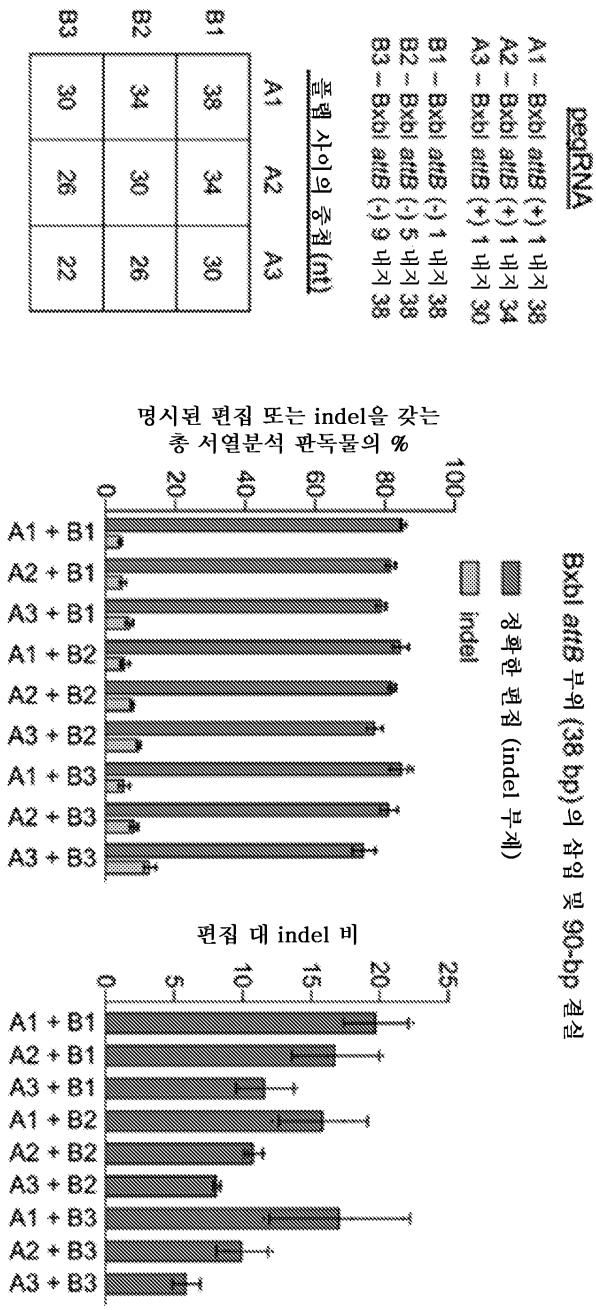
5' – GTCAACCAGTATCCCCGGTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC

CGTTATCAACTTGAAAAGTGGCACCGAGTCGGTGCAACACCATCATCACCATCATTTCCGGGATACTGG – 3'

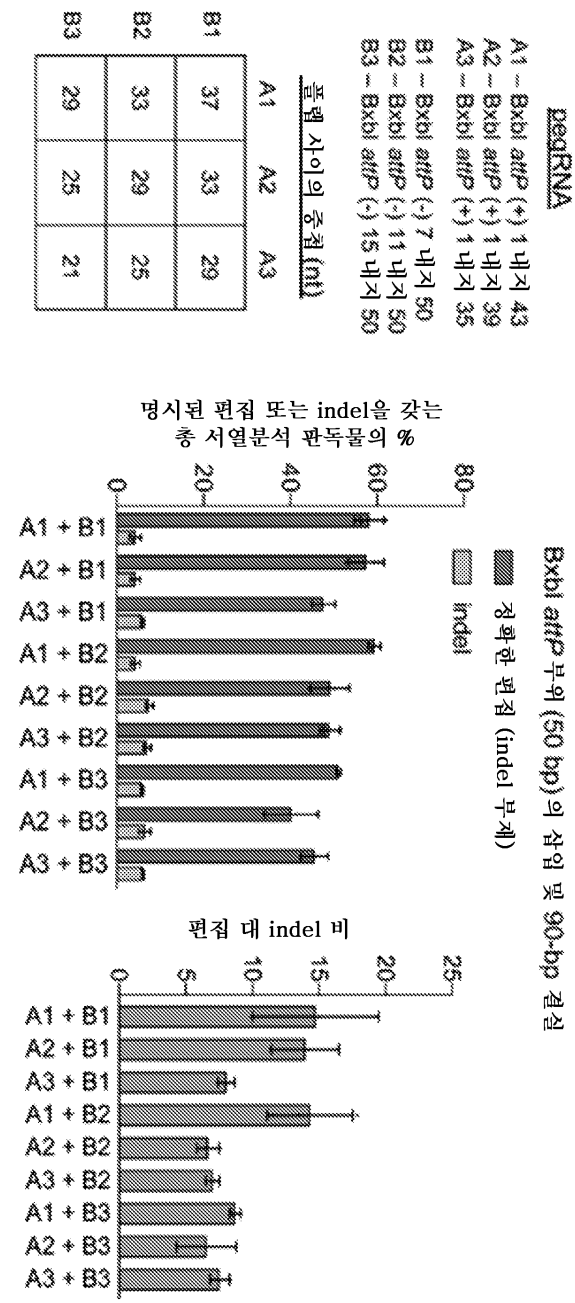
도면92

pegRNA A
gRNA 코어
RT 주형 B에 대한
일부 상보성
RT 주형
A
PBS
A
3'
5'
gRNA 스페이서
A
pegRNA B
gRNA 코어
RT 주형 A에 대한
일부 상보성
RT 주형
B
PBS
B
3'
5'
gRNA 스페이서
B
pegRNA A 및 B에서의 가능한 RT 주형의 예:
RT 주형 A
RT 주형 A에
대한 상보성
5' - CACUUCGAUUCGAUUCGUUGACUGUAGCUAGCUACAGCUCA - 3'
(SEQ ID NO: 4293)
RT 주형 B
RT 주형 A에
대한 상보성
5' - CAUCAGCAUUCGAUUCGAACGUUACGUUAGCUAGCUUCA - 3'
(SEQ ID NO: 4294)
RT 주형 A 및 B로부터의 역전사 생성물:
RT 생성물 A
RT 생성물 B에
대한 상보성
5' - TGAGCTGTAGCTAGCTACAGTCAGCATCGATCGAATCCAGTG - 3'
(SEQ ID NO: 4295)
RT 생성물 B
RT 생성물 A에
대한 상보성
5' - TGAGCTGTAGCTAGCTACAGACTGGATCGATCGATGCTGAC - 3'
(SEQ ID NO: 4296)
RT 생성물 A 및 B로부터의 생성된 플랩 플랩:
어닐링을 용이하게 하기 위한
플랩 사이의 상보적 영역
(SEQ ID NO: 4295)
5' - TGAGCTGTAGCTACAGTCAGCATCGATCGAATCCAGTG - 3'
3' - CAGTCGTAGCTAGCTAGGTCAGTGAGCTGTAGCTACA - 5'
(SEQ ID NO: 4297)

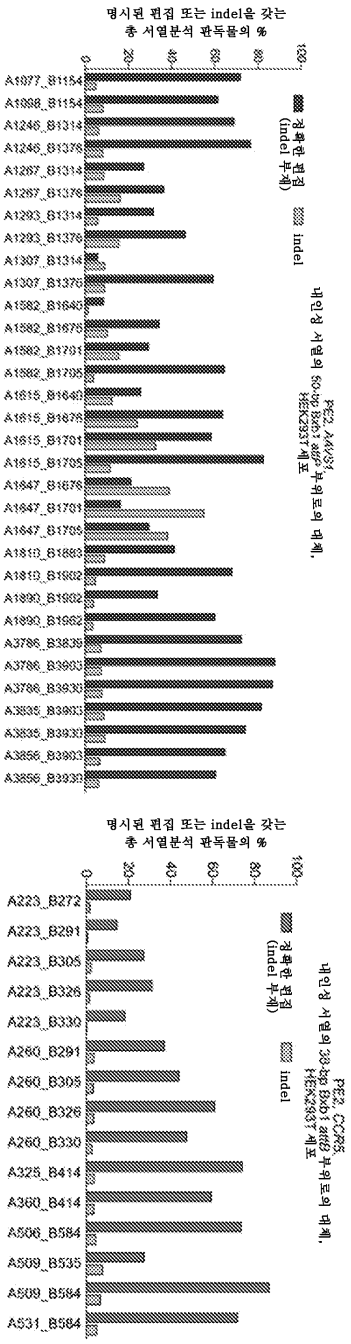
도면93a


pegRNA
A1 ~ BxbI attB (+) 1 내지 38
A2 ~ BxbI attB (+) 1 내지 34
A3 ~ BxbI attB (+) 1 내지 30
B1 ~ BxbI attB (-) 1 내지 38
B2 ~ BxbI attB (-) 5 내지 38
B3 ~ BxbI attB (-) 9 내지 38
플랩 사이의 중첩 (nt)
A1 A2 A3
B1 38 34 30
B2 34 30 26
B3 30 26 22
BxbI attB 부위 (38 bp)의 삽입 및 90-bp 결실
정확한 편집 (indel 부재)
indel
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0 20 40 60 80 100
A1 + B1
A2 + B1
A3 + B1
A1 + B2
A2 + B2
A3 + B2
A1 + B3
A2 + B3
A3 + B3
편집 대 indel 비
0 5 10 15 20 25
A1 + B1
A2 + B1
A3 + B1
A1 + B2
A2 + B2
A3 + B2
A1 + B3
A2 + B3
A3 + B3

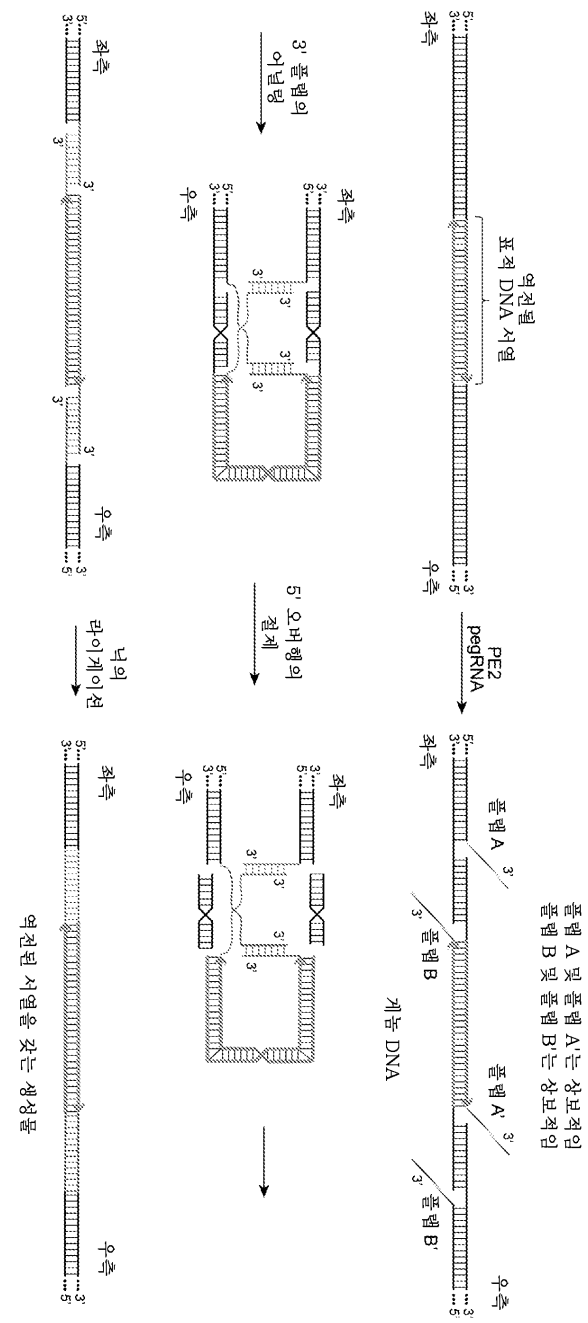
도면93b


pegRNA
A1 – BxbI attP (+) 1 내지 43
A2 – BxbI attP (+) 1 내지 39
A3 – BxbI attP (+) 1 내지 35
B1 – BxbI attP (-) 7 내지 50
B2 – BxbI attP (-) 11 내지 50
B3 – BxbI attP (-) 15 내지 50
플랩 사이의 중첩 (nt)
A1 A2 A3
B1 37 33 29
B2 33 29 25
B3 29 25 21
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0 20 40 60 80
A1 + B1
A2 + B1
A3 + B1
A1 + B2
A2 + B2
A3 + B2
A1 + B3
A2 + B3
A3 + B3
정확한 편집 (indel 부재)
indel
BxbI attP 부위 (50 bp)의 삽입 및 90-bp 결실
편집 대 indel 비
0 5 10 15 20 25
A1 + B1
A2 + B1
A3 + B1
A1 + B2
A2 + B2
A3 + B2
A1 + B3
A2 + B3
A3 + B3

도면94


명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0
20
40
60
80
100
A1077_B1154
A1088_B1154
A1246_B1314
A1246_B1376
A1267_B1314
A1267_B1376
A1283_B1314
A1283_B1376
A1307_B1314
A1307_B1376
A1582_B1645
A1582_B1676
A1582_B1701
A1582_B1705
A1615_B1645
A1615_B1676
A1615_B1701
A1615_B1705
A1647_B1676
A1647_B1701
A1647_B1705
A1810_B1883
A1810_B1902
A1890_B1902
A1890_B1962
A3786_B3839
A3786_B3903
A3786_B3930
A3835_B3903
A3835_B3930
A3856_B3903
A3856_B3930
정확한 편집
(indel 누계)
indel
PE2, A6V3X, 5'-3' 방향의 pegRNA 유효 프로브 대체,
서열의 50-3bp 하류의 HEK2937 세포
명시된 편집 또는 indel을 갖는
총 서열분석 판독물의 %
0
20
40
60
80
100
A223_B272
A223_B291
A223_B305
A223_B326
A223_B330
A260_B291
A260_B305
A260_B326
A260_B330
A325_B414
A360_B414
A506_B584
A509_B535
A509_B584
A531_B584
정확한 편집
(indel 누계)
indel
PE2, CCR5, 5'-3' 방향의 pegRNA 유효 프로브 대체,
서열의 28-29bp 하류의 HEK2937 세포

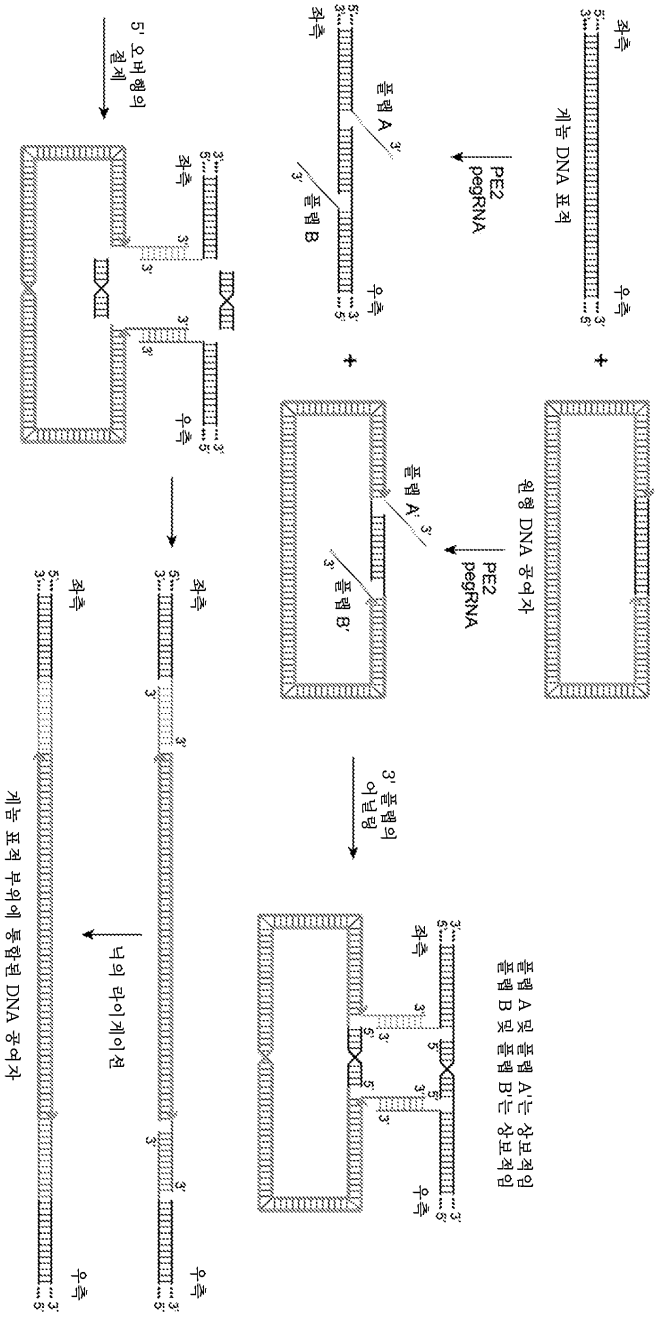
도면95


좌측
우측
역전된 표적 DNA 서열
PE2
pegRNA
플랩 A
플랩 B
게놈 DNA
플랩 A' 및 플랩 B'는 상보적임
3' 플랩의 어닐링
5' 오버행의 절제
닉의 라이게이션
역전된 서열을 갖는 생성물

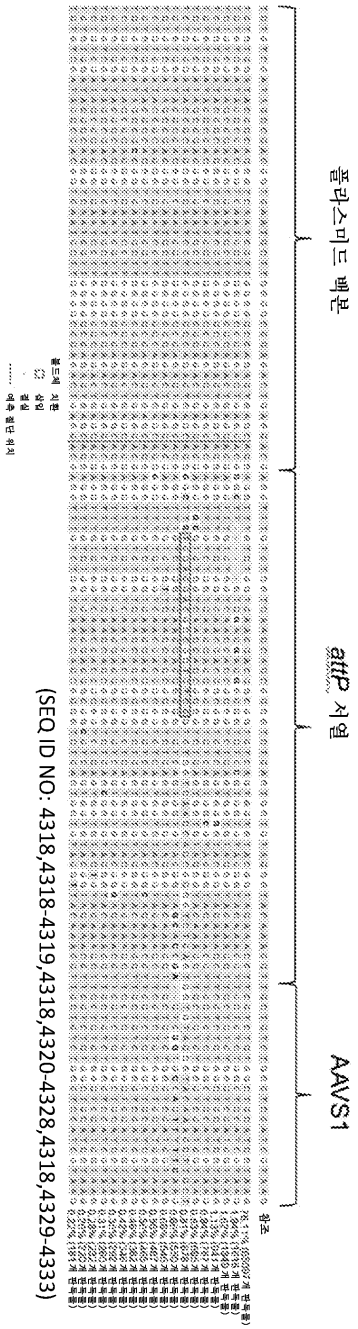
도면96

PCR 앰플리콘의 HTS: (SEQ ID NO: 4298,4298-4307)
예상 접합부 A(attP 부위를 함유함)에 정렬된 판독물 - 정렬된 총 판독물의 90%, 8-10% indel
(SEQ ID NO: 4308,4308-4317)
예상 접합부 B(attB 부위를 함유함)에 정렬된 판독물 - 정렬된 총 판독물의 93%, 2.5% indel

도면97a

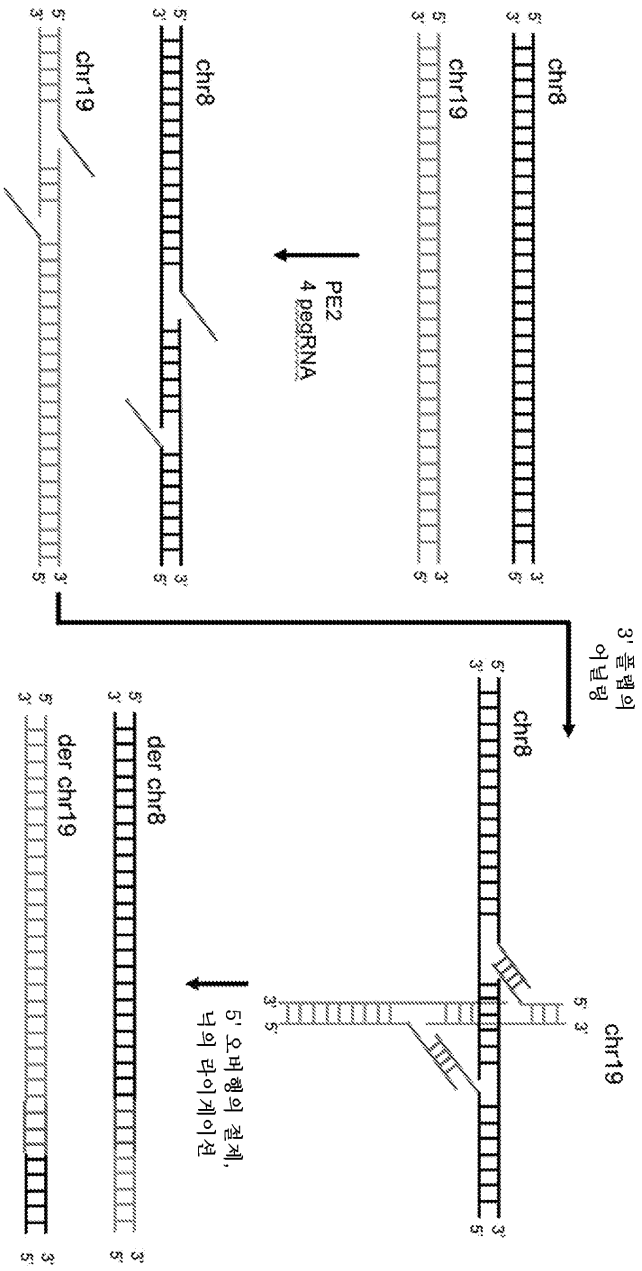

좌측
게놈 DNA 표적
우측
플랩 A 3'
3'
플랩 B
PE2
pegRNA
플랩 A' 3'
3'
플랩 B'
원형 DNA 공여자
3' 플랩의
어닐링
플랩 A 및 플랩 A'는 상보적임
플랩 B 및 플랩 B'는 상보적임
5' 오버행의
절제
넉의 라이게이션
게놈 표적 부위에 통합된 DNA 공여자

도면97b



(SEQ ID NO: 4318,4318-4319,4318,4320-4328,4318,4329-4333)

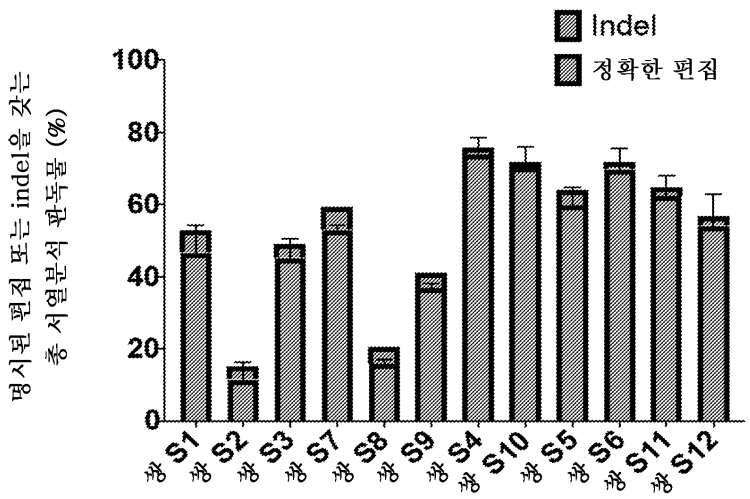
도면98a


chr8
chr19
PE2
4 pegRNA
chr8
chr19
3' 플랩의
어닐링
chr8
chr19
5' 오버행의 절제,
닉의 라이게이션
der chr8
der chr19

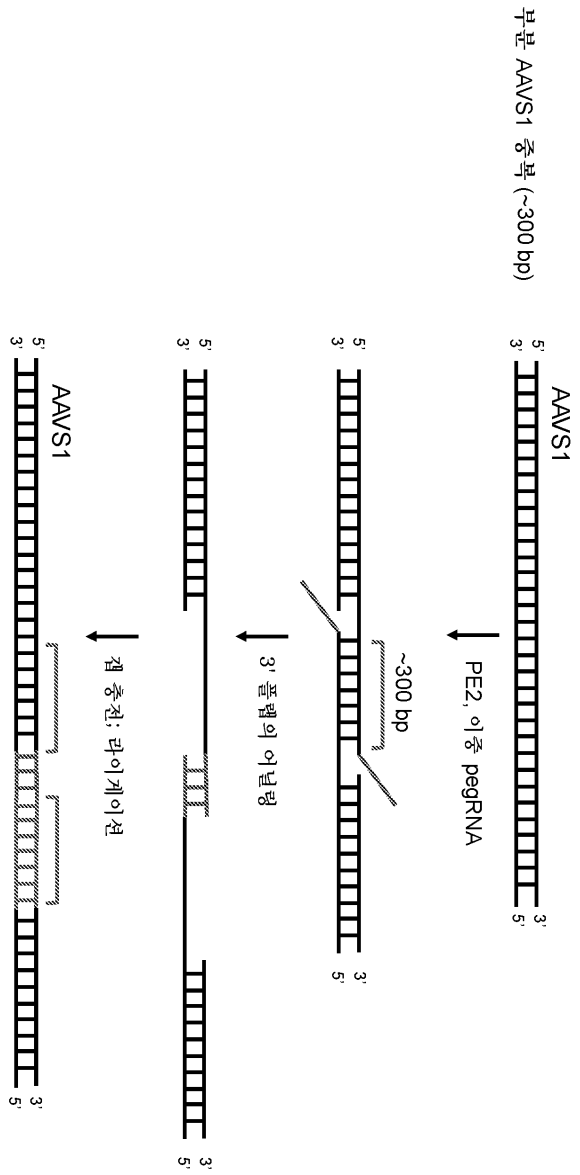
도면98b

PCR 앰플리콘의 HTS (pegRNA 세트 #5 rep2):
예상 접합부 1 (His6 태그를 함유함)에 정렬된 판독물 - 정렬된 총 판독물의 99.5%
(SEQ ID NO: 4334,4334-4336)
spacer
pegRNA
참조
PCR 앰플리콘의 HTS (pegRNA 세트 #5 rep2):
예상 접합부 2 (22nt 무작위 서열을 함유함)에 정렬된 판독물 - 정렬된 총 판독물의 97.9%
(SEQ ID NO: 4337,4337-4341)

도면99


Indel
정확한 편집
명시된 편집 또는 indel을 갖는
총 서열분석 판독물 (%)
100
80
60
40
20
0
S1
S2
S3
S7
S8
S9
S4
S10
S5
S6
S11
S12

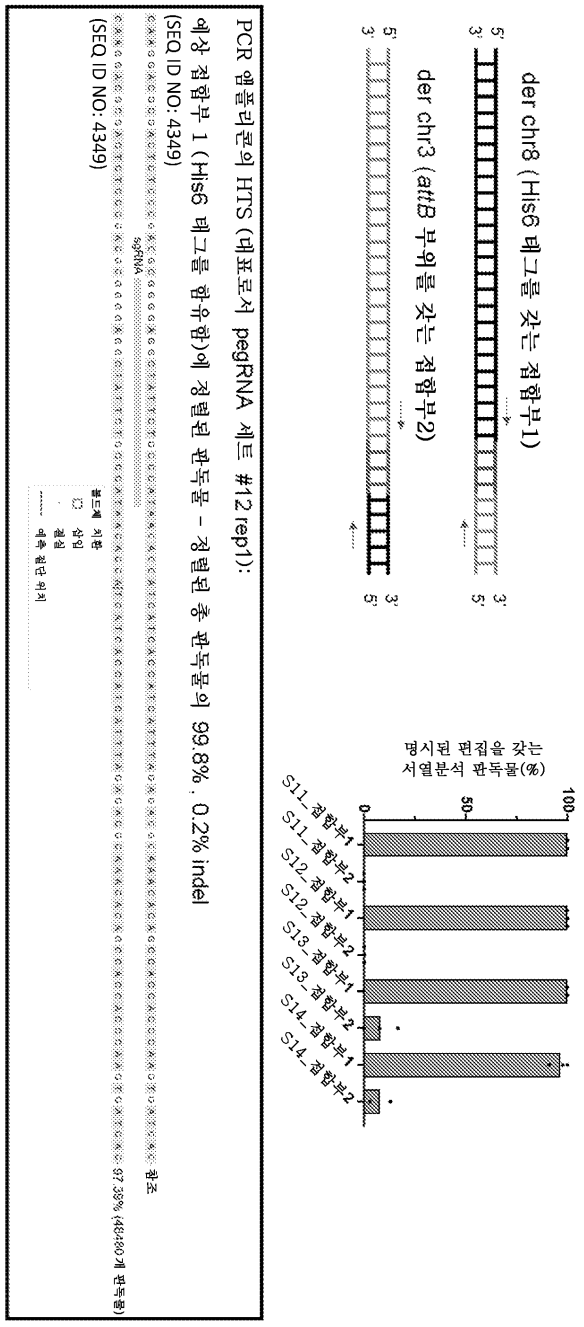
도면100a


부분 AAVS1 중복 (~300 bp)
AAVS1
PE2, 이중 pegRNA
~300 bp
3' 플랩의 어닐링
갭 충전; 라이게이션
AAVS1
5'
3'

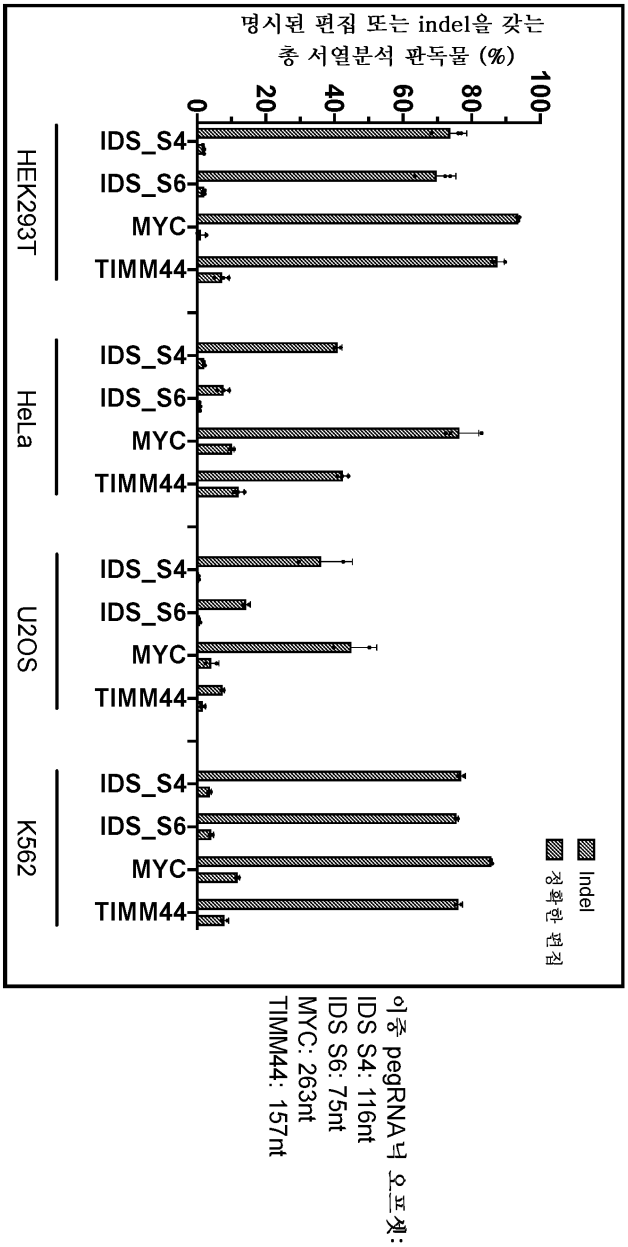
도면100b

(SEQ ID NO: 4337,4337-4341)

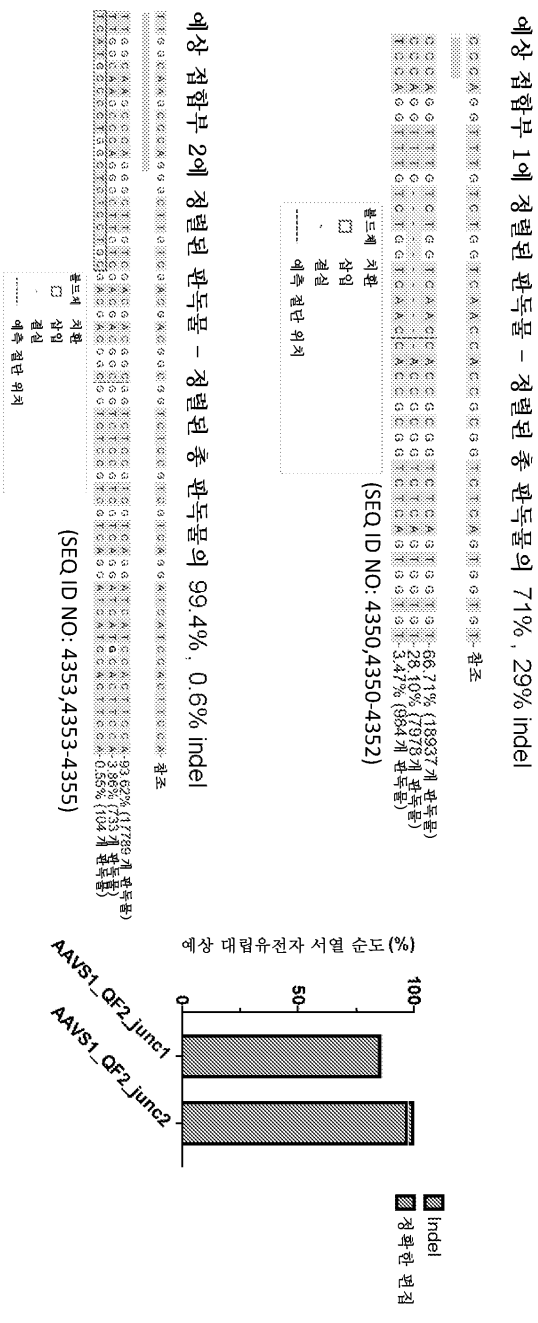
도면101


der chr8 (His6 태그를 갖는 접합부1)
5'
3'
3'
5'
der chr3 (attB 부위를 갖는 접합부2)
5'
3'
3'
5'
명시된 편집을 갖는
서열분석 판독물(%)
100
50
0
S11_접합부1
S11_접합부2
S12_접합부1
S12_접합부2
S13_접합부1
S13_접합부2
S14_접합부1
S14_접합부2
PCR 앰플리콘의 HTS (대표로서 pegRNA 세트 #12 rep1):
예상 접합부 1 (His6 태그를 함유함)에 정렬된 판독물 - 정렬된 총 판독물의 99.8%, 0.2% indel
(SEQ ID NO: 4349)
sgRNA
참조
97.39% (46460개 판독물)
(SEQ ID NO: 4349)
블드체 치환
삽입
결실
예측 절단 위치

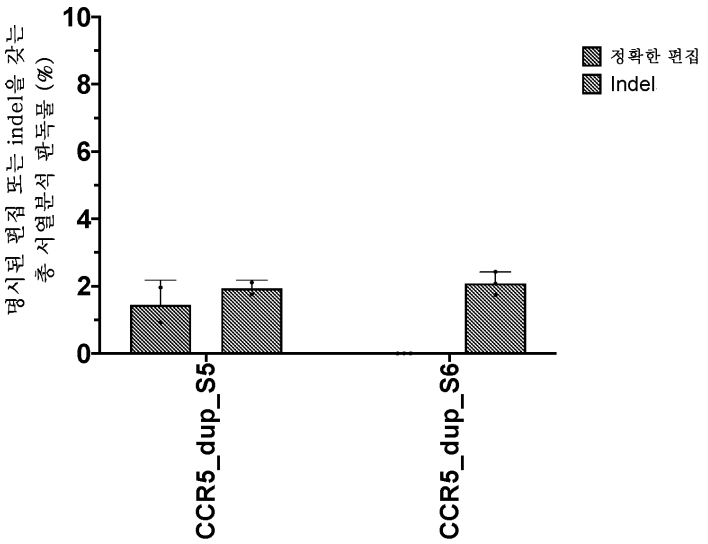
도면102a


명시된 편집 또는 indel을 갖는
총 서열분석 판독물 (%)
0
20
40
60
80
100
HEK293T
HeLa
U2OS
K562
IDS_S4
IDS_S6
MYC
TIMM44
Indel
정확한 편집
이중 pegRNA닉 오프셋:
IDS S4: 116nt
IDS S6: 75nt
MYC: 263nt
TIMM44: 157nt

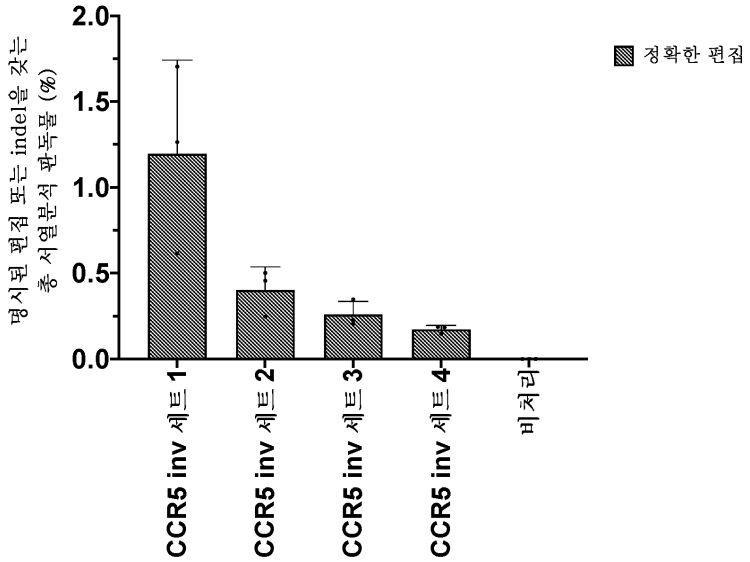
도면102b


예상 접합부 1에 정렬된 판독물 – 정렬된 총 판독물의 71% , 29% indel
참조
66.71% (18937개 판독물)
28.10% (7978개 판독물)
3.47% (984개 판독물)
(SEQ ID NO: 4350,4350-4352)
폴드체 처환
삽입
결실
예측 절단 위치
예상 접합부 2에 정렬된 판독물 – 정렬된 총 판독물의 99.4% , 0.6% indel
참조
93.62% (17289개 판독물)
3.88% (733개 판독물)
0.55% (104개 판독물)
(SEQ ID NO: 4353,4353-4355)
폴드체 처환
삽입
결실
예측 절단 위치
예상 대립유전자 서열 순도(%)
0
50
100
AAVS1_QF2_junc1
AAVS1_QF2_junc2
Indel
정확한 편집

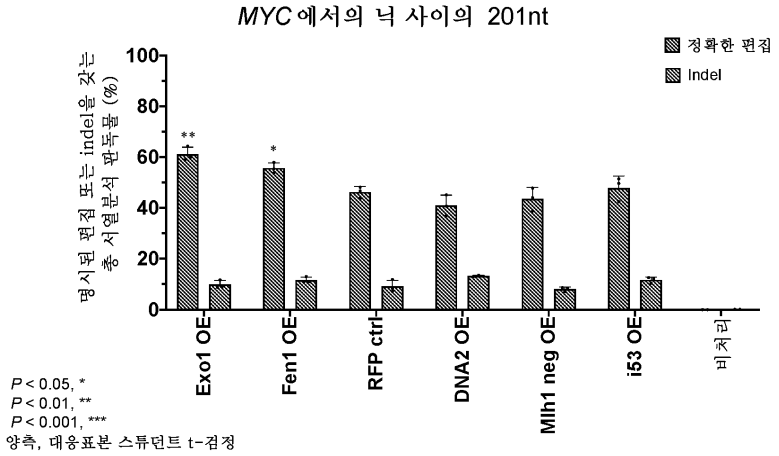
도면103a



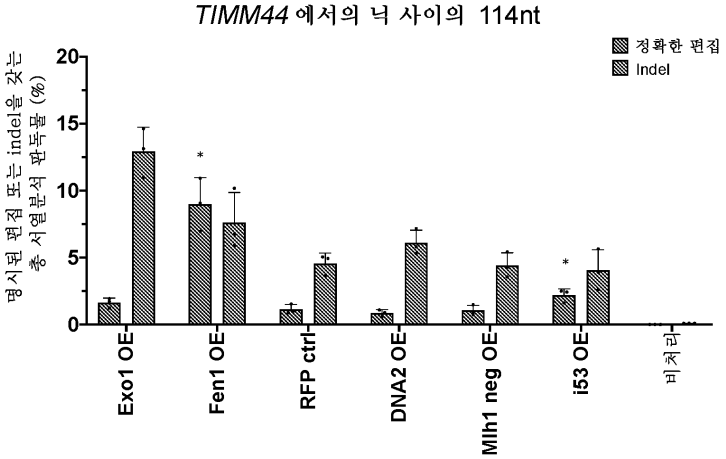
도면103b



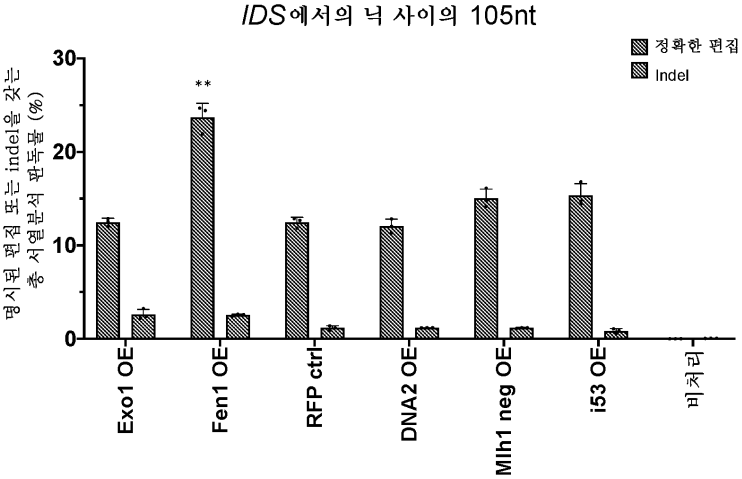
도면104a



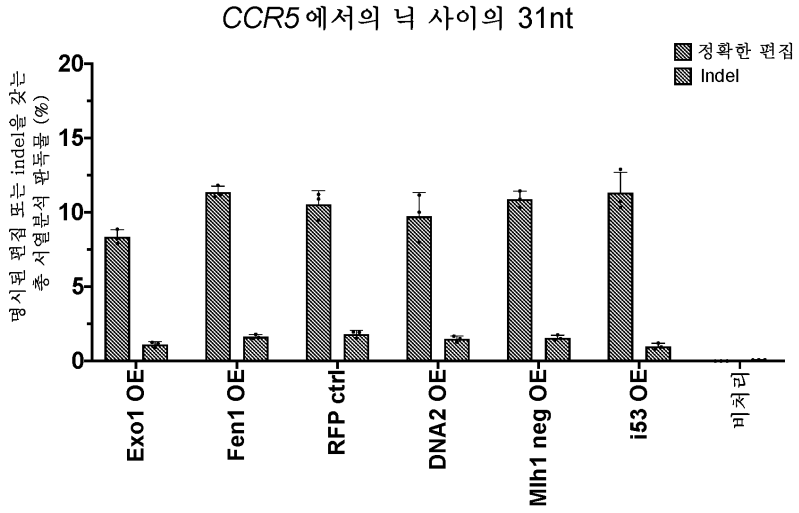
도면104b



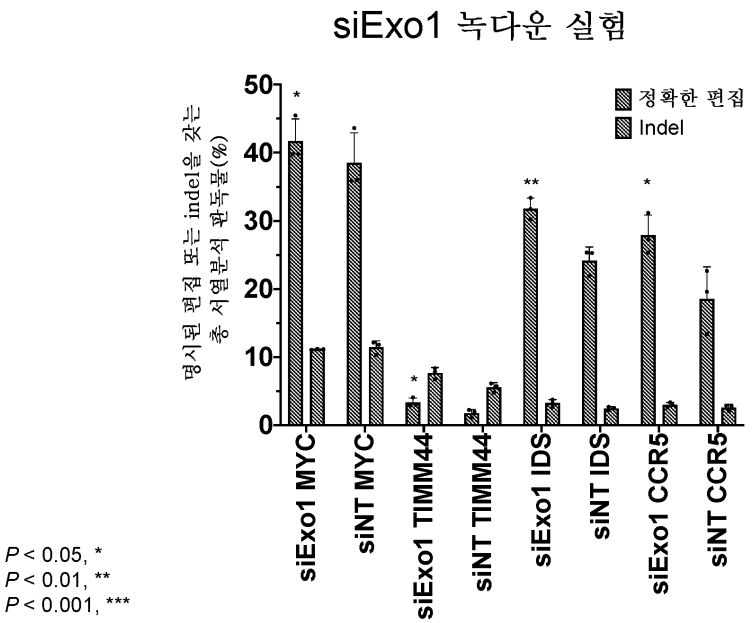
도면104c



도면104d

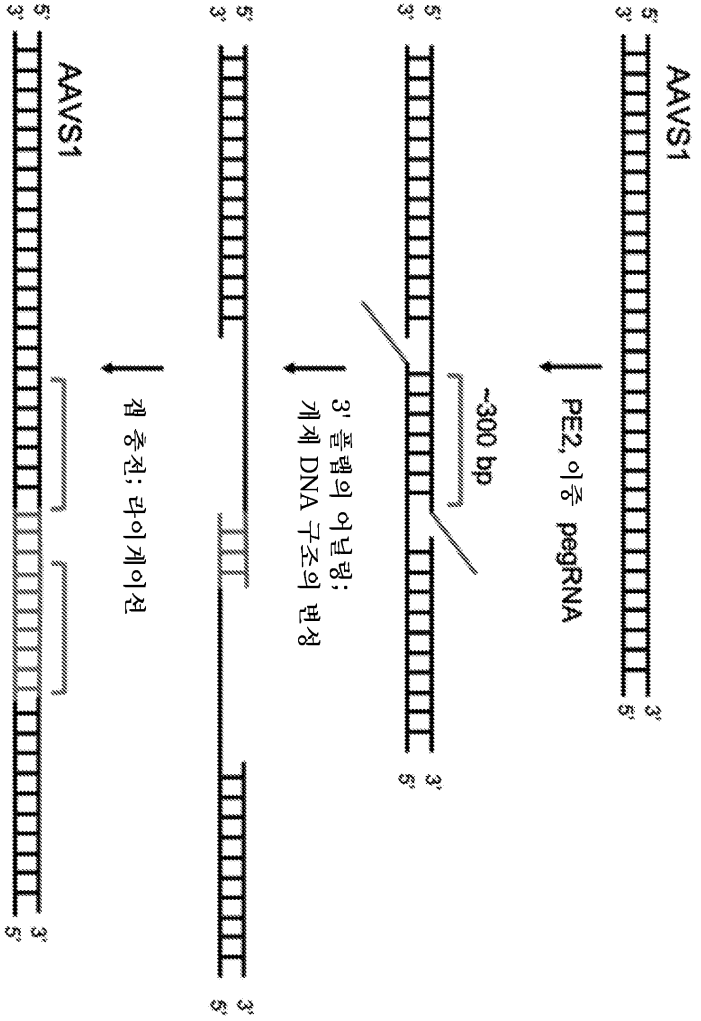


도면104e



P < 0.05, *
P < 0.01, **
P < 0.001, ***

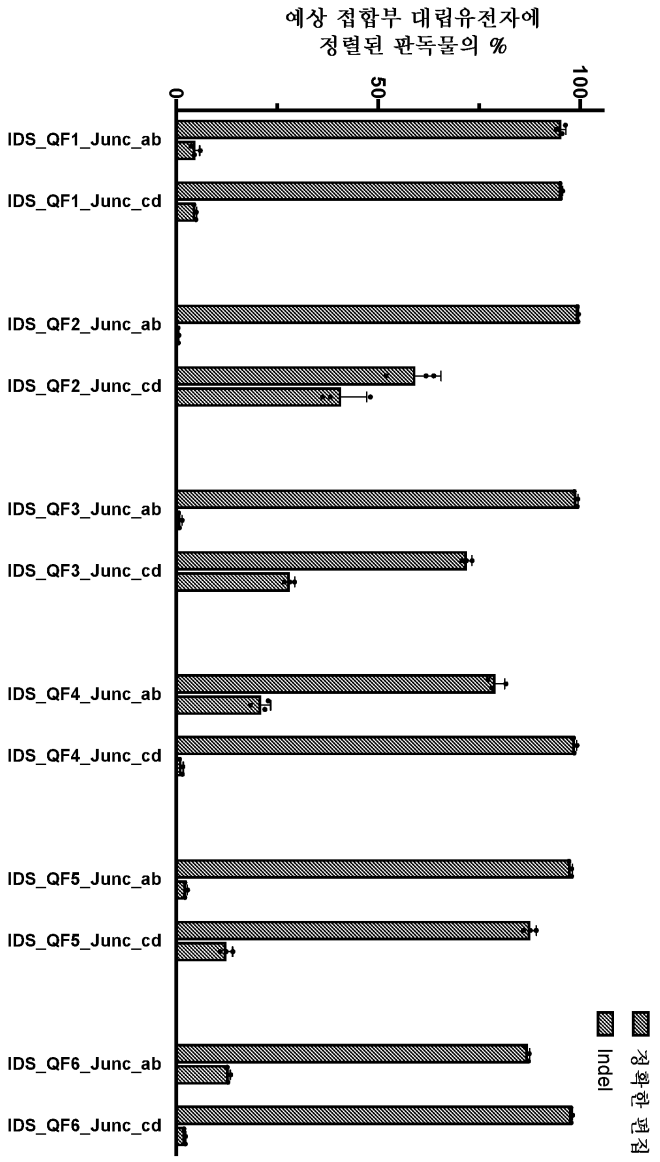
도면105a


AAVS1
PE2, 이중 pegRNA
~300 bp
3' 플랩의 어닐링;
게제 DNA 구조의 변성
갭 충전; 라이게이션
AAVS1

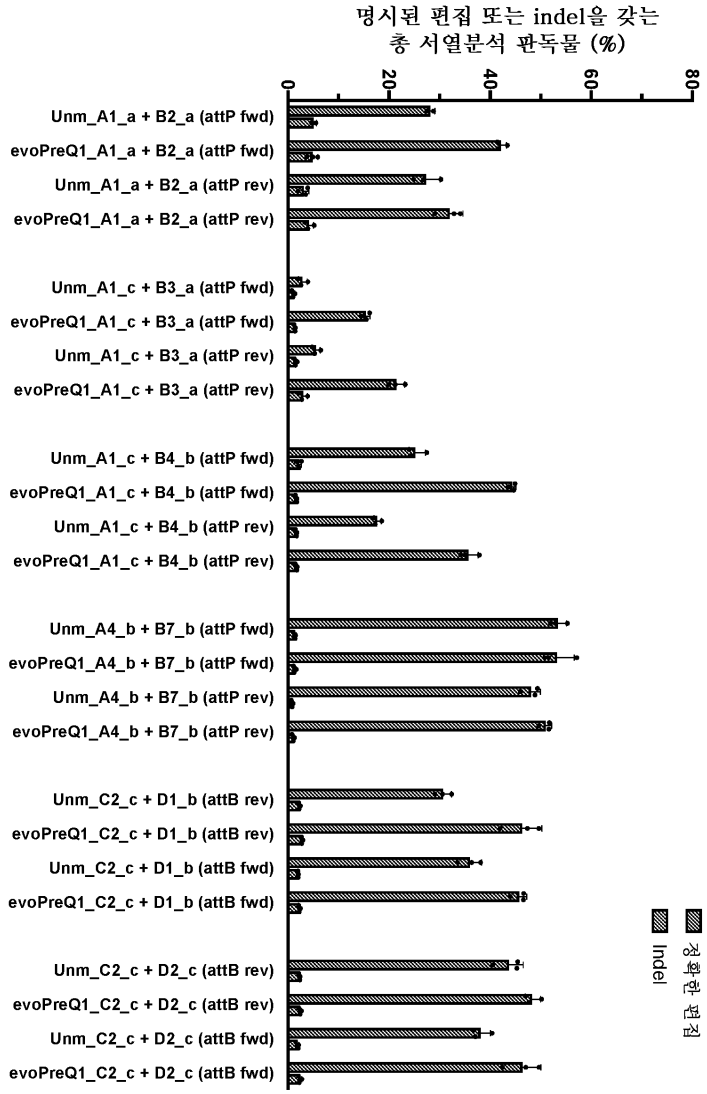
도면105b

(SEQ ID NO: 4342-4348)

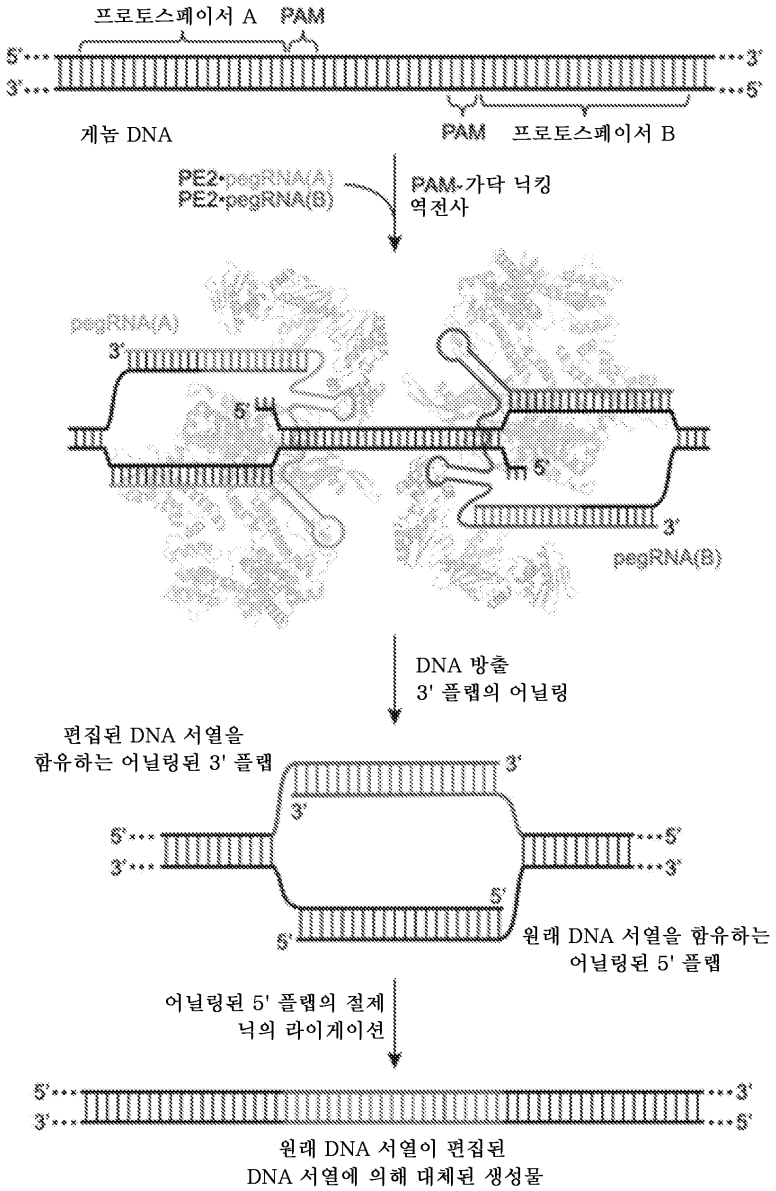
도면106


예상 접합부 대립유전자에
정렬된 판독물의 %
0
50
100
IDS_QF1_Junc_ab
IDS_QF1_Junc_cd
IDS_QF2_Junc_ab
IDS_QF2_Junc_cd
IDS_QF3_Junc_ab
IDS_QF3_Junc_cd
IDS_QF4_Junc_ab
IDS_QF4_Junc_cd
IDS_QF5_Junc_ab
IDS_QF5_Junc_cd
IDS_QF6_Junc_ab
IDS_QF6_Junc_cd
Indel
정확한 편집

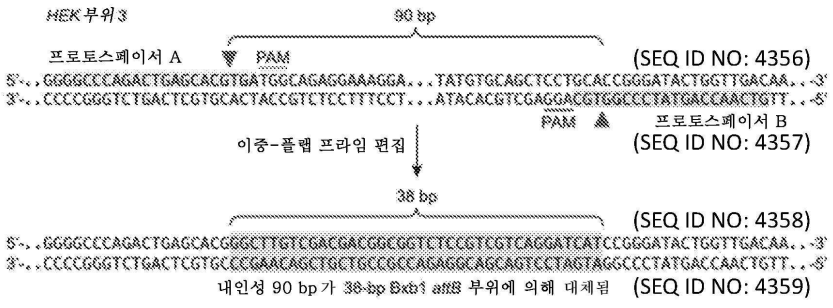
도면107


명시된 편집 또는 indel을 갖는
총 서열분석 판독물 (%)
0
20
40
60
80
Unm_A1_a + B2_a (attP fwd)
evoPreQ1_A1_a + B2_a (attP fwd)
Unm_A1_a + B2_a (attP rev)
evoPreQ1_A1_a + B2_a (attP rev)
Unm_A1_c + B3_a (attP fwd)
evoPreQ1_A1_c + B3_a (attP fwd)
Unm_A1_c + B3_a (attP rev)
evoPreQ1_A1_c + B3_a (attP rev)
Unm_A1_c + B4_b (attP fwd)
evoPreQ1_A1_c + B4_b (attP fwd)
Unm_A1_c + B4_b (attP rev)
evoPreQ1_A1_c + B4_b (attP rev)
Unm_A4_b + B7_b (attP fwd)
evoPreQ1_A4_b + B7_b (attP fwd)
Unm_A4_b + B7_b (attP rev)
evoPreQ1_A4_b + B7_b (attP rev)
Unm_C2_c + D1_b (attB rev)
evoPreQ1_C2_c + D1_b (attB rev)
Unm_C2_c + D1_b (attB fwd)
evoPreQ1_C2_c + D1_b (attB fwd)
Unm_C2_c + D2_c (attB rev)
evoPreQ1_C2_c + D2_c (attB rev)
Unm_C2_c + D2_c (attB fwd)
evoPreQ1_C2_c + D2_c (attB fwd)
Indel
정확한 편집

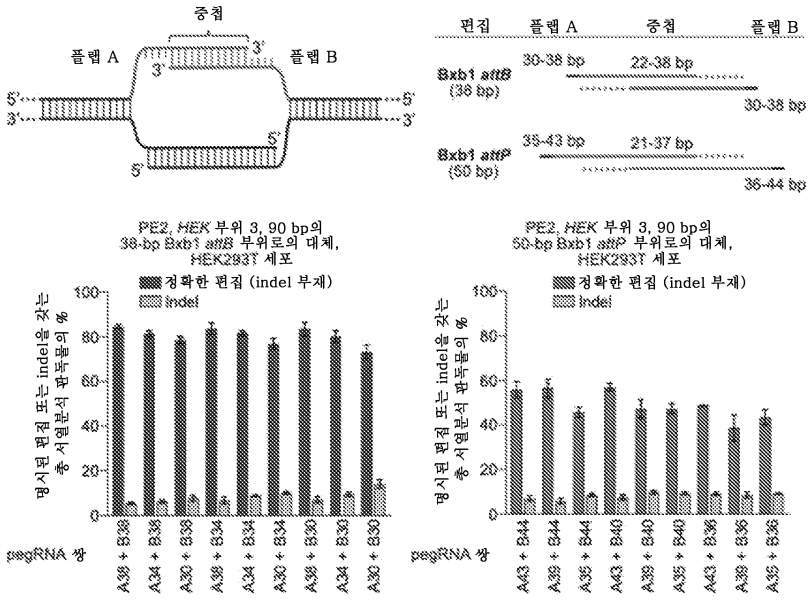
도면108a



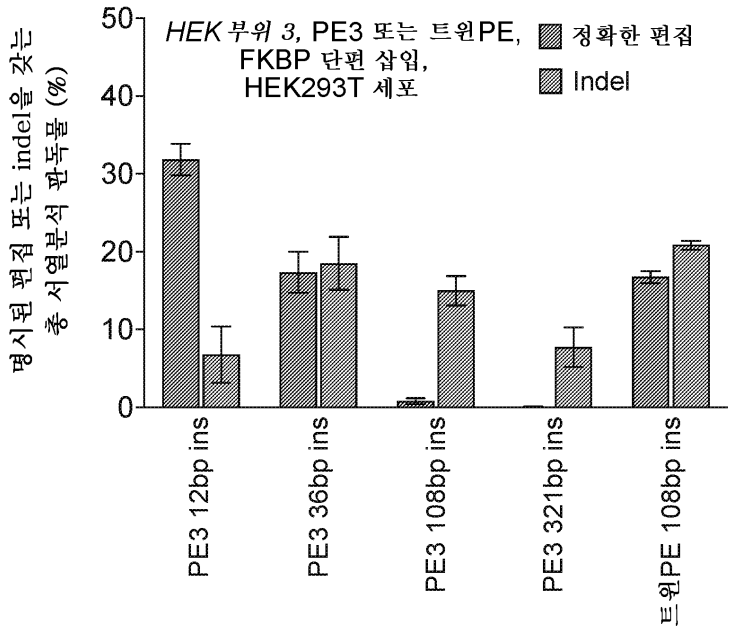
도면108b



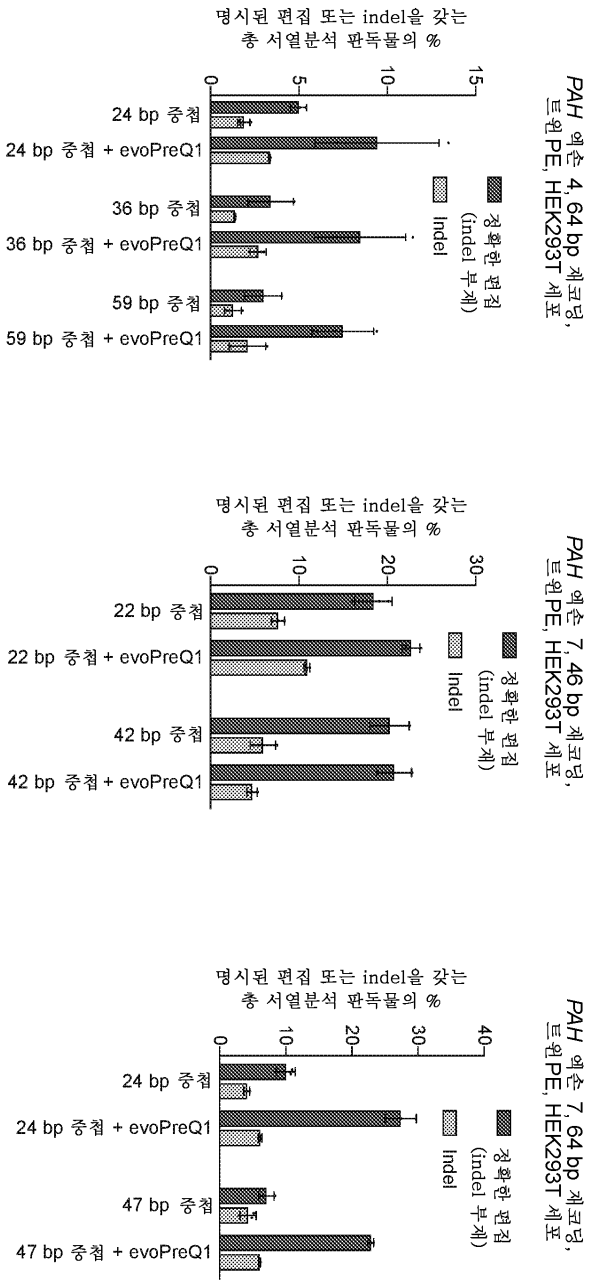
도면108c



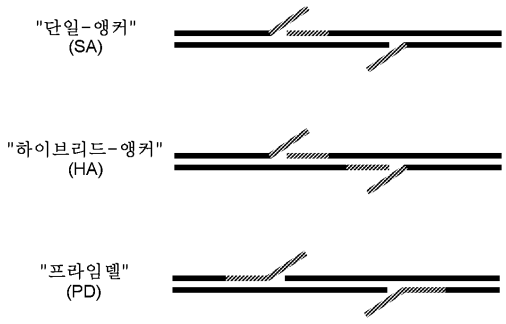
도면109a



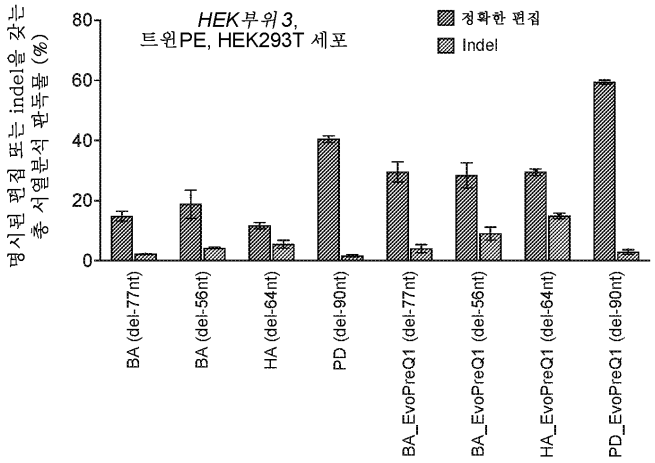
도면109b



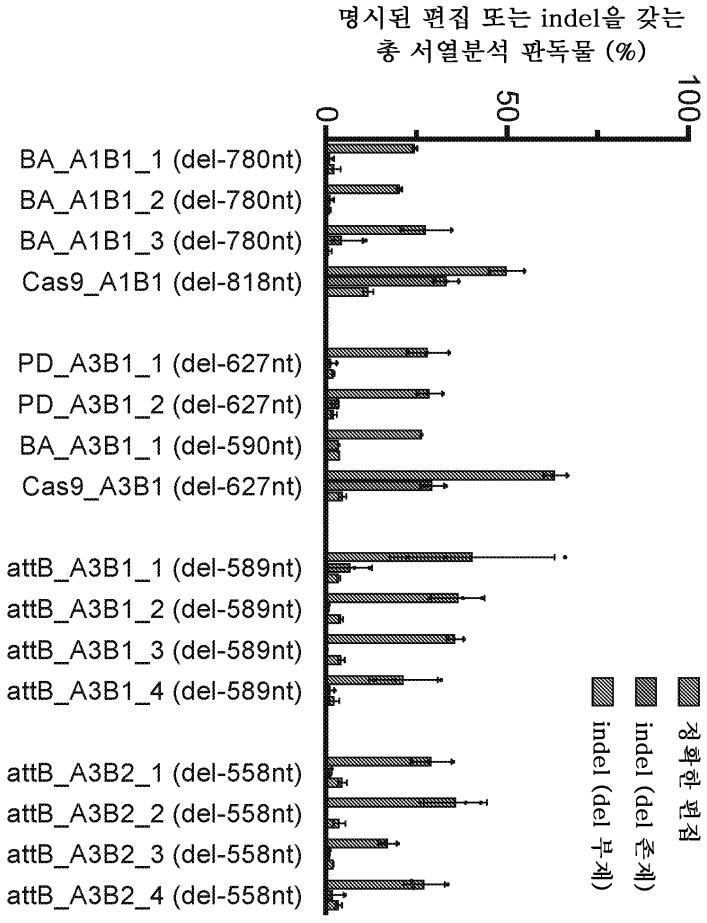
도면109c



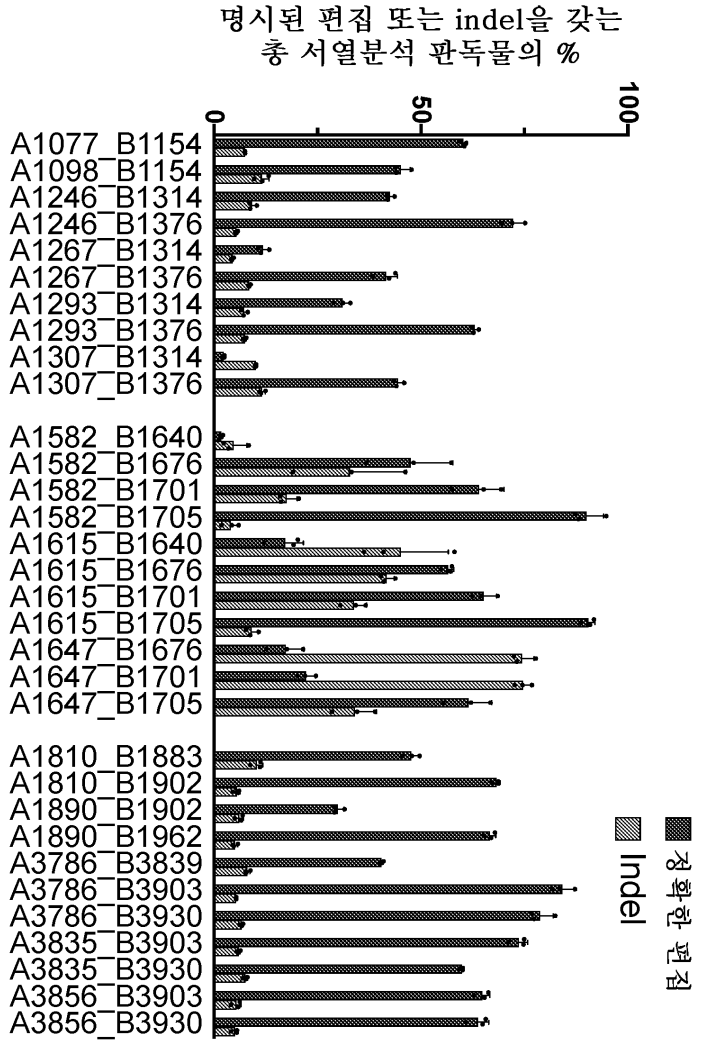
도면109d



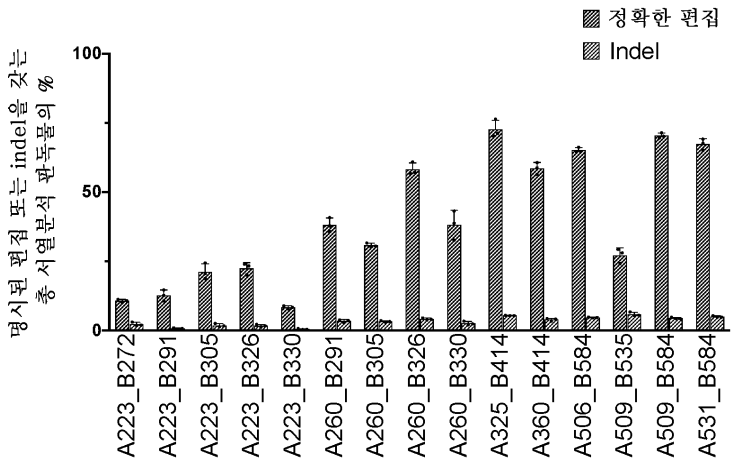
도면109e



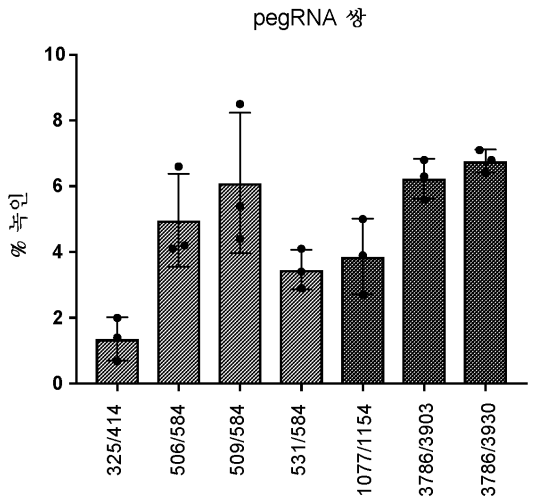
도면110a



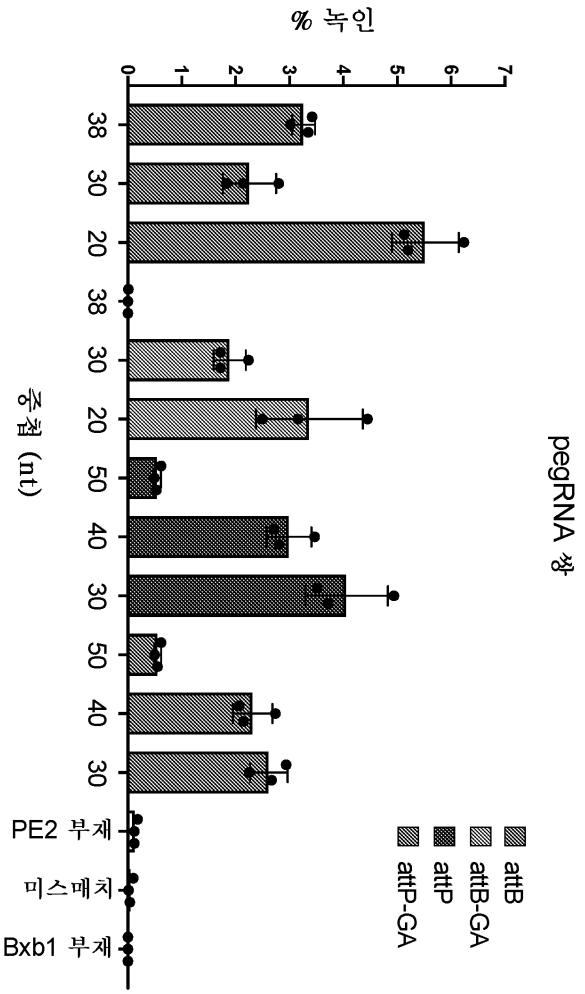
도면110b



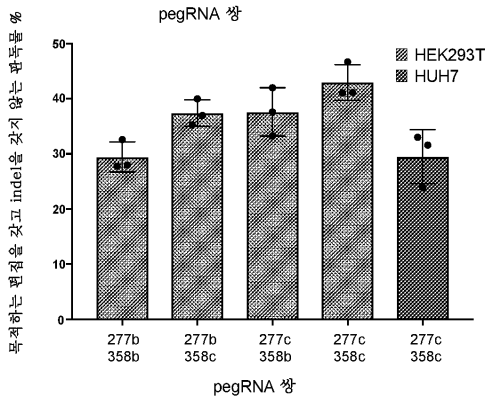
도면110c



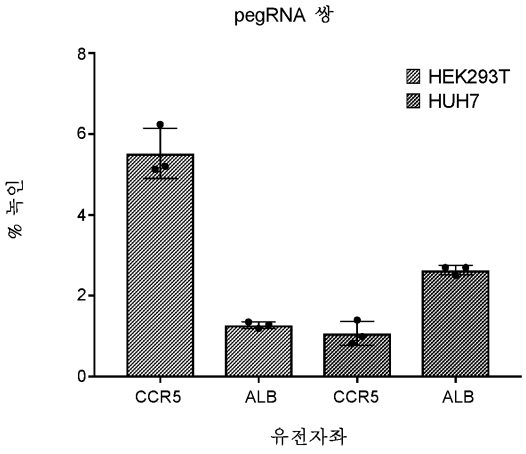
도면110d



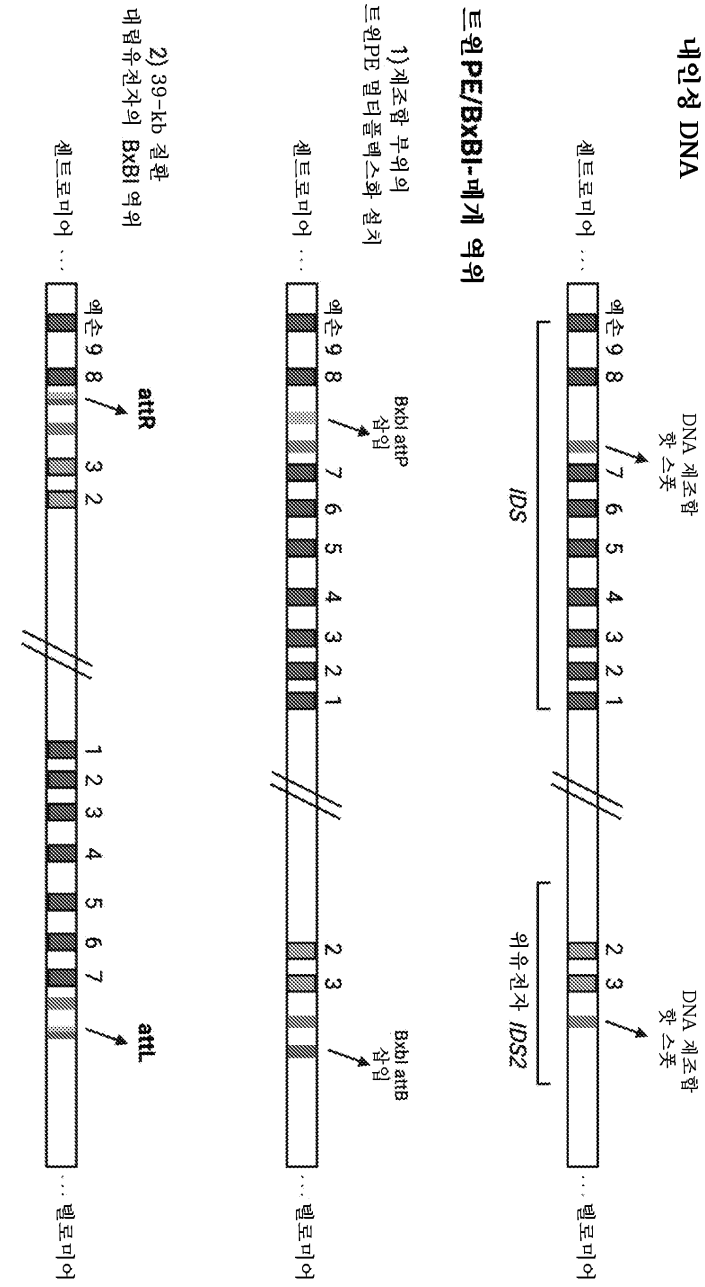
도면110e



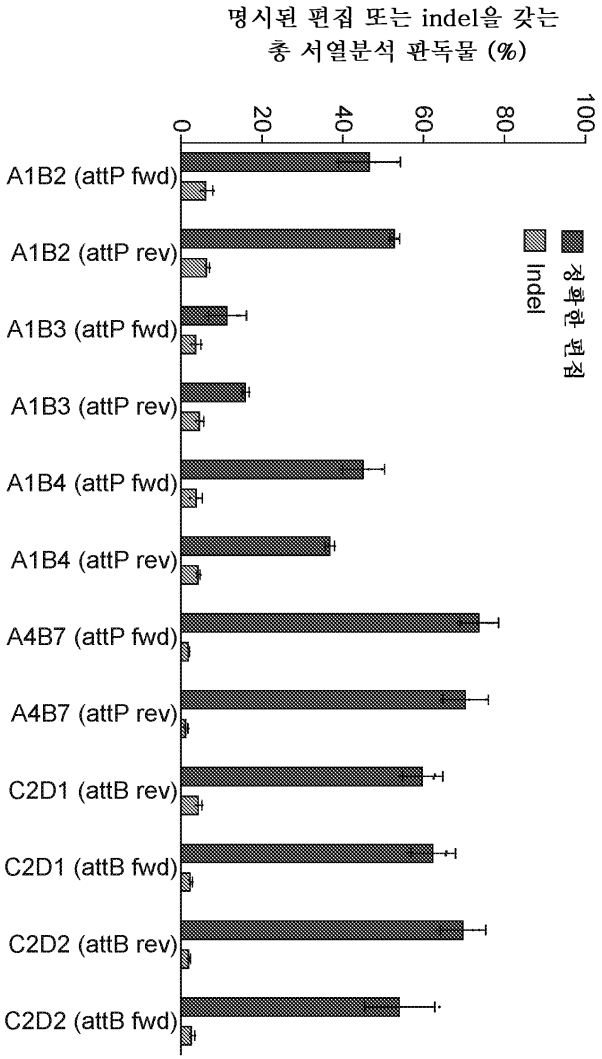
도면110f



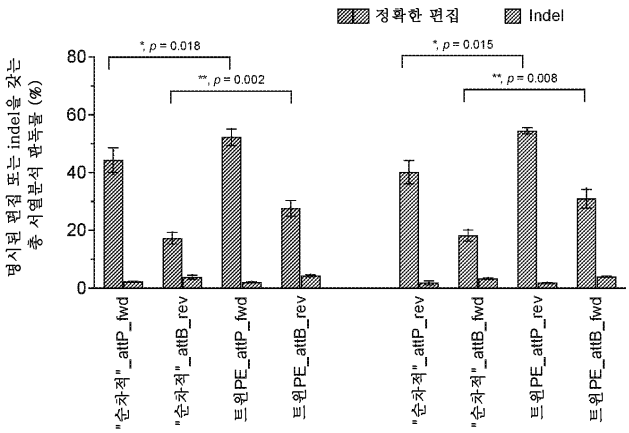
도면111a


내인성 DNA
셀트로미어
엑손 9 8 7 6 5 4 3 2 1
DNA 제조합 핫 스폿
IDS
위유전자 IDS2
2 3
DNA 제조합 핫 스폿
텔로미어
트윈PE/BxBI-매개 역위
1)제조합 부위의 트윈PE 멀티플렉스화 설치
셀트로미어
엑손 9 8 7 6 5 4 3 2 1
Bxb1 attP 삽입
2 3
Bxb1 attB 삽입
텔로미어
2) 39-kb 질환 대립유전자의 BxBI 역위
셀트로미어
엑손 9 8 3 2
attR
1 2 3 4 5 6 7
attL
텔로미어

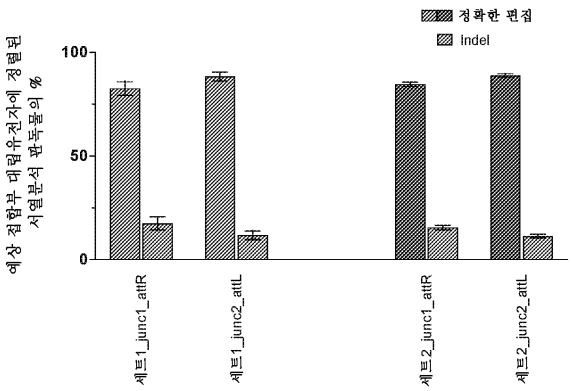
도면111b



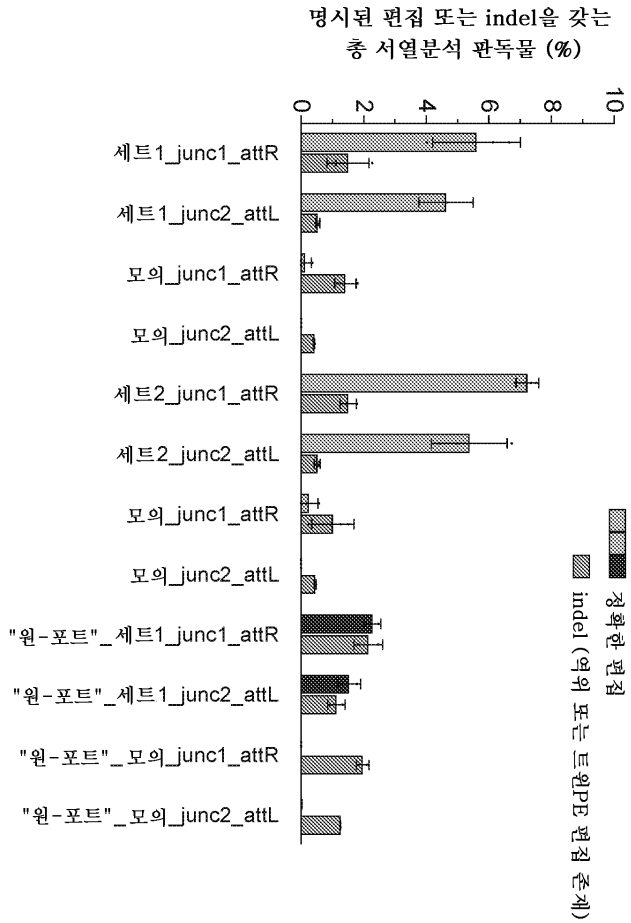
도면111c



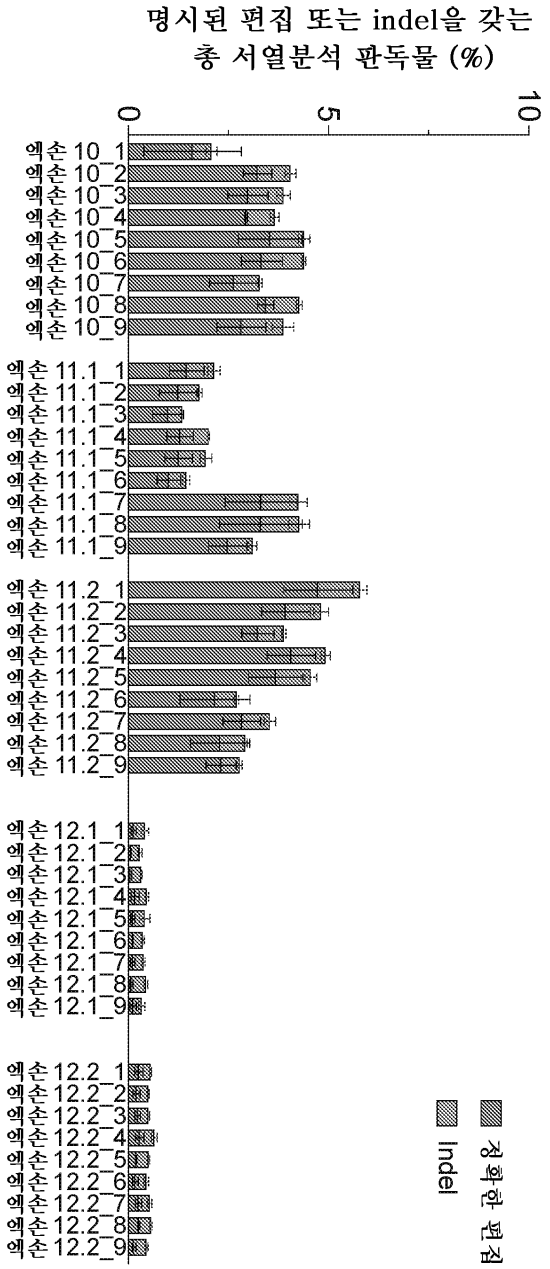
도면111d



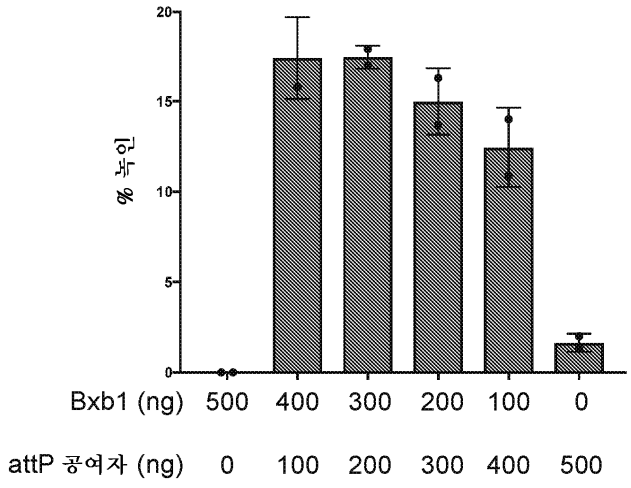
도면111e



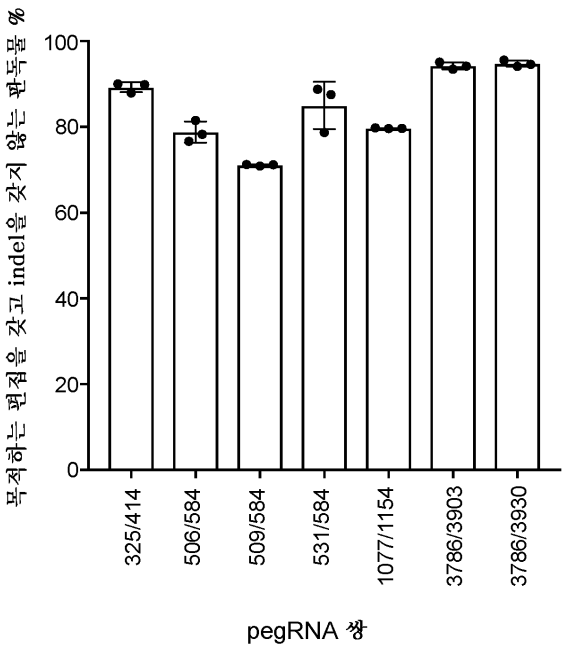
도면112


명시된 편집 또는 indel을 갖는
총 서열분석 판독물 (%)
0
5
10
엑손 10_1
엑손 10_2
엑손 10_3
엑손 10_4
엑손 10_5
엑손 10_6
엑손 10_7
엑손 10_8
엑손 10_9
엑손 11.1_1
엑손 11.1_2
엑손 11.1_3
엑손 11.1_4
엑손 11.1_5
엑손 11.1_6
엑손 11.1_7
엑손 11.1_8
엑손 11.1_9
엑손 11.2_1
엑손 11.2_2
엑손 11.2_3
엑손 11.2_4
엑손 11.2_5
엑손 11.2_6
엑손 11.2_7
엑손 11.2_8
엑손 11.2_9
엑손 12.1_1
엑손 12.1_2
엑손 12.1_3
엑손 12.1_4
엑손 12.1_5
엑손 12.1_6
엑손 12.1_7
엑손 12.1_8
엑손 12.1_9
엑손 12.2_1
엑손 12.2_2
엑손 12.2_3
엑손 12.2_4
엑손 12.2_5
엑손 12.2_6
엑손 12.2_7
엑손 12.2_8
엑손 12.2_9
Indel
정확한 편집

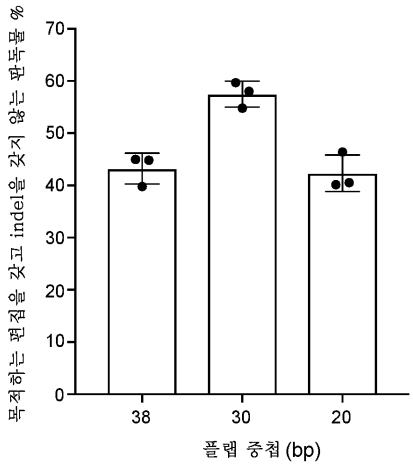
도면113


% 녹인
20
15
10
5
0
Bxb1 (ng) 500 400 300 200 100 0
attP 공여자 (ng) 0 100 200 300 400 500

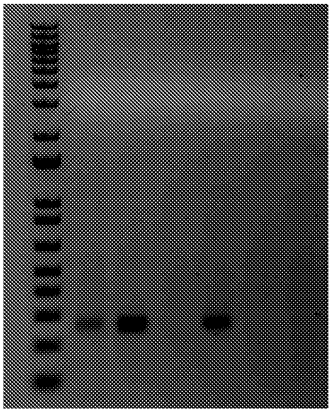
도면114


목적하는 편집을 갖고 indel을 갖지 않는 판독물 %
100
80
60
40
20
0
325/414
506/584
509/584
531/584
1077/1154
3786/3903
3786/3930
pegRNA 쌍

도면115a

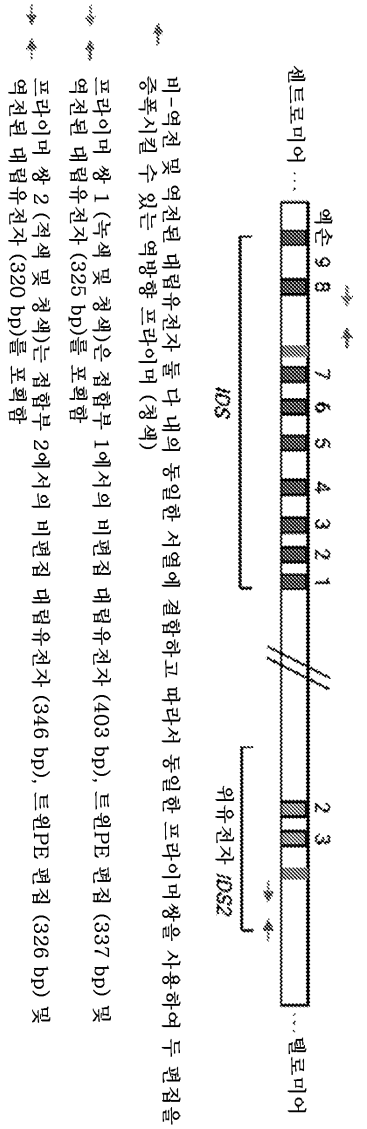


도면115b



1: 래더
2: 38-bp 중첩
3: 30-bp 중첩
4: 20-bp 중첩
5: PE2 부재
6: Bxb1 부재
7: 미스매치

도면116



서 열 목 록

SEQUENCE LISTING

<110> The Broad Institute, Inc.

President and Fellows of Harvard College

<120> METHODS AND COMPOSITIONS FOR SIMULTANEOUS EDITING OF BOTH STRANDS OF A TARGET DOUBLE-STRANDED NUCLEOTIDE SEQUENCE

<130> B1195.70091WO00

<140> Not Yet Assigned

<141> Concurrently Herewith

<150> US 63/022,397

<151> 2020-05-08

<150> US 63/116,785

<151> 2020-11-20

<160> 4359

<170> PatentIn version 3.5

<210> 1

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1

Phe Leu Asn Cys Cys Pro Gly Cys Cys Met Glu Pro

1 5 10

<210> 2

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2

Asp Tyr Lys Asp Asp Asp Asp Lys

1 5

<210> 3

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3

Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr

1 5 10

<210> 4

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 4

Glu Glu Leu Leu Ser Lys Asn Tyr His Leu Glu Asn Glu Val Ala Arg
1 5 10 15

Leu Lys Lys

<210> 5

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 5

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> 6

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223>

Synthetic

<400> 6

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 7

<211> 380

<212> PRT

<213> Homo sapiens

<400> 7

Met Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser
1 5 10 15

Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala
20 25 30

Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln
35 40 45

Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu
50 55 60
Met Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys
65 70 75 80
Pro Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu
85 90 95
Leu Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln
100 105 110

Gln Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys
115 120 125
Arg Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu
130 135 140
Leu Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu
145 150 155 160
Ala Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala Ala
165 170 175

Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg
180 185 190
His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His
195 200 205
Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val
210 215 220
Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly
225 230 235 240

Ile Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile
245 250 255
Glu Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu
260 265 270
Asn Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val
275 280 285
Leu Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu

290 295 300

Glu Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg
305 310 315 320
Ile Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr
325 330 335
Gln Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser
340 345 350
Ala Lys Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala
355 360 365

Lys Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys
370 375 380

<210> 8
<211> 413
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 8

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15
His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val
20 25 30
Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45
Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60
Ala Ile Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80
Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95
Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr
115 120 125
Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140
Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160
Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175
Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190
Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205
Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220
Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240
Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255
Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270
His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285
Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu

290 295 300
Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320
His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335
Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly
340 345 350
Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 9

<211> 413

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 9

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr

1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Val Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60

Ala Ile Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly

65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr

115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr

130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg

145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser

180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp

195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met

210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile

260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu

290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320

His Leu Tyr Ser Met Lys Tyr Thr Asn Val Val Pro Leu Tyr Asp Leu

325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys
405 410

<210> 10

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 10

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp
35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met
100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Pro Tyr Ser Glu Tyr
115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg

145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205
Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met

210 215 220
Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu
225 230 235 240
Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255
Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270
His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285
Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300
Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320
His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335
Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350
Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365
His Asp Met Leu Ala Glu Gly Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380
Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu
385 390 395 400
His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 11
<211> 413
<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 11

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val
20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp
35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met
100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr
115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly
340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu
385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys
405 410

<210> 12

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 12

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr

1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly

65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Pro Tyr Ser Glu Tyr

115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr

130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg

145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser

180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp

195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met

210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile

260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu

290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu

325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 13

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 13

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr

1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met
100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr
115 120 125

Asn Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175

Glu Arg Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu
225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320
His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335
Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly
340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365
His Asp Met Leu Ala Glu Gly Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380
Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu
385 390 395 400
His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys
405 410

<210> 14

<211> 413
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 14
Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15
His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val
20 25 30
Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp
35 40 45
Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60
Ala Ile Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80
Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met
100 105 110
Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr

115 120 125
Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140
Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160
Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175
Glu Arg Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser

180 185 190
Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205
Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220
Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu
225 230 235 240
Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255
Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270
His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285
Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300
Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320
His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335
Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Gly Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 15

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 15

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr

1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Val Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly

65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr

115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr

130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160
Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175
Glu Arg Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190
Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220
Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu
225 230 235 240
Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255
Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285
Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300
Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320
His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly
340 345 350
Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365
His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu
385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys
405 410

<210> 16
<211> 7
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 16
Pro Lys Lys Lys Arg Lys Val
1 5

<210> 17
<211> 30
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 17
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys
20 25 30

<210> 18
<211> 1368
<212> PRT
<213> Streptococcus pyogenes
<400> 18
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 19

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 19

atggataaaa aatatagcat tggcctggat attggcacca acagcgtggg ctgggcggtg 60

attaccgatg aatataaagt gccgagcaaa aaatttaaag tgctgggcaa caccgatcgc 120

catagcatta aaaaaaacct gattggcgcg ctgctgtttg atagcggcga aaccgcggaa 180

gcgacccgcc tgaaacgcac cgcgcgccgc cgctataccc gccgcaaaaa ccgcatttgc 240

tatctgcagg aaatttttag caacgaaatg gcgaaagtgg atgatagctt ttttcatcgc 300

ctggaagaaa gctttctggt ggaagaagat aaaaaacatg aacgccatcc gatttttggc 360

aacattgtgg atgaagtggc gtatcatgaa aaatatccga ccatttatca tctgcgcaaa 420

aaactggtgg atagcaccga taaagcggat ctgcgcctga tttatctggc gctggcgcat 480

atgattaaat ttcgcggcca ttttctgatt gaaggcgatc tgaacccgga taacagcgat 540

gtggataaac tgtttattca gctggtgcag acctataacc agctgtttga agaaaacccg 600

attaacgcga gcggcgtgga tgcgaaagcg attctgagcg cgcgcctgag caaaagccgc 660

cgcctggaaa acctgattgc gcagctgccg ggcgaaaaaa aaaacggcct gtttggcaac 720

ctgattgcgc tgagcctggg cctgaccccg aactttaaaa gcaactttga tctggcggaa 780

gatgcgaaac tgcagctgag caaagatacc tatgatgatg atctggataa cctgctggcg 840
cagattggcg atcagtatgc ggatctgttt ctggcggcga aaaacctgag cgatgcgatt 900
ctgctgagcg atattctgcg cgtgaacacc gaaattacca aagcgccgct gagcgcgagc 960
atgattaaac gctatgatga acatcatcag gatctgaccc tgctgaaagc gctggtgcgc 1020
cagcagctgc cggaaaaata taaagaaatt ttttttgatc agagcaaaaa cggctatgcg 1080
ggctatattg atggcggcgc gagccaggaa gaattttata aatttattaa accgattctg 1140
gaaaaaatgg atggcaccga agaactgctg gtgaaactga accgcgaaga tctgctgcgc 1200

aaacagcgca cctttgataa cggcagcatt ccgcatcaga ttcatctggg cgaactgcat 1260
gcgattctgc gccgccagga agatttttat ccgtttctga aagataaccg cgaaaaaatt 1320
gaaaaaattc tgacctttcg cattccgtat tatgtgggcc cgctggcgcg cggcaacagc 1380
cgctttgcgt ggatgacccg caaaagcgaa gaaaccatta ccccgtggaa ctttgaagaa 1440
gtggtggata aaggcgcgag cgcgcagagc tttattgaac gcatgaccaa ctttgataaa 1500
aacctgccga acgaaaaagt gctgccgaaa catagcctgc tgtatgaata ttttaccgtg 1560
tataacgaac tgaccaaagt gaaatatgtg accgaaggca tgcgcaaacc ggcgtttctg 1620

agcggcgaac agaaaaaagc gattgtggat ctgctgttta aaaccaaccg caaagtgacc 1680
gtgaaacagc tgaaagaaga ttattttaaa aaaattgaat gctttgatag cgtggaaatt 1740
agcggcgtgg aagatcgctt taacgcgagc ctgggcacct atcatgatct gctgaaaatt 1800
attaaagata aagattttct ggataacgaa gaaaacgaag atattctgga agatattgtg 1860
ctgaccctga ccctgtttga agatcgcgaa atgattgaag aacgcctgaa aacctatgcg 1920
catctgtttg atgataaagt gatgaaacag ctgaaacgcc gccgctatac cggctggggc 1980
cgcctgagcc gcaaactgat taacggcatt cgcgataaac agagcggcaa aaccattctg 2040

gattttctga aaagcgatgg ctttgcgaac cgcaacttta tgcagctgat tcatgatgat 2100
agcctgacct ttaaagaaga tattcagaaa gcgcaggtga gcggccaggg cgatagcctg 2160
catgaacata ttgcgaacct ggcgggcagc ccggcgatta aaaaaggcat tctgcagacc 2220
gtgaaagtgg tggatgaact ggtgaaagtg atgggccgcc ataaaccgga aaacattgtg 2280
attgaaatgg cgcgcgaaaa ccagaccacc cagaaaggcc agaaaaacag ccgcgaacgc 2340
atgaaacgca ttgaagaagg cattaaagaa ctgggcagcc agattctgaa agaacatccg 2400
gtggaaaaca cccagctgca gaacgaaaaa ctgtatctgt attatctgca gaacggccgc 2460

gatatgtatg tggatcagga actggatatt aaccgcctga gcgattatga tgtggatcat 2520

attgtgccgc agagctttct gaaagatgat agcattgata acaaagtgct gacccgcagc 2580

gataaaaacc gcggcaaaag cgataacgtg ccgagcgaag aagtggtgaa aaaaatgaaa 2640

aactattggc gccagctgct gaacgcgaaa ctgattaccc agcgcaaatt tgataacctg 2700

accaaagcgg aacgcggcgg cctgagcgaa ctggataaag cgggctttat taaacgccag 2760

ctggtggaaa cccgccagat taccaaacat gtggcgcaga ttctggatag ccgcatgaac 2820

accaaatatg atgaaaacga taaactgatt cgcgaagtga aagtgattac cctgaaaagc 2880

aaactggtga gcgattttcg caaagatttt cagttttata aagtgcgcga aattaacaac 2940

tatcatcatg cgcatgatgc gtatctgaac gcggtggtgg gcaccgcgct gattaaaaaa 3000

tatccgaaac tggaaagcga atttgtgtat ggcgattata aagtgtatga tgtgcgcaaa 3060

atgattgcga aaagcgaaca ggaaattggc aaagcgaccg cgaaatattt tttttatagc 3120

aacattatga actttttaa aaccgaaatt accctggcga acggcgaaat tcgcaaacgc 3180

ccgctgattg aaaccaacgg cgaaaccggc gaaattgtgt gggataaagg ccgcgatttt 3240

gcgaccgtgc gcaaagtgct gagcatgccg caggtgaaca ttgtgaaaaa aaccgaagtg 3300

cagaccggcg gctttagcaa agaaagcatt ctgccgaaac gcaacagcga taaactgatt 3360

gcgcgcaaaa aagattggga tccgaaaaaa tatggcggct ttgatagccc gaccgtggcg 3420

tatagcgtgc tggtggtggc gaaagtggaa aaaggcaaaa gcaaaaaact gaaaagcgtg 3480

aaagaactgc tgggcattac cattatggaa cgcagcagct ttgaaaaaaa cccgattgat 3540

tttctggaag cgaaaggcta taaagaagtg aaaaaagatc tgattattaa actgccgaaa 3600

tatagcctgt ttgaactgga aaacggccgc aaacgcatgc tggcgagcgc gggcgaactg 3660

cagaaaggca acgaactggc gctgccgagc aaatatgtga actttctgta tctggcgagc 3720

cattatgaaa aactgaaagg cagcccggaa gataacgaac agaaacagct gtttgtggaa 3780

cagcataaac attatctgga tgaaattatt gaacagatta gcgaatttag caaacgcgtg 3840

attctggcgg atgcgaacct ggataaagtg ctgagcgcgt ataacaaaca tcgcgataaa 3900

ccgattcgcg aacaggcgga aaacattatt catctgttta ccctgaccaa cctgggcgcg 3960

ccggcggcgt ttaaatattt tgataccacc attgatcgca aacgctatac cagcaccaaa 4020

gaagtgctgg atgcgaccct gattcatcag agcattaccg gcctgtatga aacccgcatt 4080

gatctgagcc agctgggcgg cgat 4104

<210> 20

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 20

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa atctacaatc aattatttga agaaaaccct 600

attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat 720

ctcattgctt tgtcattggg attgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatagt gaaataacta aggctcccct atcagcttca 960

atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agataggggg atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta 2160

catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggtat tttacagact 2220

gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt 2280

gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg 2340

aaacgaatcg aagaaggtat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt 2400

gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctacaaaa tggaagagac 2460

atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt 2520

gttccacaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat 2580

aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac 2640

tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg 2700

aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg 2760

gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact 2820

aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa 2880

ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac 2940

catcatgccc atgatgcgta tctaaatgcc gtcgttggaa ctgctttgat taagaaatat 3000

ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg 3060

attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat 3120

atcatgaact tcttcaaaac agaaattaca cttgcaaatg gagagattcg caaacgccct 3180

ctaatcgaaa ctaatgggga aactggagaa attgtctggg ataaagggcg agattttgcc 3240

acagtgcgca aagtattgtc catgccccaa gtcaatattg tcaagaaaac agaagtacag 3300

acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360

cgtaaaaaag actgggatcc aaaaaaatat ggtggttttg atagtccaac ggtagcttat 3420

tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480

gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt 3540

ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600

agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660

aaaggaaatg agctggctct gccaagcaaa tatgtgaatt ttttatattt agctagtcat 3720

tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780

cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840

ttagcagatg ccaatttaga taaagttctt agtgcatata acaaacatag agacaaacca 3900

atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc 3960

gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc tacaaaagaa 4020

gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat 4080

ttgagtcagc taggaggtga ctga 4104

<210> 21

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 21

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu

705 710 715 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 22

<211> 4212

<212> DNA

<213> Streptococcus pyogenes

<400> 22

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60

ataaccgatg aatacaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120

cattcgatta aaaagaatct tatcggtgcc ctcctattcg atagtggcga aacggcagag 180

gcgactcgcc tgaaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240

tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt 300

ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcaccc catctttgga 360

aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa 420

aagctagttg actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480

atgataaagt tccgtgggca ctttctcatt gagggtgatc taaatccgga caactcggat 540

gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct 600

ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga 660

cggctagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac 720

cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa 780

gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca 840

caaattggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc 900

ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca 960

atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc cctagtccgt 1020

cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca 1080

ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta 1140

gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcgaaga tctactgcga 1200

aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat 1260

gctatactta gaaggcagga ggatttttat ccgttcctca aagacaatcg tgaaaagatt 1320

gagaaaatcc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct 1380

cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa 1440

gttgtcgata aaggtgcgtc agctcaatcg ttcatcgaga ggatgaccaa ctttgacaag 1500

aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagtg 1560

tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta 1620

agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca 1680

gttaagcaat tgaaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc 1740

tccggggtag aagatcgatt taatgcgtca cttggtacgt atcatgacct cctaaagata 1800

attaaagata aggacttcct ggataacgaa gagaatgaag atatcttaga agatatagtg 1860

ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgct 1920

cacctgttcg acgataaggt tatgaaacag ttaaagaggc gtcgctatac gggctgggga 1980

cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc 2040

gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac 2100

tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg 2160

cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaagggcat actccagaca 2220

gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaaccgga aaacattgta 2280

atcgagatgg cacgcgaaaa tcaaacgact cagaaggggc aaaaaaacag tcgagagcgg 2340

atgaagagaa tagaagaggg tattaaagaa ctgggcagcc agatcttaaa ggagcatcct 2400

gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg 2460

gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac 2520

attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgct tacacgctcg 2580

gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag 2640

aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagtt cgataactta 2700

actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccggatttat taaacgtcag 2760

ctcgtggaaa cccgccaaat cacaaagcat gttgcacaga tactagattc ccgaatgaat 2820

acgaaatacg acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880

aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940

taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000

tacccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060

atgatcgcga aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120

aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180

cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240

gcgacggtga gaaaagtttt gtccatgccc caagtcaaca tagtaaagaa aactgaggtg 3300

cagaccggag ggttttcaaa ggaatcgatt cttccaaaaa ggaatagtga taagctcatc 3360

gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420
tattctgtcc tagtagtggc aaaagttgag aagggaaaat ccaagaaact gaagtcagtc 3480
aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540
ttccttgagg cgaaaggtta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600
tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660
caaaaggggа acgaactcgc actaccgtct aaatacgtga atttcctgta tttagcgtcc 3720
cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact ttttgttgag 3780

cagcacaaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840
atcctagctg atgccaatct ggacaaagta ttaagcgcat acaacaagca cagggataaa 3900
cccatacgtg agcaggcgga aaatattatc catttgttta ctcttaccaa cctcggcgct 3960
ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020
gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080
gatttgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140
tacaaagacc atgacggtga ttataaagat catgacatcg attacaagga tgacgatgac 4200

aaggctgcag ga 4212

<210> 23
<211> 1404
<212> PRT
<213> Streptococcus pyogenes
<400> 23

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

Gly Ser Pro Lys Lys Lys Arg Lys Val Ser Ser Asp Tyr Lys Asp
1370 1375 1380
His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp Tyr Lys Asp Asp
1385 1390 1395
Asp Asp Lys Ala Ala Gly
1400

<210> 24
<211> 4107
<212> DNA
<213> Streptococcus pyogenes
<400> 24

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60
atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300
cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600
attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720

ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca 960

atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta 2160

catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact 2220

gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt 2280

attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt 2340

atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct 2400

gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460

gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520

attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580

gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640

aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700

acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa 2760

ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820

actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct 2880

aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940

taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000

tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060

atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120

aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180

cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240

gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300

cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360

gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420

tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480

aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540

tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600

tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660

caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttttata tttagctagt 3720

cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780

cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840

attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa 3900

ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960

cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020

gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080

gatttgagtc agctaggagg tgactga 4107

<210> 25

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 25

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 26
<211> 1378
<212> PRT
<213> Lactobacillus fermentum
<400> 26

Met Lys Glu Tyr His Ile Gly Leu Asp Ile Gly Thr Ser Ser Ile Gly
1 5 10 15
Trp Ala Val Thr Asp Ser Gln Phe Lys Leu Met Arg Ile Lys Gly Lys
20 25 30
Thr Ala Ile Gly Val Arg Leu Phe Glu Glu Gly Lys Thr Ala Ala Glu
35 40 45
Arg Arg Thr Phe Arg Thr Thr Arg Arg Arg Leu Lys Arg Arg Lys Trp

50 55 60

Arg Leu His Tyr Leu Asp Glu Ile Phe Ala Pro His Leu Gln Glu Val

65 70 75 80

Asp Glu Asn Phe Leu Arg Arg Leu Lys Gln Ser Asn Ile His Pro Glu

85 90 95

Asp Pro Thr Lys Asn Gln Ala Phe Ile Gly Lys Leu Leu Phe Pro Asp

100 105 110

Leu Leu Lys Lys Asn Glu Arg Gly Tyr Pro Thr Leu Ile Lys Met Arg

115 120 125

Asp Glu Leu Pro Val Glu Gln Arg Ala His Tyr Pro Val Met Asn Ile

130 135 140

Tyr Lys Leu Arg Glu Ala Met Ile Asn Glu Asp Arg Gln Phe Asp Leu

145 150 155 160

Arg Glu Val Tyr Leu Ala Val His His Ile Val Lys Tyr Arg Gly His

165 170 175

Phe Leu Asn Asn Ala Ser Val Asp Lys Phe Lys Val Gly Arg Ile Asp

180 185 190

Phe Asp Lys Ser Phe Asn Val Leu Asn Glu Ala Tyr Glu Glu Leu Gln

195 200 205

Asn Gly Glu Gly Ser Phe Thr Ile Glu Pro Ser Lys Val Glu Lys Ile

210 215 220

Gly Gln Leu Leu Leu Asp Thr Lys Met Arg Lys Leu Asp Arg Gln Lys

225 230 235 240

Ala Val Ala Lys Leu Leu Glu Val Lys Val Ala Asp Lys Glu Glu Thr

245 250 255

Lys Arg Asn Lys Gln Ile Ala Thr Ala Met Ser Lys Leu Val Leu Gly

260 265 270

Tyr Lys Ala Asp Phe Ala Thr Val Ala Met Ala Asn Gly Asn Glu Trp

275 280 285

Lys Ile Asp Leu Ser Ser Glu Thr Ser Glu Asp Glu Ile Glu Lys Phe

290 295 300

Arg Glu Glu Leu Ser Asp Ala Gln Asn Asp Ile Leu Thr Glu Ile Thr
305 310 315 320
Ser Leu Phe Ser Gln Ile Met Leu Asn Glu Ile Val Pro Asn Gly Met
325 330 335
Ser Ile Ser Glu Ser Met Met Asp Arg Tyr Trp Thr His Glu Arg Gln

340 345 350
Leu Ala Glu Val Lys Glu Tyr Leu Ala Thr Gln Pro Ala Ser Ala Arg
355 360 365
Lys Glu Phe Asp Gln Val Tyr Asn Lys Tyr Ile Gly Gln Ala Pro Lys
370 375 380
Glu Arg Gly Phe Asp Leu Glu Lys Gly Leu Lys Lys Ile Leu Ser Lys
385 390 395 400
Lys Glu Asn Trp Lys Glu Ile Asp Glu Leu Leu Lys Ala Gly Asp Phe

405 410 415
Leu Pro Lys Gln Arg Thr Ser Ala Asn Gly Val Ile Pro His Gln Met
420 425 430
His Gln Gln Glu Leu Asp Arg Ile Ile Glu Lys Gln Ala Lys Tyr Tyr
435 440 445
Pro Trp Leu Ala Thr Glu Asn Pro Ala Thr Gly Glu Arg Asp Arg His
450 455 460
Gln Ala Lys Tyr Glu Leu Asp Gln Leu Val Ser Phe Arg Ile Pro Tyr

465 470 475 480
Tyr Val Gly Pro Leu Val Thr Pro Glu Val Gln Lys Ala Thr Ser Gly
485 490 495
Ala Lys Phe Ala Trp Ala Lys Arg Lys Glu Asp Gly Glu Ile Thr Pro
500 505 510
Trp Asn Leu Trp Asp Lys Ile Asp Arg Ala Glu Ser Ala Glu Ala Phe
515 520 525
Ile Lys Arg Met Thr Val Lys Asp Thr Tyr Leu Leu Asn Glu Asp Val

530 535 540
Leu Pro Ala Asn Ser Leu Leu Tyr Gln Lys Tyr Asn Val Leu Asn Glu

545 550 555 560

Leu Asn Asn Val Arg Val Asn Gly Arg Arg Leu Ser Val Gly Ile Lys

565 570 575

Gln Asp Ile Tyr Thr Glu Leu Phe Lys Lys Lys Lys Thr Val Lys Ala

580 585 590

Ser Asp Val Ala Ser Leu Val Met Ala Lys Thr Arg Gly Val Asn Lys

595 600 605

Pro Ser Val Glu Gly Leu Ser Asp Pro Lys Lys Phe Asn Ser Asn Leu

610 615 620

Ala Thr Tyr Leu Asp Leu Lys Ser Ile Val Gly Asp Lys Val Asp Asp

625 630 635 640

Asn Arg Tyr Gln Thr Asp Leu Glu Asn Ile Ile Glu Trp Arg Ser Val

645 650 655

Phe Glu Asp Gly Glu Ile Phe Ala Asp Lys Leu Thr Glu Val Glu Trp

660 665 670

Leu Thr Asp Glu Gln Arg Ser Ala Leu Val Lys Lys Arg Tyr Lys Gly

675 680 685

Trp Gly Arg Leu Ser Lys Lys Leu Leu Thr Gly Ile Val Asp Glu Asn

690 695 700

Gly Gln Arg Ile Ile Asp Leu Met Trp Asn Thr Asp Gln Asn Phe Lys

705 710 715 720

Glu Ile Val Asp Gln Pro Val Phe Lys Glu Gln Ile Asp Gln Leu Asn

725 730 735

Gln Lys Ala Ile Thr Asn Asp Gly Met Thr Leu Arg Glu Arg Val Glu

740 745 750

Ser Val Leu Asp Asp Ala Tyr Thr Ser Pro Gln Asn Lys Lys Ala Ile

755 760 765

Trp Gln Val Val Arg Val Val Glu Asp Ile Val Lys Ala Val Gly Asn

770 775 780

Ala Pro Lys Ser Ile Ser Ile Glu Phe Ala Arg Asn Glu Gly Asn Lys

785 790 795 800

Gly Glu Ile Thr Arg Ser Arg Arg Thr Gln Leu Gln Lys Leu Phe Glu
805 810 815
Asp Gln Ala His Glu Leu Val Lys Asp Thr Ser Leu Thr Glu Glu Leu
820 825 830
Glu Lys Ala Pro Asp Leu Ser Asp Arg Tyr Tyr Phe Tyr Phe Thr Gln
835 840 845
Gly Gly Lys Asp Met Tyr Thr Gly Asp Pro Ile Asn Phe Asp Glu Ile

850 855 860
Ser Thr Lys Tyr Asp Ile Asp His Ile Leu Pro Gln Ser Phe Val Lys
865 870 875 880
Asp Asn Ser Leu Asp Asn Arg Val Leu Thr Ser Arg Lys Glu Asn Asn
885 890 895
Lys Lys Ser Asp Gln Val Pro Ala Lys Leu Tyr Ala Ala Lys Met Lys
900 905 910
Pro Tyr Trp Asn Gln Leu Leu Lys Gln Gly Leu Ile Thr Gln Arg Lys

915 920 925
Phe Glu Asn Leu Thr Lys Asp Val Asp Gln Asn Ile Lys Tyr Arg Ser
930 935 940
Leu Gly Phe Val Lys Arg Gln Leu Val Glu Thr Arg Gln Val Ile Lys
945 950 955 960
Leu Thr Ala Asn Ile Leu Gly Ser Met Tyr Gln Glu Ala Gly Thr Glu
965 970 975
Ile Ile Glu Thr Arg Ala Gly Leu Thr Lys Gln Leu Arg Glu Glu Phe

980 985 990
Asp Leu Pro Lys Val Arg Glu Val Asn Asp Tyr His His Ala Val Asp
995 1000 1005
Ala Tyr Leu Thr Thr Phe Ala Gly Gln Tyr Leu Asn Arg Arg Tyr
1010 1015 1020
Pro Lys Leu Arg Ser Phe Phe Val Tyr Gly Glu Tyr Met Lys Phe
1025 1030 1035
Lys His Gly Ser Asp Leu Lys Leu Arg Asn Phe Asn Phe Phe His

1040 1045 1050

Glu Leu Met Glu Gly Asp Lys Ser Gln Gly Lys Val Val Asp Gln

1055 1060 1065

Gln Thr Gly Glu Leu Ile Thr Thr Arg Asp Glu Val Ala Lys Ser

1070 1075 1080

Phe Asp Arg Leu Leu Asn Met Lys Tyr Met Leu Val Ser Lys Glu

1085 1090 1095

Val His Asp Arg Ser Asp Gln Leu Tyr Gly Ala Thr Ile Val Thr

1100 1105 1110

Ala Lys Glu Ser Gly Lys Leu Thr Ser Pro Ile Glu Ile Lys Lys

1115 1120 1125

Asn Arg Leu Val Asp Leu Tyr Gly Ala Tyr Thr Asn Gly Thr Ser

1130 1135 1140

Ala Phe Met Thr Ile Ile Lys Phe Thr Gly Asn Lys Pro Lys Tyr

1145 1150 1155

Lys Val Ile Gly Ile Pro Thr Thr Ser Ala Ala Ser Leu Lys Arg

1160 1165 1170

Ala Gly Lys Pro Gly Ser Glu Ser Tyr Asn Gln Glu Leu His Arg

1175 1180 1185

Ile Ile Lys Ser Asn Pro Lys Val Lys Lys Gly Phe Glu Ile Val

1190 1195 1200

Val Pro His Val Ser Tyr Gly Gln Leu Ile Val Asp Gly Asp Cys

1205 1210 1215

Lys Phe Thr Leu Ala Ser Pro Thr Val Gln His Pro Ala Thr Gln

1220 1225 1230

Leu Val Leu Ser Lys Lys Ser Leu Glu Thr Ile Ser Ser Gly Tyr

1235 1240 1245

Lys Ile Leu Lys Asp Lys Pro Ala Ile Ala Asn Glu Arg Leu Ile

1250 1255 1260

Arg Val Phe Asp Glu Val Val Gly Gln Met Asn Arg Tyr Phe Thr

1265 1270 1275

Ile Phe Asp Gln Arg Ser Asn Arg Gln Lys Val Ala Asp Ala Arg

1280 1285 1290

Asp Lys Phe Leu Ser Leu Pro Thr Glu Ser Lys Tyr Glu Gly Ala

1295 1300 1305

Lys Lys Val Gln Val Gly Lys Thr Glu Val Ile Thr Asn Leu Leu

1310 1315 1320

Met Gly Leu His Ala Asn Ala Thr Gln Gly Asp Leu Lys Val Leu

1325 1330 1335

Gly Leu Ala Thr Phe Gly Phe Phe Gln Ser Thr Thr Gly Leu Ser

1340 1345 1350

Leu Ser Glu Asp Thr Met Ile Val Tyr Gln Ser Pro Thr Gly Leu

1355 1360 1365

Phe Glu Arg Arg Ile Cys Leu Lys Asp Ile

1370 1375

<210> 27

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 27

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 28

<211> 1053

<212> PRT

<213> Staphylococcus aureus

<400> 28

Met Gly Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser

1 5 10 15

Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala

20 25 30

Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg

35 40 45

Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg

50 55 60

Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp

65 70 75 80

His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly

85 90 95

Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His

100 105 110

Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp

115 120 125

Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys

130 135 140

Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys

145 150 155 160

Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp
165 170 175
Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His
180 185 190
Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr
195 200 205
Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp
210 215 220

Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr
225 230 235 240
Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu
245 250 255
Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu
260 265 270
Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val
275 280 285

Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile
290 295 300
Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly
305 310 315 320
Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile
325 330 335
Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile
340 345 350

Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu
355 360 365
Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile
370 375 380
Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala
385 390 395 400

Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile
405 410 415

Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser
420 425 430

Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser
435 440 445

Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala
450 455 460

Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala
465 470 475 480

Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln
485 490 495

Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr
500 505 510

Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His
515 520 525

Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu
530 535 540

Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile
545 550 555 560

Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val
565 570 575

Lys Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr
580 585 590

Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His
595 600 605

Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys
610 615 620

Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys
625 630 635 640

Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly

645 650 655

Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val

660 665 670

Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys

675 680 685

Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu

690 695 700

Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys

705 710 715 720

Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu

725 730 735

Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys

740 745 750

Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys

755 760 765

Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Lys Leu

770 775 780

Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr

785 790 795 800

Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys

805 810 815

Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His

820 825 830

His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr

835 840 845

Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn

850 855 860

Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys

865 870 875 880

Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp

885 890 895

Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro
900 905 910
Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr
915 920 925

Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn
930 935 940
Ser Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln
945 950 955 960
Ala Glu Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile Asn
965 970 975
Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg
980 985 990

Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn
995 1000 1005
Met Asn Asp Lys Arg Pro Pro His Ile Ile Lys Thr Ile Ala Ser
1010 1015 1020
Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn
1025 1030 1035
Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys
1040 1045 1050

<210> 29
<211> 1409
<212> PRT
<213> Streptococcus thermophilus
<400> 29
Met Leu Phe Asn Lys Cys Ile Ile Ile Ser Ile Asn Leu Asp Phe Ser
1 5 10 15
Asn Lys Glu Lys Cys Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile
20 25 30
Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val
35 40 45

Pro Ser Lys Lys Met Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile

50 55 60

Lys Lys Asn Leu Leu Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala

65 70 75 80

Glu Gly Arg Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg

85 90 95

Arg Asn Arg Ile Leu Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala

100 105 110

Thr Leu Asp Asp Ala Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val

115 120 125

Pro Asp Asp Lys Arg Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val

130 135 140

Glu Glu Lys Val Tyr His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg

145 150 155 160

Lys Tyr Leu Ala Asp Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr

165 170 175

Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu

180 185 190

Gly Glu Phe Asn Ser Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp

195 200 205

Phe Leu Asp Thr Tyr Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu

210 215 220

Asn Ser Lys Gln Leu Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu

225 230 235 240

Glu Lys Lys Asp Arg Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser

245 250 255

Gly Ile Phe Ser Glu Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp

260 265 270

Phe Arg Lys Cys Phe Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser

275 280 285

Lys Glu Ser Tyr Asp Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly

290 295 300

Asp Asp Tyr Ser Asp Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala

305 310 315 320

Ile Leu Leu Ser Gly Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala

325 330 335

Pro Leu Ser Ser Ala Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp

340 345 350

Leu Ala Leu Leu Lys Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr

355 360 365

Asn Glu Val Phe Lys Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile

370 375 380

Asp Gly Lys Thr Asn Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu

385 390 395 400

Leu Ala Glu Phe Glu Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg

405 410 415

Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro

420 425 430

Tyr Gln Ile His Leu Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala

435 440 445

Lys Phe Tyr Pro Phe Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile

450 455 460

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn

465 470 475 480

Ser Asp Phe Ala Trp Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro

485 490 495

Trp Asn Phe Glu Asp Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe

500 505 510

Ile Asn Arg Met Thr Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val

515 520 525

Leu Pro Lys His Ser Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu

530 535 540

Leu Thr Lys Val Arg Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe
545 550 555 560
Leu Asp Ser Lys Gln Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp

565 570 575
Lys Arg Lys Val Thr Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile
580 585 590
Tyr Gly Tyr Asp Gly Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn
595 600 605
Ser Ser Leu Ser Thr Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys
610 615 620
Glu Phe Leu Asp Asp Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile

625 630 635 640
His Thr Leu Thr Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu
645 650 655
Ser Lys Phe Glu Asn Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser
660 665 670
Arg Arg His Tyr Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn
675 680 685
Gly Ile Arg Asp Glu Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile

690 695 700
Asp Asp Gly Ile Ser Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
705 710 715 720
Ala Leu Ser Phe Lys Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp
725 730 735
Glu Asp Lys Gly Asn Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser
740 745 750
Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu

755 760 765
Leu Val Lys Val Met Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu
770 775 780
Met Ala Arg Glu Asn Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln

785 790 795 800

Gln Arg Leu Lys Arg Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys

805 810 815

Ile Leu Lys Glu Asn Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn

820 825 830

Ala Leu Gln Asn Asp Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys

835 840 845

Asp Met Tyr Thr Gly Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr

850 855 860

Asp Ile Asp His Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile

865 870 875 880

Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp

885 890 895

Asp Phe Pro Ser Leu Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr

900 905 910

Gln Leu Leu Lys Ser Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu

915 920 925

Thr Lys Ala Glu Arg Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe

930 935 940

Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala

945 950 955 960

Arg Leu Leu Asp Glu Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg

965 970 975

Ala Val Arg Thr Val Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser

980 985 990

Gln Phe Arg Lys Asp Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp

995 1000 1005

Phe His His Ala His Asp Ala Tyr Leu Asn Ala Val Ile Ala Ser

1010 1015 1020

Ala Leu Leu Lys Lys Tyr Pro Lys Leu Glu Pro Glu Phe Val Tyr

1025 1030 1035

Gly Asp Tyr Pro Lys Tyr Asn Ser Phe Arg Glu Arg Lys Ser Ala

1040 1045 1050

Thr Glu Lys Val Tyr Phe Tyr Ser Asn Ile Met Asn Ile Phe Lys

1055 1060 1065

Lys Ser Ile Ser Leu Ala Asp Gly Arg Val Ile Glu Arg Pro Leu

1070 1075 1080

Ile Glu Val Asn Glu Glu Thr Gly Glu Ser Val Trp Asn Lys Glu

1085 1090 1095

Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser Tyr Pro Gln Val

1100 1105 1110

Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly Leu Asp Arg

1115 1120 1125

Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser Lys Pro

1130 1135 1140

Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr Leu

1145 1150 1155

Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe

1160 1165 1170

Ala Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys

1175 1180 1185

Ile Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg

1190 1195 1200

Ile Asn Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly

1205 1210 1215

Tyr Lys Asp Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu

1220 1225 1230

Phe Glu Leu Ser Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu

1235 1240 1245

Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Gln Ile

1250 1255 1260

Phe Leu Ser Gln Lys Phe Val Lys Leu Leu Tyr His Ala Lys Arg

1265 1270 1275

Ile Ser Asn Thr Ile Asn Glu Asn His Arg Lys Tyr Val Glu Asn

1280 1285 1290

His Lys Lys Glu Phe Glu Glu Leu Phe Tyr Tyr Ile Leu Glu Phe

1295 1300 1305

Asn Glu Asn Tyr Val Gly Ala Lys Lys Asn Gly Lys Leu Leu Asn

1310 1315 1320

Ser Ala Phe Gln Ser Trp Gln Asn His Ser Ile Asp Glu Leu Cys

1325 1330 1335

Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg Lys Gly Leu Phe

1340 1345 1350

Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu Phe Leu Gly

1355 1360 1365

Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser Leu Leu

1370 1375 1380

Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr Glu

1385 1390 1395

Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly

1400 1405

<210> 30

<211> 1402

<212> PRT

<213> Lactobacillus crispatus

<400> 30

Met Lys Ile Lys Asn Tyr Asn Leu Ala Leu Thr Pro Ser Thr Ser Ala

1 5 10 15

Val Gly His Val Glu Val Asp Asp Asp Leu Asn Ile Leu Glu Pro Val

20 25 30

His His Gln Lys Ala Ile Gly Val Ala Lys Phe Gly Glu Gly Glu Thr

35 40 45

Ala Glu Ala Arg Arg Leu Ala Arg Ser Ala Arg Arg Thr Thr Lys Arg

50 55 60

Arg Ala Asn Arg Ile Asn His Tyr Phe Asn Glu Ile Met Lys Pro Glu
65 70 75 80
Ile Asp Lys Val Asp Pro Leu Met Phe Asp Arg Ile Lys Gln Ala Gly

85 90 95
Leu Ser Pro Leu Asp Glu Arg Lys Glu Phe Arg Thr Val Ile Phe Asp
100 105 110
Arg Pro Asn Ile Ala Ser Tyr Tyr His Asn Gln Phe Pro Thr Ile Trp
115 120 125
His Leu Gln Lys Tyr Leu Met Ile Thr Asp Glu Lys Ala Asp Ile Arg
130 135 140
Leu Ile Tyr Trp Ala Leu His Ser Leu Leu Lys His Arg Gly His Phe

145 150 155 160
Phe Asn Thr Thr Pro Met Ser Gln Phe Lys Pro Gly Lys Leu Asn Leu
165 170 175
Lys Asp Asp Met Leu Ala Leu Asp Asp Tyr Asn Asp Leu Glu Gly Leu
180 185 190
Ser Phe Ala Val Ala Asn Ser Pro Glu Ile Glu Lys Val Ile Lys Asp
195 200 205
Arg Ser Met His Lys Lys Glu Lys Ile Ala Glu Leu Lys Lys Leu Ile

210 215 220
Val Asn Asp Val Pro Asp Lys Asp Leu Ala Lys Arg Asn Asn Lys Ile
225 230 235 240
Ile Thr Gln Ile Val Asn Ala Ile Met Gly Asn Ser Phe His Leu Asn
245 250 255
Phe Ile Phe Asp Met Asp Leu Asp Lys Leu Thr Ser Lys Ala Trp Ser
260 265 270
Phe Lys Leu Asp Asp Pro Glu Leu Asp Thr Lys Phe Asp Ala Ile Ser

275 280 285
Gly Ser Met Thr Asp Asn Gln Ile Gly Ile Phe Glu Thr Leu Gln Lys
290 295 300
Ile Tyr Ser Ala Ile Ser Leu Leu Asp Ile Leu Asn Gly Ser Ser Asn

305 310 315 320

Val Val Asp Ala Lys Asn Ala Leu Tyr Asp Lys His Lys Arg Asp Leu

325 330 335

Asn Leu Tyr Phe Lys Phe Leu Asn Thr Leu Pro Asp Glu Ile Ala Lys

340 345 350

Thr Leu Lys Ala Gly Tyr Thr Leu Tyr Ile Gly Asn Arg Lys Lys Asp

355 360 365

Leu Leu Ala Ala Arg Lys Leu Leu Lys Val Asn Val Ala Lys Asn Phe

370 375 380

Ser Gln Asp Asp Phe Tyr Lys Leu Ile Asn Lys Glu Leu Lys Ser Ile

385 390 395 400

Asp Lys Gln Gly Leu Gln Thr Arg Phe Ser Glu Lys Val Gly Glu Leu

405 410 415

Val Ala Gln Asn Asn Phe Leu Pro Val Gln Arg Ser Ser Asp Asn Val

420 425 430

Phe Ile Pro Tyr Gln Leu Asn Ala Ile Thr Phe Asn Lys Ile Leu Glu

435 440 445

Asn Gln Gly Lys Tyr Tyr Asp Phe Leu Val Lys Pro Asn Pro Ala Lys

450 455 460

Lys Asp Arg Lys Asn Ala Pro Tyr Glu Leu Ser Gln Leu Met Gln Phe

465 470 475 480

Thr Ile Pro Tyr Tyr Val Gly Pro Leu Val Thr Pro Glu Glu Gln Val

485 490 495

Lys Ser Gly Ile Pro Lys Thr Ser Arg Phe Ala Trp Met Val Arg Lys

500 505 510

Asp Asn Gly Ala Ile Thr Pro Trp Asn Phe Tyr Asp Lys Val Asp Ile

515 520 525

Glu Ala Thr Ala Asp Lys Phe Ile Lys Arg Ser Ile Ala Lys Asp Ser

530 535 540

Tyr Leu Leu Ser Glu Leu Val Leu Pro Lys His Ser Leu Leu Tyr Glu

545 550 555 560

Lys Tyr Glu Val Phe Asn Glu Leu Ser Asn Val Ser Leu Asp Gly Lys
565 570 575
Lys Leu Ser Gly Gly Val Lys Gln Ile Leu Phe Asn Glu Val Phe Lys
580 585 590
Lys Thr Asn Lys Val Asn Thr Ser Arg Ile Leu Lys Ala Leu Ala Lys

595 600 605
His Asn Ile Pro Gly Ser Lys Ile Thr Gly Leu Ser Asn Pro Glu Glu
610 615 620
Phe Thr Ser Ser Leu Gln Thr Tyr Asn Ala Trp Lys Lys Tyr Phe Pro
625 630 635 640
Asn Gln Ile Asp Asn Phe Ala Tyr Gln Gln Asp Leu Glu Lys Met Ile
645 650 655
Glu Trp Ser Thr Val Phe Glu Asp His Lys Ile Leu Ala Lys Lys Leu

660 665 670
Asp Glu Ile Glu Trp Leu Asp Asp Asp Gln Lys Lys Phe Val Ala Asn
675 680 685
Thr Arg Leu Arg Gly Trp Gly Arg Leu Ser Lys Arg Leu Leu Thr Gly
690 695 700
Leu Lys Asp Asn Tyr Gly Lys Ser Ile Met Gln Arg Leu Glu Thr Thr
705 710 715 720
Lys Ala Asn Phe Gln Gln Ile Val Tyr Lys Pro Glu Phe Arg Glu Gln

725 730 735
Ile Asp Lys Ile Ser Gln Ala Ala Ala Lys Asn Gln Ser Leu Glu Asp
740 745 750
Ile Leu Ala Asn Ser Tyr Thr Ser Pro Ser Asn Arg Lys Ala Ile Arg
755 760 765
Lys Thr Met Ser Val Val Asp Glu Tyr Ile Lys Leu Asn His Gly Lys
770 775 780
Glu Pro Asp Lys Ile Phe Leu Met Phe Gln Arg Ser Glu Gln Glu Lys

785 790 795 800
Gly Lys Gln Thr Glu Ala Arg Ser Lys Gln Leu Asn Arg Ile Leu Ser

805 810 815

Gln Leu Lys Ala Asp Lys Ser Ala Asn Lys Leu Phe Ser Lys Gln Leu

820 825 830

Ala Asp Glu Phe Ser Asn Ala Ile Lys Lys Ser Lys Tyr Lys Leu Asn

835 840 845

Asp Lys Gln Tyr Phe Tyr Phe Gln Gln Leu Gly Arg Asp Ala Leu Thr

850 855 860

Gly Glu Val Ile Asp Tyr Asp Glu Leu Tyr Lys Tyr Thr Val Leu His

865 870 875 880

Ile Ile Pro Arg Ser Lys Leu Thr Asp Asp Ser Gln Asn Asn Lys Val

885 890 895

Leu Thr Lys Tyr Lys Ile Val Asp Gly Ser Val Ala Leu Lys Phe Gly

900 905 910

Asn Ser Tyr Ser Asp Ala Leu Gly Met Pro Ile Lys Ala Phe Trp Thr

915 920 925

Glu Leu Asn Arg Leu Lys Leu Ile Pro Lys Gly Lys Leu Leu Asn Leu

930 935 940

Thr Thr Asp Phe Ser Thr Leu Asn Lys Tyr Gln Arg Asp Gly Tyr Ile

945 950 955 960

Ala Arg Gln Leu Val Glu Thr Gln Gln Ile Val Lys Leu Leu Ala Thr

965 970 975

Ile Met Gln Ser Arg Phe Lys His Thr Lys Ile Ile Glu Val Arg Asn

980 985 990

Ser Gln Val Ala Asn Ile Arg Tyr Gln Phe Asp Tyr Phe Arg Ile Lys

995 1000 1005

Asn Leu Asn Glu Tyr Tyr Arg Gly Phe Asp Ala Tyr Leu Ala Ala

1010 1015 1020

Val Val Gly Thr Tyr Leu Tyr Lys Val Tyr Pro Lys Ala Arg Arg

1025 1030 1035

Leu Phe Val Tyr Gly Gln Tyr Leu Lys Pro Lys Lys Thr Asn Gln

1040 1045 1050

Glu Asn Gln Asp Met His Leu Asp Ser Glu Lys Lys Ser Gln Gly
1055 1060 1065
Phe Asn Phe Leu Trp Asn Leu Leu Tyr Gly Lys Gln Asp Gln Ile
1070 1075 1080
Phe Val Asn Gly Thr Asp Val Ile Ala Phe Asn Arg Lys Asp Leu
1085 1090 1095
Ile Thr Lys Met Asn Thr Val Tyr Asn Tyr Lys Ser Gln Lys Ile

1100 1105 1110
Ser Leu Ala Ile Asp Tyr His Asn Gly Ala Met Phe Lys Ala Thr
1115 1120 1125
Leu Phe Pro Arg Asn Asp Arg Asp Thr Ala Lys Thr Arg Lys Leu
1130 1135 1140
Ile Pro Lys Lys Lys Asp Tyr Asp Thr Asp Ile Tyr Gly Gly Tyr
1145 1150 1155
Thr Ser Asn Val Asp Gly Tyr Met Leu Leu Ala Glu Ile Ile Lys

1160 1165 1170
Arg Asp Gly Asn Lys Gln Tyr Gly Phe Tyr Gly Val Pro Ser Arg
1175 1180 1185
Leu Val Ser Glu Leu Asp Thr Leu Lys Lys Thr Arg Tyr Thr Glu
1190 1195 1200
Tyr Glu Glu Lys Leu Lys Glu Ile Ile Lys Pro Glu Leu Gly Val
1205 1210 1215
Asp Leu Lys Lys Ile Lys Lys Ile Lys Ile Leu Lys Asn Lys Val

1220 1225 1230
Pro Phe Asn Gln Val Ile Ile Asp Lys Gly Ser Lys Phe Phe Ile
1235 1240 1245
Thr Ser Thr Ser Tyr Arg Trp Asn Tyr Arg Gln Leu Ile Leu Ser
1250 1255 1260
Ala Glu Ser Gln Gln Thr Leu Met Asp Leu Val Val Asp Pro Asp
1265 1270 1275
Phe Ser Asn His Lys Ala Arg Lys Asp Ala Arg Lys Asn Ala Asp

1280 1285 1290

Glu Arg Leu Ile Lys Val Tyr Glu Glu Ile Leu Tyr Gln Val Lys

1295 1300 1305

Asn Tyr Met Pro Met Phe Val Glu Leu His Arg Cys Tyr Glu Lys

1310 1315 1320

Leu Val Asp Ala Gln Lys Thr Phe Lys Ser Leu Lys Ile Ser Asp

1325 1330 1335

Lys Ala Met Val Leu Asn Gln Ile Leu Ile Leu Leu His Ser Asn

1340 1345 1350

Ala Thr Ser Pro Val Leu Glu Lys Leu Gly Tyr His Thr Arg Phe

1355 1360 1365

Thr Leu Gly Lys Lys His Asn Leu Ile Ser Glu Asn Ala Val Leu

1370 1375 1380

Val Thr Gln Ser Ile Thr Gly Leu Lys Glu Asn His Val Ser Ile

1385 1390 1395

Lys Gln Met Leu

1400

<210> 31

<211

> 1385

<212> PRT

<213> Thermoactinomyces candidus

<400> 31

Met Thr Asn Glu Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Ser Ser

1 5 10 15

Ile Gly Phe Ala Val Val Asn Asp Asn Asn Arg Val Ile Arg Val Lys

20 25 30

Gly Lys Asn Ala Ile Gly Val Arg Leu Phe Asp Glu Gly Lys Ala Ala

35 40 45

Ala Asp Arg Arg Ser Phe Arg Thr Thr Arg Arg Ser Phe Arg Thr Thr

50 55 60

Arg Arg Arg Leu Ser Arg Arg Arg Trp Arg Leu Lys Leu Leu Arg Glu

65 70 75 80

Ile Phe Asp Ala Tyr Ile Thr Pro Val Asp Glu Ala Phe Phe Ile Arg
85 90 95
Leu Lys Glu Ser Asn Leu Ser Pro Lys Asp Ser Lys Lys Gln Tyr Ser
100 105 110
Gly Asp Ile Leu Phe Asn Asp Arg Ser Asp Lys Asp Phe Tyr Glu Lys

115 120 125
Tyr Pro Thr Ile Tyr His Leu Arg Asn Ala Leu Met Thr Glu His Arg
130 135 140
Lys Phe Asp Val Arg Glu Ile Tyr Leu Ala Ile His His Ile Met Lys
145 150 155 160
Phe Arg Gly His Phe Leu Asn Ala Thr Pro Ala Asn Asn Phe Lys Val
165 170 175
Gly Arg Leu Asn Leu Glu Glu Lys Phe Glu Glu Leu Asn Asp Ile Tyr

180 185 190
Gln Arg Val Phe Pro Asp Glu Ser Ile Glu Phe Arg Thr Asp Asn Leu
195 200 205
Glu Gln Ile Lys Glu Val Leu Leu Asp Asn Lys Arg Ser Arg Ala Asp
210 215 220
Arg Gln Arg Thr Leu Val Ser Asp Ile Tyr Gln Ser Ser Glu Asp Lys
225 230 235 240
Asp Ile Glu Lys Arg Asn Lys Ala Val Ala Thr Glu Ile Leu Lys Ala

245 250 255
Ser Leu Gly Asn Lys Ala Lys Leu Asn Val Ile Thr Asn Val Glu Val
260 265 270
Asp Lys Glu Ala Ala Lys Glu Trp Ser Ile Thr Phe Asp Ser Glu Ser
275 280 285
Ile Asp Asp Asp Leu Ala Lys Ile Glu Gly Gln Met Thr Asp Asp Gly
290 295 300
His Glu Ile Ile Glu Val Leu Arg Ser Leu Tyr Ser Gly Ile Thr Leu

305 310 315 320
Ser Ala Ile Val Pro Glu Asn His Thr Leu Ser Gln Ser Met Val Ala

325 330 335

Lys Tyr Asp Leu His Lys Asp His Leu Lys Leu Phe Lys Lys Leu Ile

340 345 350

Asn Gly Met Thr Asp Thr Lys Lys Ala Lys Asn Leu Arg Ala Ala Tyr

355 360 365

Asp Gly Tyr Ile Asp Gly Val Lys Gly Lys Val Leu Pro Gln Glu Asp

370 375 380

Phe Tyr Lys Gln Val Gln Val Asn Leu Asp Asp Ser Ala Glu Ala Asn

385 390 395 400

Glu Ile Gln Thr Tyr Ile Asp Gln Asp Ile Phe Met Pro Lys Gln Arg

405 410 415

Thr Lys Ala Asn Gly Ser Ile Pro His Gln Leu Gln Gln Gln Glu Leu

420 425 430

Asp Gln Ile Ile Glu Asn Gln Lys Ala Tyr Tyr Pro Trp Leu Ala Glu

435 440 445

Leu Asn Pro Asn Pro Asp Lys Lys Arg Gln Gln Leu Ala Lys Tyr Lys

450 455 460

Leu Asp Glu Leu Val Thr Phe Arg Val Pro Tyr Tyr Val Gly Pro Met

465 470 475 480

Ile Thr Ala Lys Asp Gln Lys Asn Gln Ser Gly Ala Glu Phe Ala Trp

485 490 495

Met Ile Arg Lys Glu Pro Gly Asn Ile Thr Pro Trp Asn Phe Asp Gln

500 505 510

Lys Val Asp Arg Met Ala Thr Ala Asn Gln Phe Ile Lys Arg Met Thr

515 520 525

Thr Thr Asp Thr Tyr Leu Leu Gly Glu Asp Val Leu Pro Ala Gln Ser

530 535 540

Leu Leu Tyr Gln Lys Phe Glu Val Leu Asn Glu Leu Asn Lys Ile Arg

545 550 555 560

Ile Asp His Lys Pro Ile Ser Ile Glu Gln Lys Gln Gln Ile Phe Asn

565 570 575

Asp Leu Phe Lys Gln Phe Lys Asn Val Thr Ile Lys His Leu Gln Asp
580 585 590
Tyr Leu Val Ser Gln Gly Gln Tyr Ser Lys Arg Pro Leu Ile Glu Gly
595 600 605
Leu Ala Asp Glu Lys Arg Phe Asn Ser Ser Leu Ser Thr Tyr Ser Asp
610 615 620
Leu Cys Gly Ile Phe Gly Ala Lys Leu Val Glu Glu Asn Asp Arg Gln

625 630 635 640
Glu Asp Leu Glu Lys Ile Ile Glu Trp Ser Thr Ile Phe Glu Asp Lys
645 650 655
Lys Ile Tyr Arg Ala Lys Leu Asn Asp Leu Thr Trp Leu Thr Asp Asp
660 665 670
Gln Lys Glu Lys Leu Ala Thr Lys Arg Tyr Gln Gly Trp Gly Arg Leu
675 680 685
Ser Arg Lys Leu Leu Val Gly Leu Lys Asn Ser Glu His Arg Asn Ile

690 695 700
Met Asp Ile Leu Trp Ile Thr Asn Glu Asn Phe Met Gln Ile Gln Ala
705 710 715 720
Glu Pro Asp Phe Ala Lys Leu Val Thr Asp Ala Asn Lys Gly Met Leu
725 730 735
Glu Lys Thr Asp Ser Gln Asp Val Ile Asn Asp Leu Tyr Thr Ser Pro
740 745 750
Gln Asn Lys Lys Ala Ile Arg Gln Ile Leu Leu Val Val His Asp Ile

755 760 765
Gln Asn Ala Met His Gly Gln Ala Pro Ala Lys Ile His Val Glu Phe
770 775 780
Ala Arg Gly Glu Glu Arg Asn Pro Arg Arg Ser Val Gln Arg Gln Arg
785 790 795 800
Gln Val Glu Ala Ala Tyr Glu Lys Val Ser Asn Glu Leu Val Ser Ala
805 810 815
Lys Val Arg Gln Glu Phe Lys Glu Ala Ile Asn Asn Lys Arg Asp Phe

820 825 830

Lys Asp Arg Leu Phe Leu Tyr Phe Met Gln Gly Gly Ile Asp Ile Tyr

835 840 845

Thr Gly Lys Gln Leu Asn Ile Asp Gln Leu Ser Ser Tyr Gln Ile Asp

850 855 860

His Ile Leu Pro Gln Ala Phe Val Lys Asp Asp Ser Leu Thr Asn Arg

865 870 875 880

Val Leu Thr Asn Glu Asn Gln Val Lys Ala Asp Ser Val Pro Ile Asp

885 890 895

Ile Phe Gly Lys Lys Met Leu Ser Val Trp Gly Arg Met Lys Asp Gln

900 905 910

Gly Leu Ile Ser Lys Gly Lys Tyr Arg Asn Leu Thr Met Asn Pro Glu

915 920 925

Asn Ile Ser Ala His Thr Glu Asn Gly Phe Ile Asn Arg Gln Leu Val

930 935 940

Glu Thr Arg Gln Val Ile Lys Leu Ala Val Asn Ile Leu Ala Asp Glu

945 950 955 960

Tyr Gly Asp Ser Thr Gln Ile Ile Ser Val Lys Ala Asp Leu Ser His

965 970 975

Gln Met Arg Glu Asp Phe Glu Leu Leu Lys Asn Arg Asp Val Asn Asp

980 985 990

Tyr His His Ala Phe Asp Ala Tyr Leu Ala Ala Phe Ile Gly Asn Tyr

995 1000 1005

Leu Leu Lys Arg Tyr Pro Lys Leu Glu Ser Tyr Phe Val Tyr Gly

1010 1015 1020

Asp Phe Lys Lys Phe Thr Gln Lys Glu Thr Lys Met Arg Arg Phe

1025 1030 1035

Asn Phe Ile Tyr Asp Leu Lys His Cys Asp Gln Val Val Asn Lys

1040 1045 1050

Glu Thr Gly Glu Ile Leu Trp Thr Lys Asp Glu Asp Ile Lys Tyr

1055 1060 1065

Ile Arg His Leu Phe Ala Tyr Lys Lys Ile Leu Val Ser His Glu

1070 1075 1080

Val Arg Glu Lys Arg Gly Ala Leu Tyr Asn Gln Thr Ile Tyr Lys

1085 1090 1095

Ala Lys Asp Asp Lys Gly Ser Gly Gln Glu Ser Lys Lys Leu Ile

1100 1105 1110

Arg Ile Lys Asp Asp Lys Glu Thr Lys Ile Tyr Gly Gly Tyr Ser

1115 1120 1125

Gly Lys Ser Leu Ala Tyr Met Thr Ile Val Gln Ile Thr Lys Lys

1130 1135 1140

Asn Lys Val Ser Tyr Arg Val Ile Gly Ile Pro Thr Leu Ala Leu

1145 1150 1155

Ala Arg Leu Asn Lys Leu Glu Asn Asp Ser Thr Glu Asn Asn Gly

1160 1165 1170

Glu Leu Tyr Lys Ile Ile Lys Pro Gln Phe Thr His Tyr Lys Val

1175 1180 1185

Asp Lys Lys Asn Gly Glu Ile Ile Glu Thr Thr Asp Asp Phe Lys

1190 1195 1200

Ile Val Val Ser Lys Val Arg Phe Gln Gln Leu Ile Asp Asp Ala

1205 1210 1215

Gly Gln Phe Phe Met Leu Ala Ser Asp Thr Tyr Lys Asn Asn Ala

1220 1225 1230

Gln Gln Leu Val Ile Ser Asn Asn Ala Leu Lys Ala Ile Asn Asn

1235 1240 1245

Thr Asn Ile Thr Asp Cys Pro Arg Asp Asp Leu Glu Arg Leu Asp

1250 1255 1260

Asn Leu Arg Leu Asp Ser Ala Phe Asp Glu Ile Val Lys Lys Met

1265 1270 1275

Asp Lys Tyr Phe Ser Ala Tyr Asp Ala Asn Asn Phe Arg Glu Lys

1280 1285 1290

Ile Arg Asn Ser Asn Leu Ile Phe Tyr Gln Leu Pro Val Glu Asp

1295 1300 1305

Gln Trp Glu Asn Asn Lys Ile Thr Glu Leu Gly Lys Arg Thr Val

1310 1315 1320

Leu Thr Arg Ile Leu Gln Gly Leu His Ala Asn Ala Thr Thr Thr

1325 1330 1335

Asp Met Ser Ile Phe Lys Ile Lys Thr Pro Phe Gly Gln Leu Arg

1340 1345 1350

Gln Arg Ser Gly Ile Ser Leu Ser Glu Asn Ala Gln Leu Ile Tyr

1355 1360 1365

Gln Ser Pro Thr Gly Leu Phe Glu Arg Arg Val Gln Leu Asn Lys

1370 1375 1380

Ile Lys

1385

<210> 32

<211> 1367

<212> PRT

<213> Fusobacterium nucleatum

<400> 32

Met Lys Lys Gln Lys Phe Ser Asp Tyr Tyr Leu Gly Phe Asp Ile Gly

1 5 10 15

Thr Asn Ser Val Gly Trp Cys Val Thr Asp Leu Asp Tyr Asn Val Leu

20 25 30

Arg Phe Asn Lys Lys Asp Met Trp Gly Ser Arg Leu Phe Glu Glu Ala

35 40 45

Lys Thr Ala Ala Glu Arg Arg Val Gln Arg Asn Ser Arg Arg Arg Leu

50 55 60

Lys Arg Arg Lys Trp Arg Leu Asn Leu Leu Glu Glu Ile Phe Ser Asn

65 70 75 80

Glu Ile Leu Lys Ile Asp Ser Asn Phe Phe Arg Arg Leu Lys Glu Ser

85 90 95

Ser Leu Trp Leu Glu Asp Lys Ser Ser Lys Glu Lys Phe Thr Leu Phe

100 105 110

Asn Asp Asp Asn Tyr Lys Asp Tyr Asp Phe Tyr Lys Gln Tyr Pro Thr
115 120 125
Ile Phe His Leu Arg Asn Glu Leu Ile Lys Asn Pro Glu Lys Lys Asp
130 135 140
Ile Arg Leu Val Tyr Leu Ala Ile His Ser Ile Phe Lys Ser Arg Gly
145 150 155 160
His Phe Leu Phe Glu Gly Gln Asn Leu Lys Glu Ile Lys Asn Phe Glu
165 170 175

Thr Leu Tyr Asn Asn Leu Ile Ala Phe Leu Glu Asp Asn Gly Ile Asn
180 185 190
Lys Ile Ile Asp Lys Asn Asn Ile Glu Lys Leu Glu Lys Ile Val Cys
195 200 205
Asp Ser Lys Lys Gly Leu Lys Asp Lys Glu Lys Glu Phe Lys Glu Ile
210 215 220
Phe Asn Ser Asp Lys Gln Leu Val Ala Ile Phe Lys Leu Ser Val Gly
225 230 235 240

Ser Ser Val Ser Leu Asn Asp Leu Phe Asp Thr Asp Glu Tyr Lys Lys
245 250 255
Gly Glu Val Glu Lys Glu Lys Ile Ser Phe Arg Glu Gln Ile Tyr Glu
260 265 270
Asp Asp Lys Pro Ile Tyr Tyr Ser Ile Leu Gly Glu Lys Ile Glu Leu
275 280 285
Leu Asp Ile Ala Lys Thr Phe Tyr Asp Phe Met Val Leu Asn Asn Ile
290 295 300

Leu Ala Asp Ser Gln Tyr Ile Ser Glu Ala Lys Val Lys Leu Tyr Glu
305 310 315 320
Glu His Lys Lys Asp Leu Lys Asn Leu Lys Tyr Ile Ile Arg Lys Tyr
325 330 335
Asn Lys Gly Asn Tyr Asp Lys Leu Phe Lys Asp Lys Asn Glu Asn Asn
340 345 350

Tyr Ser Ala Tyr Ile Gly Leu Asn Lys Glu Lys Ser Lys Lys Glu Val
355 360 365

Ile Glu Lys Ser Arg Leu Lys Ile Asp Asp Leu Ile Lys Asn Ile Lys
370 375 380

Gly Tyr Leu Pro Lys Val Glu Glu Ile Glu Glu Lys Asp Lys Ala Ile
385 390 395 400

Phe Asn Lys Ile Leu Asn Lys Ile Glu Leu Lys Thr Ile Leu Pro Lys
405 410 415

Gln Arg Ile Ser Asp Asn Gly Thr Leu Pro Tyr Gln Ile His Glu Ala
420 425 430

Glu Leu Glu Lys Ile Leu Glu Asn Gln Ser Lys Tyr Tyr Asp Phe Leu
435 440 445

Asn Tyr Glu Glu Asn Gly Ile Ile Thr Lys Asp Lys Leu Leu Met Thr
450 455 460

Phe Lys Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Asn Ser Tyr His
465 470 475 480

Lys Asp Lys Gly Gly Asn Ser Trp Ile Val Arg Lys Glu Glu Gly Lys
485 490 495

Ile Leu Pro Trp Asn Phe Glu Gln Lys Val Asp Ile Glu Lys Ser Ala
500 505 510

Glu Glu Phe Ile Lys Arg Met Thr Asn Lys Cys Thr Tyr Leu Asn Gly
515 520 525

Glu Asp Val Ile Pro Lys Asp Thr Phe Leu Tyr Ser Glu Tyr Val Ile
530 535 540

Leu Asn Glu Leu Asn Lys Val Gln Val Asn Asp Glu Phe Leu Asn Glu
545 550 555 560

Glu Asn Lys Arg Lys Ile Ile Asp Glu Leu Phe Lys Glu Asn Lys Lys
565 570 575

Val Ser Glu Lys Lys Phe Lys Glu Tyr Leu Leu Val Lys Gln Ile Val
580 585 590

Asp Gly Thr Ile Glu Leu Lys Gly Val Lys Asp Ser Phe Asn Ser Asn

595 600 605

Tyr Ile Ser Tyr Ile Arg Phe Lys Asp Ile Phe Gly Glu Lys Leu Asn

610 615 620

Leu Asp Ile Tyr Lys Glu Ile Ser Glu Lys Ser Ile Leu Trp Lys Cys

625 630 635 640

Leu Tyr Gly Asp Asp Lys Lys Ile Phe Glu Lys Lys Ile Lys Asn Glu

645 650 655

Tyr Gly Asp Ile Leu Thr Lys Asp Glu Ile Lys Lys Ile Asn Thr Phe

660 665 670

Lys Phe Asn Asn Trp Gly Arg Leu Ser Glu Lys Leu Leu Thr Gly Ile

675 680 685

Glu Phe Ile Asn Leu Glu Thr Gly Glu Cys Tyr Ser Ser Val Met Asp

690 695 700

Ala Leu Arg Arg Thr Asn Tyr Asn Leu Met Glu Leu Leu Ser Ser Lys

705 710 715 720

Phe Thr Leu Gln Glu Ser Ile Asn Asn Glu Asn Lys Glu Met Asn Glu

725 730 735

Ala Ser Tyr Arg Asp Leu Ile Glu Glu Ser Tyr Val Ser Pro Ser Leu

740 745 750

Lys Arg Ala Ile Phe Gln Thr Leu Lys Ile Tyr Glu Glu Ile Arg Lys

755 760 765

Ile Thr Gly Arg Val Pro Lys Lys Val Phe Ile Glu Met Ala Arg Gly

770 775 780

Gly Asp Glu Ser Met Lys Asn Lys Lys Ile Pro Ala Arg Gln Glu Gln

785 790 795 800

Leu Lys Lys Leu Tyr Asp Ser Cys Gly Asn Asp Ile Ala Asn Phe Ser

805 810 815

Ile Asp Ile Lys Glu Met Lys Asn Ser Leu Ile Ser Tyr Asp Asn Asn

820 825 830

Ser Leu Arg Gln Lys Lys Leu Tyr Leu Tyr Tyr Leu Gln Phe Gly Lys

835 840 845

Cys Met Tyr Thr Gly Arg Glu Ile Asp Leu Asp Arg Leu Leu Gln Asn
850 855 860
Asn Asp Thr Tyr Asp Ile Asp His Ile Tyr Pro Arg Ser Lys Val Ile
865 870 875 880

Lys Asp Asp Ser Phe Asp Asn Leu Val Leu Val Leu Lys Asn Glu Asn
885 890 895
Ala Glu Lys Ser Asn Glu Tyr Pro Val Lys Lys Glu Ile Gln Glu Lys
900 905 910
Met Lys Ser Phe Trp Arg Phe Leu Lys Glu Lys Asn Phe Ile Ser Asp
915 920 925
Glu Lys Tyr Lys Arg Leu Thr Gly Lys Asp Asp Phe Glu Leu Arg Gly
930 935 940

Phe Met Ala Arg Gln Leu Val Asn Val Arg Gln Thr Thr Lys Glu Val
945 950 955 960
Gly Lys Ile Leu Gln Gln Ile Glu Pro Glu Ile Lys Ile Val Tyr Ser
965 970 975
Lys Ala Glu Ile Ala Ser Ser Phe Arg Glu Met Phe Asp Phe Ile Lys
980 985 990
Val Arg Glu Leu Asn Asp Thr His His Ala Lys Asp Ala Tyr Leu Asn
995 1000 1005

Ile Val Ala Gly Asn Val Tyr Asn Thr Lys Phe Thr Glu Lys Pro
1010 1015 1020
Tyr Arg Tyr Leu Gln Glu Ile Lys Glu Asn Tyr Asp Val Lys Lys
1025 1030 1035
Ile Tyr Asn Tyr Asp Ile Lys Asn Ala Trp Asp Lys Glu Asn Ser
1040 1045 1050
Leu Glu Ile Val Lys Lys Asn Met Glu Lys Asn Thr Val Asn Ile
1055 1060 1065

Thr Arg Phe Ile Lys Glu Lys Lys Gly Gln Leu Phe Asp Leu Asn
1070 1075 1080
Pro Ile Lys Lys Gly Glu Thr Ser Asn Glu Ile Ile Ser Ile Lys

1085 1090 1095

Pro Lys Val Tyr Asn Gly Lys Asp Asp Lys Leu Asn Glu Lys Tyr

1100 1105 1110

Gly Tyr Tyr Lys Ser Leu Asn Pro Ala Tyr Phe Leu Tyr Val Glu

1115 1120 1125

His Lys Glu Lys Asn Lys Arg Ile Lys Ser Phe Glu Arg Val Asn

1130 1135 1140

Leu Val Asp Val Asn Asn Ile Lys Asp Glu Lys Ser Leu Val Lys

1145 1150 1155

Tyr Leu Ile Glu Asn Lys Lys Leu Val Glu Pro Arg Val Ile Lys

1160 1165 1170

Lys Val Tyr Lys Arg Gln Val Ile Leu Ile Asn Asp Tyr Pro Tyr

1175 1180 1185

Ser Ile Val Thr Leu Asp Ser Asn Lys Leu Met Asp Phe Glu Asn

1190 1195 1200

Leu Lys Pro Leu Phe Leu Glu Asn Lys Tyr Glu Lys Ile Leu Lys

1205 1210 1215

Asn Val Ile Lys Phe Leu Glu Asp Asn Gln Gly Lys Ser Glu Glu

1220 1225 1230

Asn Tyr Lys Phe Ile Tyr Leu Lys Lys Lys Asp Arg Tyr Glu Lys

1235 1240 1245

Asn Glu Thr Leu Glu Ser Val Lys Asp Arg Tyr Asn Leu Glu Phe

1250 1255 1260

Asn Glu Met Tyr Asp Lys Phe Leu Glu Lys Leu Asp Ser Lys Asp

1265 1270 1275

Tyr Lys Asn Tyr Met Asn Asn Lys Lys Tyr Gln Glu Leu Leu Asp

1280 1285 1290

Val Lys Glu Lys Phe Ile Lys Leu Asn Leu Phe Asp Lys Ala Phe

1295 1300 1305

Thr Leu Lys Ser Phe Leu Asp Leu Phe Asn Arg Lys Thr Met Ala

1310 1315 1320

Asp Phe Ser Lys Val Gly Leu Thr Lys Tyr Leu Gly Lys Ile Gln

1325 1330 1335

Lys Ile Ser Ser Asn Val Leu Ser Lys Asn Glu Leu Tyr Leu Leu

1340 1345 1350

Glu Glu Ser Val Thr Gly Leu Phe Val Lys Lys Ile Lys Leu

1355 1360 1365

<210> 33

<211> 1286

<212> PRT

<213> Enterococcus cecorum

<400> 33

Arg Arg Lys Gln Arg Ile Gln Ile Leu Gln Glu Leu Leu Gly Glu Glu

1 5 10 15

Val Leu Lys Thr Asp Pro Gly Phe Phe His Arg Met Lys Glu Ser Arg

20 25 30

Tyr Val Val Glu Asp Lys Arg Thr Leu Asp Gly Lys Gln Val Glu Leu

35 40 45

Pro Tyr Ala Leu Phe Val Asp Lys Asp Tyr Thr Asp Lys Glu Tyr Tyr

50 55 60

Lys Gln Phe Pro Thr Ile Asn His Leu Ile Val Tyr Leu Met Thr Thr

65 70 75 80

Ser Asp Thr Pro Asp Ile Arg Leu Val Tyr Leu Ala Leu His Tyr Tyr

85 90 95

Met Lys Asn Arg Gly Asn Phe Leu His Ser Gly Asp Ile Asn Asn Val

100 105 110

Lys Asp Ile Asn Asp Ile Leu Glu Gln Leu Asp Asn Val Leu Glu Thr

115 120 125

Phe Leu Asp Gly Trp Asn Leu Lys Leu Lys Ser Tyr Val Glu Asp Ile

130 135 140

Lys Asn Ile Tyr Asn Arg Asp Leu Gly Arg Gly Glu Arg Lys Lys Ala

145 150 155 160

Phe Val Asn Thr Leu Gly Ala Lys Thr Lys Ala Glu Lys Ala Phe Cys

165 170 175

Ser Leu Ile Ser Gly Gly Ser Thr Asn Leu Ala Glu Leu Phe Asp Asp

180 185 190

Ser Ser Leu Lys Glu Ile Glu Thr Pro Lys Ile Glu Phe Ala Ser Ser

195 200 205

Ser Leu Glu Asp Lys Ile Asp Gly Ile Gln Glu Ala Leu Glu Asp Arg

210 215 220

Phe Ala Val Ile Glu Ala Ala Lys Arg Leu Tyr Asp Trp Lys Thr Leu

225 230 235 240

Thr Asp Ile Leu Gly Asp Ser Ser Ser Leu Ala Glu Ala Arg Val Asn

245 250 255

Ser Tyr Gln Met His His Glu Gln Leu Leu Glu Leu Lys Ser Leu Val

260 265 270

Lys Glu Tyr Leu Asp Arg Lys Val Phe Gln Glu Val Phe Val Ser Leu

275 280 285

Asn Val Ala Asn Asn Tyr Pro Ala Tyr Ile Gly His Thr Lys Ile Asn

290 295 300

Gly Lys Lys Lys Glu Leu Glu Val Lys Arg Thr Lys Arg Asn Asp Phe

305 310 315 320

Tyr Ser Tyr Val Lys Lys Gln Val Ile Glu Pro Ile Lys Lys Lys Val

325 330 335

Ser Asp Glu Ala Val Leu Thr Lys Leu Ser Glu Ile Glu Ser Leu Ile

340 345 350

Glu Val Asp Lys Tyr Leu Pro Leu Gln Val Asn Ser Asp Asn Gly Val

355 360 365

Ile Pro Tyr Gln Val Lys Leu Asn Glu Leu Thr Arg Ile Phe Asp Asn

370 375 380

Leu Glu Asn Arg Ile Pro Val Leu Arg Glu Asn Arg Asp Lys Ile Ile

385 390 395 400

Lys Thr Phe Lys Phe Arg Ile Pro Tyr Tyr Val Gly Ser Leu Asn Gly

405 410 415

Val Val Lys Asn Gly Lys Cys Thr Asn Trp Met Val Arg Lys Glu Glu
420 425 430
Gly Lys Ile Tyr Pro Trp Asn Phe Glu Asp Lys Val Asp Leu Glu Ala

435 440 445
Ser Ala Glu Gln Phe Ile Arg Arg Met Thr Asn Lys Cys Thr Tyr Leu
450 455 460
Val Asn Glu Asp Val Leu Pro Lys Tyr Ser Leu Leu Tyr Ser Lys Tyr
465 470 475 480
Leu Val Leu Ser Glu Leu Asn Asn Leu Arg Ile Asp Gly Arg Pro Leu
485 490 495
Asp Val Lys Ile Lys Gln Asp Ile Tyr Glu Asn Val Phe Lys Lys Asn

500 505 510
Arg Lys Val Thr Leu Lys Lys Ile Lys Lys Tyr Leu Leu Lys Glu Gly
515 520 525
Ile Ile Thr Asp Asp Asp Glu Leu Ser Gly Leu Ala Asp Asp Val Lys
530 535 540
Ser Ser Leu Thr Ala Tyr Arg Asp Phe Lys Glu Lys Leu Gly His Leu
545 550 555 560
Asp Leu Ser Glu Ala Gln Met Glu Asn Ile Ile Leu Asn Ile Thr Leu

565 570 575
Phe Gly Asp Asp Lys Lys Leu Leu Lys Lys Arg Leu Ala Ala Leu Tyr
580 585 590
Pro Phe Ile Asp Asp Lys Ser Leu Asn Arg Ile Ala Thr Leu Asn Tyr
595 600 605
Arg Asp Trp Gly Arg Leu Ser Glu Arg Phe Leu Ser Gly Ile Thr Ser
610 615 620
Val Asp Gln Glu Thr Gly Glu Leu Arg Thr Ile Ile Gln Cys Met Tyr

625 630 635 640
Glu Thr Gln Ala Asn Leu Met Gln Leu Leu Ala Glu Pro Tyr His Phe
645 650 655
Val Glu Ala Ile Glu Lys Glu Asn Pro Lys Val Asp Leu Glu Ser Ile

660 665 670

Ser Tyr Arg Ile Val Asn Asp Leu Tyr Val Ser Pro Ala Val Lys Arg

675 680 685

Gln Ile Trp Gln Thr Leu Leu Val Ile Lys Asp Ile Lys Gln Val Met

690 695 700

Lys His Asp Pro Glu Arg Ile Phe Ile Glu Met Ala Arg Glu Lys Gln

705 710 715 720

Glu Ser Lys Lys Thr Lys Ser Arg Lys Gln Val Leu Ser Glu Val Tyr

725 730 735

Lys Lys Ala Lys Glu Tyr Glu His Leu Phe Glu Lys Leu Asn Ser Leu

740 745 750

Thr Glu Glu Gln Leu Arg Ser Lys Lys Ile Tyr Leu Tyr Phe Thr Gln

755 760 765

Leu Gly Lys Cys Met Tyr Ser Gly Glu Pro Ile Asp Phe Glu Asn Leu

770 775 780

Val Ser Ala Asn Ser Asn Tyr Asp Ile Asp His Ile Tyr Pro Gln Ser

785 790 795 800

Lys Thr Ile Asp Asp Ser Phe Asn Asn Ile Val Leu Val Lys Lys Ser

805 810 815

Leu Asn Ala Tyr Lys Ser Asn His Tyr Pro Ile Asp Lys Asn Ile Arg

820 825 830

Asp Asn Glu Lys Val Lys Thr Leu Trp Asn Thr Leu Val Ser Lys Gly

835 840 845

Leu Ile Thr Lys Glu Lys Tyr Glu Arg Leu Ile Arg Ser Thr Pro Phe

850 855 860

Ser Asp Glu Glu Leu Ala Gly Phe Ile Ala Arg Gln Leu Val Glu Thr

865 870 875 880

Arg Gln Ser Thr Lys Ala Val Ala Glu Ile Leu Ser Asn Trp Phe Pro

885 890 895

Glu Ser Glu Ile Val Tyr Ser Lys Ala Lys Asn Val Ser Asn Phe Arg

900 905 910

Gln Asp Phe Glu Ile Leu Lys Val Arg Glu Leu Asn Asp Cys His His

915 920 925

Ala His Asp Ala Tyr Leu Asn Ile Val Val Gly Asn Ala Tyr His Thr

930 935 940

Lys Phe Thr Asn Ser Pro Tyr Arg Phe Ile Lys Asn Lys Ala Asn Gln

945 950 955 960

Glu Tyr Asn Leu Arg Lys Leu Leu Gln Lys Val Asn Lys Ile Glu Ser

965 970 975

Asn Gly Val Val Ala Trp Val Gly Gln Ser Glu Asn Asn Pro Gly Thr

980 985 990

Ile Ala Thr Val Lys Lys Val Ile Arg Arg Asn Thr Val Leu Ile Ser

995 1000 1005

Arg Met Val Lys Glu Val Asp Gly Gln Leu Phe Asp Leu Thr Leu

1010 1015 1020

Met Lys Lys Gly Lys Gly Gln Val Pro Ile Lys Ser Ser Asp Glu

1025 1030 1035

Arg Leu Thr Asp Ile Ser Lys Tyr Gly Gly Tyr Asn Lys Ala Thr

1040 1045 1050

Gly Ala Tyr Phe Thr Phe Val Lys Ser Lys Lys Arg Gly Lys Val

1055 1060 1065

Val Arg Ser Phe Glu Tyr Val Pro Leu His Leu Ser Lys Gln Phe

1070 1075 1080

Glu Asn Asn Asn Glu Leu Leu Lys Glu Tyr Ile Glu Lys Asp Arg

1085 1090 1095

Gly Leu Thr Asp Val Glu Ile Leu Ile Pro Lys Val Leu Ile Asn

1100 1105 1110

Ser Leu Phe Arg Tyr Asn Gly Ser Leu Val Arg Ile Thr Gly Arg

1115 1120 1125

Gly Asp Thr Arg Leu Leu Leu Val His Glu Gln Pro Leu Tyr Val

1130 1135 1140

Ser Asn Ser Phe Val Gln Gln Leu Lys Ser Val Ser Ser Tyr Lys

1145 1150 1155

Leu Lys Lys Ser Glu Asn Asp Asn Ala Lys Leu Thr Lys Thr Ala

1160 1165 1170

Thr Glu Lys Leu Ser Asn Ile Asp Glu Leu Tyr Asp Gly Leu Leu

1175 1180 1185

Arg Lys Leu Asp Leu Pro Ile Tyr Ser Tyr Trp Phe Ser Ser Ile

1190 1195 1200

Lys Glu Tyr Leu Val Glu Ser Arg Thr Lys Tyr Ile Lys Leu Ser

1205 1210 1215

Ile Glu Glu Lys Ala Leu Val Ile Phe Glu Ile Leu His Leu Phe

1220 1225 1230

Gln Ser Asp Ala Gln Val Pro Asn Leu Lys Ile Leu Gly Leu Ser

1235 1240 1245

Thr Lys Pro Ser Arg Ile Arg Ile Gln Lys Asn Leu Lys Asp Thr

1250 1255 1260

Asp Lys Met Ser Ile Ile His Gln Ser Pro Ser Gly Ile Phe Glu

1265 1270 1275

His Glu Ile Glu Leu Thr Ser Leu

1280 1285

<210> 34

<211> 1312

<212> PRT

<213> Anaerostipes hadrus

<400> 34

Met Gln Asn Gly Phe Leu Gly Ile Thr Val Ser Ser Glu Gln Val Gly

1 5 10 15

Trp Ala Val Thr Asn Pro Lys Tyr Glu Leu Glu Arg Ala Ser Arg Lys

20 25 30

Asp Leu Trp Gly Val Arg Leu Phe Asp Lys Ala Glu Thr Ala Glu Asp

35 40 45

Arg Arg Met Phe Arg Thr Asn Arg Arg Leu Asn Gln Arg Lys Lys Asn

50 55 60

Arg Ile His Tyr Leu Arg Asp Ile Phe His Glu Glu Val Asn Gln Lys
65 70 75 80
Asp Pro Asn Phe Phe Gln Gln Leu Asp Glu Ser Asn Phe Cys Glu Asp

85 90 95
Asp Arg Thr Val Glu Phe Asn Phe Asp Thr Asn Leu Tyr Lys Asn Gln
100 105 110
Phe Pro Thr Val Tyr His Leu Arg Lys Tyr Leu Met Glu Thr Lys Asp
115 120 125
Lys Pro Asp Ile Arg Leu Val Tyr Leu Ala Phe Ser Lys Phe Met Lys
130 135 140
Asn Arg Gly His Phe Leu Tyr Lys Gly Asn Leu Gly Glu Val Met Asp

145 150 155 160
Phe Glu Asn Ser Met Lys Gly Phe Cys Glu Ser Leu Glu Lys Phe Asn
165 170 175
Ile Asp Phe Pro Thr Leu Ser Asp Glu Gln Val Lys Glu Val Arg Asp
180 185 190
Ile Leu Cys Asp His Lys Ile Ala Lys Thr Val Lys Lys Lys Asn Ile
195 200 205
Ile Thr Ile Thr Lys Val Lys Ser Lys Thr Ala Lys Ala Trp Ile Gly

210 215 220
Leu Phe Cys Gly Cys Ser Val Pro Val Lys Val Leu Phe Gln Asp Ile
225 230 235 240
Asp Glu Glu Ile Val Thr Asp Pro Glu Lys Ile Ser Phe Glu Asp Ala
245 250 255
Ser Tyr Asp Asp Tyr Ile Ala Asn Ile Glu Lys Gly Val Gly Ile Tyr
260 265 270
Tyr Glu Ala Ile Val Ser Ala Lys Met Leu Phe Asp Trp Ser Ile Leu

275 280 285
Asn Glu Ile Leu Gly Asp His Gln Leu Leu Ser Asp Ala Met Ile Ala
290 295 300
Glu Tyr Asn Lys His His Asp Asp Leu Lys Arg Leu Gln Lys Ile Ile

305 310 315 320

Lys Gly Thr Gly Ser Arg Glu Leu Tyr Gln Asp Ile Phe Ile Asn Asp

325 330 335

Val Ser Gly Asn Tyr Val Cys Tyr Val Gly His Ala Lys Thr Met Ser

340 345 350

Ser Ala Asp Gln Lys Gln Phe Tyr Thr Phe Leu Lys Asn Arg Leu Lys

355 360 365

Asn Val Asn Gly Ile Ser Ser Glu Asp Ala Glu Trp Ile Asp Thr Glu

370 375 380

Ile Lys Asn Gly Thr Leu Leu Pro Lys Gln Thr Lys Arg Asp Asn Ser

385 390 395 400

Val Ile Pro His Gln Leu Gln Leu Arg Glu Phe Glu Leu Ile Leu Asp

405 410 415

Asn Met Gln Glu Met Tyr Pro Phe Leu Lys Glu Asn Arg Glu Lys Leu

420 425 430

Leu Lys Ile Phe Asn Phe Val Ile Pro Tyr Tyr Val Gly Pro Leu Lys

435 440 445

Gly Val Val Arg Lys Gly Glu Ser Thr Asn Trp Met Val Pro Lys Lys

450 455 460

Asp Gly Val Ile His Pro Trp Asn Phe Asp Glu Met Val Asp Lys Glu

465 470 475 480

Ala Ser Ala Glu Cys Phe Ile Ser Arg Met Thr Gly Asn Cys Ser Tyr

485 490 495

Leu Phe Asn Glu Lys Val Leu Pro Lys Asn Ser Leu Leu Tyr Glu Thr

500 505 510

Phe Glu Val Leu Asn Glu Leu Asn Pro Leu Lys Ile Asn Gly Glu Pro

515 520 525

Ile Ser Val Glu Leu Lys Gln Arg Ile Tyr Glu Gln Leu Phe Leu Thr

530 535 540

Gly Lys Lys Val Thr Lys Lys Ser Leu Thr Lys Tyr Leu Ile Lys Asn

545 550 555 560

Gly Tyr Asp Lys Asp Ile Glu Leu Ser Gly Ile Asp Asn Glu Phe His
565 570 575
Ser Asn Leu Lys Ser His Ile Asp Phe Glu Asp Tyr Asp Asn Leu Ser
580 585 590
Asp Glu Glu Val Glu Gln Ile Ile Leu Arg Ile Thr Val Phe Glu Asp
595 600 605
Lys Gln Leu Leu Lys Asp Tyr Leu Asn Arg Glu Phe Val Lys Leu Ser
610 615 620
Glu Asp Glu Arg Lys Gln Ile Cys Ser Leu Ser Tyr Lys Gly Trp Gly
625 630 635 640
Asn Leu Ser Glu Met Leu Leu Asn Gly Ile Thr Val Thr Asp Ser Asn
645 650 655
Gly Val Glu Val Ser Val Met Asp Met Leu Trp Asn Thr Asn Leu Asn
660 665 670
Leu Met Gln Ile Leu Ser Lys Lys Tyr Gly Tyr Lys Ala Glu Ile Glu
675 680 685
His Tyr Asn Lys Glu His Glu Lys Thr Ile Tyr Asn Arg Glu Asp Leu
690 695 700
Met Asp Tyr Leu Asn Ile Pro Pro Ala Gln Arg Arg Lys Val Asn Gln
705 710 715 720
Leu Ile Thr Ile Val Lys Ser Leu Lys Lys Thr Tyr Gly Val Pro Asn
725 730 735
Lys Ile Phe Phe Lys Ile Ser Arg Glu His Gln Asp Asp Pro Lys Arg
740 745 750
Thr Ser Ser Arg Lys Glu Gln Leu Lys Tyr Leu Tyr Lys Ser Leu Lys
755 760 765
Ser Glu Asp Glu Lys His Leu Met Lys Glu Leu Asp Glu Leu Asn Asp
770 775 780
His Glu Leu Ser Asn Asp Lys Val Tyr Leu Tyr Phe Leu Gln Lys Gly
785 790 795 800
Arg Cys Ile Tyr Ser Gly Lys Lys Leu Asn Leu Ser Arg Leu Arg Lys

805 810 815

Ser Asn Tyr Gln Asn Asp Ile Asp Tyr Ile Tyr Pro Leu Ser Ala Val

820 825 830

Asn Asp Arg Ser Met Asn Asn Lys Val Leu Thr Gly Ile Gln Glu Asn

835 840 845

Arg Ala Asp Lys Tyr Thr Tyr Phe Pro Val Asp Ser Glu Ile Gln Lys

850 855 860

Lys Met Lys Gly Phe Trp Met Glu Leu Val Leu Gln Gly Phe Met Thr

865 870 875 880

Lys Glu Lys Tyr Phe Arg Leu Ser Arg Glu Asn Asp Phe Ser Lys Ser

885 890 895

Glu Leu Val Ser Phe Ile Glu Arg Glu Ile Ser Asp Asn Gln Gln Ser

900 905 910

Gly Arg Met Ile Ala Ser Val Leu Gln Tyr Tyr Phe Pro Glu Ser Lys

915 920 925

Ile Val Phe Val Lys Glu Lys Leu Ile Ser Ser Phe Lys Arg Asp Phe

930 935 940

His Leu Ile Ser Ser Tyr Gly His Asn His Leu Gln Ala Ala Lys Asp

945 950 955 960

Ala Tyr Ile Thr Ile Val Val Gly Asn Val Tyr His Thr Lys Phe Thr

965 970 975

Met Asp Pro Ala Ile Tyr Phe Lys Asn His Lys Arg Lys Asp Tyr Asp

980 985 990

Leu Asn Arg Leu Phe Leu Glu Asn Ile Ser Arg Asp Gly Gln Ile Ala

995 1000 1005

Trp Glu Ser Gly Pro Tyr Gly Ser Ile Gln Thr Val Arg Lys Glu

1010 1015 1020

Tyr Ala Gln Asn His Ile Ala Val Thr Lys Arg Val Val Glu Val

1025 1030 1035

Lys Gly Gly Leu Phe Lys Gln Met Pro Leu Lys Lys Gly His Gly

1040 1045 1050

Glu Tyr Pro Leu Lys Thr Asn Asp Pro Arg Phe Gly Asn Ile Ala

1055 1060 1065

Gln Tyr Gly Gly Tyr Thr Asn Val Thr Gly Ser Tyr Phe Val Leu

1070 1075 1080

Val Glu Ser Met Glu Lys Gly Lys Lys Arg Ile Ser Leu Glu Tyr

1085 1090 1095

Val Pro Val Tyr Leu His Glu Arg Leu Glu Asp Asp Pro Gly His

1100 1105 1110

Lys Leu Leu Lys Glu Tyr Leu Val Asp His Arg Lys Leu Asn His

1115 1120 1125

Pro Lys Ile Leu Leu Ala Lys Val Arg Lys Asn Ser Leu Leu Lys

1130 1135 1140

Ile Asp Gly Phe Tyr Tyr Arg Leu Asn Gly Arg Ser Gly Asn Ala

1145 1150 1155

Leu Ile Leu Thr Asn Ala Val Glu Leu Ile Met Asp Asp Trp Gln

1160 1165 1170

Thr Lys Thr Ala Asn Lys Ile Ser Gly Tyr Met Lys Arg Arg Ala

1175 1180 1185

Ile Asp Lys Lys Ala Arg Val Tyr Gln Asn Glu Phe His Ile Gln

1190 1195 1200

Glu Leu Glu Gln Leu Tyr Asp Phe Tyr Leu Asp Lys Leu Lys Asn

1205 1210 1215

Gly Val Tyr Lys Asn Arg Lys Asn Asn Gln Ala Glu Leu Ile His

1220 1225 1230

Asn Glu Lys Glu Gln Phe Met Glu Leu Lys Thr Glu Asp Gln Cys

1235 1240 1245

Val Leu Leu Thr Glu Ile Lys Lys Leu Phe Val Cys Ser Pro Met

1250 1255 1260

Gln Ala Asp Leu Thr Leu Ile Gly Gly Ser Lys His Thr Gly Met

1265 1270 1275

Ile Ala Met Ser Ser Asn Val Thr Lys Ala Asp Phe Ala Val Ile

1280 1285 1290

Ala Glu Asp Pro Leu Gly Leu Arg Asn Lys Val Ile Tyr Ser His

1295 1300 1305

Lys Gly Glu Lys

1310

<210> 35

<211> 1344

<212> PRT

<213> Kandleria vitulina

<400> 35

Met Ser Gln Asn Asn Asn Lys Ile Tyr Asn Ile Gly Leu Asp Ile Gly

1 5 10 15

Asp Ala Ser Val Gly Trp Ala Val Val Asp Glu His Tyr Asn Leu Leu

20 25 30

Lys Arg His Gly Lys His Met Trp Gly Ser Arg Leu Phe Thr Gln Ala

35 40 45

Asn Thr Ala Val Glu Arg Arg Ser Ser Arg Ser Thr Arg Arg Arg Tyr

50 55 60

Asn Lys Arg Arg Glu Arg Ile Arg Leu Leu Arg Glu Ile Met Glu Asp

65 70 75 80

Met Val Leu Asp Val Asp Pro Thr Phe Phe Ile Arg Leu Ala Asn Val

85 90 95

Ser Phe Leu Asp Gln Glu Asp Lys Lys Asp Tyr Leu Lys Glu Asn Tyr

100 105 110

His Ser Asn Tyr Asn Leu Phe Ile Asp Lys Asp Phe Asn Asp Lys Thr

115 120 125

Tyr Tyr Asp Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys His Leu Cys

130 135 140

Glu Ser Lys Glu Lys Glu Asp Pro Arg Leu Ile Tyr Leu Ala Leu His

145 150 155 160

His Ile Val Lys Tyr Arg Gly Asn Phe Leu Tyr Glu Gly Gln Lys Phe

165 170 175

Ser Met Asp Val Ser Asn Ile Glu Asp Lys Met Ile Asp Val Leu Arg
180 185 190
Gln Phe Asn Glu Ile Asn Leu Phe Glu Tyr Val Glu Asp Arg Lys Lys
195 200 205
Ile Asp Glu Val Leu Asn Val Leu Lys Glu Pro Leu Ser Lys Lys His

210 215 220
Lys Ala Glu Lys Ala Phe Ala Leu Phe Asp Thr Thr Lys Asp Asn Lys
225 230 235 240
Ala Ala Tyr Lys Glu Leu Cys Ala Ala Leu Ala Gly Asn Lys Phe Asn
245 250 255
Val Thr Lys Met Leu Lys Glu Ala Glu Leu His Asp Glu Asp Glu Lys
260 265 270
Asp Ile Ser Phe Lys Phe Ser Asp Ala Thr Phe Asp Asp Ala Phe Val

275 280 285
Glu Lys Gln Pro Leu Leu Gly Asp Cys Val Glu Phe Ile Asp Leu Leu
290 295 300
His Asp Ile Tyr Ser Trp Val Glu Leu Gln Asn Ile Leu Gly Ser Ala
305 310 315 320
His Thr Ser Glu Pro Ser Ile Ser Ala Ala Met Ile Gln Arg Tyr Glu
325 330 335
Asp His Lys Asn Asp Leu Lys Leu Leu Lys Asp Val Ile Arg Lys Tyr

340 345 350
Leu Pro Lys Lys Tyr Phe Glu Val Phe Arg Asp Glu Lys Ser Lys Lys
355 360 365
Asn Asn Tyr Cys Asn Tyr Ile Asn His Pro Ser Lys Thr Pro Val Asp
370 375 380
Glu Phe Tyr Lys Tyr Ile Lys Lys Leu Ile Glu Lys Ile Asp Asp Pro
385 390 395 400
Asp Val Lys Thr Ile Leu Asn Lys Ile Glu Leu Glu Ser Phe Met Leu

405 410 415
Lys Gln Asn Ser Arg Thr Asn Gly Ala Val Pro Tyr Gln Met Gln Leu

420 425 430

Asp Glu Leu Asn Lys Ile Leu Glu Asn Gln Ser Val Tyr Tyr Ser Asp

435 440 445

Leu Lys Asp Asn Glu Asp Lys Ile Arg Ser Ile Leu Thr Phe Arg Ile

450 455 460

Pro Tyr Tyr Phe Gly Pro Leu Asn Ile Thr Lys Asp Arg Gln Phe Asp

465 470 475 480

Trp Ile Ile Lys Lys Glu Gly Lys Glu Asn Glu Arg Ile Leu Pro Trp

485 490 495

Asn Ala Asn Glu Ile Val Asp Val Asp Lys Thr Ala Asp Glu Phe Ile

500 505 510

Lys Arg Met Arg Asn Phe Cys Thr Tyr Phe Pro Asp Glu Pro Val Met

515 520 525

Ala Lys Asn Ser Leu Thr Val Ser Lys Tyr Glu Val Leu Asn Glu Ile

530 535 540

Asn Lys Leu Arg Ile Asn Asp His Leu Ile Lys Arg Asp Met Lys Asp

545 550 555 560

Lys Met Leu His Thr Leu Phe Met Asp His Lys Ser Ile Ser Ala Asn

565 570 575

Ala Met Lys Lys Trp Leu Val Lys Asn Gln Tyr Phe Ser Asn Thr Asp

580 585 590

Asp Ile Lys Ile Glu Gly Phe Gln Lys Glu Asn Ala Cys Ser Thr Ser

595 600 605

Leu Thr Pro Trp Ile Asp Phe Thr Lys Ile Phe Gly Lys Ile Asn Glu

610 615 620

Ser Asn Tyr Asp Phe Ile Glu Lys Ile Ile Tyr Asp Val Thr Val Phe

625 630 635 640

Glu Asp Lys Lys Ile Leu Arg Arg Arg Leu Lys Lys Glu Tyr Asp Leu

645 650 655

Asp Glu Glu Lys Ile Lys Lys Ile Leu Lys Leu Lys Tyr Ser Gly Trp

660 665 670

Ser Arg Leu Ser Lys Lys Leu Leu Ser Gly Ile Lys Thr Lys Tyr Lys
675 680 685
Asp Ser Thr Arg Thr Pro Glu Thr Val Leu Glu Val Met Glu Arg Thr
690 695 700
Asn Met Asn Leu Met Gln Val Ile Asn Asp Glu Lys Leu Gly Phe Lys
705 710 715 720
Lys Thr Ile Asp Asp Ala Asn Ser Thr Ser Val Ser Gly Lys Phe Ser

725 730 735
Tyr Ala Glu Val Gln Glu Leu Ala Gly Ser Pro Ala Ile Lys Arg Gly
740 745 750
Ile Trp Gln Ala Leu Leu Ile Val Asp Glu Ile Lys Lys Ile Met Lys
755 760 765
His Glu Pro Ala His Val Tyr Ile Glu Phe Ala Arg Asn Glu Asp Glu
770 775 780
Lys Glu Arg Lys Asp Ser Phe Val Asn Gln Met Leu Lys Leu Tyr Lys

785 790 795 800
Asp Tyr Asp Phe Glu Asp Glu Thr Glu Lys Glu Ala Asn Lys His Leu
805 810 815
Lys Gly Glu Asp Ala Lys Ser Lys Ile Arg Ser Glu Arg Leu Lys Leu
820 825 830
Tyr Tyr Thr Gln Met Gly Lys Cys Met Tyr Thr Gly Lys Ser Leu Asp
835 840 845
Ile Asp Arg Leu Asp Thr Tyr Gln Val Asp His Ile Val Pro Gln Ser

850 855 860
Leu Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Val Leu Ser Ser
865 870 875 880
Glu Asn Gln Arg Lys Leu Asp Asp Leu Val Ile Pro Ser Ser Ile Arg
885 890 895
Asn Lys Met Tyr Gly Phe Trp Glu Lys Leu Phe Asn Asn Lys Ile Ile
900 905 910
Ser Pro Lys Lys Phe Tyr Ser Leu Ile Lys Thr Glu Phe Asn Glu Lys

915 920 925

Asp Gln Glu Arg Phe Ile Asn Arg Gln Ile Val Glu Thr Arg Gln Ile

930 935 940

Thr Lys His Val Ala Gln Ile Ile Asp Asn His Tyr Glu Asn Thr Lys

945 950 955 960

Val Val Thr Val Arg Ala Asp Leu Ser His Gln Phe Arg Glu Arg Tyr

965 970 975

His Ile Tyr Lys Asn Arg Asp Ile Asn Asp Phe His His Ala His Asp

980 985 990

Ala Tyr Ile Ala Thr Ile Leu Gly Thr Tyr Ile Gly His Arg Phe Glu

995 1000 1005

Ser Leu Asp Ala Lys Tyr Ile Tyr Gly Glu Tyr Lys Arg Ile Phe

1010 1015 1020

Arg Asn Gln Lys Asn Lys Gly Lys Glu Met Lys Lys Asn Asn Asp

1025 1030 1035

Gly Phe Ile Leu Asn Ser Met Arg Asn Ile Tyr Ala Asp Lys Asp

1040 1045 1050

Thr Gly Glu Ile Val Trp Asp Pro Asn Tyr Ile Asp Arg Ile Lys

1055 1060 1065

Lys Cys Phe Tyr Tyr Lys Asp Cys Phe Val Thr Lys Lys Leu Glu

1070 1075 1080

Glu Asn Asn Gly Thr Phe Phe Asn Val Thr Val Leu Pro Asn Asp

1085 1090 1095

Thr Asn Ser Asp Lys Asp Asn Thr Leu Ala Thr Val Pro Val Asn

1100 1105 1110

Lys Tyr Arg Ser Asn Val Asn Lys Tyr Gly Gly Phe Ser Gly Val

1115 1120 1125

Asn Ser Phe Ile Val Ala Ile Lys Gly Lys Lys Lys Lys Gly Lys

1130 1135 1140

Lys Val Ile Glu Val Asn Lys Leu Thr Gly Ile Pro Leu Met Tyr

1145 1150 1155

Lys Asn Ala Asp Glu Glu Ile Lys Ile Asn Tyr Leu Lys Gln Ala

1160 1165 1170

Glu Asp Leu Glu Glu Val Gln Ile Gly Lys Glu Ile Leu Lys Asn

1175 1180 1185

Gln Leu Ile Glu Lys Asp Gly Gly Leu Tyr Tyr Ile Val Ala Pro

1190 1195 1200

Thr Glu Ile Ile Asn Ala Lys Gln Leu Ile Leu Asn Glu Ser Gln

1205 1210 1215

Thr Lys Leu Val Cys Glu Ile Tyr Lys Ala Met Lys Tyr Lys Asn

1220 1225 1230

Tyr Asp Asn Leu Asp Ser Glu Lys Ile Ile Asp Leu Tyr Arg Leu

1235 1240 1245

Leu Ile Asn Lys Met Glu Leu Tyr Tyr Pro Glu Tyr Arg Lys Gln

1250 1255 1260

Leu Val Lys Lys Phe Glu Asp Arg Tyr Glu Gln Leu Lys Val Ile

1265 1270 1275

Ser Ile Glu Glu Lys Cys Asn Ile Ile Lys Gln Ile Leu Ala Thr

1280 1285 1290

Leu His Cys Asn Ser Ser Ile Gly Lys Ile Met Tyr Ser Asp Phe

1295 1300 1305

Lys Ile Ser Thr Thr Ile Gly Arg Leu Asn Gly Arg Thr Ile Ser

1310 1315 1320

Leu Asp Asp Ile Ser Phe Ile Ala Glu Ser Pro Thr Gly Met Tyr

1325 1330 1335

Ser Lys Lys Tyr Lys Leu

1340

<210> 36

<211> 1288

<212> PRT

<213> Enterococcus faecalis

<400> 36

Met Arg Leu Phe Glu Glu Gly His Thr Ala Glu Asp Arg Arg Leu Lys
1 5 10 15
Arg Thr Ala Arg Arg Arg Ile Ser Arg Arg Arg Asn Arg Leu Arg Tyr
20 25 30
Leu Gln Ala Phe Phe Glu Glu Ala Met Thr Asp Leu Asp Glu Asn Phe
35 40 45
Phe Ala Arg Leu Gln Glu Ser Phe Leu Val Pro Glu Asp Lys Lys Trp

50 55 60
His Arg His Pro Ile Phe Ala Lys Leu Glu Asp Glu Val Ala Tyr His
65 70 75 80
Glu Thr Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp Ser
85 90 95
Ser Glu Gln Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Ile
100 105 110
Val Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Lys Leu Ser Thr Glu

115 120 125
Asn Thr Ser Val Lys Asp Gln Phe Gln Gln Phe Met Val Ile Tyr Asn
130 135 140
Gln Thr Phe Val Asn Gly Glu Ser Arg Leu Val Ser Ala Pro Leu Pro
145 150 155 160
Glu Ser Val Leu Ile Glu Glu Glu Leu Thr Glu Lys Ala Ser Arg Thr
165 170 175
Lys Lys Ser Glu Lys Val Leu Gln Gln Phe Pro Gln Glu Lys Ala Asn

180 185 190
Gly Leu Phe Gly Gln Phe Leu Lys Leu Met Val Gly Asn Lys Ala Asp
195 200 205
Phe Lys Lys Val Phe Gly Leu Glu Glu Glu Ala Lys Ile Thr Tyr Ala
210 215 220
Ser Glu Ser Tyr Glu Glu Asp Leu Glu Gly Ile Leu Ala Lys Val Gly
225 230 235 240
Asp Glu Tyr Ser Asp Val Phe Leu Ala Ala Lys Asn Val Tyr Asp Ala

245 250 255

Val Glu Leu Ser Thr Ile Leu Ala Asp Ser Asp Lys Lys Ser His Ala

260 265 270

Lys Leu Ser Ser Ser Met Ile Val Arg Phe Thr Glu His Gln Glu Asp

275 280 285

Leu Lys Lys Phe Lys Arg Phe Ile Arg Glu Asn Cys Pro Asp Glu Tyr

290 295 300

Asp Asn Leu Phe Lys Asn Glu Gln Lys Asp Gly Tyr Ala Gly Tyr Ile

305 310 315 320

Ala His Ala Gly Lys Val Ser Gln Leu Lys Phe Tyr Gln Tyr Val Lys

325 330 335

Lys Ile Ile Gln Asp Ile Ala Gly Ala Glu Tyr Phe Leu Glu Lys Ile

340 345 350

Ala Gln Glu Asn Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Val

355 360 365

Ile Pro His Gln Ile His Leu Ala Glu Leu Gln Ala Ile Ile His Arg

370 375 380

Gln Ala Ala Tyr Tyr Pro Phe Leu Lys Glu Asn Gln Glu Lys Ile Glu

385 390 395 400

Gln Leu Val Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ser Lys

405 410 415

Gly Asp Ala Ser Thr Phe Ala Trp Leu Lys Arg Gln Ser Glu Glu Pro

420 425 430

Ile Arg Pro Trp Asn Leu Gln Glu Thr Val Asp Leu Asp Gln Ser Ala

435 440 445

Thr Ala Phe Ile Glu Arg Met Thr Asn Phe Asp Thr Tyr Leu Pro Ser

450 455 460

Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Lys Phe Met Val

465 470 475 480

Phe Asn Glu Leu Thr Lys Ile Ser Tyr Thr Asp Asp Arg Gly Ile Lys

485 490 495

Ala Asn Phe Ser Gly Lys Glu Lys Glu Lys Ile Phe Asp Tyr Leu Phe

500 505 510

Lys Thr Arg Arg Lys Val Lys Lys Lys Asp Ile Ile Gln Phe Tyr Arg

515 520 525

Asn Glu Tyr Asn Thr Glu Ile Val Thr Leu Ser Gly Leu Glu Glu Asp

530 535 540

Gln Phe Asn Ala Ser Phe Ser Thr Tyr Gln Asp Leu Leu Lys Cys Gly

545 550 555 560

Leu Thr Arg Ala Glu Leu Asp His Pro Asp Asn Ala Glu Lys Leu Glu

565 570 575

Asp Ile Ile Lys Ile Leu Thr Ile Phe Glu Asp Arg Gln Arg Ile Arg

580 585 590

Thr Gln Leu Ser Thr Phe Lys Gly Gln Phe Ser Ala Glu Val Leu Lys

595 600 605

Lys Leu Glu Arg Lys His Tyr Thr Gly Trp Gly Arg Leu Ser Lys Lys

610 615 620

Leu Ile Asn Gly Ile Tyr Asp Lys Glu Ser Gly Lys Thr Ile Leu Asp

625 630 635 640

Tyr Leu Val Lys Asp Asp Gly Val Ser Lys His Tyr Asn Arg Asn Phe

645 650 655

Met Gln Leu Ile Asn Asp Ser Gln Leu Ser Phe Lys Asn Ala Ile Gln

660 665 670

Lys Ala Gln Ser Ser Glu His Glu Glu Thr Leu Ser Glu Thr Val Asn

675 680 685

Glu Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Tyr Gln Ser Leu

690 695 700

Lys Ile Val Asp Glu Leu Val Ala Ile Met Gly Tyr Ala Pro Lys Arg

705 710 715 720

Ile Val Val Glu Met Ala Arg Glu Asn Gln Thr Thr Ser Thr Gly Lys

725 730 735

Arg Arg Ser Ile Gln Arg Leu Lys Ile Val Glu Lys Ala Met Ala Glu

740 745 750

Ile Gly Ser Asn Leu Leu Lys Glu Gln Pro Thr Thr Asn Glu Gln Leu

755 760 765

Arg Asp Thr Arg Leu Phe Leu Tyr Tyr Met Gln Asn Gly Lys Asp Met

770 775 780

Tyr Thr Gly Asp Glu Leu Ser Leu His Arg Leu Ser His Tyr Asp Ile

785 790 795 800

Asp His Ile Ile Pro Gln Ser Phe Met Lys Asp Asp Ser Leu Asp Asn

805 810 815

Leu Val Leu Val Gly Ser Thr Glu Asn Arg Gly Lys Ser Asp Asp Val

820 825 830

Pro Ser Lys Glu Val Val Lys Asp Met Lys Ala Tyr Trp Glu Lys Leu

835 840 845

Tyr Ala Ala Gly Leu Ile Ser Gln Arg Lys Phe Gln Arg Leu Thr Lys

850 855 860

Gly Glu Gln Gly Gly Leu Thr Leu Glu Asp Lys Ala His Phe Ile Gln

865 870 875 880

Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys Asn Val Ala Gly Ile

885 890 895

Leu Asp Gln Arg Tyr Asn Ala Lys Ser Lys Glu Lys Lys Val Gln Ile

900 905 910

Ile Thr Leu Lys Ala Ser Leu Thr Ser Gln Phe Arg Ser Ile Phe Gly

915 920 925

Leu Tyr Lys Val Arg Glu Val Asn Asp Tyr His His Gly Gln Asp Ala

930 935 940

Tyr Leu Asn Cys Val Val Ala Thr Thr Leu Leu Lys Val Tyr Pro Asn

945 950 955 960

Leu Ala Pro Glu Phe Val Tyr Gly Glu Tyr Pro Lys Phe Gln Thr Phe

965 970 975

Lys Glu Asn Lys Ala Thr Ala Lys Ala Ile Ile Tyr Thr Asn Leu Leu

980 985 990

Arg Phe Phe Thr Glu Asp Glu Pro Arg Phe Thr Lys Asp Gly Glu Ile

995 1000 1005

Leu Trp Ser Asn Ser Tyr Leu Lys Thr Ile Lys Lys Glu Leu Asn

1010 1015 1020

Tyr His Gln Met Asn Ile Val Lys Lys Val Glu Val Gln Lys Gly

1025 1030 1035

Gly Phe Ser Lys Glu Ser Ile Lys Pro Lys Gly Pro Ser Asn Lys

1040 1045 1050

Leu Ile Pro Val Lys Asn Gly Leu Asp Pro Gln Lys Tyr Gly Gly

1055 1060 1065

Phe Asp Ser Pro Val Val Ala Tyr Thr Val Leu Phe Thr His Glu

1070 1075 1080

Lys Gly Lys Lys Pro Leu Ile Lys Gln Glu Ile Leu Gly Ile Thr

1085 1090 1095

Ile Met Glu Lys Thr Arg Phe Glu Gln Asn Pro Ile Leu Phe Leu

1100 1105 1110

Glu Glu Lys Gly Phe Leu Arg Pro Arg Val Leu Met Lys Leu Pro

1115 1120 1125

Lys Tyr Thr Leu Tyr Glu Phe Pro Glu Gly Arg Arg Arg Leu Leu

1130 1135 1140

Ala Ser Ala Lys Glu Ala Gln Lys Gly Asn Gln Met Val Leu Pro

1145 1150 1155

Glu His Leu Leu Thr Leu Leu Tyr His Ala Lys Gln Cys Leu Leu

1160 1165 1170

Pro Asn Gln Ser Glu Ser Leu Ala Tyr Val Glu Gln His Gln Pro

1175 1180 1185

Glu Phe Gln Glu Ile Leu Glu Arg Val Val Asp Phe Ala Glu Val

1190 1195 1200

His Thr Leu Ala Lys Ser Lys Val Gln Gln Ile Val Lys Leu Phe

1205 1210 1215

Glu Ala Asn Gln Thr Ala Asp Val Lys Glu Ile Ala Ala Ser Phe

1220 1225 1230

Ile Gln Leu Met Gln Phe Asn Ala Met Gly Ala Pro Ser Thr Phe

1235 1240 1245

Lys Phe Phe Gln Lys Asp Ile Glu Arg Ala Arg Tyr Thr Ser Ile

1250 1255 1260

Lys Glu Ile Phe Asp Ala Thr Ile Ile Tyr Gln Ser Pro Thr Gly

1265 1270 1275

Leu Tyr Glu Thr Arg Arg Lys Val Val Asp

1280 1285

<210> 37

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 37

Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly

1 5 10 15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val

20 25 30

Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser

35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln

50 55 60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser

65 70 75 80

Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser

85 90 95

Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala

100 105 110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly

115 120 125

Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu

130 135 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp

145 150 155 160

Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr Val

165 170 175

Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu

180 185 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg

195 200 205

Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp

210 215 220

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro

225 230 235 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn

245 250 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu

260 265 270

Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys

275 280 285

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val

290 295 300

Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro

305 310 315 320

Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala

325 330 335

Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys

340 345 350

Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr

355 360 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn

370 375 380

Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn

385 390 395 400

Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile

405 410 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln

420 425 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val

435 440 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile

450 455 460

Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu

465 470 475 480

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg

485 490 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly

500 505 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met

515 520 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp

530 535 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg

545 550 555 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln

565 570 575

Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser

580 585 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu

595 600 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr

610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe

625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met
645 650 655
Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val

660 665 670
Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
675 680 685
Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
690 695 700
Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705 710 715 720
Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln

725 730 735
Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
740 745 750
Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
755 760 765
Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn
770 775 780
Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile

785 790 795 800
Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys
805 810 815
Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
820 825 830
Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
835 840 845
Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu

850 855 860
Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865 870 875 880
Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr

885 890 895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg

900 905 910

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys

915 920 925

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys

930 935 940

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu

945 950 955 960

Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu

965 970 975

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu

980 985 990

Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn

995 1000 1005

Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr

1010 1015 1020

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr

1025 1030 1035

Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly

1040 1045 1050

<210> 38

<211> 1082

<212> PRT

<213> Geobacillus thermodenitrificans

<400> 38

Met Lys Tyr Lys Ile Gly Leu Asp Ile Gly Ile Thr Ser Ile Gly Trp

1 5 10 15

Ala Val Ile Asn Leu Asp Ile Pro Arg Ile Glu Asp Leu Gly Val Arg

20 25 30

Ile Phe Asp Arg Ala Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu

35 40 45

Pro Arg Arg Leu Ala Arg Ser Ala Arg Arg Arg Leu Arg Arg Arg Lys
50 55 60
His Arg Leu Glu Arg Ile Arg Arg Leu Phe Val Arg Glu Gly Ile Leu
65 70 75 80
Thr Lys Glu Glu Leu Asn Lys Leu Phe Glu Lys Lys His Glu Ile Asp
85 90 95
Val Trp Gln Leu Arg Val Glu Ala Leu Asp Arg Lys Leu Asn Asn Asp
100 105 110

Glu Leu Ala Arg Ile Leu Leu His Leu Ala Lys Arg Arg Gly Phe Arg
115 120 125
Ser Asn Arg Lys Ser Glu Arg Thr Asn Lys Glu Asn Ser Thr Met Leu
130 135 140
Lys His Ile Glu Glu Asn Gln Ser Ile Leu Ser Ser Tyr Arg Thr Val
145 150 155 160
Ala Glu Met Val Val Lys Asp Pro Lys Phe Ser Leu His Lys Arg Asn
165 170 175

Lys Glu Asp Asn Tyr Thr Asn Thr Val Ala Arg Asp Asp Leu Glu Arg
180 185 190
Glu Ile Lys Leu Ile Phe Ala Lys Gln Arg Glu Tyr Gly Asn Ile Val
195 200 205
Cys Thr Glu Ala Phe Glu His Glu Tyr Ile Ser Ile Trp Ala Ser Gln
210 215 220
Arg Pro Phe Ala Ser Lys Asp Asp Ile Glu Lys Lys Val Gly Phe Cys
225 230 235 240

Thr Phe Glu Pro Lys Glu Lys Arg Ala Pro Lys Ala Thr Tyr Thr Phe
245 250 255
Gln Ser Phe Thr Val Trp Glu His Ile Asn Lys Leu Arg Leu Val Ser
260 265 270
Pro Gly Gly Ile Arg Ala Leu Thr Asp Asp Glu Arg Arg Leu Ile Tyr
275 280 285

Lys Gln Ala Phe His Lys Asn Lys Ile Thr Phe His Asp Val Arg Thr
290 295 300

Leu Leu Asn Leu Pro Asp Asp Thr Arg Phe Lys Gly Leu Leu Tyr Asp
305 310 315 320
Arg Asn Thr Thr Leu Lys Glu Asn Glu Lys Val Arg Phe Leu Glu Leu
325 330 335
Gly Ala Tyr His Lys Ile Arg Lys Ala Ile Asp Ser Val Tyr Gly Lys
340 345 350
Gly Ala Ala Lys Ser Phe Arg Pro Ile Asp Phe Asp Thr Phe Gly Tyr
355 360 365

Ala Leu Thr Met Phe Lys Asp Asp Thr Asp Ile Arg Ser Tyr Leu Arg
370 375 380
Asn Glu Tyr Glu Gln Asn Gly Lys Arg Met Glu Asn Leu Ala Asp Lys
385 390 395 400
Val Tyr Asp Glu Glu Leu Ile Glu Glu Leu Leu Asn Leu Ser Phe Ser
405 410 415
Lys Phe Gly His Leu Ser Leu Lys Ala Leu Arg Asn Ile Leu Pro Tyr
420 425 430

Met Glu Gln Gly Glu Val Tyr Ser Thr Ala Cys Glu Arg Ala Gly Tyr
435 440 445
Thr Phe Thr Gly Pro Lys Lys Lys Gln Lys Thr Val Leu Leu Pro Asn
450 455 460
Ile Pro Pro Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln Ala
465 470 475 480
Arg Lys Val Val Asn Ala Ile Ile Lys Lys Tyr Gly Ser Pro Val Ser
485 490 495

Ile His Ile Glu Leu Ala Arg Glu Leu Ser Gln Ser Phe Asp Glu Arg
500 505 510
Arg Lys Met Gln Lys Glu Gln Glu Gly Asn Arg Lys Lys Asn Glu Thr
515 520 525
Ala Ile Arg Gln Leu Val Glu Tyr Gly Leu Thr Leu Asn Pro Thr Gly

530 535 540

Leu Asp Ile Val Lys Phe Lys Leu Trp Ser Glu Gln Asn Gly Lys Cys
545 550 555 560

Ala Tyr Ser Leu Gln Pro Ile Glu Ile Glu Arg Leu Leu Glu Pro Gly
565 570 575

Tyr Thr Glu Val Asp His Val Ile Pro Tyr Ser Arg Ser Leu Asp Asp
580 585 590

Ser Tyr Thr Asn Lys Val Leu Val Leu Thr Lys Glu Asn Arg Glu Lys
595 600 605

Gly Asn Arg Thr Pro Ala Glu Tyr Leu Gly Leu Gly Ser Glu Arg Trp
610 615 620

Gln Gln Phe Glu Thr Phe Val Leu Thr Asn Lys Gln Phe Ser Lys Lys
625 630 635 640

Lys Arg Asp Arg Leu Leu Arg Leu His Tyr Asp Glu Asn Glu Glu Asn
645 650 655

Glu Phe Lys Asn Arg Asn Leu Asn Asp Thr Arg Tyr Ile Ser Arg Phe
660 665 670

Leu Ala Asn Phe Ile Arg Glu His Leu Lys Phe Ala Asp Ser Asp Asp
675 680 685

Lys Gln Lys Val Tyr Thr Val Asn Gly Arg Ile Thr Ala His Leu Arg
690 695 700

Ser Arg Trp Asn Phe Asn Lys Asn Arg Glu Glu Ser Asn Leu His His
705 710 715 720

Ala Val Asp Ala Ala Ile Val Ala Cys Thr Thr Pro Ser Asp Ile Ala
725 730 735

Arg Val Thr Ala Phe Tyr Gln Arg Arg Glu Gln Asn Lys Glu Leu Ser
740 745 750

Lys Lys Thr Asp Pro Gln Phe Pro Gln Pro Trp Pro His Phe Ala Asp
755 760 765

Glu Leu Gln Ala Arg Leu Ser Lys Asn Pro Lys Glu Ser Ile Lys Ala
770 775 780

Leu Asn Leu Gly Asn Tyr Asp Asn Glu Lys Leu Glu Ser Leu Gln Pro
785 790 795 800
Val Phe Val Ser Arg Met Pro Lys Arg Ser Ile Thr Gly Ala Ala His
805 810 815

Gln Glu Thr Leu Arg Arg Tyr Ile Gly Ile Asp Glu Arg Ser Gly Lys
820 825 830
Ile Gln Thr Val Val Lys Lys Lys Leu Ser Glu Ile Gln Leu Asp Lys
835 840 845
Thr Gly His Phe Pro Met Tyr Gly Lys Glu Ser Asp Pro Arg Thr Tyr
850 855 860
Glu Ala Ile Arg Gln Arg Leu Leu Glu His Asn Asn Asp Pro Lys Lys
865 870 875 880

Ala Phe Gln Glu Pro Leu Tyr Lys Pro Lys Lys Asn Gly Glu Leu Gly
885 890 895
Pro Ile Ile Arg Thr Ile Lys Ile Ile Asp Thr Thr Asn Gln Val Ile
900 905 910
Pro Leu Asn Asp Gly Lys Thr Val Ala Tyr Asn Ser Asn Ile Val Arg
915 920 925
Val Asp Val Phe Glu Lys Asp Gly Lys Tyr Tyr Cys Val Pro Ile Tyr
930 935 940

Thr Ile Asp Met Met Lys Gly Ile Leu Pro Asn Lys Ala Ile Glu Pro
945 950 955 960
Asn Lys Pro Tyr Ser Glu Trp Lys Glu Met Thr Glu Asp Tyr Thr Phe
965 970 975
Arg Phe Ser Leu Tyr Pro Asn Asp Leu Ile Arg Ile Glu Phe Pro Arg
980 985 990
Glu Lys Thr Ile Lys Thr Ala Val Gly Glu Glu Ile Lys Ile Lys Asp
995 1000 1005

Leu Phe Ala Tyr Tyr Gln Thr Ile Asp Ser Ser Asn Gly Gly Leu
1010 1015 1020
Ser Leu Val Ser His Asp Asn Asn Phe Ser Leu Arg Ser Ile Gly

1025 1030 1035

Ser Arg Thr Leu Lys Arg Phe Glu Lys Tyr Gln Val Asp Val Leu

1040 1045 1050

Gly Asn Ile Tyr Lys Val Arg Gly Glu Lys Arg Val Gly Val Ala

1055 1060 1065

Ser Ser Ser His Ser Lys Ala Gly Glu Thr Ile Arg Pro Leu

1070 1075 1080

<210> 39

<211> 1375

<212> PRT

<213> Streptococcus canis

<400> 39

Met Glu Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asn Arg Lys Ser Ile Lys Lys Asn Leu Met

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Arg

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ala Asn Glu Met Ala Lys Leu Asp Asp Ser

85 90 95

Phe Phe Gln Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

Asn Glu Arg His Pro Ile Phe Gly Asn Leu Ala Asp Glu Val Ala Tyr

115 120 125

His Arg Asn Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp

130 135 140

Ser Pro Glu Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Ile Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Lys Leu Asn Ala

165 170 175

Glu Asn Ser Asp Val Ala Lys Leu Phe Tyr Gln Leu Ile Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Ser Pro Leu Asp Glu Ile Glu Val Asp Ala

195 200 205

Lys Gly Ile Leu Ser Ala Arg Leu Ser Lys Ser Lys Arg Leu Glu Lys

210 215 220

Leu Ile Ala Val Phe Pro Asn Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Ile Ile Ala Leu Ala Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Thr Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Glu Leu Leu Gly Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Ser Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Ser Asn Ser Glu Val Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Val Lys Arg Tyr Asp Glu His His Gln Asp Leu Ala Leu Leu Lys

325 330 335

Thr Leu Val Arg Gln Gln Phe Pro Glu Lys Tyr Ala Glu Ile Phe Lys

340 345 350

Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Val Gly Ile Gly Ile Lys

355 360 365

His Arg Lys Arg Thr Thr Lys Leu Ala Thr Gln Glu Glu Phe Tyr Lys

370 375 380

Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Ala Glu Glu Leu Leu

385 390 395 400

Ala Lys Leu Asn Arg Asp Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp

405 410 415

Asn Gly Ser Ile Pro His Gln Ile His Leu Lys Glu Leu His Ala Ile

420 425 430

Leu Arg Arg Gln Glu Glu Phe Tyr Pro Phe Leu Lys Glu Asn Arg Glu

435 440 445

Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro

450 455 460

Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Leu Thr Arg Lys Ser Glu

465 470 475 480

Glu Ala Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala

485 490 495

Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Glu Gln Leu

500 505 510

Pro Asn Lys Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe

515 520 525

Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Arg Met

530 535 540

Arg Lys Pro Glu Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp

545 550 555 560

Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu

565 570 575

Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ile Gly

580 585 590

Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu

595 600 605

Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp

610 615 620

Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu

625 630 635 640

Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys

645 650 655

Val Met Lys Gln Leu Lys Arg Arg His Tyr Thr Gly Trp Gly Arg Leu
660 665 670
Ser Arg Lys Met Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr

675 680 685
Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ser Asn Arg Asn Phe Met
690 695 700
Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Glu Ile Glu Lys
705 710 715 720
Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu Gln Ile Ala Asp
725 730 735
Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys

740 745 750
Ile Val Asp Glu Leu Val Lys Val Met Gly His Lys Pro Glu Asn Ile
755 760 765
Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Thr Lys Gly Leu Gln
770 775 780
Gln Ser Arg Glu Arg Lys Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu
785 790 795 800
Glu Ser Gln Ile Leu Lys Glu Asn Pro Val Glu Asn Thr Gln Leu Gln

805 810 815
Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr
820 825 830
Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
835 840 845
His Ile Val Pro Gln Ser Phe Ile Lys Asp Asp Ser Ile Asp Asn Lys
850 855 860
Val Leu Thr Arg Ser Val Glu Asn Arg Gly Lys Ser Asp Asn Val Pro

865 870 875 880
Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu
885 890 895
Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala

900 905 910

Glu Arg Gly Gly Leu Ser Glu Ala Asp Lys Ala Gly Phe Ile Lys Arg

915 920 925

Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Ile Leu

930 935 940

Asp Ser Arg Met Asn Thr Lys Arg Asp Lys Asn Asp Lys Pro Ile Arg

945 950 955 960

Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg

965 970 975

Lys Asp Phe Gln Leu Tyr Lys Val Arg Asp Ile Asn Asn Tyr His His

980 985 990

Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys

995 1000 1005

Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys

1010 1015 1020

Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile

1025 1030 1035

Gly Lys Ala Thr Ala Lys Arg Phe Phe Tyr Ser Asn Ile Met Asn

1040 1045 1050

Phe Phe Lys Thr Glu Val Lys Leu Ala Asn Gly Glu Ile Arg Lys

1055 1060 1065

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Val Val Trp

1070 1075 1080

Asn Lys Glu Lys Asp Phe Ala Thr Val Arg Lys Val Leu Ala Met

1085 1090 1095

Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly

1100 1105 1110

Phe Ser Lys Glu Ser Ile Leu Ser Lys Arg Glu Ser Ala Lys Leu

1115 1120 1125

Ile Pro Arg Lys Lys Gly Trp Asp Thr Arg Lys Tyr Gly Gly Phe

1130 1135 1140

Gly Ser Pro Thr Val Ala Tyr Ser Ile Leu Val Val Ala Lys Val

1145 1150 1155

Glu Lys Gly Lys Ala Lys Lys Leu Lys Ser Val Lys Val Leu Val

1160 1165 1170

Gly Ile Thr Ile Met Glu Lys Gly Ser Tyr Glu Lys Asp Pro Ile

1175 1180 1185

Gly Phe Leu Glu Ala Lys Gly Tyr Lys Asp Ile Lys Lys Glu Leu

1190 1195 1200

Ile Phe Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly

1205 1210 1215

Arg Arg Arg Met Leu Ala Ser Ala Thr Glu Leu Gln Lys Ala Asn

1220 1225 1230

Glu Leu Val Leu Pro Gln His Leu Val Arg Leu Leu Tyr Tyr Thr

1235 1240 1245

Gln Asn Ile Ser Ala Thr Thr Gly Ser Asn Asn Leu Gly Tyr Ile

1250 1255 1260

Glu Gln His Arg Glu Glu Phe Lys Glu Ile Phe Glu Lys Ile Ile

1265 1270 1275

Asp Phe Ser Glu Lys Tyr Ile Leu Lys Asn Lys Val Asn Ser Asn

1280 1285 1290

Leu Lys Ser Ser Phe Asp Glu Gln Phe Ala Val Ser Asp Ser Ile

1295 1300 1305

Leu Leu Ser Asn Ser Phe Val Ser Leu Leu Lys Tyr Thr Ser Phe

1310 1315 1320

Gly Ala Ser Gly Gly Phe Thr Phe Leu Asp Leu Asp Val Lys Gln

1325 1330 1335

Gly Arg Leu Arg Tyr Gln Thr Val Thr Glu Val Leu Asp Ala Thr

1340 1345 1350

Leu Ile Tyr Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Thr Asp

1355 1360 1365

Leu Ser Gln Leu Gly Gly Asp

1370 1375

<210> 40

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<220><221> misc_feature

<222> (10)..(10)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (840)..(840)

<223> Xaa can be any naturally occurring amino acid

<400> 40

Met Asp Lys Lys Tyr Ser Ile Gly Leu Xaa Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp Xaa Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365
<210> 41

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 41

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 42

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<220><221> misc_feature

<222> (10)..(10)

<223> Xaa can be any naturally occurring amino acid

<400> 42

Met Asp Lys Lys Tyr Ser Ile Gly Leu Xaa Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 43
<211> 1368
<212> PRT
<213> Streptococcus pyogenes
<220><221> misc_feature
<222> (762)..(762)
<223> Xaa can be any naturally occurring amino acid
<400> 43

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Xaa Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 44

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<220><221> misc_feature

<222> (983)..(983)

<223> Xaa can be any naturally occurring amino acid

<400> 44

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His Xaa Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 45
<211> 1368
<212> PRT
<213> Streptococcus pyogenes
<220><221> misc_feature
<222> (986)..(986)

<223> Xaa can be any naturally occurring amino acid
<400> 45
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Xaa Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 46
<211> 1368
<212> PRT
<213> Streptococcus pyogenes

<400> 46
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 47

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 47

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 48

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 48

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His Ala Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 49

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 49

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Ala Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 50

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<220><221> misc_feature

<222> (840)..(840)

<223> Xaa can be any naturally occurring amino acid

<400> 50

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp Xaa Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365
<210> 51
<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 51

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 52

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<220><221> misc_feature

<222> (864)..(864)

<223> Xaa can be any naturally occurring amino acid

<400> 52

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Xaa
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 53
<211> 1368
<212> PRT
<213> Streptococcus pyogenes

<400> 53
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Ala

850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 54
<211> 1367
<212> PRT
<213> Streptococcus pyogenes
<220><221> misc_feature
<222> (839)..(839)
<223> Xaa can be any naturally occurring amino acid
<400> 54
Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp Xaa Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 55

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 55

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 56

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<220><221> misc_feature

<222> (863)..(863)

<223> Xaa can be any naturally occurring amino acid

<400> 56

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Xaa Gly

850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 57

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 57

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Ala Gly
850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 58
<211> 1053
<212> PRT
<213> Staphylococcus aureus
<400> 58

Met Gly Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser
1 5 10 15
Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala
20 25 30
Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg

35 40 45
Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg
50 55 60
Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp
65 70 75 80
His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly

85 90 95

Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His

100 105 110

Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp

115 120 125

Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys

130 135 140

Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys

145 150 155 160

Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp

165 170 175

Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His

180 185 190

Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr

195 200 205

Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp

210 215 220

Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr

225 230 235 240

Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu

245 250 255

Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu

260 265 270

Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val

275 280 285

Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile

290 295 300

Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly

305 310 315 320

Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile

325 330 335

Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile
340 345 350
Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu

355 360 365
Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile
370 375 380
Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala
385 390 395 400
Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile
405 410 415
Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser

420 425 430
Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser
435 440 445
Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala
450 455 460
Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala
465 470 475 480
Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln

485 490 495
Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr
500 505 510
Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His
515 520 525
Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu
530 535 540
Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile

545 550 555 560
Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val
565 570 575
Lys Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr

580 585 590

Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His

595 600 605

Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys

610 615 620

Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys

625 630 635 640

Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly

645 650 655

Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val

660 665 670

Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys

675 680 685

Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu

690 695 700

Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys

705 710 715 720

Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu

725 730 735

Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys

740 745 750

Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys

755 760 765

Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Lys Leu

770 775 780

Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr

785 790 795 800

Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys

805 810 815

Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His

820 825 830

His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr
835 840 845
Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn
850 855 860
Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys

865 870 875 880
Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp
885 890 895
Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro
900 905 910
Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr
915 920 925
Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn

930 935 940
Ser Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln
945 950 955 960
Ala Glu Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile Asn
965 970 975
Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg
980 985 990
Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn

995 1000 1005
Met Asn Asp Lys Arg Pro Pro His Ile Ile Lys Thr Ile Ala Ser
1010 1015 1020
Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn
1025 1030 1035
Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys
1040 1045 1050
<210> 59
<211> 1082
<212> PRT

<213> Neisseria meningitidis

<400> 59

Met Ala Ala Phe Lys Pro Asn Ser Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15
Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30
Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45
Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu
50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65 70 75 80
Arg Thr Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asn
85 90 95
Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
100 105 110
Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
115 120 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
130 135 140
Lys Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys
145 150 155 160
Gly Val Ala Gly Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr
165 170 175
Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile
180 185 190

Arg Asn Gln Arg Ser Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
195 200 205
Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
210 215 220
Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met

225 230 235 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly

245 250 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr

260 265 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile

275 280 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr

290 295 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala

305 310 315 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg

325 330 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala

340 345 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys

355 360 365

Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr

370 375 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys

385 390 395 400

Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser

405 410 415

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val

420 425 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile

435 440 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu

450 455 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala

465 470 475 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
485 490 495
Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
500 505 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
515 520 525
Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
530 535 540
Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545 550 555 560
Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
565 570 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp Ala Ala Leu Pro Phe
580 585 590
Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
595 600 605
Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn
610 615 620
Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625 630 635 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
645 650 655
Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
660 665 670
Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
675 680 685
Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
690 695 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705 710 715 720
Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala

725 730 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala

740 745 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln

755 760 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met

770 775 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala

785 790 795 800

Asp Thr Leu Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser

805 810 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg

820 825 830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys

835 840 845

Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu

850 855 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg

865 870 875 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys

885 890 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys

900 905 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val

915 920 925

Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn

930 935 940

Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr

945 950 955 960

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp

965 970 975

Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
980 985 990
Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu
995 1000 1005
Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys
1010 1015 1020

His Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp
1025 1030 1035
His Lys Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys
1040 1045 1050
Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys
1055 1060 1065
Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1070 1075 1080

<210> 60
<211> 984
<212> PRT
<213> Campylobacter jejuni
<400> 60
Met Ala Arg Ile Leu Ala Phe Asp Ile Gly Ile Ser Ser Ile Gly Trp
1 5 10 15
Ala Phe Ser Glu Asn Asp Glu Leu Lys Asp Cys Gly Val Arg Ile Phe
20 25 30
Thr Lys Val Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu Pro Arg
35 40 45
Arg Leu Ala Arg Ser Ala Arg Lys Arg Leu Ala Arg Arg Lys Ala Arg

50 55 60
Leu Asn His Leu Lys His Leu Ile Ala Asn Glu Phe Lys Leu Asn Tyr
65 70 75 80
Glu Asp Tyr Gln Ser Phe Asp Glu Ser Leu Ala Lys Ala Tyr Lys Gly
85 90 95
Ser Leu Ile Ser Pro Tyr Glu Leu Arg Phe Arg Ala Leu Asn Glu Leu

100 105 110

Leu Ser Lys Gln Asp Phe Ala Arg Val Ile Leu His Ile Ala Lys Arg

115 120 125

Arg Gly Tyr Asp Asp Ile Lys Asn Ser Asp Asp Lys Glu Lys Gly Ala

130 135 140

Ile Leu Lys Ala Ile Lys Gln Asn Glu Glu Lys Leu Ala Asn Tyr Gln

145 150 155 160

Ser Val Gly Glu Tyr Leu Tyr Lys Glu Tyr Phe Gln Lys Phe Lys Glu

165 170 175

Asn Ser Lys Glu Phe Thr Asn Val Arg Asn Lys Lys Glu Ser Tyr Glu

180 185 190

Arg Cys Ile Ala Gln Ser Phe Leu Lys Asp Glu Leu Lys Leu Ile Phe

195 200 205

Lys Lys Gln Arg Glu Phe Gly Phe Ser Phe Ser Lys Lys Phe Glu Glu

210 215 220

Glu Val Leu Ser Val Ala Phe Tyr Lys Arg Ala Leu Lys Asp Phe Ser

225 230 235 240

His Leu Val Gly Asn Cys Ser Phe Phe Thr Asp Glu Lys Arg Ala Pro

245 250 255

Lys Asn Ser Pro Leu Ala Phe Met Phe Val Ala Leu Thr Arg Ile Ile

260 265 270

Asn Leu Leu Asn Asn Leu Lys Asn Thr Glu Gly Ile Leu Tyr Thr Lys

275 280 285

Asp Asp Leu Asn Ala Leu Leu Asn Glu Val Leu Lys Asn Gly Thr Leu

290 295 300

Thr Tyr Lys Gln Thr Lys Lys Leu Leu Gly Leu Ser Asp Asp Tyr Glu

305 310 315 320

Phe Lys Gly Glu Lys Gly Thr Tyr Phe Ile Glu Phe Lys Lys Tyr Lys

325 330 335

Glu Phe Ile Lys Ala Leu Gly Glu His Asn Leu Ser Gln Asp Asp Leu

340 345 350

Asn Glu Ile Ala Lys Asp Ile Thr Leu Ile Lys Asp Glu Ile Lys Leu
355 360 365
Lys Lys Ala Leu Ala Lys Tyr Asp Leu Asn Gln Asn Gln Ile Asp Ser

370 375 380
Leu Ser Lys Leu Glu Phe Lys Asp His Leu Asn Ile Ser Phe Lys Ala
385 390 395 400
Leu Lys Leu Val Thr Pro Leu Met Leu Glu Gly Lys Lys Tyr Asp Glu
405 410 415
Ala Cys Asn Glu Leu Asn Leu Lys Val Ala Ile Asn Glu Asp Lys Lys
420 425 430
Asp Phe Leu Pro Ala Phe Asn Glu Thr Tyr Tyr Lys Asp Glu Val Thr

435 440 445
Asn Pro Val Val Leu Arg Ala Ile Lys Glu Tyr Arg Lys Val Leu Asn
450 455 460
Ala Leu Leu Lys Lys Tyr Gly Lys Val His Lys Ile Asn Ile Glu Leu
465 470 475 480
Ala Arg Glu Val Gly Lys Asn His Ser Gln Arg Ala Lys Ile Glu Lys
485 490 495
Glu Gln Asn Glu Asn Tyr Lys Ala Lys Lys Asp Ala Glu Leu Glu Cys

500 505 510
Glu Lys Leu Gly Leu Lys Ile Asn Ser Lys Asn Ile Leu Lys Leu Arg
515 520 525
Leu Phe Lys Glu Gln Lys Glu Phe Cys Ala Tyr Ser Gly Glu Lys Ile
530 535 540
Lys Ile Ser Asp Leu Gln Asp Glu Lys Met Leu Glu Ile Asp His Ile
545 550 555 560
Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr Met Asn Lys Val Leu

565 570 575
Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu
580 585 590
Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala

595 600 605

Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr

610 615 620

Lys Asp Lys Glu Gln Lys Asn Phe Lys Asp Arg Asn Leu Asn Asp Thr

625 630 635 640

Arg Tyr Ile Ala Arg Leu Val Leu Asn Tyr Thr Lys Asp Tyr Leu Asp

645 650 655

Phe Leu Pro Leu Ser Asp Asp Glu Asn Thr Lys Leu Asn Asp Thr Gln

660 665 670

Lys Gly Ser Lys Val His Val Glu Ala Lys Ser Gly Met Leu Thr Ser

675 680 685

Ala Leu Arg His Thr Trp Gly Phe Ser Ala Lys Asp Arg Asn Asn His

690 695 700

Leu His His Ala Ile Asp Ala Val Ile Ile Ala Tyr Ala Asn Asn Ser

705 710 715 720

Ile Val Lys Ala Phe Ser Asp Phe Lys Lys Glu Gln Glu Ser Asn Ser

725 730 735

Ala Glu Leu Tyr Ala Lys Lys Ile Ser Glu Leu Asp Tyr Lys Asn Lys

740 745 750

Arg Lys Phe Phe Glu Pro Phe Ser Gly Phe Arg Gln Lys Val Leu Asp

755 760 765

Lys Ile Asp Glu Ile Phe Val Ser Lys Pro Glu Arg Lys Lys Pro Ser

770 775 780

Gly Ala Leu His Glu Glu Thr Phe Arg Lys Glu Glu Glu Phe Tyr Gln

785 790 795 800

Ser Tyr Gly Gly Lys Glu Gly Val Leu Lys Ala Leu Glu Leu Gly Lys

805 810 815

Ile Arg Lys Val Asn Gly Lys Ile Val Lys Asn Gly Asp Met Phe Arg

820 825 830

Val Asp Ile Phe Lys His Lys Lys Thr Asn Lys Phe Tyr Ala Val Pro

835 840 845

Ile Tyr Thr Met Asp Phe Ala Leu Lys Val Leu Pro Asn Lys Ala Val
850 855 860
Ala Arg Ser Lys Lys Gly Glu Ile Lys Asp Trp Ile Leu Met Asp Glu
865 870 875 880
Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile

885 890 895
Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe
900 905 910
Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe
915 920 925
Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
930 935 940
Lys Glu Val Ile Ala Lys Ser Ile Gly Ile Gln Asn Leu Lys Val Phe

945 950 955 960
Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
965 970 975
Arg Gln Arg Glu Asp Phe Lys Lys
980

<210> 61
<211> 1087
<212> PRT
<213> Geobacillus stearothermophilus
<400> 61
Met Arg Tyr Lys Ile Gly Leu Asp Ile Gly Ile Thr Ser Val Gly Trp
1 5 10 15
Ala Val Met Asn Leu Asp Ile Pro Arg Ile Glu Asp Leu Gly Val Arg

20 25 30
Ile Phe Asp Arg Ala Glu Asn Pro Gln Thr Gly Glu Ser Leu Ala Leu
35 40 45
Pro Arg Arg Leu Ala Arg Ser Ala Arg Arg Arg Leu Arg Arg Arg Lys
50 55 60
His Arg Leu Glu Arg Ile Arg Arg Leu Val Ile Arg Glu Gly Ile Leu

65 70 75 80

Thr Lys Glu Glu Leu Asp Lys Leu Phe Glu Glu Lys His Glu Ile Asp

85 90 95

Val Trp Gln Leu Arg Val Glu Ala Leu Asp Arg Lys Leu Asn Asn Asp

100 105 110

Glu Leu Ala Arg Val Leu Leu His Leu Ala Lys Arg Arg Gly Phe Lys

115 120 125

Ser Asn Arg Lys Ser Glu Arg Ser Asn Lys Glu Asn Ser Thr Met Leu

130 135 140

Lys His Ile Glu Glu Asn Arg Ala Ile Leu Ser Ser Tyr Arg Thr Val

145 150 155 160

Gly Glu Met Ile Val Lys Asp Pro Lys Phe Ala Leu His Lys Arg Asn

165 170 175

Lys Gly Glu Asn Tyr Thr Asn Thr Ile Ala Arg Asp Asp Leu Glu Arg

180 185 190

Glu Ile Arg Leu Ile Phe Ser Lys Gln Arg Glu Phe Gly Asn Met Ser

195 200 205

Cys Thr Glu Glu Phe Glu Asn Glu Tyr Ile Thr Ile Trp Ala Ser Gln

210 215 220

Arg Pro Val Ala Ser Lys Asp Asp Ile Glu Lys Lys Val Gly Phe Cys

225 230 235 240

Thr Phe Glu Pro Lys Glu Lys Arg Ala Pro Lys Ala Thr Tyr Thr Phe

245 250 255

Gln Ser Phe Ile Ala Trp Glu His Ile Asn Lys Leu Arg Leu Ile Ser

260 265 270

Pro Ser Gly Ala Arg Gly Leu Thr Asp Glu Glu Arg Arg Leu Leu Tyr

275 280 285

Glu Gln Ala Phe Gln Lys Asn Lys Ile Thr Tyr His Asp Ile Arg Thr

290 295 300

Leu Leu His Leu Pro Asp Asp Thr Tyr Phe Lys Gly Ile Val Tyr Asp

305 310 315 320

Arg Gly Glu Ser Arg Lys Gln Asn Glu Asn Ile Arg Phe Leu Glu Leu
325 330 335
Asp Ala Tyr His Gln Ile Arg Lys Ala Val Asp Lys Val Tyr Gly Lys

340 345 350
Gly Lys Ser Ser Ser Phe Leu Pro Ile Asp Phe Asp Thr Phe Gly Tyr
355 360 365
Ala Leu Thr Leu Phe Lys Asp Asp Ala Asp Ile His Ser Tyr Leu Arg
370 375 380
Asn Glu Tyr Glu Gln Asn Gly Lys Arg Met Pro Asn Leu Ala Asn Lys
385 390 395 400
Val Tyr Asp Asn Glu Leu Ile Glu Glu Leu Leu Asn Leu Ser Phe Thr

405 410 415
Lys Phe Gly His Leu Ser Leu Lys Ala Leu Arg Ser Ile Leu Pro Tyr
420 425 430
Met Glu Gln Gly Glu Val Tyr Ser Ser Ala Cys Glu Arg Ala Gly Tyr
435 440 445
Thr Phe Thr Gly Pro Lys Lys Lys Gln Lys Thr Met Leu Leu Pro Asn
450 455 460
Ile Pro Pro Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln Ala

465 470 475 480
Arg Lys Val Val Asn Ala Ile Ile Lys Lys Tyr Gly Ser Pro Val Ser
485 490 495
Ile His Ile Glu Leu Ala Arg Asp Leu Ser Gln Thr Phe Asp Glu Arg
500 505 510
Arg Lys Thr Lys Lys Glu Gln Asp Glu Asn Arg Lys Lys Asn Glu Thr
515 520 525
Ala Ile Arg Gln Leu Met Glu Tyr Gly Leu Thr Leu Asn Pro Thr Gly

530 535 540
His Asp Ile Val Lys Phe Lys Leu Trp Ser Glu Gln Asn Gly Arg Cys
545 550 555 560
Ala Tyr Ser Leu Gln Pro Ile Glu Ile Glu Arg Leu Leu Glu Pro Gly

565 570 575

Tyr Val Glu Val Asp His Val Ile Pro Tyr Ser Arg Ser Leu Asp Asp

580 585 590

Ser Tyr Thr Asn Lys Val Leu Val Leu Thr Arg Glu Asn Arg Glu Lys

595 600 605

Gly Asn Arg Ile Pro Ala Glu Tyr Leu Gly Val Gly Thr Glu Arg Trp

610 615 620

Gln Gln Phe Glu Thr Phe Val Leu Thr Asn Lys Gln Phe Ser Lys Lys

625 630 635 640

Lys Arg Asp Arg Leu Leu Arg Leu His Tyr Asp Glu Asn Glu Glu Thr

645 650 655

Glu Phe Lys Asn Arg Asn Leu Asn Asp Thr Arg Tyr Ile Ser Arg Phe

660 665 670

Phe Ala Asn Phe Ile Arg Glu His Leu Lys Phe Ala Glu Ser Asp Asp

675 680 685

Lys Gln Lys Val Tyr Thr Val Asn Gly Arg Val Thr Ala His Leu Arg

690 695 700

Ser Arg Trp Glu Phe Asn Lys Asn Arg Glu Glu Ser Asp Leu His His

705 710 715 720

Ala Val Asp Ala Val Ile Val Ala Cys Thr Thr Pro Ser Asp Ile Ala

725 730 735

Lys Val Thr Ala Phe Tyr Gln Arg Arg Glu Gln Asn Lys Glu Leu Ala

740 745 750

Lys Lys Thr Glu Pro His Phe Pro Gln Pro Trp Pro His Phe Ala Asp

755 760 765

Glu Leu Arg Ala Arg Leu Ser Lys His Pro Lys Glu Ser Ile Lys Ala

770 775 780

Leu Asn Leu Gly Asn Tyr Asp Asp Gln Lys Leu Glu Ser Leu Gln Pro

785 790 795 800

Val Phe Val Ser Arg Met Pro Lys Arg Ser Val Thr Gly Ala Ala His

805 810 815

Gln Glu Thr Leu Arg Arg Tyr Val Gly Ile Asp Glu Arg Ser Gly Lys
820 825 830
Ile Gln Thr Val Val Lys Thr Lys Leu Ser Glu Ile Lys Leu Asp Ala
835 840 845
Ser Gly His Phe Pro Met Tyr Gly Lys Glu Ser Asp Pro Arg Thr Tyr

850 855 860
Glu Ala Ile Arg Gln Arg Leu Leu Glu His Asn Asn Asp Pro Lys Lys
865 870 875 880
Ala Phe Gln Glu Pro Leu Tyr Lys Pro Lys Lys Asn Gly Glu Pro Gly
885 890 895
Pro Val Ile Arg Thr Val Lys Ile Ile Asp Thr Lys Asn Gln Val Ile
900 905 910
Pro Leu Asn Asp Gly Lys Thr Val Ala Tyr Asn Ser Asn Ile Val Arg

915 920 925
Val Asp Val Phe Glu Lys Asp Gly Lys Tyr Tyr Cys Val Pro Val Tyr
930 935 940
Thr Met Asp Ile Met Lys Gly Ile Leu Pro Asn Lys Ala Ile Glu Pro
945 950 955 960
Asn Lys Pro Tyr Ser Glu Trp Lys Glu Met Thr Glu Asp Tyr Thr Phe
965 970 975
Arg Phe Ser Leu Tyr Pro Asn Asp Leu Ile Arg Ile Glu Leu Pro Arg

980 985 990
Glu Lys Thr Val Lys Thr Ala Ala Gly Glu Glu Ile Asn Val Lys Asp
995 1000 1005
Val Phe Val Tyr Tyr Lys Thr Ile Asp Ser Ala Asn Gly Gly Leu
1010 1015 1020
Glu Leu Ile Ser His Asp His Arg Phe Ser Leu Arg Gly Val Gly
1025 1030 1035
Ser Arg Thr Leu Lys Arg Phe Glu Lys Tyr Gln Val Asp Val Leu

1040 1045 1050
Gly Asn Ile Tyr Lys Val Arg Gly Glu Lys Arg Val Gly Leu Ala

1055 1060 1065

Ser Ser Ala His Ser Lys Pro Gly Lys Thr Ile Arg Pro Leu Gln

1070 1075 1080

Ser Thr Arg Asp

1085

<210> 62

<211> 1228

<212> PRT

<213> Unknown

<220><223> Lachnospiraceae bacterium

<400> 62

Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr

1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp

20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys

35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp

50 55 60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu

65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn

85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn

100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu

115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe

130 135 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn

145 150 155 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile

165 170 175

Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys

180 185 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys

195 200 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe

210 215 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile

225 230 235 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn

245 250 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys

260 265 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser

275 280 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe

290 295 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys

305 310 315 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile

325 330 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe

340 345 350

Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp

355 360 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp

370 375 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu

385 390 395 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu

405 410 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
420 425 430
Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
435 440 445
Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys

450 455 460
Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465 470 475 480
Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
485 490 495
Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
500 505 510
Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro

515 520 525
Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
530 535 540
Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545 550 555 560
Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
565 570 575
Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met

580 585 590
Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro
595 600 605
Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
610 615 620
Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys
625 630 635 640
Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn

645 650 655
Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu

660 665 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys

675 680 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile

690 695 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His

705 710 715 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile

725 730 735

Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys

740 745 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys

755 760 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr

770 775 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile

785 790 795 800

Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val

805 810 815

Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Asp

820 825 830

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys Gly

835 840 845

Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn

850 855 860

Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys Glu

865 870 875 880

Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn Ile

885 890 895

Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys

900 905 910

Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu Asn

915 920 925

Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr Gln

930 935 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys

945 950 955 960

Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile

965 970 975

Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe

980 985 990

Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser Thr

995 1000 1005

Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala Asp

1010 1015 1020

Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro

1025 1030 1035

Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys Asn Phe Ser

1040 1045 1050

Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr Ser Tyr

1055 1060 1065

Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val

1070 1075 1080

Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu

1085 1090 1095

Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala

1100 1105 1110

Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met

1115 1120 1125

Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly

1130 1135 1140

Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp

1145 1150 1155

Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala

1160 1165 1170

Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala

1175 1180 1185

Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp

1190 1195 1200

Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn Lys Glu Trp

1205 1210 1215

Leu Glu Tyr Ala Gln Thr Ser Val Lys His

1220 1225

<210> 63

<211> 1108

<212> PRT

<213> Unknown

<220><223> Bacillushisashii

<400> 63

Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val

1 5 10 15

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala

20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu

35 40 45

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu

50 55 60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn

65 70 75 80

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg

85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu

100 105 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn

115 120 125

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp

130 135 140

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys

145 150 155 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly

165 170 175

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp

180 185 190

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg

195 200 205

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

210 215 220

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

225 230 235 240

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

245 250 255

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg

260 265 270

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu

275 280 285

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

290 295 300

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys

305 310 315 320

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr

325 330 335

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

340 345 350

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

355 360 365

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
370 375 380
His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
385 390 395 400
Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
405 410 415

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
420 425 430
Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
435 440 445
Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
450 455 460
Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
465 470 475 480

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
485 490 495
Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
500 505 510
Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
515 520 525
Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
530 535 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
545 550 555 560
Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
565 570 575
Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
580 585 590
Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
595 600 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

610 615 620

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625 630 635 640

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
645 650 655

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
660 665 670

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
675 680 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
690 695 700

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
705 710 715 720

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
725 730 735

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
740 745 750

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
755 760 765

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
770 775 780

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
785 790 795 800

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
805 810 815

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
820 825 830

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
835 840 845

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
850 855 860

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
865 870 875 880
Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
885 890 895
Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
900 905 910
Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
915 920 925

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
930 935 940
Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
945 950 955 960
Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
965 970 975
Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp
980 985 990

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu
995 1000 1005
Lys Asp Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys Ile
1010 1015 1020
Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser
1025 1030 1035
Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly
1040 1045 1050

Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro
1055 1060 1065
Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu
1070 1075 1080
Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr
1085 1090 1095
Ile Glu Asp Asp Ser Ser Lys Gln Ser Met
1100 1105

<210> 64

<211> 1307

<212>

PRT

<213> Unknown

<220><223> Acidaminococcus sp. (strain BV3L6)

<400> 64

Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln

50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile

115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg

180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe

195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn

210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val

225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp

245 250 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu

260 265 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn

275 280 285

Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro

290 295 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu

305 310 315 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr

325 330 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu

340 345 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His

355 360 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr

370 375 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys

385 390 395 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu

405 410 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser

420 425 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala

435 440 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
450 455 460
Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465 470 475 480
Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
485 490 495
Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser

500 505 510
Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
515 520 525
Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
530 535 540
Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545 550 555 560
Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys

565 570 575
Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
580 585 590
Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
595 600 605
Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
610 615 620
Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys

625 630 635 640
Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
645 650 655
Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
660 665 670
Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
675 680 685
Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr

690 695 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His

705 710 715 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu

725 730 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys

740 745 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu

755 760 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln

770 775 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His

785 790 795 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr

805 810 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His

820 825 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn

835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe

850 855 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln

865 870 875 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu

885 890 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg

900 905 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu

915 920 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu

930 935 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val

945 950 955 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile

965 970 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu

980 985 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu

995 1000 1005

Lys Ala Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu

1010 1015 1020

Asn Cys Leu Val Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly

1025 1030 1035

Val Leu Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala

1040 1045 1050

Lys Met Gly Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro

1055 1060 1065

Tyr Thr Ser Lys Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe

1070 1075 1080

Val Trp Lys Thr Ile Lys Asn His Glu Ser Arg Lys His Phe Leu

1085 1090 1095

Glu Gly Phe Asp Phe Leu His Tyr Asp Val Lys Thr Gly Asp Phe

1100 1105 1110

Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly

1115 1120 1125

Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn

1130 1135 1140

Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys

1145 1150 1155

Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr

1160 1165 1170

Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu

1175 1180 1185

Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu

1190 1195 1200

Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu

1205 1210 1215

Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly

1220 1225 1230

Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys

1235 1240 1245

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp

1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu

1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile

1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn

1295 1300 1305

<210> 65

<211> 1307

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 65

Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln

20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys

35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln

50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80
Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95
Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110
Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
130 135 140
Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145 150 155 160
Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
165 170 175
Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
195 200 205
Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
210 215 220
Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225 230 235 240
Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
245 250 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
260 265 270
Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
275 280 285
Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
290 295 300
Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu

305 310 315 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
325 330 335
Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
340 345 350
Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
355 360 365
Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
370 375 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385 390 395 400
Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
405 410 415
Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
420 425 430
Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
435 440 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
450 455 460
Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465 470 475 480
Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
485 490 495
Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
500 505 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
515 520 525
Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
530 535 540
Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545 550 555 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
565 570 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
580 585 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
595 600 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
610 615 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625 630 635 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
645 650 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
660 665 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
675 680 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
690 695 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705 710 715 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
725 730 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
740 745 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
755 760 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
770 775 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785 790 795 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr

805 810 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His

820 825 830

Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn

835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe

850 855 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln

865 870 875 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu

885 890 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg

900 905 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu

915 920 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu

930 935 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val

945 950 955 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile

965 970 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu

980 985 990

Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu

995 1000 1005

Lys Ala Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu

1010 1015 1020

Asn Cys Leu Val Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly

1025 1030 1035

Val Leu Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala

1040 1045 1050

Lys Met Gly Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro
1055 1060 1065
Tyr Thr Ser Lys Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe
1070 1075 1080

Val Trp Lys Thr Ile Lys Asn His Glu Ser Arg Lys His Phe Leu
1085 1090 1095
Glu Gly Phe Asp Phe Leu His Tyr Asp Val Lys Thr Gly Asp Phe
1100 1105 1110
Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly
1115 1120 1125
Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn
1130 1135 1140

Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys
1145 1150 1155
Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr
1160 1165 1170
Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu
1175 1180 1185
Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu
1190 1195 1200

Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu
1205 1210 1215
Ile Arg Ser Val Leu Gln Met Ala Asn Ser Asn Ala Ala Thr Gly
1220 1225 1230
Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys
1235 1240 1245
Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp
1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275
Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile

1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn

1295 1300 1305

<210> 66

<211> 1259

<212> PRT

<213> Unknown

<220><223> Lachnospiraceae bacterium GAM79

<400> 66

Met Asn Tyr Lys Thr Gly Leu Glu Asp Phe Ile Gly Lys Glu Ser Leu

1 5 10 15

Ser Lys Thr Leu Arg Asn Ala Leu Ile Pro Thr Glu Ser Thr Lys Ile

20 25 30

His Met Glu Glu Met Gly Val Ile Arg Asp Asp Glu Leu Arg Ala Glu

35 40 45

Lys Gln Gln Glu Leu Lys Glu Ile Met Asp Asp Tyr Tyr Arg Thr Phe

50 55 60

Ile Glu Glu Lys Leu Gly Gln Ile Gln Gly Ile Gln Trp Asn Ser Leu

65 70 75 80

Phe Gln Lys Met Glu Glu Thr Met Glu Asp Ile Ser Val Arg Lys Asp

85 90 95

Leu Asp Lys Ile Gln Asn Glu Lys Arg Lys Glu Ile Cys Cys Tyr Phe

100 105 110

Thr Ser Asp Lys Arg Phe Lys Asp Leu Phe Asn Ala Lys Leu Ile Thr

115 120 125

Asp Ile Leu Pro Asn Phe Ile Lys Asp Asn Lys Glu Tyr Thr Glu Glu

130 135 140

Glu Lys Ala Glu Lys Glu Gln Thr Arg Val Leu Phe Gln Arg Phe Ala

145 150 155 160

Thr Ala Phe Thr Asn Tyr Phe Asn Gln Arg Arg Asn Asn Phe Ser Glu

165 170 175

Asp Asn Ile Ser Thr Ala Ile Ser Phe Arg Ile Val Asn Glu Asn Ser

180 185 190

Glu Ile His Leu Gln Asn Met Arg Ala Phe Gln Arg Ile Glu Gln Gln

195 200 205

Tyr Pro Glu Glu Val Cys Gly Met Glu Glu Glu Tyr Lys Asp Met Leu

210 215 220

Gln Glu Trp Gln Met Lys His Ile Tyr Ser Val Asp Phe Tyr Asp Arg

225 230 235 240

Glu Leu Thr Gln Pro Gly Ile Glu Tyr Tyr Asn Gly Ile Cys Gly Lys

245 250 255

Ile Asn Glu His Met Asn Gln Phe Cys Gln Lys Asn Arg Ile Asn Lys

260 265 270

Asn Asp Phe Arg Met Lys Lys Leu His Lys Gln Ile Leu Cys Lys Lys

275 280 285

Ser Ser Tyr Tyr Glu Ile Pro Phe Arg Phe Glu Ser Asp Gln Glu Val

290 295 300

Tyr Asp Ala Leu Asn Glu Phe Ile Lys Thr Met Lys Lys Lys Glu Ile

305 310 315 320

Ile Arg Arg Cys Val His Leu Gly Gln Glu Cys Asp Asp Tyr Asp Leu

325 330 335

Gly Lys Ile Tyr Ile Ser Ser Asn Lys Tyr Glu Gln Ile Ser Asn Ala

340 345 350

Leu Tyr Gly Ser Trp Asp Thr Ile Arg Lys Cys Ile Lys Glu Glu Tyr

355 360 365

Met Asp Ala Leu Pro Gly Lys Gly Glu Lys Lys Glu Glu Lys Ala Glu

370 375 380

Ala Ala Ala Lys Lys Glu Glu Tyr Arg Ser Ile Ala Asp Ile Asp Lys

385 390 395 400

Ile Ile Ser Leu Tyr Gly Ser Glu Met Asp Arg Thr Ile Ser Ala Lys

405 410 415

Lys Cys Ile Thr Glu Ile Cys Asp Met Ala Gly Gln Ile Ser Ile Asp

420 425 430

Pro Leu Val Cys Asn Ser Asp Ile Lys Leu Leu Gln Asn Lys Glu Lys
435 440 445
Thr Thr Glu Ile Lys Thr Ile Leu Asp Ser Phe Leu His Val Tyr Gln

450 455 460
Trp Gly Gln Thr Phe Ile Val Ser Asp Ile Ile Glu Lys Asp Ser Tyr
465 470 475 480
Phe Tyr Ser Glu Leu Glu Asp Val Leu Glu Asp Phe Glu Gly Ile Thr
485 490 495
Thr Leu Tyr Asn His Val Arg Ser Tyr Val Thr Gln Lys Pro Tyr Ser
500 505 510
Thr Val Lys Phe Lys Leu His Phe Gly Ser Pro Thr Leu Ala Asn Gly

515 520 525
Trp Ser Gln Ser Lys Glu Tyr Asp Asn Asn Ala Ile Leu Leu Met Arg
530 535 540
Asp Gln Lys Phe Tyr Leu Gly Ile Phe Asn Val Arg Asn Lys Pro Asp
545 550 555 560
Lys Gln Ile Ile Lys Gly His Glu Lys Glu Glu Lys Gly Asp Tyr Lys
565 570 575
Lys Met Ile Tyr Asn Leu Leu Pro Gly Pro Ser Lys Met Leu Pro Lys

580 585 590
Val Phe Ile Thr Ser Arg Ser Gly Gln Glu Thr Tyr Lys Pro Ser Lys
595 600 605
His Ile Leu Asp Gly Tyr Asn Glu Lys Arg His Ile Lys Ser Ser Pro
610 615 620
Lys Phe Asp Leu Gly Tyr Cys Trp Asp Leu Ile Asp Tyr Tyr Lys Glu
625 630 635 640
Cys Ile His Lys His Pro Asp Trp Lys Asn Tyr Asp Phe His Phe Ser

645 650 655
Asp Thr Lys Asp Tyr Glu Asp Ile Ser Gly Phe Tyr Arg Glu Val Glu
660 665 670
Met Gln Gly Tyr Gln Ile Lys Trp Thr Tyr Ile Ser Ala Asp Glu Ile

675 680 685

Gln Lys Leu Asp Glu Lys Gly Gln Ile Phe Leu Phe Gln Ile Tyr Asn

690 695 700

Lys Asp Phe Ser Val His Ser Thr Gly Lys Asp Asn Leu His Thr Met

705 710 715 720

Tyr Leu Lys Asn Leu Phe Ser Glu Glu Asn Leu Lys Asp Ile Val Leu

725 730 735

Lys Leu Asn Gly Glu Ala Glu Leu Phe Phe Arg Lys Ala Ser Ile Lys

740 745 750

Thr Pro Ile Val His Lys Lys Gly Ser Val Leu Val Asn Arg Ser Tyr

755 760 765

Thr Gln Thr Val Gly Asn Lys Glu Ile Arg Val Ser Ile Pro Glu Glu

770 775 780

Tyr Tyr Thr Glu Ile Tyr Asn Tyr Leu Asn His Ile Gly Lys Gly Lys

785 790 795 800

Leu Ser Ser Glu Ala Gln Arg Tyr Leu Asp Glu Gly Lys Ile Lys Ser

805 810 815

Phe Thr Ala Thr Lys Asp Ile Val Lys Asn Tyr Arg Tyr Cys Cys Asp

820 825 830

His Tyr Phe Leu His Leu Pro Ile Thr Ile Asn Phe Lys Ala Lys Ser

835 840 845

Asp Val Ala Val Asn Glu Arg Thr Leu Ala Tyr Ile Ala Lys Lys Glu

850 855 860

Asp Ile His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr

865 870 875 880

Ile Ser Val Val Asp Val His Gly Asn Ile Arg Glu Gln Arg Ser Phe

885 890 895

Asn Ile Val Asn Gly Tyr Asp Tyr Gln Gln Lys Leu Lys Asp Arg Glu

900 905 910

Lys Ser Arg Asp Ala Ala Arg Lys Asn Trp Glu Glu Ile Glu Lys Ile

915 920 925

Lys Glu Leu Lys Glu Gly Tyr Leu Ser Met Val Ile His Tyr Ile Ala

930 935 940

Gln Leu Val Val Lys Tyr Asn Ala Val Val Ala Met Glu Asp Leu Asn

945 950 955 960

Tyr Gly Phe Lys Thr Gly Arg Phe Lys Val Glu Arg Gln Val Tyr Gln

965 970 975

Lys Phe Glu Thr Met Leu Ile Glu Lys Leu His Tyr Leu Val Phe Lys

980 985 990

Asp Arg Glu Val Cys Glu Glu Gly Gly Val Leu Arg Gly Tyr Gln Leu

995 1000 1005

Thr Tyr Ile Pro Glu Ser Leu Lys Lys Val Gly Lys Gln Cys Gly

1010 1015 1020

Phe Ile Phe Tyr Val Pro Ala Gly Tyr Thr Ser Lys Ile Asp Pro

1025 1030 1035

Thr Thr Gly Phe Val Asn Leu Phe Ser Phe Lys Asn Leu Thr Asn

1040 1045 1050

Arg Glu Ser Arg Gln Asp Phe Val Gly Lys Phe Asp Glu Ile Arg

1055 1060 1065

Tyr Asp Arg Asp Lys Lys Met Phe Glu Phe Ser Phe Asp Tyr Asn

1070 1075 1080

Asn Tyr Ile Lys Lys Gly Thr Ile Leu Ala Ser Thr Lys Trp Lys

1085 1090 1095

Val Tyr Thr Asn Gly Thr Arg Leu Lys Arg Ile Val Val Asn Gly

1100 1105 1110

Lys Tyr Thr Ser Gln Ser Met Glu Val Glu Leu Thr Asp Ala Met

1115 1120 1125

Glu Lys Met Leu Gln Arg Ala Gly Ile Glu Tyr His Asp Gly Lys

1130 1135 1140

Asp Leu Lys Gly Gln Ile Val Glu Lys Gly Ile Glu Ala Glu Ile

1145 1150 1155

Ile Asp Ile Phe Arg Leu Thr Val Gln Met Arg Asn Ser Arg Ser

1160 1165 1170

Glu Ser Glu Asp Arg Glu Tyr Asp Arg Leu Ile Ser Pro Val Leu

1175 1180 1185

Asn Asp Lys Gly Glu Phe Phe Asp Thr Ala Thr Ala Asp Lys Thr

1190 1195 1200

Leu Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr Cys Ile Ala Leu

1205 1210 1215

Lys Gly Leu Tyr Glu Val Lys Gln Ile Lys Glu Asn Trp Lys Glu

1220 1225 1230

Asn Glu Gln Phe Pro Arg Asn Lys Leu Val Gln Asp Asn Lys Thr

1235 1240 1245

Trp Phe Asp Phe Met Gln Lys Lys Arg Tyr Leu

1250 1255

<210> 67

<211> 1306

<212> PRT

<213> Prevotella copri

<400> 67

Met Ala Lys Asn Phe Glu Asp Phe Lys Arg Leu Tyr Ser Leu Ser Lys

1 5 10 15

Thr Leu Arg Phe Glu Ala Lys Pro Ile Gly Ala Thr Leu Asp Asn Ile

20 25 30

Val Lys Ser Gly Leu Leu Asp Glu Asp Glu His Arg Ala Ala Ser Tyr

35 40 45

Val Lys Val Lys Lys Leu Ile Asp Glu Tyr His Lys Val Phe Ile Asp

50 55 60

Arg Val Leu Asp Asp Gly Cys Leu Pro Leu Glu Asn Lys Gly Asn Asn

65 70 75 80

Asn Ser Leu Ala Glu Tyr Tyr Glu Ser Tyr Val Ser Arg Ala Gln Asp

85 90 95

Glu Asp Ala Lys Lys Lys Phe Lys Glu Ile Gln Gln Asn Leu Arg Ser

100 105 110

Val Ile Ala Lys Lys Leu Thr Glu Asp Lys Ala Tyr Ala Asn Leu Phe
115 120 125
Gly Asn Lys Leu Ile Glu Ser Tyr Lys Asp Lys Glu Asp Lys Lys Lys

130 135 140
Ile Ile Asp Ser Asp Leu Ile Gln Phe Ile Asn Thr Ala Glu Ser Thr
145 150 155 160
Gln Leu Asp Ser Met Ser Gln Asp Glu Ala Lys Glu Leu Val Lys Glu
165 170 175
Phe Trp Gly Phe Val Thr Tyr Phe Tyr Gly Phe Phe Asp Asn Arg Lys
180 185 190
Asn Met Tyr Thr Ala Glu Glu Lys Ser Thr Gly Ile Ala Tyr Arg Leu

195 200 205
Val Asn Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Glu Ala Phe Asn
210 215 220
Arg Ala Ile Thr Arg Pro Glu Ile Gln Glu Asn Met Gly Val Leu Tyr
225 230 235 240
Ser Asp Phe Ser Glu Tyr Leu Asn Val Glu Ser Ile Gln Glu Met Phe
245 250 255
Gln Leu Asp Tyr Tyr Asn Met Leu Leu Thr Gln Lys Gln Ile Asp Val

260 265 270
Tyr Asn Ala Ile Ile Gly Gly Lys Thr Asp Asp Glu His Asp Val Lys
275 280 285
Ile Lys Gly Ile Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Gln His Lys
290 295 300
Asp Asp Lys Leu Pro Lys Leu Lys Ala Leu Phe Lys Gln Ile Leu Ser
305 310 315 320
Asp Arg Asn Ala Ile Ser Trp Leu Pro Glu Glu Phe Asn Ser Asp Gln

325 330 335
Glu Val Leu Asn Ala Ile Lys Asp Cys Tyr Glu Arg Leu Ala Glu Asn
340 345 350
Val Leu Gly Asp Lys Val Leu Lys Ser Leu Leu Gly Ser Leu Ala Asp

355 360 365

Tyr Ser Leu Asp Gly Ile Phe Ile Arg Asn Asp Leu Gln Leu Thr Asp

370 375 380

Ile Ser Gln Lys Met Phe Gly Asn Trp Gly Val Ile Gln Asn Ala Ile

385 390 395 400

Met Gln Asn Ile Lys Arg Val Ala Pro Ala Arg Lys His Lys Glu Ser

405 410 415

Glu Glu Asp Tyr Glu Lys Arg Ile Ala Gly Ile Phe Lys Lys Ala Asp

420 425 430

Ser Phe Ser Ile Ser Tyr Ile Asn Asp Cys Leu Asn Glu Ala Asp Pro

435 440 445

Asn Asn Ala Tyr Phe Val Glu Asn Tyr Phe Ala Thr Phe Gly Ala Val

450 455 460

Asn Thr Pro Thr Met Gln Arg Glu Asn Leu Phe Ala Leu Val Gln Asn

465 470 475 480

Ala Tyr Thr Glu Val Ala Ala Leu Leu His Ser Asp Tyr Pro Thr Val

485 490 495

Lys His Leu Ala Gln Asp Lys Ala Asn Val Ser Lys Ile Lys Ala Leu

500 505 510

Leu Asp Ala Ile Lys Ser Leu Gln His Phe Val Lys Pro Leu Leu Gly

515 520 525

Lys Gly Asp Glu Ser Asp Lys Asp Glu Arg Phe Tyr Gly Glu Leu Ala

530 535 540

Ser Leu Trp Ala Glu Leu Asp Thr Val Thr Pro Leu Tyr Asn Met Ile

545 550 555 560

Arg Asn Tyr Met Thr Arg Lys Pro Tyr Ser Gln Lys Lys Ile Lys Leu

565 570 575

Asn Phe Glu Asn Pro Gln Leu Leu Gly Gly Trp Asp Ala Asn Lys Glu

580 585 590

Lys Asp Tyr Ala Thr Ile Ile Leu Arg Arg Asn Gly Leu Tyr Tyr Leu

595 600 605

Ala Ile Met Asp Lys Asp Ser Arg Lys Leu Leu Gly Lys Ala Met Pro
610 615 620
Ser Asp Gly Glu Cys Tyr Glu Lys Met Val Tyr Lys Phe Phe Lys Asp
625 630 635 640
Val Thr Thr Met Ile Pro Lys Cys Ser Thr Gln Leu Lys Asp Val Gln

645 650 655
Ala Tyr Phe Lys Val Asn Thr Asp Asp Tyr Val Leu Asn Ser Lys Ala
660 665 670
Phe Asn Lys Pro Leu Thr Ile Thr Lys Glu Val Phe Asp Leu Asn Asn
675 680 685
Val Leu Tyr Gly Lys Tyr Lys Lys Phe Gln Lys Gly Tyr Leu Thr Ala
690 695 700
Thr Gly Asp Asn Val Gly Tyr Thr His Ala Val Asn Val Trp Ile Lys

705 710 715 720
Phe Cys Met Asp Phe Leu Asn Ser Tyr Asp Ser Thr Cys Ile Tyr Asp
725 730 735
Phe Ser Ser Leu Lys Pro Glu Ser Tyr Leu Ser Leu Asp Ala Phe Tyr
740 745 750
Gln Asp Ala Asn Leu Leu Leu Tyr Lys Leu Ser Phe Ala Arg Ala Ser
755 760 765
Val Ser Tyr Ile Asn Gln Leu Val Glu Glu Gly Lys Met Tyr Leu Phe

770 775 780
Gln Ile Tyr Asn Lys Asp Phe Ser Glu Tyr Ser Lys Gly Thr Pro Asn
785 790 795 800
Met His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn Leu Ala
805 810 815
Asp Val Val Tyr Lys Leu Asn Gly Gln Ala Glu Met Phe Tyr Arg Lys
820 825 830
Lys Ser Ile Glu Asn Thr His Pro Thr His Pro Ala Asn His Pro Ile

835 840 845
Leu Asn Lys Asn Lys Asp Asn Lys Lys Lys Glu Ser Leu Phe Asp Tyr

850 855 860

Asp Leu Ile Lys Asp Arg Arg Tyr Thr Val Asp Lys Phe Met Phe His

865 870 875 880

Val Pro Ile Thr Met Asn Phe Lys Ser Val Gly Ser Glu Asn Ile Asn

885 890 895

Gln Asp Val Lys Ala Tyr Leu Arg His Ala Asp Asp Met His Ile Ile

900 905 910

Gly Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Leu Val Val Ile Asp

915 920 925

Leu Gln Gly Asn Ile Lys Glu Gln Tyr Ser Leu Asn Glu Ile Val Asn

930 935 940

Glu Tyr Asn Gly Asn Thr Tyr His Thr Asn Tyr His Asp Leu Leu Asp

945 950 955 960

Val Arg Glu Glu Glu Arg Leu Lys Ala Arg Gln Ser Trp Gln Thr Ile

965 970 975

Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Ile His

980 985 990

Lys Ile Thr Gln Leu Met Val Arg Tyr His Ala Ile Val Val Leu Glu

995 1000 1005

Asp Leu Ser Lys Gly Phe Met Arg Ser Arg Gln Lys Val Glu Lys

1010 1015 1020

Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn

1025 1030 1035

Tyr Leu Val Asp Lys Lys Thr Asp Val Ser Thr Pro Gly Gly Leu

1040 1045 1050

Leu Asn Ala Tyr Gln Leu Thr Cys Lys Ser Asp Ser Ser Gln Lys

1055 1060 1065

Leu Gly Lys Gln Ser Gly Phe Leu Phe Tyr Ile Pro Ala Trp Asn

1070 1075 1080

Thr Ser Lys Ile Asp Pro Val Thr Gly Phe Val Asn Leu Leu Asp

1085 1090 1095

Thr His Ser Leu Asn Ser Lys Glu Lys Ile Lys Ala Phe Phe Ser
1100 1105 1110
Lys Phe Asp Ala Ile Arg Tyr Asn Lys Asp Lys Lys Trp Phe Glu
1115 1120 1125
Phe Asn Leu Asp Tyr Asp Lys Phe Gly Lys Lys Ala Glu Asp Thr
1130 1135 1140
Arg Thr Lys Trp Thr Leu Cys Thr Arg Gly Met Arg Ile Asp Thr

1145 1150 1155
Phe Arg Asn Lys Glu Lys Asn Ser Gln Trp Asp Asn Gln Glu Val
1160 1165 1170
Asp Leu Thr Thr Glu Met Lys Ser Leu Leu Glu His Tyr Tyr Ile
1175 1180 1185
Asp Ile His Gly Asn Leu Lys Asp Ala Ile Ser Ala Gln Thr Asp
1190 1195 1200
Lys Ala Phe Phe Thr Gly Leu Leu His Ile Leu Lys Leu Thr Leu

1205 1210 1215
Gln Met Arg Asn Ser Ile Thr Gly Thr Glu Thr Asp Tyr Leu Val
1220 1225 1230
Ser Pro Val Ala Asp Glu Asn Gly Ile Phe Tyr Asp Ser Arg Ser
1235 1240 1245
Cys Gly Asn Gln Leu Pro Glu Asn Ala Asp Ala Asn Gly Ala Tyr
1250 1255 1260
Asn Ile Ala Arg Lys Gly Leu Met Leu Ile Glu Gln Ile Lys Asn

1265 1270 1275
Ala Glu Asp Leu Asn Asn Val Lys Phe Asp Ile Ser Asn Lys Ala
1280 1285 1290
Trp Leu Asn Phe Ala Gln Gln Lys Pro Tyr Lys Asn Gly
1295 1300 1305

<210> 68

<211> 1142

<212> PRT

<213> Eubacterium rectale

<400> 68

Met Phe Ser Ala Lys Leu Ile Ser Asp Ile Leu Pro Glu Phe Val Ile
1 5 10 15

His Asn Asn Asn Tyr Ser Ala Ser Glu Lys Glu Glu Lys Thr Gln Val
20 25 30

Ile Lys Leu Phe Ser Arg Phe Ala Thr Ser Phe Lys Asp Tyr Phe Lys
35 40 45

Asn Arg Ala Asn Cys Phe Ser Ala Asn Asp Ile Ser Ser Ser Ser Cys
50 55 60

His Arg Ile Val Asn Asp Asn Ala Glu Ile Phe Phe Ser Asn Ala Leu
65 70 75 80

Val Tyr Arg Arg Ile Val Lys Asn Leu Ser Asn Asp Asp Ile Asn Lys
85 90 95

Ile Ser Gly Asp Met Lys Asp Ser Leu Lys Glu Met Ser Leu Glu Glu
100 105 110

Ile Tyr Ser Tyr Glu Lys Tyr Gly Glu Phe Ile Thr Gln Glu Gly Ile
115 120 125

Ser Phe Tyr Asn Asp Ile Cys Gly Lys Val Asn Leu Phe Met Asn Leu
130 135 140

Tyr Cys Gln Lys Asn Lys Glu Asn Lys Asn Leu Tyr Lys Leu Arg Lys
145 150 155 160

Leu His Lys Gln Ile Leu Cys Ile Ala Asp Thr Ser Tyr Glu Val Pro
165 170 175

Tyr Lys Phe Glu Ser Asp Glu Glu Val Tyr Gln Ser Val Asn Gly Phe
180 185 190

Leu Asp Asn Ile Ser Ser Lys His Ile Val Glu Arg Leu Arg Lys Ile
195 200 205

Gly Glu Asn Tyr Asn Gly Tyr Asn Leu Asp Lys Ile Tyr Ile Val Ser
210 215 220

Lys Phe Tyr Glu Ser Val Ser Gln Lys Thr Tyr Arg Asp Trp Glu Thr
225 230 235 240

Ile Asn Thr Ala Leu Glu Ile His Tyr Asn Asn Ile Leu Pro Gly Asn
245 250 255
Gly Lys Ser Lys Ala Asp Lys Val Lys Lys Ala Val Lys Asn Asp Leu
260 265 270

Gln Lys Ser Ile Thr Glu Ile Asn Glu Leu Val Ser Asn Tyr Lys Leu
275 280 285
Cys Pro Asp Asp Asn Ile Lys Ala Glu Thr Tyr Ile His Glu Ile Ser
290 295 300
His Ile Leu Asn Asn Phe Glu Ala Gln Glu Leu Lys Tyr Asn Pro Glu
305 310 315 320
Ile His Leu Val Glu Ser Glu Leu Lys Ala Ser Glu Leu Lys Asn Val
325 330 335

Leu Asp Val Ile Met Asn Ala Phe His Trp Cys Ser Val Phe Met Thr
340 345 350
Glu Glu Leu Val Asp Lys Asp Asn Asn Phe Tyr Ala Glu Leu Glu Glu
355 360 365
Ile Tyr Asp Glu Ile Tyr Pro Val Ile Ser Leu Tyr Asn Leu Val Arg
370 375 380
Asn Tyr Val Thr Gln Lys Pro Tyr Ser Thr Lys Lys Ile Lys Leu Asn
385 390 395 400

Phe Gly Ile Pro Thr Leu Ala Asp Gly Trp Ser Lys Ser Lys Glu Tyr
405 410 415
Ser Asn Asn Ala Ile Ile Leu Met Arg Asp Asn Leu Tyr Tyr Leu Gly
420 425 430
Ile Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Glu Gly Asn
435 440 445
Thr Ser Glu Asn Lys Gly Asp Tyr Lys Lys Met Ile Tyr Asn Leu Leu
450 455 460

Pro Gly Pro Asn Lys Met Ile Pro Lys Val Phe Leu Ser Ser Lys Thr
465 470 475 480
Gly Val Glu Thr Tyr Lys Pro Ser Ala Tyr Ile Leu Glu Gly Tyr Lys

485 490 495

Gln Asn Lys His Leu Lys Ser Ser Lys Asp Phe Asp Ile Thr Phe Cys

500 505 510

His Asp Leu Ile Asp Tyr Phe Lys Asn Cys Ile Ala Ile His Pro Glu

515 520 525

Trp Lys Asn Phe Gly Phe Asp Phe Ser Asp Thr Ser Thr Tyr Glu Asp

530 535 540

Ile Ser Gly Phe Tyr Arg Glu Val Glu Leu Gln Gly Tyr Lys Ile Asp

545 550 555 560

Trp Thr Tyr Ile Ser Glu Lys Asp Ile Asp Leu Leu Gln Glu Lys Gly

565 570 575

Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Lys Lys Ser

580 585 590

Ser Gly Asn Asp Asn Leu His Thr Met Tyr Leu Lys Asn Leu Phe Ser

595 600 605

Glu Glu Asn Leu Lys Asp Ile Val Leu Lys Leu Asn Gly Glu Ala Glu

610 615 620

Ile Phe Phe Arg Lys Ser Ser Ile Lys Asn Pro Ile Ile His Lys Lys

625 630 635 640

Gly Ser Ile Leu Val Asn Arg Thr Tyr Glu Ala Glu Glu Lys Asp Gln

645 650 655

Phe Gly Asn Ile Gln Ile Val Arg Lys Thr Ile Pro Glu Asn Ile Tyr

660 665 670

Gln Glu Leu Tyr Lys Tyr Phe Asn Asp Lys Ser Asp Lys Glu Leu Ser

675 680 685

Asp Glu Ala Ala Lys Leu Lys Asn Val Val Gly His His Glu Ala Ala

690 695 700

Thr Asn Ile Val Lys Asp Tyr Arg Tyr Thr Tyr Asp Lys Tyr Phe Leu

705 710 715 720

His Met Pro Ile Thr Ile Asn Phe Lys Ala Asn Lys Thr Ser Phe Ile

725 730 735

Asn Asp Arg Ile Leu Gln Tyr Ile Ala Lys Glu Lys Asp Leu His Val
740 745 750

Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Val Ser Val Ile
755 760 765

Asp Thr Cys Gly Asn Ile Val Glu Gln Lys Ser Phe Asn Ile Val Asn
770 775 780

Gly Tyr Asp Tyr Gln Ile Lys Leu Lys Gln Gln Glu Gly Ala Arg Gln
785 790 795 800

Ile Ala Arg Lys Glu Trp Lys Glu Ile Gly Lys Ile Lys Glu Ile Lys
805 810 815

Glu Gly Tyr Leu Ser Leu Val Ile His Glu Ile Ser Lys Met Val Ile
820 825 830

Lys Tyr Asn Ala Ile Ile Ala Met Glu Asp Leu Ser Tyr Gly Phe Lys
835 840 845

Lys Gly Arg Phe Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Thr
850 855 860

Met Leu Ile Asn Lys Leu Asn Tyr Leu Val Phe Lys Asp Ile Ser Ile
865 870 875 880

Thr Glu Asn Gly Gly Leu Leu Lys Gly Tyr Gln Leu Thr Tyr Ile Pro
885 890 895

Asp Lys Leu Lys Asn Val Gly His Gln Cys Gly Cys Ile Phe Tyr Val
900 905 910

Pro Ala Ala Tyr Thr Ser Lys Ile Asp Pro Thr Thr Gly Phe Val Asn
915 920 925

Ile Phe Lys Phe Lys Asp Leu Thr Val Asp Ala Lys Arg Glu Phe Ile
930 935 940

Lys Lys Phe Asp Ser Ile Arg Tyr Asp Ser Asp Lys Asn Leu Phe Cys
945 950 955 960

Phe Thr Phe Asp Tyr Asn Asn Phe Ile Thr Gln Asn Thr Val Met Ser
965 970 975

Lys Ser Ser Trp Ser Val Tyr Thr Tyr Gly Val Arg Ile Lys Arg Arg

980 985 990

Phe Val Asn Gly Arg Phe Ser Asn Glu Ser Asp Thr Ile Asp Ile Thr

995 1000 1005

Lys Asp Met Glu Lys Thr Leu Glu Met Thr Asp Ile Asn Trp Arg

1010 1015 1020

Asp Gly His Asp Leu Arg Gln Asp Ile Ile Asp Tyr Glu Ile Val

1025 1030 1035

Gln His Ile Phe Glu Ile Phe Lys Leu Thr Val Gln Met Arg Asn

1040 1045 1050

Ser Leu Ser Glu Leu Glu Asp Arg Asp Tyr Asp Arg Leu Ile Ser

1055 1060 1065

Pro Val Leu Asn Glu Asn Asn Ile Phe Tyr Asp Ser Ala Lys Ala

1070 1075 1080

Gly Asp Ala Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Cys

1085 1090 1095

Ile Ala Leu Lys Gly Leu Tyr Glu Ile Lys Gln Ile Thr Glu Asn

1100 1105 1110

Trp Lys Glu Asp Gly Lys Phe Ser Arg Asp Lys Leu Lys Ile Ser

1115 1120 1125

Asn Lys Asp Trp Phe Asp Phe Ile Gln Asn Lys Arg Tyr Leu

1130 1135 1140

<210> 69

<211> 1259

<212> PRT

<213> Unknown

<220><223> Clostridium sp. AF34-10BH

<400> 69

Met Asn Tyr Lys Thr Gly Leu Glu Asp Phe Ile Gly Lys Glu Ser Leu

1 5 10 15

Ser Lys Thr Leu Arg Asn Ala Leu Ile Pro Thr Glu Ser Thr Lys Ile

20 25 30

His Met Glu Glu Met Gly Val Ile Arg Asp Asp Glu Leu Arg Ala Glu

35 40 45

Lys Gln Gln Glu Leu Lys Glu Ile Met Asp Asp Tyr Tyr Arg Ala Phe

50 55 60

Ile Glu Glu Lys Leu Gly Gln Ile Gln Gly Ile Gln Trp Asn Ser Leu

65 70 75 80

Phe Gln Lys Met Glu Glu Thr Met Glu Asp Ile Ser Val Arg Lys Asp

85 90 95

Leu Asp Lys Ile Gln Asn Glu Lys Arg Lys Glu Ile Cys Cys Tyr Phe

100 105 110

Thr Ser Asp Lys Arg Phe Lys Asp Leu Phe Asn Ala Lys Leu Ile Thr

115 120 125

Asp Ile Leu Pro Asn Phe Ile Lys Asp Asn Lys Glu Tyr Thr Glu Glu

130 135 140

Glu Lys Ala Glu Lys Glu Gln Thr Arg Val Leu Phe Gln Arg Phe Ala

145 150 155 160

Thr Ala Phe Thr Asn Tyr Phe Asn Gln Arg Arg Asn Asn Phe Ser Glu

165 170 175

Asp Asn Ile Ser Thr Ala Ile Ser Phe Arg Ile Val Asn Glu Asn Ser

180 185 190

Glu Ile His Leu Gln Asn Met Arg Ala Phe Gln Arg Ile Glu Gln Gln

195 200 205

Tyr Pro Glu Glu Val Cys Gly Met Glu Glu Glu Tyr Lys Asp Met Leu

210 215 220

Gln Glu Trp Gln Met Lys His Ile Tyr Leu Val Asp Phe Tyr Asp Arg

225 230 235 240

Val Leu Thr Gln Pro Gly Ile Glu Tyr Tyr Asn Gly Ile Cys Gly Lys

245 250 255

Ile Asn Glu His Met Asn Gln Phe Cys Gln Lys Asn Arg Ile Asn Lys

260 265 270

Asn Asp Phe Arg Met Lys Lys Leu His Lys Gln Ile Leu Cys Lys Lys

275 280 285

Ser Ser Tyr Tyr Glu Ile Pro Phe Arg Phe Glu Ser Asp Gln Glu Val

290 295 300

Tyr Asp Ala Leu Asn Glu Phe Ile Lys Thr Met Lys Glu Lys Glu Ile

305 310 315 320

Ile Cys Arg Cys Val His Leu Gly Gln Lys Cys Asp Asp Tyr Asp Leu

325 330 335

Gly Lys Ile Tyr Ile Ser Ser Asn Lys Tyr Glu Gln Ile Ser Asn Ala

340 345 350

Leu Tyr Gly Ser Trp Asp Thr Ile Arg Lys Cys Ile Lys Glu Glu Tyr

355 360 365

Met Asp Ala Leu Pro Gly Lys Gly Glu Lys Lys Glu Glu Lys Ala Glu

370 375 380

Ala Ala Ala Lys Lys Glu Glu Tyr Arg Ser Ile Ala Asp Ile Asp Lys

385 390 395 400

Ile Ile Ser Leu Tyr Gly Ser Glu Met Asp Arg Thr Ile Ser Ala Lys

405 410 415

Lys Cys Ile Thr Glu Ile Cys Asp Met Ala Gly Gln Ile Ser Thr Asp

420 425 430

Pro Leu Val Cys Asn Ser Asp Ile Lys Leu Leu Gln Asn Lys Glu Lys

435 440 445

Thr Thr Glu Ile Lys Thr Ile Leu Asp Ser Phe Leu His Val Tyr Gln

450 455 460

Trp Gly Gln Thr Phe Ile Val Ser Asp Ile Ile Glu Lys Asp Ser Tyr

465 470 475 480

Phe Tyr Ser Glu Leu Glu Asp Val Leu Glu Asp Phe Glu Gly Ile Thr

485 490 495

Thr Leu Tyr Asn His Val Arg Ser Tyr Val Thr Gln Lys Pro Tyr Ser

500 505 510

Thr Val Lys Phe Lys Leu His Phe Gly Ser Pro Thr Leu Ala Asn Gly

515 520 525

Trp Ser Gln Ser Lys Glu Tyr Asp Asn Asn Ala Ile Leu Leu Met Arg

530 535 540

Asp Gln Lys Phe Tyr Leu Gly Ile Phe Asn Val Arg Asn Lys Pro Asp

545 550 555 560

Lys Gln Ile Ile Lys Gly His Glu Lys Glu Glu Lys Gly Asp Tyr Lys

565 570 575

Lys Met Ile Tyr Asn Leu Leu Pro Gly Pro Ser Lys Met Leu Pro Lys

580 585 590

Val Phe Ile Thr Ser Arg Ser Gly Gln Glu Thr Tyr Lys Pro Ser Lys

595 600 605

His Ile Leu Asp Gly Tyr Asn Glu Lys Arg His Ile Lys Ser Ser Pro

610 615 620

Lys Phe Asp Leu Gly Tyr Cys Trp Asp Leu Ile Asp Tyr Tyr Lys Glu

625 630 635 640

Cys Ile His Lys His Pro Asp Trp Lys Asn Tyr Asp Phe His Phe Ser

645 650 655

Asp Thr Lys Asp Tyr Glu Asp Ile Ser Gly Phe Tyr Arg Glu Val Glu

660 665 670

Met Gln Gly Tyr Gln Ile Lys Trp Thr Tyr Ile Ser Ala Asp Glu Ile

675 680 685

Gln Lys Leu Asp Glu Lys Gly Gln Ile Phe Leu Phe Gln Ile Tyr Asn

690 695 700

Lys Asp Phe Ser Val His Ser Thr Gly Lys Asp Asn Leu His Thr Met

705 710 715 720

Tyr Leu Lys Asn Leu Phe Ser Glu Glu Asn Leu Lys Asp Ile Val Leu

725 730 735

Lys Leu Asn Gly Glu Ala Glu Leu Phe Phe Arg Lys Ala Ser Ile Lys

740 745 750

Thr Pro Val Val His Lys Lys Gly Ser Val Leu Val Asn Arg Ser Tyr

755 760 765

Thr Gln Thr Val Gly Asp Lys Glu Ile Arg Val Ser Ile Pro Glu Glu

770 775 780

Tyr Tyr Thr Glu Ile Tyr Asn Tyr Leu Asn His Ile Gly Arg Gly Lys
785 790 795 800
Leu Ser Thr Glu Ala Gln Arg Tyr Leu Glu Glu Arg Lys Ile Lys Ser
805 810 815
Phe Thr Ala Thr Lys Asp Ile Val Lys Asn Tyr Arg Tyr Cys Cys Asp
820 825 830
His Tyr Phe Leu His Leu Pro Ile Thr Ile Asn Phe Lys Ala Lys Ser

835 840 845
Asp Ile Ala Val Asn Glu Arg Thr Leu Ala Tyr Ile Ala Lys Lys Glu
850 855 860
Asp Ile His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr
865 870 875 880
Ile Ser Val Val Asp Val His Gly Asn Ile Arg Glu Gln Arg Ser Phe
885 890 895
Asn Ile Val Asn Gly Tyr Asp Tyr Gln Gln Lys Leu Lys Asp Arg Glu

900 905 910
Lys Ser Arg Asp Ala Ala Arg Lys Asn Trp Glu Glu Ile Glu Lys Ile
915 920 925
Lys Glu Leu Lys Glu Gly Tyr Leu Ser Met Val Ile His Tyr Ile Ala
930 935 940
Gln Leu Val Val Lys Tyr Asn Ala Val Val Ala Met Glu Asp Leu Asn
945 950 955 960
Tyr Gly Phe Lys Thr Gly Arg Phe Lys Val Glu Arg Gln Val Tyr Gln

965 970 975
Lys Phe Glu Thr Met Leu Ile Glu Lys Leu His Tyr Leu Val Phe Lys
980 985 990
Asp Arg Glu Val Cys Glu Glu Gly Gly Val Leu Arg Gly Tyr Gln Leu
995 1000 1005
Thr Tyr Ile Pro Glu Ser Leu Lys Lys Val Gly Lys Gln Cys Gly
1010 1015 1020
Phe Ile Phe Tyr Val Pro Ala Gly Tyr Thr Ser Lys Ile Asp Pro

1025 1030 1035

Thr Thr Gly Phe Val Asn Leu Phe Ser Phe Lys Asn Leu Thr Asn

1040 1045 1050

Arg Glu Ser Arg Gln Asp Phe Val Gly Lys Phe Asp Glu Ile Arg

1055 1060 1065

Tyr Asp Arg Asp Lys Lys Met Phe Glu Phe Ser Phe Asp Tyr Asn

1070 1075 1080

Asn Tyr Ile Lys Lys Gly Thr Met Leu Ala Ser Thr Lys Trp Lys

1085 1090 1095

Val Tyr Thr Asn Gly Thr Arg Leu Lys Arg Ile Val Val Asn Gly

1100 1105 1110

Lys Tyr Thr Ser Gln Ser Met Glu Val Glu Leu Thr Asp Ala Met

1115 1120 1125

Glu Lys Met Leu Gln Arg Ala Gly Ile Glu Tyr His Asp Gly Lys

1130 1135 1140

Asp Leu Lys Gly Gln Ile Val Glu Lys Gly Ile Glu Ala Glu Ile

1145 1150 1155

Ile Asp Ile Phe Arg Leu Thr Val Gln Met Arg Asn Ser Arg Ser

1160 1165 1170

Glu Ser Glu Asp Arg Glu Tyr Asp Arg Leu Ile Ser Pro Val Leu

1175 1180 1185

Asn Asp Lys Gly Glu Phe Phe Asp Thr Ala Thr Ala Asp Lys Thr

1190 1195 1200

Leu Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr Cys Ile Ala Leu

1205 1210 1215

Lys Gly Leu Tyr Glu Val Lys Gln Ile Lys Glu Asn Trp Lys Glu

1220 1225 1230

Asn Glu Gln Phe Pro Arg Asn Lys Leu Val Gln Asp Asn Lys Thr

1235 1240 1245

Trp Phe Asp Phe Met Gln Lys Lys Arg Tyr Leu

1250 1255

<210> 70

<211> 1108

<212> PRT

<213> Unknown

<220><223> Bacillushisashii

<

400> 70

Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
1 5 10 15

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
35 40 45

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
50 55 60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
65 70 75 80

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
100 105 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
115 120 125

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
130 135 140

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
145 150 155 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
165 170 175

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
180 185 190

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
195 200 205

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
210 215 220
Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
225 230 235 240
Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
245 250 255

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
260 265 270
Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
275 280 285
Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
290 295 300
Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
305 310 315 320

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
325 330 335
Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
340 345 350
Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
355 360 365
Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
370 375 380

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
385 390 395 400
Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
405 410 415
Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
420 425 430
Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
435 440 445

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His

450 455 460

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr

465 470 475 480

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

485 490 495

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met

500 505 510

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys

515 520 525

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu

530 535 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

545 550 555 560

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly

565 570 575

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys

580 585 590

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu

595 600 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

610 615 620

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu

625 630 635 640

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

645 650 655

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

660 665 670

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

675 680 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

690 695 700

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
705 710 715 720
Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
725 730 735
Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
740 745 750
Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
755 760 765

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
770 775 780
Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
785 790 795 800
Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
805 810 815
Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
820 825 830

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
835 840 845
Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
850 855 860
Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
865 870 875 880
Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
885 890 895

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
900 905 910
Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
915 920 925
Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
930 935 940
Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln

945 950 955 960

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
965 970 975
Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp
980 985 990
Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu
995 1000 1005
Lys Asp Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys Ile
1010 1015 1020

Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser
1025 1030 1035
Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly
1040 1045 1050
Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro
1055 1060 1065
Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu
1070 1075 1080

Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr
1085 1090 1095
Ile Glu Asp Asp Ser Ser Lys Gln Ser Met
1100 1105

<210> 71
<211> 1465
<212> PRT
<213> Thermomonas hydrothermalis
<400> 71
Met Ser Glu Lys Thr Thr Gln Arg Ala Tyr Thr Leu Arg Leu Asn Arg
1 5 10 15
Ala Ser Gly Glu Cys Ala Val Cys Gln Asn Asn Ser Cys Asp Cys Trp
20 25 30

His Asp Ala Leu Trp Ala Thr His Lys Ala Val Asn Arg Gly Ala Lys

35 40 45

Ala Phe Gly Asp Trp Leu Leu Thr Leu Arg Gly Gly Leu Cys His Thr

50 55 60

Leu Val Glu Met Glu Val Pro Ala Lys Gly Asn Asn Pro Pro Gln Arg

65 70 75 80

Pro Thr Asp Gln Glu Arg Arg Asp Arg Arg Val Leu Leu Ala Leu Ser

85 90 95

Trp Leu Ser Val Glu Asp Glu His Gly Ala Pro Lys Glu Phe Ile Val

100 105 110

Ala Thr Gly Arg Asp Ser Ala Asp Asp Arg Ala Lys Lys Val Glu Glu

115 120 125

Lys Leu Arg Glu Ile Leu Glu Lys Arg Asp Phe Gln Glu His Glu Ile

130 135 140

Asp Ala Trp Leu Gln Asp Cys Gly Pro Ser Leu Lys Ala His Ile Arg

145 150 155 160

Glu Asp Ala Val Trp Val Asn Arg Arg Ala Leu Phe Asp Ala Ala Val

165 170 175

Glu Arg Ile Lys Thr Leu Thr Trp Glu Glu Ala Trp Asp Phe Leu Glu

180 185 190

Pro Phe Phe Gly Thr Gln Tyr Phe Ala Gly Ile Gly Asp Gly Lys Asp

195 200 205

Lys Asp Asp Ala Glu Gly Pro Ala Arg Gln Gly Glu Lys Ala Lys Asp

210 215 220

Leu Val Gln Lys Ala Gly Gln Trp Leu Ser Ala Arg Phe Gly Ile Gly

225 230 235 240

Thr Gly Ala Asp Phe Met Ser Met Ala Glu Ala Tyr Glu Lys Ile Ala

245 250 255

Lys Trp Ala Ser Gln Ala Gln Asn Gly Asp Asn Gly Lys Ala Thr Ile

260 265 270

Glu Lys Leu Ala Cys Ala Leu Arg Pro Ser Glu Pro Pro Thr Leu Asp

275 280 285

Thr Val Leu Lys Cys Ile Ser Gly Pro Gly His Lys Ser Ala Thr Arg
290 295 300
Glu Tyr Leu Lys Thr Leu Asp Lys Lys Ser Thr Val Thr Gln Glu Asp
305 310 315 320
Leu Asn Gln Leu Arg Lys Leu Ala Asp Glu Asp Ala Arg Asn Cys Arg
325 330 335
Lys Lys Val Gly Lys Lys Gly Lys Lys Pro Trp Ala Asp Glu Val Leu
340 345 350

Lys Asp Val Glu Asn Ser Cys Glu Leu Thr Tyr Leu Gln Asp Asn Ser
355 360 365
Pro Ala Arg His Arg Glu Phe Ser Val Met Leu Asp His Ala Ala Arg
370 375 380
Arg Val Ser Met Ala His Ser Trp Ile Lys Lys Ala Glu Gln Arg Arg
385 390 395 400
Arg Gln Phe Glu Ser Asp Ala Gln Lys Leu Lys Asn Leu Gln Glu Arg
405 410 415

Ala Pro Ser Ala Val Glu Trp Leu Asp Arg Phe Cys Glu Ser Arg Ser
420 425 430
Met Thr Thr Gly Ala Asn Thr Gly Ser Gly Tyr Arg Ile Arg Lys Arg
435 440 445
Ala Ile Glu Gly Trp Ser Tyr Val Val Gln Ala Trp Ala Glu Ala Ser
450 455 460
Cys Asp Thr Glu Asp Lys Arg Ile Ala Ala Ala Arg Lys Val Gln Ala
465 470 475 480

Asp Pro Glu Ile Glu Lys Phe Gly Asp Ile Gln Leu Phe Glu Ala Leu
485 490 495
Ala Ala Asp Glu Ala Ile Cys Val Trp Arg Asp Gln Glu Gly Thr Gln
500 505 510
Asn Pro Ser Ile Leu Ile Asp Tyr Val Thr Gly Lys Thr Ala Glu His
515 520 525
Asn Gln Lys Arg Phe Lys Val Pro Ala Tyr Arg His Pro Asp Glu Leu

530 535 540

Arg His Pro Val Phe Cys Asp Phe Gly Asn Ser Arg Trp Ser Ile Gln
545 550 555 560
Phe Ala Ile His Lys Glu Ile Arg Asp Arg Asp Lys Gly Ala Lys Gln
565 570 575
Asp Thr Arg Gln Leu Gln Asn Arg His Gly Leu Lys Met Arg Leu Trp
580 585 590
Asn Gly Arg Ser Met Thr Asp Val Asn Leu His Trp Ser Ser Lys Arg
595 600 605

Leu Thr Ala Asp Leu Ala Leu Asp Gln Asn Pro Asn Pro Asn Pro Thr
610 615 620
Glu Val Thr Arg Ala Asp Arg Leu Gly Arg Ala Ala Ser Ser Ala Phe
625 630 635 640
Asp His Val Lys Ile Lys Asn Val Phe Asn Glu Lys Glu Trp Asn Gly
645 650 655
Arg Leu Gln Ala Pro Arg Ala Glu Leu Asp Arg Ile Ala Lys Leu Glu
660 665 670

Glu Gln Gly Lys Thr Glu Gln Ala Glu Lys Leu Arg Lys Arg Leu Arg
675 680 685
Trp Tyr Val Ser Phe Ser Pro Cys Leu Ser Pro Ser Gly Pro Phe Ile
690 695 700
Val Tyr Ala Gly Gln His Asn Ile Gln Pro Lys Arg Ser Gly Gln Tyr
705 710 715 720
Ala Pro His Ala Gln Ala Asn Lys Gly Arg Ala Arg Leu Ala Gln Leu
725 730 735

Ile Leu Ser Arg Leu Pro Asp Leu Arg Ile Leu Ser Val Asp Leu Gly
740 745 750
His Arg Phe Ala Ala Ala Cys Ala Val Trp Glu Thr Leu Ser Ser Asp
755 760 765
Ala Phe Arg Arg Glu Ile Gln Gly Leu Asn Val Leu Ala Gly Gly Ser
770 775 780

Gly Glu Gly Asp Leu Phe Leu His Val Glu Met Thr Gly Asp Asp Gly
785 790 795 800

Lys Arg Arg Thr Val Val Tyr Arg Arg Ile Gly Pro Asp Gln Leu Leu
805 810 815

Asp Asn Thr Pro His Pro Ala Pro Trp Ala Arg Leu Asp Arg Gln Phe
820 825 830

Leu Ile Lys Leu Gln Gly Glu Asp Glu Gly Val Arg Glu Ala Ser Asn
835 840 845

Glu Glu Leu Trp Thr Val His Lys Leu Glu Val Glu Val Gly Arg Thr
850 855 860

Val Pro Leu Ile Asp Arg Met Val Arg Ser Gly Phe Gly Lys Thr Glu
865 870 875 880

Lys Gln Lys Glu Arg Leu Lys Lys Leu Arg Glu Leu Gly Trp Ile Ser
885 890 895

Ala Met Pro Asn Glu Pro Ser Ala Glu Thr Asp Glu Lys Glu Gly Glu
900 905 910

Ile Arg Ser Ile Ser Arg Ser Val Asp Glu Leu Met Ser Ser Ala Leu
915 920 925

Gly Thr Leu Arg Leu Ala Leu Lys Arg His Gly Asn Arg Ala Arg Ile
930 935 940

Ala Phe Ala Met Thr Ala Asp Tyr Lys Pro Met Pro Gly Gly Gln Lys
945 950 955 960

Tyr Tyr Phe His Glu Ala Lys Glu Ala Ser Lys Asn Asp Asp Glu Thr
965 970 975

Lys Arg Arg Asp Asn Gln Ile Glu Phe Leu Gln Asp Ala Leu Ser Leu
980 985 990

Trp His Asp Leu Phe Ser Ser Pro Asp Trp Glu Asp Asn Glu Ala Lys
995 1000 1005

Lys Leu Trp Gln Asn His Ile Ala Thr Leu Pro Asn Tyr Gln Thr
1010 1015 1020

Pro Glu Glu Ile Ser Ala Glu Leu Lys Arg Val Glu Arg Asn Lys

1025 1030 1035

Lys Arg Lys Glu Asn Arg Asp Lys Leu Arg Thr Ala Ala Lys Ala

1040 1045 1050

Leu Ala Glu Asn Asp Gln Leu Arg Gln His Leu His Asp Thr Trp

1055 1060 1065

Lys Glu Arg Trp Glu Ser Asp Asp Gln Gln Trp Lys Glu Arg Leu

1070 1075 1080

Arg Ser Leu Lys Asp Trp Ile Phe Pro Arg Gly Lys Ala Glu Asp

1085 1090 1095

Asn Pro Ser Ile Arg His Val Gly Gly Leu Ser Ile Thr Arg Ile

1100 1105 1110

Asn Thr Ile Ser Gly Leu Tyr Gln Ile Leu Lys Ala Phe Lys Met

1115 1120 1125

Arg Pro Glu Pro Asp Asp Leu Arg Lys Asn Ile Pro Gln Lys Gly

1130 1135 1140

Asp Asp Glu Leu Glu Asn Phe Asn Arg Arg Leu Leu Glu Ala Arg

1145 1150 1155

Asp Arg Leu Arg Glu Gln Arg Val Lys Gln Leu Ala Ser Arg Ile

1160 1165 1170

Ile Glu Ala Ala Leu Gly Val Gly Arg Ile Lys Ile Pro Lys Asn

1175 1180 1185

Gly Lys Leu Pro Lys Arg Pro Arg Thr Thr Val Asp Thr Pro Cys

1190 1195 1200

His Ala Val Val Ile Glu Ser Leu Lys Thr Tyr Arg Pro Asp Asp

1205 1210 1215

Leu Arg Thr Arg Arg Glu Asn Arg Gln Leu Met Gln Trp Ser Ser

1220 1225 1230

Ala Lys Val Arg Lys Tyr Leu Lys Glu Gly Cys Glu Leu Tyr Gly

1235 1240 1245

Leu His Phe Leu Glu Val Pro Ala Asn Tyr Thr Ser Arg Gln Cys

1250 1255 1260

Ser Arg Thr Gly Leu Pro Gly Ile Arg Cys Asp Asp Val Pro Thr
1265 1270 1275
Gly Asp Phe Leu Lys Ala Pro Trp Trp Arg Arg Ala Ile Asn Thr
1280 1285 1290

Ala Arg Glu Lys Asn Gly Gly Asp Ala Lys Asp Arg Phe Leu Val
1295 1300 1305
Asp Leu Tyr Asp His Leu Asn Asn Leu Gln Ser Lys Gly Glu Ala
1310 1315 1320
Leu Pro Ala Thr Val Arg Val Pro Arg Gln Gly Gly Asn Leu Phe
1325 1330 1335
Ile Ala Gly Ala Gln Leu Asp Asp Thr Asn Lys Glu Arg Arg Ala
1340 1345 1350

Ile Gln Ala Asp Leu Asn Ala Ala Ala Asn Ile Gly Leu Arg Ala
1355 1360 1365
Leu Leu Asp Pro Asp Trp Arg Gly Arg Trp Trp Tyr Val Pro Cys
1370 1375 1380
Lys Asp Gly Thr Ser Glu Pro Ala Leu Asp Arg Ile Glu Gly Ser
1385 1390 1395
Thr Ala Phe Asn Asp Val Arg Ser Leu Pro Thr Gly Asp Asn Ser
1400 1405 1410

Ser Arg Arg Ala Pro Arg Glu Ile Glu Asn Leu Trp Arg Asp Pro
1415 1420 1425
Ser Gly Asp Ser Leu Glu Ser Gly Thr Trp Ser Pro Thr Arg Ala
1430 1435 1440
Tyr Trp Asp Thr Val Gln Ser Arg Val Ile Glu Leu Leu Arg Arg
1445 1450 1455
His Ala Gly Leu Pro Thr Ser
1460 1465
<210> 72
<211> 1090
<212> PRT

<213> Laceyella sacchari

<400> 72

Met Ser Ile Arg Ser Phe Lys Leu Lys Leu Lys Thr Lys Ser Gly Val
1 5 10 15
Asn Ala Glu Gln Leu Arg Arg Gly Leu Trp Arg Thr His Gln Leu Ile
20 25 30
Asn Asp Gly Ile Ala Tyr Tyr Met Asn Trp Leu Val Leu Leu Arg Gln
35 40 45
Glu Asp Leu Phe Ile Arg Asn Lys Glu Thr Asn Glu Ile Glu Lys Arg
50 55 60

Ser Lys Glu Glu Ile Gln Ala Val Leu Leu Glu Arg Val His Lys Gln
65 70 75 80
Gln Gln Arg Asn Gln Trp Ser Gly Glu Val Asp Glu Gln Thr Leu Leu
85 90 95
Gln Ala Leu Arg Gln Leu Tyr Glu Glu Ile Val Pro Ser Val Ile Gly
100 105 110
Lys Ser Gly Asn Ala Ser Leu Lys Ala Arg Phe Phe Leu Gly Pro Leu
115 120 125

Val Asp Pro Asn Asn Lys Thr Thr Lys Asp Val Ser Lys Ser Gly Pro
130 135 140
Thr Pro Lys Trp Lys Lys Met Lys Asp Ala Gly Asp Pro Asn Trp Val
145 150 155 160
Gln Glu Tyr Glu Lys Tyr Met Ala Glu Arg Gln Thr Leu Val Arg Leu
165 170 175
Glu Glu Met Gly Leu Ile Pro Leu Phe Pro Met Tyr Thr Asp Glu Val
180 185 190

Gly Asp Ile His Trp Leu Pro Gln Ala Ser Gly Tyr Thr Arg Thr Trp
195 200 205
Asp Arg Asp Met Phe Gln Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu
210 215 220
Ser Trp Asn Arg Arg Val Arg Glu Arg Arg Ala Gln Phe Glu Lys Lys

225 230 235 240

Thr His Asp Phe Ala Ser Arg Phe Ser Glu Ser Asp Val Gln Trp Met

245 250 255

Asn Lys Leu Arg Glu Tyr Glu Ala Gln Gln Glu Lys Ser Leu Glu Glu

260 265 270

Asn Ala Phe Ala Pro Asn Glu Pro Tyr Ala Leu Thr Lys Lys Ala Leu

275 280 285

Arg Gly Trp Glu Arg Val Tyr His Ser Trp Met Arg Leu Asp Ser Ala

290 295 300

Ala Ser Glu Glu Ala Tyr Trp Gln Glu Val Ala Thr Cys Gln Thr Ala

305 310 315 320

Met Arg Gly Glu Phe Gly Asp Pro Ala Ile Tyr Gln Phe Leu Ala Gln

325 330 335

Lys Glu Asn His Asp Ile Trp Arg Gly Tyr Pro Glu Arg Val Ile Asp

340 345 350

Phe Ala Glu Leu Asn His Leu Gln Arg Glu Leu Arg Arg Ala Lys Glu

355 360 365

Asp Ala Thr Phe Thr Leu Pro Asp Ser Val Asp His Pro Leu Trp Val

370 375 380

Arg Tyr Glu Ala Pro Gly Gly Thr Asn Ile His Gly Tyr Asp Leu Val

385 390 395 400

Gln Asp Thr Lys Arg Asn Leu Thr Leu Ile Leu Asp Lys Phe Ile Leu

405 410 415

Pro Asp Glu Asn Gly Ser Trp His Glu Val Lys Lys Val Pro Phe Ser

420 425 430

Leu Ala Lys Ser Lys Gln Phe His Arg Gln Val Trp Leu Gln Glu Glu

435 440 445

Gln Lys Gln Lys Lys Arg Glu Val Val Phe Tyr Asp Tyr Ser Thr Asn

450 455 460

Leu Pro His Leu Gly Thr Leu Ala Gly Ala Lys Leu Gln Trp Asp Arg

465 470 475 480

Asn Phe Leu Asn Lys Arg Thr Gln Gln Gln Ile Glu Glu Thr Gly Glu
485 490 495
Ile Gly Lys Val Phe Phe Asn Ile Ser Val Asp Val Arg Pro Ala Val
500 505 510

Glu Val Lys Asn Gly Arg Leu Gln Asn Gly Leu Gly Lys Ala Leu Thr
515 520 525
Val Leu Thr His Pro Asp Gly Thr Lys Ile Val Thr Gly Trp Lys Ala
530 535 540
Glu Gln Leu Glu Lys Trp Val Gly Glu Ser Gly Arg Val Ser Ser Leu
545 550 555 560
Gly Leu Asp Ser Leu Ser Glu Gly Leu Arg Val Met Ser Ile Asp Leu
565 570 575

Gly Gln Arg Thr Ser Ala Thr Val Ser Val Phe Glu Ile Thr Lys Glu
580 585 590
Ala Pro Asp Asn Pro Tyr Lys Phe Phe Tyr Gln Leu Glu Gly Thr Glu
595 600 605
Met Phe Ala Val His Gln Arg Ser Phe Leu Leu Ala Leu Pro Gly Glu
610 615 620
Asn Pro Pro Gln Lys Ile Lys Gln Met Arg Glu Ile Arg Trp Lys Glu
625 630 635 640

Arg Asn Arg Ile Lys Gln Gln Val Asp Gln Leu Ser Ala Ile Leu Arg
645 650 655
Leu His Lys Lys Val Asn Glu Asp Glu Arg Ile Gln Ala Ile Asp Lys
660 665 670
Leu Leu Gln Lys Val Ala Ser Trp Gln Leu Asn Glu Glu Ile Ala Thr
675 680 685
Ala Trp Asn Gln Ala Leu Ser Gln Leu Tyr Ser Lys Ala Lys Glu Asn
690 695 700

Asp Leu Gln Trp Asn Gln Ala Ile Lys Asn Ala His His Gln Leu Glu
705 710 715 720
Pro Val Val Gly Lys Gln Ile Ser Leu Trp Arg Lys Asp Leu Ser Thr

725 730 735

Gly Arg Gln Gly Ile Ala Gly Leu Ser Leu Trp Ser Ile Glu Glu Leu

740 745 750

Glu Ala Thr Lys Lys Leu Leu Thr Arg Trp Ser Lys Arg Ser Arg Glu

755 760 765

Pro Gly Val Val Lys Arg Ile Glu Arg Phe Glu Thr Phe Ala Lys Gln

770 775 780

Ile Gln His His Ile Asn Gln Val Lys Glu Asn Arg Leu Lys Gln Leu

785 790 795 800

Ala Asn Leu Ile Val Met Thr Ala Leu Gly Tyr Lys Tyr Asp Gln Glu

805 810 815

Gln Lys Lys Trp Ile Glu Val Tyr Pro Ala Cys Gln Val Val Leu Phe

820 825 830

Glu Asn Leu Arg Ser Tyr Arg Phe Ser Phe Glu Arg Ser Arg Arg Glu

835 840 845

Asn Lys Lys Leu Met Glu Trp Ser His Arg Ser Ile Pro Lys Leu Val

850 855 860

Gln Met Gln Gly Glu Leu Phe Gly Leu Gln Val Ala Asp Val Tyr Ala

865 870 875 880

Ala Tyr Ser Ser Arg Tyr His Gly Arg Thr Gly Ala Pro Gly Ile Arg

885 890 895

Cys His Ala Leu Thr Glu Ala Asp Leu Arg Asn Glu Thr Asn Ile Ile

900 905 910

His Glu Leu Ile Glu Ala Gly Phe Ile Lys Glu Glu His Arg Pro Tyr

915 920 925

Leu Gln Gln Gly Asp Leu Val Pro Trp Ser Gly Gly Glu Leu Phe Ala

930 935 940

Thr Leu Gln Lys Pro Tyr Asp Asn Pro Arg Ile Leu Thr Leu His Ala

945 950 955 960

Asp Ile Asn Ala Ala Gln Asn Ile Gln Lys Arg Phe Trp His Pro Ser

965 970 975

Met Trp Phe Arg Val Asn Cys Glu Ser Val Met Glu Gly Glu Ile Val
980 985 990
Thr Tyr Val Pro Lys Asn Lys Thr Val His Lys Lys Gln Gly Lys Thr
995 1000 1005
Phe Arg Phe Val Lys Val Glu Gly Ser Asp Val Tyr Glu Trp Ala
1010 1015 1020

Lys Trp Ser Lys Asn Arg Asn Lys Asn Thr Phe Ser Ser Ile Thr
1025 1030 1035
Glu Arg Lys Pro Pro Ser Ser Met Ile Leu Phe Arg Asp Pro Ser
1040 1045 1050
Gly Thr Phe Phe Lys Glu Gln Glu Trp Val Glu Gln Lys Thr Phe
1055 1060 1065
Trp Gly Lys Val Gln Ser Met Ile Gln Ala Tyr Met Lys Lys Thr
1070 1075 1080

Ile Val Gln Arg Met Glu Glu
1085 1090
<210> 73
<211> 1194
<212> PRT
<213> Dsulfonatronum thiodismutans
<400> 73
Met Val Leu Gly Arg Lys Asp Asp Thr Ala Glu Leu Arg Arg Ala Leu
1 5 10 15
Trp Thr Thr His Glu His Val Asn Leu Ala Val Ala Glu Val Glu Arg
20 25 30
Val Leu Leu Arg Cys Arg Gly Arg Ser Tyr Trp Thr Leu Asp Arg Arg
35 40 45

Gly Asp Pro Val His Val Pro Glu Ser Gln Val Ala Glu Asp Ala Leu
50 55 60
Ala Met Ala Arg Glu Ala Gln Arg Arg Asn Gly Trp Pro Val Val Gly
65 70 75 80
Glu Asp Glu Glu Ile Leu Leu Ala Leu Arg Tyr Leu Tyr Glu Gln Ile

85 90 95

Val Pro Ser Cys Leu Leu Asp Asp Leu Gly Lys Pro Leu Lys Gly Asp

100 105 110

Ala Gln Lys Ile Gly Thr Asn Tyr Ala Gly Pro Leu Phe Asp Ser Asp

115 120 125

Thr Cys Arg Arg Asp Glu Gly Lys Asp Val Ala Cys Cys Gly Pro Phe

130 135 140

His Glu Val Ala Gly Lys Tyr Leu Gly Ala Leu Pro Glu Trp Ala Thr

145 150 155 160

Pro Ile Ser Lys Gln Glu Phe Asp Gly Lys Asp Ala Ser His Leu Arg

165 170 175

Phe Lys Ala Thr Gly Gly Asp Asp Ala Phe Phe Arg Val Ser Ile Glu

180 185 190

Lys Ala Asn Ala Trp Tyr Glu Asp Pro Ala Asn Gln Asp Ala Leu Lys

195 200 205

Asn Lys Ala Tyr Asn Lys Asp Asp Trp Lys Lys Glu Lys Asp Lys Gly

210 215 220

Ile Ser Ser Trp Ala Val Lys Tyr Ile Gln Lys Gln Leu Gln Leu Gly

225 230 235 240

Gln Asp Pro Arg Thr Glu Val Arg Arg Lys Leu Trp Leu Glu Leu Gly

245 250 255

Leu Leu Pro Leu Phe Ile Pro Val Phe Asp Lys Thr Met Val Gly Asn

260 265 270

Leu Trp Asn Arg Leu Ala Val Arg Leu Ala Leu Ala His Leu Leu Ser

275 280 285

Trp Glu Ser Trp Asn His Arg Ala Val Gln Asp Gln Ala Leu Ala Arg

290 295 300

Ala Lys Arg Asp Glu Leu Ala Ala Leu Phe Leu Gly Met Glu Asp Gly

305 310 315 320

Phe Ala Gly Leu Arg Glu Tyr Glu Leu Arg Arg Asn Glu Ser Ile Lys

325 330 335

Gln His Ala Phe Glu Pro Val Asp Arg Pro Tyr Val Val Ser Gly Arg
340 345 350
Ala Leu Arg Ser Trp Thr Arg Val Arg Glu Glu Trp Leu Arg His Gly
355 360 365

Asp Thr Gln Glu Ser Arg Lys Asn Ile Cys Asn Arg Leu Gln Asp Arg
370 375 380
Leu Arg Gly Lys Phe Gly Asp Pro Asp Val Phe His Trp Leu Ala Glu
385 390 395 400
Asp Gly Gln Glu Ala Leu Trp Lys Glu Arg Asp Cys Val Thr Ser Phe
405 410 415
Ser Leu Leu Asn Asp Ala Asp Gly Leu Leu Glu Lys Arg Lys Gly Tyr
420 425 430

Ala Leu Met Thr Phe Ala Asp Ala Arg Leu His Pro Arg Trp Ala Met
435 440 445
Tyr Glu Ala Pro Gly Gly Ser Asn Leu Arg Thr Tyr Gln Ile Arg Lys
450 455 460
Thr Glu Asn Gly Leu Trp Ala Asp Val Val Leu Leu Ser Pro Arg Asn
465 470 475 480
Glu Ser Ala Ala Val Glu Glu Lys Thr Phe Asn Val Arg Leu Ala Pro
485 490 495

Ser Gly Gln Leu Ser Asn Val Ser Phe Asp Gln Ile Gln Lys Gly Ser
500 505 510
Lys Met Val Gly Arg Cys Arg Tyr Gln Ser Ala Asn Gln Gln Phe Glu
515 520 525
Gly Leu Leu Gly Gly Ala Glu Ile Leu Phe Asp Arg Lys Arg Ile Ala
530 535 540
Asn Glu Gln His Gly Ala Thr Asp Leu Ala Ser Lys Pro Gly His Val
545 550 555 560

Trp Phe Lys Leu Thr Leu Asp Val Arg Pro Gln Ala Pro Gln Gly Trp
565 570 575
Leu Asp Gly Lys Gly Arg Pro Ala Leu Pro Pro Glu Ala Lys His Phe

580 585 590

Lys Thr Ala Leu Ser Asn Lys Ser Lys Phe Ala Asp Gln Val Arg Pro

595 600 605

Gly Leu Arg Val Leu Ser Val Asp Leu Gly Val Arg Ser Phe Ala Ala

610 615 620

Cys Ser Val Phe Glu Leu Val Arg Gly Gly Pro Asp Gln Gly Thr Tyr

625 630 635 640

Phe Pro Ala Ala Asp Gly Arg Thr Val Asp Asp Pro Glu Lys Leu Trp

645 650 655

Ala Lys His Glu Arg Ser Phe Lys Ile Thr Leu Pro Gly Glu Asn Pro

660 665 670

Ser Arg Lys Glu Glu Ile Ala Arg Arg Ala Ala Met Glu Glu Leu Arg

675 680 685

Ser Leu Asn Gly Asp Ile Arg Arg Leu Lys Ala Ile Leu Arg Leu Ser

690 695 700

Val Leu Gln Glu Asp Asp Pro Arg Thr Glu His Leu Arg Leu Phe Met

705 710 715 720

Glu Ala Ile Val Asp Asp Pro Ala Lys Ser Ala Leu Asn Ala Glu Leu

725 730 735

Phe Lys Gly Phe Gly Asp Asp Arg Phe Arg Ser Thr Pro Asp Leu Trp

740 745 750

Lys Gln His Cys His Phe Phe His Asp Lys Ala Glu Lys Val Val Ala

755 760 765

Glu Arg Phe Ser Arg Trp Arg Thr Glu Thr Arg Pro Lys Ser Ser Ser

770 775 780

Trp Gln Asp Trp Arg Glu Arg Arg Gly Tyr Ala Gly Gly Lys Ser Tyr

785 790 795 800

Trp Ala Val Thr Tyr Leu Glu Ala Val Arg Gly Leu Ile Leu Arg Trp

805 810 815

Asn Met Arg Gly Arg Thr Tyr Gly Glu Val Asn Arg Gln Asp Lys Lys

820 825 830

Gln Phe Gly Thr Val Ala Ser Ala Leu Leu His His Ile Asn Gln Leu
835 840 845
Lys Glu Asp Arg Ile Lys Thr Gly Ala Asp Met Ile Ile Gln Ala Ala
850 855 860
Arg Gly Phe Val Pro Arg Lys Asn Gly Ala Gly Trp Val Gln Val His
865 870 875 880

Glu Pro Cys Arg Leu Ile Leu Phe Glu Asp Leu Ala Arg Tyr Arg Phe
885 890 895
Arg Thr Asp Arg Ser Arg Arg Glu Asn Ser Arg Leu Met Arg Trp Ser
900 905 910
His Arg Glu Ile Val Asn Glu Val Gly Met Gln Gly Glu Leu Tyr Gly
915 920 925
Leu His Val Asp Thr Thr Glu Ala Gly Phe Ser Ser Arg Tyr Leu Ala
930 935 940

Ser Ser Gly Ala Pro Gly Val Arg Cys Arg His Leu Val Glu Glu Asp
945 950 955 960
Phe His Asp Gly Leu Pro Gly Met His Leu Val Gly Glu Leu Asp Trp
965 970 975
Leu Leu Pro Lys Asp Lys Asp Arg Thr Ala Asn Glu Ala Arg Arg Leu
980 985 990
Leu Gly Gly Met Val Arg Pro Gly Met Leu Val Pro Trp Asp Gly Gly
995 1000 1005

Glu Leu Phe Ala Thr Leu Asn Ala Ala Ser Gln Leu His Val Ile
1010 1015 1020
His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg Phe Trp
1025 1030 1035
Gly Arg Cys Gly Glu Ala Ile Arg Ile Val Cys Asn Gln Leu Ser
1040 1045 1050
Val Asp Gly Ser Thr Arg Tyr Glu Met Ala Lys Ala Pro Lys Ala
1055 1060 1065

Arg Leu Leu Gly Ala Leu Gln Gln Leu Lys Asn Gly Asp Ala Pro

1070 1075 1080

Phe His Leu Thr Ser Ile Pro Asn Ser Gln Lys Pro Glu Asn Ser

1085 1090 1095

Tyr Val Met Thr Pro Thr Asn Ala Gly Lys Lys Tyr Arg Ala Gly

1100 1105 1110

Pro Gly Glu Lys Ser Ser Gly Glu Glu Asp Glu Leu Ala Leu Asp

1115 1120 1125

Ile Val Glu Gln Ala Glu Glu Leu Ala Gln Gly Arg Lys Thr Phe

1130 1135 1140

Phe Arg Asp Pro Ser Gly Val Phe Phe Ala Pro Asp Arg Trp Leu

1145 1150 1155

Pro Ser Glu Ile Tyr Trp Ser Arg Ile Arg Arg Arg Ile Trp Gln

1160 1165 1170

Val Thr Leu Glu Arg Asn Ser Ser Gly Arg Gln Glu Arg Ala Glu

1175 1180 1185

Met Asp Glu Met Pro Tyr

1190

<210> 74

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 74

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys

20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys

35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu

50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser

65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95
Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser

325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met

340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys

355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln

370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr

385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala

405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn

420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala

435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn

450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys

485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp

515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His

530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His

545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val

565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu

915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile

930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu

995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 75

<211> 1082

<212> PRT

<213> Geobacillus thermodenitrificans

<400> 75

Met Lys Tyr Lys Ile Gly Leu Asp Ile Gly Ile Thr Ser Ile Gly Trp
1 5 10 15

Ala Val Ile Asn Leu Asp Ile Pro Arg Ile Glu Asp Leu Gly Val Arg
20 25 30

Ile Phe Asp Arg Ala Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu
35 40 45

Pro Arg Arg Leu Ala Arg Ser Ala Arg Arg Arg Leu Arg Arg Arg Lys
50 55 60

His Arg Leu Glu Arg Ile Arg Arg Leu Phe Val Arg Glu Gly Ile Leu
65 70 75 80

Thr Lys Glu Glu Leu Asn Lys Leu Phe Glu Lys Lys His Glu Ile Asp
85 90 95

Val Trp Gln Leu Arg Val Glu Ala Leu Asp Arg Lys Leu Asn Asn Asp
100 105 110

Glu Leu Ala Arg Ile Leu Leu His Leu Ala Lys Arg Arg Gly Phe Arg
115 120 125

Ser Asn Arg Lys Ser Glu Arg Thr Asn Lys Glu Asn Ser Thr Met Leu
130 135 140

Lys His Ile Glu Glu Asn Gln Ser Ile Leu Ser Ser Tyr Arg Thr Val
145 150 155 160

Ala Glu Met Val Val Lys Asp Pro Lys Phe Ser Leu His Lys Arg Asn
165 170 175

Lys Glu Asp Asn Tyr Thr Asn Thr Val Ala Arg Asp Asp Leu Glu Arg
180 185 190

Glu Ile Lys Leu Ile Phe Ala Lys Gln Arg Glu Tyr Gly Asn Ile Val
195 200 205

Cys Thr Glu Ala Phe Glu His Glu Tyr Ile Ser Ile Trp Ala Ser Gln
210 215 220
Arg Pro Phe Ala Ser Lys Asp Asp Ile Glu Lys Lys Val Gly Phe Cys
225 230 235 240
Thr Phe Glu Pro Lys Glu Lys Arg Ala Pro Lys Ala Thr Tyr Thr Phe
245 250 255

Gln Ser Phe Thr Val Trp Glu His Ile Asn Lys Leu Arg Leu Val Ser
260 265 270
Pro Gly Gly Ile Arg Ala Leu Thr Asp Asp Glu Arg Arg Leu Ile Tyr
275 280 285
Lys Gln Ala Phe His Lys Asn Lys Ile Thr Phe His Asp Val Arg Thr
290 295 300
Leu Leu Asn Leu Pro Asp Asp Thr Arg Phe Lys Gly Leu Leu Tyr Asp
305 310 315 320

Arg Asn Thr Thr Leu Lys Glu Asn Glu Lys Val Arg Phe Leu Glu Leu
325 330 335
Gly Ala Tyr His Lys Ile Arg Lys Ala Ile Asp Ser Val Tyr Gly Lys
340 345 350
Gly Ala Ala Lys Ser Phe Arg Pro Ile Asp Phe Asp Thr Phe Gly Tyr
355 360 365
Ala Leu Thr Met Phe Lys Asp Asp Thr Asp Ile Arg Ser Tyr Leu Arg
370 375 380

Asn Glu Tyr Glu Gln Asn Gly Lys Arg Met Glu Asn Leu Ala Asp Lys
385 390 395 400
Val Tyr Asp Glu Glu Leu Ile Glu Glu Leu Leu Asn Leu Ser Phe Ser
405 410 415
Lys Phe Gly His Leu Ser Leu Lys Ala Leu Arg Asn Ile Leu Pro Tyr
420 425 430
Met Glu Gln Gly Glu Val Tyr Ser Thr Ala Cys Glu Arg Ala Gly Tyr
435 440 445

Thr Phe Thr Gly Pro Lys Lys Lys Gln Lys Thr Val Leu Leu Pro Asn

450 455 460

Ile Pro Pro Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln Ala

465 470 475 480

Arg Lys Val Val Asn Ala Ile Ile Lys Lys Tyr Gly Ser Pro Val Ser

485 490 495

Ile His Ile Glu Leu Ala Arg Glu Leu Ser Gln Ser Phe Asp Glu Arg

500 505 510

Arg Lys Met Gln Lys Glu Gln Glu Gly Asn Arg Lys Lys Asn Glu Thr

515 520 525

Ala Ile Arg Gln Leu Val Glu Tyr Gly Leu Thr Leu Asn Pro Thr Gly

530 535 540

Leu Asp Ile Val Lys Phe Lys Leu Trp Ser Glu Gln Asn Gly Lys Cys

545 550 555 560

Ala Tyr Ser Leu Gln Pro Ile Glu Ile Glu Arg Leu Leu Glu Pro Gly

565 570 575

Tyr Thr Glu Val Asp His Val Ile Pro Tyr Ser Arg Ser Leu Asp Asp

580 585 590

Ser Tyr Thr Asn Lys Val Leu Val Leu Thr Lys Glu Asn Arg Glu Lys

595 600 605

Gly Asn Arg Thr Pro Ala Glu Tyr Leu Gly Leu Gly Ser Glu Arg Trp

610 615 620

Gln Gln Phe Glu Thr Phe Val Leu Thr Asn Lys Gln Phe Ser Lys Lys

625 630 635 640

Lys Arg Asp Arg Leu Leu Arg Leu His Tyr Asp Glu Asn Glu Glu Asn

645 650 655

Glu Phe Lys Asn Arg Asn Leu Asn Asp Thr Arg Tyr Ile Ser Arg Phe

660 665 670

Leu Ala Asn Phe Ile Arg Glu His Leu Lys Phe Ala Asp Ser Asp Asp

675 680 685

Lys Gln Lys Val Tyr Thr Val Asn Gly Arg Ile Thr Ala His Leu Arg

690 695 700

Ser Arg Trp Asn Phe Asn Lys Asn Arg Glu Glu Ser Asn Leu His His
705 710 715 720
Ala Val Asp Ala Ala Ile Val Ala Cys Thr Thr Pro Ser Asp Ile Ala
725 730 735
Arg Val Thr Ala Phe Tyr Gln Arg Arg Glu Gln Asn Lys Glu Leu Ser
740 745 750
Lys Lys Thr Asp Pro Gln Phe Pro Gln Pro Trp Pro His Phe Ala Asp
755 760 765

Glu Leu Gln Ala Arg Leu Ser Lys Asn Pro Lys Glu Ser Ile Lys Ala
770 775 780
Leu Asn Leu Gly Asn Tyr Asp Asn Glu Lys Leu Glu Ser Leu Gln Pro
785 790 795 800
Val Phe Val Ser Arg Met Pro Lys Arg Ser Ile Thr Gly Ala Ala His
805 810 815
Gln Glu Thr Leu Arg Arg Tyr Ile Gly Ile Asp Glu Arg Ser Gly Lys
820 825 830

Ile Gln Thr Val Val Lys Lys Lys Leu Ser Glu Ile Gln Leu Asp Lys
835 840 845
Thr Gly His Phe Pro Met Tyr Gly Lys Glu Ser Asp Pro Arg Thr Tyr
850 855 860
Glu Ala Ile Arg Gln Arg Leu Leu Glu His Asn Asn Asp Pro Lys Lys
865 870 875 880
Ala Phe Gln Glu Pro Leu Tyr Lys Pro Lys Lys Asn Gly Glu Leu Gly
885 890 895

Pro Ile Ile Arg Thr Ile Lys Ile Ile Asp Thr Thr Asn Gln Val Ile
900 905 910
Pro Leu Asn Asp Gly Lys Thr Val Ala Tyr Asn Ser Asn Ile Val Arg
915 920 925
Val Asp Val Phe Glu Lys Asp Gly Lys Tyr Tyr Cys Val Pro Ile Tyr
930 935 940
Thr Ile Asp Met Met Lys Gly Ile Leu Pro Asn Lys Ala Ile Glu Pro

945 950 955 960

Asn Lys Pro Tyr Ser Glu Trp Lys Glu Met Thr Glu Asp Tyr Thr Phe
965 970 975
Arg Phe Ser Leu Tyr Pro Asn Asp Leu Ile Arg Ile Glu Phe Pro Arg
980 985 990
Glu Lys Thr Ile Lys Thr Ala Val Gly Glu Glu Ile Lys Ile Lys Asp
995 1000 1005
Leu Phe Ala Tyr Tyr Gln Thr Ile Asp Ser Ser Asn Gly Gly Leu
1010 1015 1020

Ser Leu Val Ser His Asp Asn Asn Phe Ser Leu Arg Ser Ile Gly
1025 1030 1035
Ser Arg Thr Leu Lys Arg Phe Glu Lys Tyr Gln Val Asp Val Leu
1040 1045 1050
Gly Asn Ile Tyr Lys Val Arg Gly Glu Lys Arg Val Gly Val Ala
1055 1060 1065
Ser Ser Ser His Ser Lys Ala Gly Glu Thr Ile Arg Pro Leu
1070 1075 1080

<210> 76
<211> 887
<212> PRT
<213> Natronobacterium gregoryi
<400> 76
Met Thr Val Ile Asp Leu Asp Ser Thr Thr Thr Ala Asp Glu Leu Thr
1 5 10 15
Ser Gly His Thr Tyr Asp Ile Ser Val Thr Leu Thr Gly Val Tyr Asp
20 25 30
Asn Thr Asp Glu Gln His Pro Arg Met Ser Leu Ala Phe Glu Gln Asp
35 40 45
Asn Gly Glu Arg Arg Tyr Ile Thr Leu Trp Lys Asn Thr Thr Pro Lys

50 55 60
Asp Val Phe Thr Tyr Asp Tyr Ala Thr Gly Ser Thr Tyr Ile Phe Thr

65 70 75 80

Asn Ile Asp Tyr Glu Val Lys Asp Gly Tyr Glu Asn Leu Thr Ala Thr

85 90 95

Tyr Gln Thr Thr Val Glu Asn Ala Thr Ala Gln Glu Val Gly Thr Thr

100 105 110

Asp Glu Asp Glu Thr Phe Ala Gly Gly Glu Pro Leu Asp His His Leu

115 120 125

Asp Asp Ala Leu Asn Glu Thr Pro Asp Asp Ala Glu Thr Glu Ser Asp

130 135 140

Ser Gly His Val Met Thr Ser Phe Ala Ser Arg Asp Gln Leu Pro Glu

145 150 155 160

Trp Thr Leu His Thr Tyr Thr Leu Thr Ala Thr Asp Gly Ala Lys Thr

165 170 175

Asp Thr Glu Tyr Ala Arg Arg Thr Leu Ala Tyr Thr Val Arg Gln Glu

180 185 190

Leu Tyr Thr Asp His Asp Ala Ala Pro Val Ala Thr Asp Gly Leu Met

195 200 205

Leu Leu Thr Pro Glu Pro Leu Gly Glu Thr Pro Leu Asp Leu Asp Cys

210 215 220

Gly Val Arg Val Glu Ala Asp Glu Thr Arg Thr Leu Asp Tyr Thr Thr

225 230 235 240

Ala Lys Asp Arg Leu Leu Ala Arg Glu Leu Val Glu Glu Gly Leu Lys

245 250 255

Arg Ser Leu Trp Asp Asp Tyr Leu Val Arg Gly Ile Asp Glu Val Leu

260 265 270

Ser Lys Glu Pro Val Leu Thr Cys Asp Glu Phe Asp Leu His Glu Arg

275 280 285

Tyr Asp Leu Ser Val Glu Val Gly His Ser Gly Arg Ala Tyr Leu His

290 295 300

Ile Asn Phe Arg His Arg Phe Val Pro Lys Leu Thr Leu Ala Asp Ile

305 310 315 320

Asp Asp Asp Asn Ile Tyr Pro Gly Leu Arg Val Lys Thr Thr Tyr Arg
325 330 335
Pro Arg Arg Gly His Ile Val Trp Gly Leu Arg Asp Glu Cys Ala Thr
340 345 350
Asp Ser Leu Asn Thr Leu Gly Asn Gln Ser Val Val Ala Tyr His Arg
355 360 365
Asn Asn Gln Thr Pro Ile Asn Thr Asp Leu Leu Asp Ala Ile Glu Ala

370 375 380
Ala Asp Arg Arg Val Val Glu Thr Arg Arg Gln Gly His Gly Asp Asp
385 390 395 400
Ala Val Ser Phe Pro Gln Glu Leu Leu Ala Val Glu Pro Asn Thr His
405 410 415
Gln Ile Lys Gln Phe Ala Ser Asp Gly Phe His Gln Gln Ala Arg Ser
420 425 430
Lys Thr Arg Leu Ser Ala Ser Arg Cys Ser Glu Lys Ala Gln Ala Phe

435 440 445
Ala Glu Arg Leu Asp Pro Val Arg Leu Asn Gly Ser Thr Val Glu Phe
450 455 460
Ser Ser Glu Phe Phe Thr Gly Asn Asn Glu Gln Gln Leu Arg Leu Leu
465 470 475 480
Tyr Glu Asn Gly Glu Ser Val Leu Thr Phe Arg Asp Gly Ala Arg Gly
485 490 495
Ala His Pro Asp Glu Thr Phe Ser Lys Gly Ile Val Asn Pro Pro Glu

500 505 510
Ser Phe Glu Val Ala Val Val Leu Pro Glu Gln Gln Ala Asp Thr Cys
515 520 525
Lys Ala Gln Trp Asp Thr Met Ala Asp Leu Leu Asn Gln Ala Gly Ala
530 535 540
Pro Pro Thr Arg Ser Glu Thr Val Gln Tyr Asp Ala Phe Ser Ser Pro
545 550 555 560
Glu Ser Ile Ser Leu Asn Val Ala Gly Ala Ile Asp Pro Ser Glu Val

565 570 575

Asp Ala Ala Phe Val Val Leu Pro Pro Asp Gln Glu Gly Phe Ala Asp

580 585 590

Leu Ala Ser Pro Thr Glu Thr Tyr Asp Glu Leu Lys Lys Ala Leu Ala

595 600 605

Asn Met Gly Ile Tyr Ser Gln Met Ala Tyr Phe Asp Arg Phe Arg Asp

610 615 620

Ala Lys Ile Phe Tyr Thr Arg Asn Val Ala Leu Gly Leu Leu Ala Ala

625 630 635 640

Ala Gly Gly Val Ala Phe Thr Thr Glu His Ala Met Pro Gly Asp Ala

645 650 655

Asp Met Phe Ile Gly Ile Asp Val Ser Arg Ser Tyr Pro Glu Asp Gly

660 665 670

Ala Ser Gly Gln Ile Asn Ile Ala Ala Thr Ala Thr Ala Val Tyr Lys

675 680 685

Asp Gly Thr Ile Leu Gly His Ser Ser Thr Arg Pro Gln Leu Gly Glu

690 695 700

Lys Leu Gln Ser Thr Asp Val Arg Asp Ile Met Lys Asn Ala Ile Leu

705 710 715 720

Gly Tyr Gln Gln Val Thr Gly Glu Ser Pro Thr His Ile Val Ile His

725 730 735

Arg Asp Gly Phe Met Asn Glu Asp Leu Asp Pro Ala Thr Glu Phe Leu

740 745 750

Asn Glu Gln Gly Val Glu Tyr Asp Ile Val Glu Ile Arg Lys Gln Pro

755 760 765

Gln Thr Arg Leu Leu Ala Val Ser Asp Val Gln Tyr Asp Thr Pro Val

770 775 780

Lys Ser Ile Ala Ala Ile Asn Gln Asn Glu Pro Arg Ala Thr Val Ala

785 790 795 800

Thr Phe Gly Ala Pro Glu Tyr Leu Ala Thr Arg Asp Gly Gly Gly Leu

805 810 815

Pro Arg Pro Ile Gln Ile Glu Arg Val Ala Gly Glu Thr Asp Ile Glu

820 825 830

Thr Leu Thr Arg Gln Val Tyr Leu Leu Ser Gln Ser His Ile Gln Val

835 840 845

His Asn Ser Thr Ala Arg Leu Pro Ile Thr Thr Ala Tyr Ala Asp Gln

850 855 860

Ala Ser Thr His Ala Thr Lys Gly Tyr Leu Val Gln Thr Gly Ala Phe

865 870 875 880

Glu Ser Asn Val Gly Phe Leu

885

<210> 77

<211> 1387

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 77

Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln

1 5 10 15

Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met

20 25 30

Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys

35 40 45

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

50 55 60

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln

65 70 75 80

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

85 90 95

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys

100 105 110

Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

115 120 125

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

130 135 140

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg

145 150 155 160

Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr

165 170 175

Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

180 185 190

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

195 200 205

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe

210 215 220

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp

225 230 235 240

Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp

245 250 255

Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala

260 265 270

Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp

275 280 285

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

290 295 300

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile

305 310 315 320

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

325 330 335

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser

340 345 350

Gln Leu Gly Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

355 360 365

Ser Gly Gly Ser Gly Gly Ser Gly Gly Asp Lys Lys Tyr Ser Ile Gly

370 375 380

Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu

385 390 395 400

Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg

405 410 415

His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly

420 425 430

Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr

435 440 445

Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn

450 455 460

Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser

465 470 475 480

Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly

485 490 495

Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr

500 505 510

His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg

515 520 525

Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe

530 535 540

Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu

545 550 555 560

Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro

565 570 575

Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu

580 585 590

Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu

595 600 605

Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu

610 615 620

Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu

625 630 635 640

Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala

645 650 655

Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu

660 665 670

Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile

675 680 685

Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His

690 695 700

His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro

705 710 715 720

Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala

725 730 735

Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile

740 745 750

Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys

755 760 765

Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly

770 775 780

Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg

785 790 795 800

Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile

805 810 815

Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala

820 825 830

Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr

835 840 845

Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala

850 855 860

Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn
865 870 875 880
Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val

885 890 895
Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys
900 905 910
Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu
915 920 925
Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr
930 935 940
Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu

945 950 955 960
Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile
965 970 975
Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu
980 985 990
Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile
995 1000 1005
Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val

1010 1015 1020
Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu
1025 1030 1035
Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys
1040 1045 1050
Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
1055 1060 1065
Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp

1070 1075 1080
Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu
1085 1090 1095
His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

1100 1105 1110

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

1115 1120 1125

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn

1130 1135 1140

Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys

1145 1150 1155

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys

1160 1165 1170

Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr

1175 1180 1185

Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu

1190 1195 1200

Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val

1205 1210 1215

Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu

1220 1225 1230

Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser

1235 1240 1245

Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu

1250 1255 1260

Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys

1265 1270 1275

Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile

1280 1285 1290

Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala

1295 1300 1305

Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp

1310 1315 1320

Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu

1325 1330 1335

Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

1340 1345 1350

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

1355 1360 1365

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu

1370 1375 1380

Phe Val Tyr Gly

1385

<210> 78

<211> 1388

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 78

Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met

1 5 10 15

Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys

20 25 30

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

35 40 45

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln

50 55 60

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

65 70 75 80

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys

85 90 95

Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

100 105 110

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

115 120 125

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg

130 135 140

Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr
145 150 155 160
Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

165 170 175
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
180 185 190
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
195 200 205
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp
210 215 220
Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp

225 230 235 240
Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala
245 250 255
Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
260 265 270
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
275 280 285
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile

290 295 300
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
305 310 315 320
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
325 330 335
Gln Leu Gly Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
340 345 350
Ser Gly Gly Ser Gly Gly Ser Gly Gly Met Asp Lys Lys Tyr Ser Ile

355 360 365
Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp
370 375 380
Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp

385 390 395 400

Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser

405 410 415

Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg

420 425 430

Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser

435 440 445

Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu

450 455 460

Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe

465 470 475 480

Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile

485 490 495

Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu

500 505 510

Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His

515 520 525

Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys

530 535 540

Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn

545 550 555 560

Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg

565 570 575

Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly

580 585 590

Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly

595 600 605

Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys

610 615 620

Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu

625 630 635 640

Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn
645 650 655
Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu
660 665 670
Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu

675 680 685
His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu
690 695 700
Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr
705 710 715 720
Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe
725 730 735
Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val

740 745 750
Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn
755 760 765
Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu
770 775 780
Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys
785 790 795 800
Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu

805 810 815
Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu
820 825 830
Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser
835 840 845
Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro
850 855 860
Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr

865 870 875 880
Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg

885 890 895

Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu

900 905 910

Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp

915 920 925

Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val

930 935 940

Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys

945 950 955 960

Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile

965 970 975

Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met

980 985 990

Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val

995 1000 1005

Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu

1010 1015 1020

Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys

1025 1030 1035

Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn

1040 1045 1050

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp

1055 1060 1065

Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu

1070 1075 1080

His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

1085 1090 1095

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

1100 1105 1110

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn

1115 1120 1125

Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys

1130 1135 1140

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys

1145 1150 1155

Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr

1160 1165 1170

Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu

1175 1180 1185

Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val

1190 1195 1200

Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu

1205 1210 1215

Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser

1220 1225 1230

Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu

1235 1240 1245

Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys

1250 1255 1260

Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile

1265 1270 1275

Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala

1280 1285 1290

Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp

1295 1300 1305

Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu

1310 1315 1320

Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

1325 1330 1335

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

1340 1345 1350

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu

1355 1360 1365

Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile

1370 1375 1380

Ala Lys Ser Glu Gln

1385

<210> 79

<211> 1387

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 79

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu

1 5 10 15

Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile

20 25 30

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser

35 40 45

Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly

50 55 60

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

65 70 75 80

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

85 90 95

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

100 105 110

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile

115 120 125

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

130 135 140

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

145 150 155 160

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser

165 170 175

Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr

180 185 190

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

195 200 205

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

210 215 220

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

225 230 235 240

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

245 250 255

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

260 265 270

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

275 280 285

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp

290 295 300

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

305 310 315 320

Asp Leu Ser Gln Leu Gly Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly

325 330 335

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Asp Lys Lys Tyr

340 345 350

Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile

355 360 365

Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn

370 375 380

Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe

385 390 395 400

Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg

405 410 415

Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile
420 425 430
Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu
435 440 445
Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro
450 455 460
Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro

465 470 475 480
Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala
485 490 495
Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg
500 505 510
Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val
515 520 525
Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu

530 535 540
Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser
545 550 555 560
Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu
565 570 575
Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser
580 585 590
Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp

595 600 605
Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn
610 615 620
Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala
625 630 635 640
Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn
645 650 655
Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr

660 665 670

Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln

675 680 685

Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn

690 695 700

Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr

705 710 715 720

Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu

725 730 735

Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe

740 745 750

Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala

755 760 765

Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg

770 775 780

Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly

785 790 795 800

Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser

805 810 815

Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly

820 825 830

Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn

835 840 845

Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr

850 855 860

Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly

865 870 875 880

Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val

885 890 895

Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys

900 905 910

Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser

915 920 925

Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu

930 935 940

Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu

945 950 955 960

Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg

965 970 975

Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp

980 985 990

Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg

995 1000 1005

Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly

1010 1015 1020

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg

1025 1030 1035

Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu

1040 1045 1050

Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

1055 1060 1065

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

1070 1075 1080

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met

1085 1090 1095

Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu

1100 1105 1110

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met

1115 1120 1125

Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu

1130 1135 1140

Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu

1145 1150 1155

Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln

1160 1165 1170

Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile

1175 1180 1185

Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val

1190 1195 1200

Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro

1205 1210 1215

Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu

1220 1225 1230

Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr

1235 1240 1245

Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe

1250 1255 1260

Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val

1265 1270 1275

Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn

1280 1285 1290

Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

1295 1300 1305

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

1310 1315 1320

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala

1325 1330 1335

Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser

1340 1345 1350

Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met

1355 1360 1365

Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr

1370 1375 1380

Phe Phe Tyr Ser

1385

<210> 80

<211> 1388

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 80

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1 5 10 15

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

20 25 30

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

35 40 45

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

50 55 60

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

65 70 75 80

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp

85 90 95

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

100 105 110

Asp Leu Ser Gln Leu Gly Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly

115 120 125

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Met Asp Lys Lys

130 135 140

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val

145 150 155 160

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly

165 170 175

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu

180 185 190

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala

195 200 205

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu

210 215 220

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg

225 230 235 240

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His

245 250 255

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr

260 265 270

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys

275 280 285

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe

290 295 300

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp

305 310 315 320

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe

325 330 335

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu

340 345 350

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln

355 360 365

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu

370 375 380

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu

385 390 395 400

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp

405 410 415

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala

420 425 430

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val

435 440 445

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg

450 455 460

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg

465 470 475 480

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys

485 490 495

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe

500 505 510

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu

515 520 525

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr

530 535 540

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His

545 550 555 560

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn

565 570 575

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val

580 585 590

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys

595 600 605

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys

610 615 620

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys

625 630 635 640

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu

645 650 655

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu

660 665 670

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile

675 680 685

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu

690 695 700

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile

705 710 715 720

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp

725 730 735

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn

740 745 750

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp

755 760 765

Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp

770 775 780

Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly

785 790 795 800

Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly

805 810 815

Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn

820 825 830

Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile

835 840 845

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile

850 855 860

Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr

865 870 875 880

Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro

885 890 895

Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys

900 905 910

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile

915 920 925

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr

930 935 940

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg

945 950 955 960

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr

965 970 975

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

980 985 990

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp

995 1000 1005

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp

1010 1015 1020

Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp

1025 1030 1035

Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

1040 1045 1050

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

1055 1060 1065

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr

1070 1075 1080

Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu

1085 1090 1095

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr

1100 1105 1110

Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr

1115 1120 1125

Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys

1130 1135 1140

Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val

1145 1150 1155

Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr

1160 1165 1170

Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr

1175 1180 1185

Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile

1190 1195 1200

Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg

1205 1210 1215

Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn

1220 1225 1230

Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu

1235 1240 1245

Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys

1250 1255 1260

Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr

1265 1270 1275

Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys

1280 1285 1290

Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile

1295 1300 1305

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu

1310 1315 1320

Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu

1325 1330 1335

Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met

1340 1345 1350

Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu

1355 1360 1365

Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu

1370 1375 1380

Lys Leu Lys Gly Ser

1385

<210> 81

<211> 1387

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 81

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1 5 10 15

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
20 25 30

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
35 40 45

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser

50 55 60

Gln Leu Gly Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
65 70 75 80

Ser Gly Gly Ser Gly Gly Ser Gly Gly Asp Lys Lys Tyr Ser Ile Gly
85 90 95

Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu
100 105 110

Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg

115 120 125

His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly
130 135 140

Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr
145 150 155 160

Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn
165 170 175

Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser

180 185 190

Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly
195 200 205

Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr

210 215 220

His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg

225 230 235 240

Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe

245 250 255

Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu

260 265 270

Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro

275 280 285

Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu

290 295 300

Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu

305 310 315 320

Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu

325 330 335

Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu

340 345 350

Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala

355 360 365

Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu

370 375 380

Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile

385 390 395 400

Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His

405 410 415

His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro

420 425 430

Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala

435 440 445

Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile

450 455 460

Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys
465 470 475 480
Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly
485 490 495
Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg

500 505 510
Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile
515 520 525
Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala
530 535 540
Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr
545 550 555 560
Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala

565 570 575
Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn
580 585 590
Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val
595 600 605
Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys
610 615 620
Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu

625 630 635 640
Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr
645 650 655
Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu
660 665 670
Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile
675 680 685
Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu

690 695 700
Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile

705 710 715 720

Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met

725 730 735

Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg

740 745 750

Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu

755 760 765

Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu

770 775 780

Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln

785 790 795 800

Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala

805 810 815

Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val

820 825 830

Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val

835 840 845

Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn

850 855 860

Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly

865 870 875 880

Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn

885 890 895

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

900 905 910

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

915 920 925

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val

930 935 940

Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser

945 950 955 960

Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn
965 970 975
Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu
980 985 990
Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln
995 1000 1005
Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu

1010 1015 1020
Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile
1025 1030 1035
Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp
1040 1045 1050
Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn
1055 1060 1065
Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1070 1075 1080
Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr
1085 1090 1095
Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser
1100 1105 1110
Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1115 1120 1125
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly

1130 1135 1140
Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly
1145 1150 1155
Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys
1160 1165 1170
Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val
1175 1180 1185
Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn

1190 1195 1200

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys

1205 1210 1215

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val

1220 1225 1230

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val

1235 1240 1245

Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu

1250 1255 1260

Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

1265 1270 1275

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu

1280 1285 1290

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu

1295 1300 1305

Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe

1310 1315 1320

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu

1325 1330 1335

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1340 1345 1350

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1355 1360 1365

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1370 1375 1380

Lys His Arg Asp

1385

<210> 82

<211> 357

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 82

Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln
1 5 10 15
Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met
20 25 30
Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
35 40 45
Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
50 55 60
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln
65 70 75 80
Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
85 90 95
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys
100 105 110
Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
115 120 125
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
130 135 140
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg
145 150 155 160
Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr
165 170 175
Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
180 185 190
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
195 200 205
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
210 215 220
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp
225 230 235 240

Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp
245 250 255
Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala
260 265 270
Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
275 280 285
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

290 295 300
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
305 310 315 320
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
325 330 335
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
340 345 350
Gln Leu Gly Gly Asp
355

<210
> 83
<211> 341
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 83
Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met
1 5 10 15
Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
20 25 30
Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
35 40 45
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln

50 55 60
Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
65 70 75 80

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys
85 90 95
Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
100 105 110
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

115 120 125
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg
130 135 140
Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr
145 150 155 160
Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
165 170 175
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

180 185 190
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
195 200 205
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp
210 215 220
Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp
225 230 235 240
Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala

245 250 255
Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
260 265 270
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
275 280 285
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
290 295 300
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

305 310 315 320
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser

325 330 335

Gln Leu Gly Gly Asp

340

<210> 84

<211> 328

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 84

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu

1 5 10 15

Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile

20 25 30

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser

35 40 45

Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly

50 55 60

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

65 70 75 80

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

85 90 95

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

100 105 110

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile

115 120 125

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

130 135 140

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

145 150 155 160

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser

165 170 175

Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr

180 185 190

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

195 200 205

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

210 215 220

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

225 230 235 240

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

245 250 255

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

260 265 270

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

275 280 285

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp

290 295 300

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

305 310 315 320

Asp Leu Ser Gln Leu Gly Gly Asp

325

<210> 85

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 85

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1 5 10 15

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

20 25 30

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

35 40 45

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

50 55 60

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

65 70 75 80

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp

85 90 95

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

100 105 110

Asp Leu Ser Gln Leu Gly Gly Asp

115 120

<210> 86

<211> 69

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 86

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1 5 10 15

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile

20 25 30

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

35 40 45

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser

50 55 60

Gln Leu Gly Gly Asp

65

<210

> 87

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 87

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Arg Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 88

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 88

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Arg Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Glu Tyr Arg Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 89
<211> 677
<212> PRT
<213> Unknown
<220><223> moloney murine leukemia virus
<400> 89
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala

195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp

325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 90

<211> 560

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<220><221> misc_feature

<222> (229)..(229)

<223> Xaa can be any naturally occurring amino acid

<400> 90

Ala Phe Pro Leu Glu Arg Pro Asp Trp Asp Tyr Thr Thr Gln Ala Gly

1 5 10 15

Arg Asn His Leu Val His Tyr Arg Gln Leu Leu Leu Ala Gly Leu Gln

20 25 30

Asn Ala Gly Arg Ser Pro Thr Asn Leu Ala Lys Val Lys Gly Ile Thr

35 40 45

Gln Gly Pro Asn Glu Ser Pro Ser Ala Phe Leu Glu Arg Leu Lys Glu

50 55 60

Ala Tyr Arg Arg Tyr Thr Pro Tyr Asp Pro Glu Asp Pro Gly Gln Glu

65 70 75 80

Thr Asn Val Ser Met Ser Phe Ile Trp Gln Ser Ala Pro Asp Ile Gly

85 90 95

Arg Lys Leu Gly Arg Leu Glu Asp Leu Lys Ser Lys Thr Leu Gly Asp

100 105 110

Leu Val Arg Glu Ala Glu Lys Ile Phe Asn Lys Arg Glu Thr Pro Glu

115 120 125

Glu Arg Glu Glu Arg Ile Arg Arg Glu Thr Glu Glu Lys Glu Glu Arg

130 135 140

Arg Arg Thr Val Asp Glu Gln Lys Glu Lys Glu Arg Asp Arg Arg Arg

145 150 155 160

His Arg Glu Met Ser Lys Leu Leu Ala Thr Val Val Ile Gly Gln Glu

165 170 175

Gln Asp Arg Gln Glu Gly Glu Arg Lys Arg Pro Gln Leu Asp Lys Asp

180 185 190

Gln Cys Ala Tyr Cys Lys Glu Lys Gly His Trp Ala Lys Asp Cys Pro

195 200 205

Lys Lys Pro Arg Gly Pro Arg Gly Pro Arg Pro Gln Thr Ser Leu Leu

210 215 220

Thr Leu Gly Asp Xaa Gly Gly Gln Gly Gln Asp Pro Pro Pro Glu Pro

225 230 235 240

Arg Ile Thr Leu Lys Val Gly Gly Gln Pro Val Thr Phe Leu Val Asp

245 250 255

Thr Gly Ala Gln His Ser Val Leu Thr Gln Asn Pro Gly Pro Leu Ser

260 265 270

Asp Lys Ser Ala Trp Val Gln Gly Ala Thr Gly Gly Lys Arg Tyr Arg

275 280 285

Trp Thr Thr Asp Arg Lys Val His Leu Ala Thr Gly Lys Val Thr His

290 295 300

Ser Phe Leu His Val Pro Asp Cys Pro Tyr Pro Leu Leu Gly Arg Asp

305 310 315 320

Leu Leu Thr Lys Leu Lys Ala Gln Ile His Phe Glu Gly Ser Gly Ala

325 330 335

Gln Val Val Gly Pro Met Gly Gln Pro Leu Gln Val Leu Thr Leu Asn

340 345 350

Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu Pro Asp Val

355 360 365

Ser Leu Gly Phe Thr Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu

370 375 380

Ser Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro

385 390 395 400

Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser

405 410 415

Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp

420 425 430

Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu

435 440 445

Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp Leu

450 455 460

Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val Pro Asn

465 470 475 480

Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln Trp Tyr Thr

485 490 495

Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu His Pro Thr

500 505 510

Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu Met Gly Ile

515 520 525

Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe Lys Asn Ser

530 535 540

Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala Asp Phe Arg

545 550 555 560

<210> 91

<211> 669

<212> PRT

<213> Unknown

<220><223> Feline leukemia virus

<400> 91

Thr Leu Gln Leu Glu Glu Glu Tyr Arg Leu Phe Glu Pro Glu Ser Thr
1 5 10 15

Gln Lys Gln Glu Met Asp Ile Trp Leu Lys Asn Phe Pro Gln Ala Trp
20 25 30

Ala Glu Thr Gly Gly Met Gly Thr Ala His Cys Gln Ala Pro Val Leu
35 40 45

Ile Gln Leu Lys Ala Thr Ala Thr Pro Ile Ser Ile Arg Gln Tyr Pro
50 55 60

Met Pro His Glu Ala Tyr Gln Gly Ile Lys Pro His Ile Arg Arg Met
65 70 75 80

Leu Asp Gln Gly Ile Leu Lys Pro Cys Gln Ser Pro Trp Asn Thr Pro
85 90 95

Leu Leu Pro Val Lys Lys Pro Gly Thr Glu Asp Tyr Arg Pro Val Gln
100 105 110

Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val
115 120 125

Pro Asn Pro Tyr Asn Leu Leu Ser Thr Leu Pro Pro Ser His Pro Trp
130 135 140

Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu His
145 150 155 160

Ser Glu Ser Gln Leu Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu Ile
165 170 175

Gly Leu Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe Lys
180 185 190

Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Ser Asp Leu Ala Asp
195 200 205

Phe Arg Val Arg Tyr Pro Ala Leu Val Leu Leu Gln Tyr Val Asp Asp
210 215 220

Leu Leu Leu Ala Ala Ala Thr Arg Thr Glu Cys Leu Glu Gly Thr Lys
225 230 235 240

Ala Leu Leu Glu Thr Leu Gly Asn Lys Gly Tyr Arg Ala Ser Ala Lys

245 250 255

Lys Ala Gln Ile Cys Leu Gln Glu Val Thr Tyr Leu Gly Tyr Ser Leu

260 265 270

Lys Asp Gly Gln Arg Trp Leu Thr Lys Ala Arg Lys Glu Ala Ile Leu

275 280 285

Ser Ile Pro Val Pro Lys Asn Ser Arg Gln Val Arg Glu Phe Leu Gly

290 295 300

Thr Ala Gly Tyr Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Leu Ala

305 310 315 320

Ala Pro Leu Tyr Pro Leu Thr Arg Pro Gly Thr Leu Phe Gln Trp Gly

325 330 335

Thr Glu Gln Gln Leu Ala Phe Glu Asp Ile Lys Lys Ala Leu Leu Ser

340 345 350

Ser Pro Ala Leu Gly Leu Pro Asp Ile Thr Lys Pro Phe Glu Leu Phe

355 360 365

Ile Asp Glu Asn Ser Gly Phe Ala Lys Gly Val Leu Val Gln Lys Leu

370 375 380

Gly Pro Trp Lys Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp Thr

385 390 395 400

Val Ala Ser Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile Ala

405 410 415

Ile Leu Val Lys Asp Ala Gly Lys Leu Thr Leu Gly Gln Pro Leu Thr

420 425 430

Ile Leu Thr Ser His Pro Val Glu Ala Leu Val Arg Gln Pro Pro Asn

435 440 445

Lys Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Met Leu Leu

450 455 460

Asp Ala Glu Arg Val His Phe Gly Pro Thr Val Ser Leu Asn Pro Ala

465 470 475 480

Thr Leu Leu Pro Leu Pro Ser Gly Gly Asn His His Asp Cys Leu Gln

485 490 495

Ile Leu Ala Glu Thr His Gly Thr Arg Pro Asp Leu Thr Asp Gln Pro
500 505 510
Leu Pro Asp Ala Asp Leu Thr Trp Tyr Thr Asp Gly Ser Ser Phe Ile
515 520 525

Arg Asn Gly Glu Arg Glu Ala Gly Ala Ala Val Thr Thr Glu Ser Glu
530 535 540
Val Ile Trp Ala Ala Pro Leu Pro Pro Gly Thr Ser Ala Gln Arg Ala
545 550 555 560
Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys
565 570 575
Leu Thr Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Thr His Val
580 585 590

His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly Lys
595 600 605
Glu Ile Lys Asn Lys Asn Glu Ile Leu Ala Leu Leu Glu Ala Leu Phe
610 615 620
Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys Gly
625 630 635 640
Asp Ser Pro Gln Ala Lys Gly Asn Arg Leu Ala Asp Asp Thr Ala Lys
645 650 655

Lys Ala Ala Thr Glu Thr His Ser Ser Leu Thr Val Leu
660 665

<210> 92
<211> 559
<212> PRT
<213> Unknown
<220><223> HIV-1
<220><221> misc_feature
<222> (280)..(280)
<223> Xaa can be any naturally occurring amino acid
<400> 92
Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met

1 5 10 15

Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys

20 25 30

Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser

35 40 45

Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys

50 55 60

Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu

65 70 75 80

Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His

85 90 95

Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly

100 105 110

Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr

115 120 125

Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr

130 135 140

Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe

145 150 155 160

Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro

165 170 175

Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp

180 185 190

Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His

195 200 205

Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu

210 215 220

Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr

225 230 235 240

Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp

245 250 255

Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro
260 265 270
Gly Ile Lys Val Arg Gln Leu Xaa Lys Leu Leu Arg Gly Thr Lys Ala
275 280 285

Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala
290 295 300
Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp
305 310 315 320
Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln
325 330 335
Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly
340 345 350

Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu
355 360 365
Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly
370 375 380
Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr
385 390 395 400
Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe
405 410 415

Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu
420 425 430
Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg
435 440 445
Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln
450 455 460
Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln
465 470 475 480

Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val
485 490 495
Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln

500 505 510

Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys

515 520 525

Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly

530 535 540

Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val

545 550 555

<210> 93

<211> 440

<212> PRT

<213> Unknown

<220><223> HIV-1

<400> 93

Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met

1 5 10 15

Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys

20 25 30

Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser

35 40 45

Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys

50 55 60

Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu

65 70 75 80

Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His

85 90 95

Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly

100 105 110

Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr

115 120 125

Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr

130 135 140

Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe

145 150 155 160

Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro

165 170 175

Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp

180 185 190

Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His

195 200 205

Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu

210 215 220

Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr

225 230 235 240

Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp

245 250 255

Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro

260 265 270

Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala

275 280 285

Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala

290 295 300

Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp

305 310 315 320

Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln

325 330 335

Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly

340 345 350

Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu

355 360 365

Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly

370 375 380

Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr

385 390 395 400

Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe
405 410 415
Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu

420 425 430
Pro Ile Val Gly Ala Glu Thr Phe
435 440

<210> 94
<211> 858
<212> PRT
<213> Unknown
<220><223> rous sarcoma virus
<400> 94
Thr Val Ala Leu His Leu Ala Ile Pro Leu Lys Trp Lys Pro Asn His
1 5 10 15
Thr Pro Val Trp Ile Asp Gln Trp Pro Leu Pro Glu Gly Lys Leu Val
20 25 30
Ala Leu Thr Gln Leu Val Glu Lys Glu Leu Gln Leu Gly His Ile Glu

35 40 45
Pro Ser Leu Ser Cys Trp Asn Thr Pro Val Phe Val Ile Arg Lys Ala
50 55 60
Ser Gly Ser Tyr Arg Leu Leu His Asp Leu Arg Ala Val Asn Ala Lys
65 70 75 80
Leu Val Pro Phe Gly Ala Val Gln Gln Gly Ala Pro Val Leu Ser Ala
85 90 95
Leu Pro Arg Gly Trp Pro Leu Met Val Leu Asp Leu Lys Asp Cys Phe

100 105 110
Phe Ser Ile Pro Leu Ala Glu Gln Asp Arg Glu Ala Phe Ala Phe Thr
115 120 125
Leu Pro Ser Val Asn Asn Gln Ala Pro Ala Arg Arg Phe Gln Trp Lys
130 135 140
Val Leu Pro Gln Gly Met Thr Cys Ser Pro Thr Ile Cys Gln Leu Ile
145 150 155 160

Val Gly Gln Ile Leu Glu Pro Leu Arg Leu Lys His Pro Ser Leu Arg

165 170 175

Met Leu His Tyr Met Asp Asp Leu Leu Leu Ala Ala Ser Ser His Asp

180 185 190

Gly Leu Glu Ala Ala Gly Glu Glu Val Ile Ser Thr Leu Glu Arg Ala

195 200 205

Gly Phe Thr Ile Ser Pro Asp Lys Val Gln Lys Glu Pro Gly Val Gln

210 215 220

Tyr Leu Gly Tyr Lys Leu Gly Ser Thr Tyr Ala Ala Pro Val Gly Leu

225 230 235 240

Val Ala Glu Pro Arg Ile Ala Thr Leu Trp Asp Val Gln Lys Leu Val

245 250 255

Gly Ser Leu Gln Trp Leu Arg Pro Ala Leu Gly Ile Pro Pro Arg Leu

260 265 270

Arg Gly Pro Phe Tyr Glu Gln Leu Arg Gly Ser Asp Pro Asn Glu Ala

275 280 285

Arg Glu Trp Asn Leu Asp Met Lys Met Ala Trp Arg Glu Ile Val Gln

290 295 300

Leu Ser Thr Thr Ala Ala Leu Glu Arg Trp Asp Pro Ala Leu Pro Leu

305 310 315 320

Glu Gly Ala Val Ala Arg Cys Glu Gln Gly Ala Ile Gly Val Leu Gly

325 330 335

Gln Gly Leu Ser Thr His Pro Arg Pro Cys Leu Trp Leu Phe Ser Thr

340 345 350

Gln Pro Thr Lys Ala Phe Thr Ala Trp Leu Glu Val Leu Thr Leu Leu

355 360 365

Ile Thr Lys Leu Arg Ala Ser Ala Val Arg Thr Phe Gly Lys Glu Val

370 375 380

Asp Ile Leu Leu Leu Pro Ala Cys Phe Arg Asp Glu Leu Pro Leu Pro

385 390 395 400

Glu Gly Ile Leu Leu Ala Leu Arg Gly Phe Ala Gly Lys Ile Arg Ser

405 410 415

Ser Asp Thr Pro Ser Ile Phe Asp Ile Ala Arg Pro Leu His Val Ser

420 425 430

Leu Lys Val Arg Val Thr Asp His Pro Val Pro Gly Pro Thr Val Phe

435 440 445

Thr Asp Ala Ser Ser Ser Thr His Lys Gly Val Val Val Trp Arg Glu

450 455 460

Gly Pro Arg Trp Glu Ile Lys Glu Ile Ala Asp Leu Gly Ala Ser Val

465 470 475 480

Gln Gln Leu Glu Ala Arg Ala Val Ala Met Ala Leu Leu Leu Trp Pro

485 490 495

Thr Thr Pro Thr Asn Val Val Thr Asp Ser Ala Phe Val Ala Lys Met

500 505 510

Leu Leu Lys Met Gly Gln Glu Gly Val Pro Ser Thr Ala Ala Ala Phe

515 520 525

Ile Leu Glu Asp Ala Leu Ser Gln Arg Ser Ala Met Ala Ala Val Leu

530 535 540

His Val Arg Ser His Ser Glu Val Pro Gly Phe Phe Thr Glu Gly Asn

545 550 555 560

Asp Val Ala Asp Ser Gln Ala Thr Phe Gln Ala Tyr Pro Leu Arg Glu

565 570 575

Ala Lys Asp Leu His Thr Ala Leu His Ile Gly Pro Arg Ala Leu Ser

580 585 590

Lys Ala Cys Asn Ile Ser Met Gln Gln Ala Arg Glu Val Val Gln Thr

595 600 605

Cys Pro His Cys Asn Ser Ala Pro Ala Leu Glu Ala Gly Val Asn Pro

610 615 620

Arg Gly Leu Gly Pro Leu Gln Ile Trp Gln Thr Asp Phe Thr Leu Glu

625 630 635 640

Pro Arg Met Ala Pro Arg Ser Trp Leu Ala Val Thr Val Asp Thr Ala

645 650 655

Ser Ser Ala Ile Val Val Thr Gln His Gly Arg Val Thr Ser Val Ala
660 665 670
Ala Gln His His Trp Ala Thr Val Ile Ala Val Leu Gly Arg Pro Lys

675 680 685
Ala Ile Lys Thr Asp Asn Gly Ser Cys Phe Thr Ser Lys Ser Thr Arg
690 695 700
Glu Trp Leu Ala Arg Trp Gly Ile Ala His Thr Thr Gly Ile Pro Gly
705 710 715 720
Asn Ser Gln Gly Gln Ala Met Val Glu Arg Ala Asn Arg Leu Leu Lys
725 730 735
Asp Lys Ile Arg Val Leu Ala Glu Gly Asp Gly Phe Met Lys Arg Ile

740 745 750
Pro Thr Ser Lys Gln Gly Glu Leu Leu Ala Lys Ala Met Tyr Ala Leu
755 760 765
Asn His Phe Glu Arg Gly Glu Asn Thr Lys Thr Pro Ile Gln Lys His
770 775 780
Trp Arg Pro Thr Val Leu Thr Glu Gly Pro Pro Val Lys Ile Arg Ile
785 790 795 800
Glu Thr Gly Glu Trp Glu Lys Gly Trp Asn Val Leu Val Trp Gly Arg

805 810 815
Gly Tyr Ala Ala Val Lys Asn Arg Asp Thr Asp Lys Val Ile Trp Val
820 825 830
Pro Ser Arg Lys Val Lys Pro Asp Ile Ala Gln Lys Asp Glu Val Thr
835 840 845
Lys Lys Asp Glu Ala Ser Pro Leu Phe Ala
850 855

<210> 95
<211> 675
<212> PRT
<213> Unknown
<220><223> cauliflower mosaic virus
<

400> 95

Met Met Asp His Leu Leu Gln Lys Thr Gln Ile Gln Asn Gln Thr Glu
1 5 10 15
Gln Val Met Asn Ile Thr Asn Pro Asn Ser Ile Tyr Ile Lys Gly Arg
20 25 30
Leu Tyr Phe Lys Gly Tyr Lys Lys Ile Glu Leu His Cys Phe Val Asp
35 40 45
Thr Gly Ala Ser Leu Cys Ile Ala Ser Lys Phe Val Ile Pro Glu Glu
50 55 60

His Trp Ile Asn Ala Glu Arg Pro Ile Met Val Lys Ile Ala Asp Gly
65 70 75 80
Ser Ser Ile Thr Ile Asn Lys Val Cys Arg Asp Ile Asp Leu Ile Ile
85 90 95
Ala Gly Glu Ile Phe His Ile Pro Thr Val Tyr Gln Gln Glu Ser Gly
100 105 110
Ile Asp Phe Ile Ile Gly Asn Asn Phe Cys Gln Leu Tyr Glu Pro Phe
115 120 125

Ile Gln Phe Thr Asp Arg Val Ile Phe Thr Lys Asp Arg Thr Tyr Pro
130 135 140
Val His Ile Ala Lys Leu Thr Arg Ala Val Arg Val Gly Thr Glu Gly
145 150 155 160
Phe Leu Glu Ser Met Lys Lys Arg Ser Lys Thr Gln Gln Pro Glu Pro
165 170 175
Val Asn Ile Ser Thr Asn Lys Ile Ala Ile Leu Ser Glu Gly Arg Arg
180 185 190

Leu Ser Glu Glu Lys Leu Phe Ile Thr Gln Gln Arg Met Gln Lys Ile
195 200 205
Glu Glu Leu Leu Glu Lys Val Cys Ser Glu Asn Pro Leu Asp Pro Asn
210 215 220
Lys Thr Lys Gln Trp Met Lys Ala Ser Ile Lys Leu Ser Asp Pro Ser
225 230 235 240
Lys Ala Ile Lys Val Lys Pro Met Lys Tyr Ser Pro Met Asp Arg Glu

245 250 255

Glu Phe Asp Lys Gln Ile Lys Glu Leu Leu Asp Leu Lys Val Ile Lys
260 265 270
Pro Ser Lys Ser Pro His Met Ala Pro Ala Phe Leu Val Asn Asn Glu
275 280 285
Ala Glu Lys Arg Arg Gly Lys Lys Arg Met Val Val Asn Tyr Lys Ala
290 295 300
Met Asn Lys Ala Thr Val Gly Asp Ala Tyr Asn Leu Pro Asn Lys Asp
305 310 315 320

Glu Leu Leu Thr Leu Ile Arg Gly Lys Lys Ile Phe Ser Ser Phe Asp
325 330 335
Cys Lys Ser Gly Phe Trp Gln Val Leu Leu Asp Gln Asp Ser Arg Pro
340 345 350
Leu Thr Ala Phe Thr Cys Pro Gln Gly His Tyr Glu Trp Asn Val Val
355 360 365
Pro Phe Gly Leu Lys Gln Ala Pro Ser Ile Phe Gln Arg His Met Asp
370 375 380

Glu Ala Phe Arg Val Phe Arg Lys Phe Cys Cys Val Tyr Val Asp Asp
385 390 395 400
Ile Leu Val Phe Ser Asn Asn Glu Glu Asp His Leu Leu His Val Ala
405 410 415
Met Ile Leu Gln Lys Cys Asn Gln His Gly Ile Ile Leu Ser Lys Lys
420 425 430
Lys Ala Gln Leu Phe Lys Lys Lys Ile Asn Phe Leu Gly Leu Glu Ile
435 440 445

Asp Glu Gly Thr His Lys Pro Gln Gly His Ile Leu Glu His Ile Asn
450 455 460
Lys Phe Pro Asp Thr Leu Glu Asp Lys Lys Gln Leu Gln Arg Phe Leu
465 470 475 480
Gly Ile Leu Thr Tyr Ala Ser Asp Tyr Ile Pro Lys Leu Ala Gln Ile
485 490 495

Arg Lys Pro Leu Gln Ala Lys Leu Lys Glu Asn Val Pro Trp Lys Trp
500 505 510

Thr Lys Glu Asp Thr Leu Tyr Met Gln Lys Val Lys Lys Asn Leu Gln
515 520 525

Gly Phe Pro Pro Leu His His Pro Leu Pro Glu Glu Lys Leu Ile Ile
530 535 540

Glu Thr Asp Ala Ser Asp Asp Tyr Trp Gly Gly Met Leu Lys Ala Ile
545 550 555 560

Lys Ile Asn Glu Gly Thr Asn Thr Glu Leu Ile Cys Arg Tyr Ala Ser
565 570 575

Gly Ser Phe Lys Ala Ala Glu Lys Asn Tyr His Ser Asn Asp Lys Glu
580 585 590

Thr Leu Ala Val Ile Asn Thr Ile Lys Lys Phe Ser Ile Tyr Leu Thr
595 600 605

Pro Val His Phe Leu Ile Arg Thr Asp Asn Thr His Phe Lys Ser Phe
610 615 620

Val Asn Leu Asn Tyr Lys Gly Asp Ser Lys Leu Gly Arg Asn Ile Arg
625 630 635 640

Trp Gln Ala Trp Leu Ser His Tyr Ser Phe Asp Val Glu His Ile Lys
645 650 655

Gly Thr Asp Asn His Phe Ala Asp Phe Leu Ser Arg Glu Phe Asn Arg
660 665 670

Val Asn Ser
675

<210> 96

<211> 366

<212> PRT

<213> Unknown

<220><223> Klebsiella pneumonia

<400> 96

Met Lys Glu Lys Ile Ser Lys Ile Asp Lys Asn Phe Tyr Thr Asp Ile

1 5 10 15

Phe Ile Lys Thr Ser Phe Gln Asn Glu Phe Glu Ala Gly Gly Val Ile
20 25 30
Pro Pro Ile Ala Lys Asn Gln Val Ser Thr Ile Ser Asn Lys Asn Lys
35 40 45
Thr Phe Tyr Ser Leu Ala His Ser Ser Pro His Tyr Ser Ile Gln Thr
50 55 60
Arg Ile Glu Lys Phe Leu Leu Lys Asn Ile Pro Leu Ser Ala Ser Ser
65 70 75 80

Phe Ala Phe Arg Lys Glu Arg Ser Tyr Leu His Tyr Leu Glu Pro His
85 90 95
Thr Gln Asn Val Lys Tyr Cys His Leu Asp Ile Val Ser Phe Phe His
100 105 110
Ser Ile Asp Val Asn Ile Val Arg Asp Thr Phe Ser Val Tyr Phe Ser
115 120 125
Asp Glu Phe Leu Val Lys Glu Lys Gln Ser Leu Leu Asp Ala Phe Met
130 135 140

Ala Ser Val Thr Leu Thr Ala Glu Leu Asp Gly Val Glu Lys Thr Phe
145 150 155 160
Ile Pro Met Gly Phe Lys Ser Ser Pro Ser Ile Ser Asn Ile Ile Phe
165 170 175
Arg Lys Ile Asp Ile Leu Ile Gln Lys Phe Cys Asp Lys Asn Lys Ile
180 185 190
Thr Tyr Thr Arg Tyr Ala Asp Asp Leu Leu Phe Ser Thr Lys Lys Glu
195 200 205

Asn Asn Ile Leu Ser Ser Thr Phe Phe Ile Asn Glu Ile Ser Ser Ile
210 215 220
Leu Ser Ile Asn Lys Phe Lys Leu Asn Lys Ser Lys Tyr Leu Tyr Lys
225 230 235 240
Glu Gly Thr Ile Ser Leu Gly Gly Tyr Val Ile Glu Asn Ile Leu Lys
245 250 255

Asp Asn Ser Ser Gly Asn Ile Arg Leu Ser Ser Ser Lys Leu Asn Pro
260 265 270

Leu Tyr Lys Ala Leu Tyr Glu Ile Lys Lys Gly Ser Ser Ser Lys His
275 280 285
Ile Cys Ile Lys Val Phe Asn Leu Lys Leu Lys Arg Phe Ile Tyr Lys
290 295 300
Lys Asn Lys Glu Lys Phe Glu Ala Lys Phe Tyr Ser Ser Gln Leu Lys
305 310 315 320
Asn Lys Leu Leu Gly Tyr Arg Ser Tyr Leu Leu Ser Phe Val Ile Phe
325 330 335

His Lys Lys Tyr Lys Cys Ile Asn Pro Ile Phe Leu Glu Lys Cys Val
340 345 350
Phe Leu Ile Ser Glu Ile Glu Ser Ile Met Asn Arg Lys Phe
355 360 365
<210> 97
<211> 263
<212> PRT
<213> Escherichia coli
<400> 97
Met Lys Ile Thr Ser Asn Asn Val Thr Ala Val Ile Asn Gly Lys Gly
1 5 10 15
Trp His Ser Ile Asn Trp Lys Lys Cys His Gln His Val Lys Thr Ile

20 25 30
Gln Thr Arg Ile Ala Lys Ala Ala Cys Asn Gln Gln Trp Arg Thr Val
35 40 45
Gly Arg Leu Gln Arg Leu Leu Val Arg Ser Phe Ser Ala Arg Ala Leu
50 55 60
Ala Val Lys Arg Val Thr Glu Asn Ser Gly Arg Lys Thr Pro Gly Val
65 70 75 80
Asp Gly Gln Ile Trp Ser Thr Pro Glu Ser Lys Trp Glu Ala Ile Phe

85 90 95

Lys Leu Arg Arg Lys Gly Tyr Lys Pro Leu Pro Leu Lys Arg Val Phe
100 105 110
Ile Pro Lys Ser Asn Gly Lys Lys Arg Pro Leu Gly Ile Pro Val Met
115 120 125
Leu Asp Arg Ala Met Gln Ala Leu His Leu Leu Gly Leu Glu Pro Val
130 135 140
Ser Glu Thr Asn Ala Asp His Asn Ser Tyr Gly Phe Arg Pro Ala Arg

145 150 155 160
Cys Thr Ala Asp Ala Ile Gln Gln Val Cys Asn Met Tyr Ser Ser Arg
165 170 175
Asn Ala Ser Lys Trp Val Leu Glu Gly Asp Ile Lys Gly Cys Phe Glu
180 185 190
His Ile Ser His Glu Trp Leu Leu Glu Asn Ile Pro Met Asp Lys Gln
195 200 205
Ile Leu Arg Asn Trp Leu Lys Ala Gly Ile Ile Glu Lys Ser Ile Phe

210 215 220
Ser Lys Thr Leu Ser Gly Thr Pro Gln Gly Gly Ile Ile Ser Pro Val
225 230 235 240
Leu Ala Asn Met Ala Leu Asp Gly Leu Glu Arg Leu Leu Gln Asn Arg
245 250 255
Phe Gly Arg Asn Arg Leu Ile
260

<210> 98
<211> 438
<212> PRT
<213> Streptococcus pyogenes
<400> 98
Met Ser Lys Ile Lys Ile Asn Tyr Glu Lys Tyr His Ile Lys Pro Phe

1 5 10 15
Pro His Phe Asp Gln Arg Ile Lys Val Asn Lys Lys Val Lys Glu Asn
20 25 30
Leu Gln Asn Pro Phe Tyr Ile Ala Ala His Ser Phe Tyr Pro Phe Ile

35 40 45

His Tyr Lys Lys Ile Ser Tyr Lys Phe Lys Asn Gly Thr Leu Ser Ser

50 55 60

Pro Lys Glu Arg Asp Ile Phe Tyr Ser Gly His Met Asp Gly Tyr Ile

65 70 75 80

Tyr Lys His Tyr Gly Glu Ile Leu Asn His Lys Tyr Asn Asn Thr Cys

85 90 95

Ile Gly Lys Gly Ile Asp His Val Ser Leu Ala Tyr Arg Asn Asn Lys

100 105 110

Met Gly Lys Ser Asn Ile His Phe Ala Ala Glu Val Ile Asn Phe Ile

115 120 125

Ser Glu Gln Gln Gln Ala Phe Ile Phe Val Ser Asp Phe Ser Ser Tyr

130 135 140

Phe Asp Ser Leu Asp His Ala Ile Leu Lys Glu Lys Leu Ile Glu Val

145 150 155 160

Leu Glu Glu Gln Asp Lys Leu Ser Lys Asp Trp Trp Asn Val Phe Lys

165 170 175

His Ile Thr Arg Tyr Asn Trp Val Glu Lys Glu Glu Val Ile Ser Asp

180 185 190

Leu Glu Cys Thr Lys Glu Lys Ile Ala Arg Asp Lys Lys Ser Arg Glu

195 200 205

Arg Tyr Tyr Thr Pro Ala Glu Phe Arg Glu Phe Arg Lys Arg Val Asn

210 215 220

Ile Lys Ser Asn Asp Thr Gly Val Gly Ile Pro Gln Gly Thr Ala Ile

225 230 235 240

Ser Ala Val Leu Ala Asn Val Tyr Ala Ile Asp Leu Asp Gln Lys Leu

245 250 255

Asn Gln Tyr Ala Leu Lys Tyr Gly Gly Ile Tyr Arg Arg Tyr Ser Asp

260 265 270

Asp Ile Ile Met Val Leu Pro Met Thr Ser Asp Gly Gln Asp Pro Ser

275 280 285

Asn Asp His Val Ser Phe Ile Lys Ser Val Val Lys Arg Asn Lys Val
290 295 300
Thr Met Gly Asp Ser Lys Thr Ser Val Leu Tyr Tyr Ala Asn Asn Asn
305 310 315 320
Ile Tyr Glu Asp Tyr Gln Arg Lys Arg Glu Ser Lys Met Asp Tyr Leu

325 330 335
Gly Phe Ser Phe Asp Gly Met Thr Val Lys Ile Arg Glu Lys Ser Leu
340 345 350
Phe Lys Tyr Tyr His Arg Thr Tyr Lys Lys Ile Asn Ser Ile Asn Trp
355 360 365
Ala Ser Val Lys Lys Glu Lys Lys Val Gly Arg Lys Lys Leu Tyr Leu
370 375 380
Leu Tyr Ser His Leu Gly Arg Asn Tyr Lys Gly His Gly Asn Phe Ile

385 390 395 400
Ser Tyr Cys Lys Lys Ala His Ala Val Phe Glu Gly Asn Lys Lys Ile
405 410 415
Glu Ser Leu Ile Asn Gln Gln Ile Lys Arg His Trp Lys Lys Ile Gln
420 425 430
Lys Arg Leu Val Asp Val
435

<210> 99
<211> 426
<212> PRT
<213> Eubacterium rectale
<400> 99
Asp Thr Ser Asn Leu Met Glu Gln Ile Leu Ser Ser Asp Asn Leu Asn

1 5 10 15
Arg Ala Tyr Leu Gln Val Val Arg Asn Lys Gly Ala Glu Gly Val Asp
20 25 30
Gly Met Lys Tyr Thr Glu Leu Lys Glu His Leu Ala Lys Asn Gly Glu
35 40 45
Thr Ile Lys Gly Gln Leu Arg Thr Arg Lys Tyr Lys Pro Gln Pro Ala

50 55 60

Arg Arg Val Glu Ile Pro Lys Pro Asp Gly Gly Val Arg Asn Leu Gly

65 70 75 80

Val Pro Thr Val Thr Asp Arg Phe Ile Gln Gln Ala Ile Ala Gln Val

85 90 95

Leu Thr Pro Ile Tyr Glu Glu Gln Phe His Asp His Ser Tyr Gly Phe

100 105 110

Arg Pro Asn Arg Cys Ala Gln Gln Ala Ile Leu Thr Ala Leu Asn Ile

115 120 125

Met Asn Asp Gly Asn Asp Trp Ile Val Asp Ile Asp Leu Glu Lys Phe

130 135 140

Phe Asp Thr Val Asn His Asp Lys Leu Met Thr Leu Ile Gly Arg Thr

145 150 155 160

Ile Lys Asp Gly Asp Val Ile Ser Ile Val Arg Lys Tyr Leu Val Ser

165 170 175

Gly Ile Met Ile Asp Asp Glu Tyr Glu Asp Ser Ile Val Gly Thr Pro

180 185 190

Gln Gly Gly Asn Leu Ser Pro Leu Leu Ala Asn Ile Met Leu Asn Glu

195 200 205

Leu Asp Lys Glu Met Glu Lys Arg Gly Leu Asn Phe Val Arg Tyr Ala

210 215 220

Asp Asp Cys Ile Ile Met Val Gly Ser Glu Met Ser Ala Asn Arg Val

225 230 235 240

Met Arg Asn Ile Ser Arg Phe Ile Glu Glu Lys Leu Gly Leu Lys Val

245 250 255

Asn Met Thr Lys Ser Lys Val Asp Arg Pro Ser Gly Leu Lys Tyr Leu

260 265 270

Gly Phe Gly Phe Tyr Phe Asp Pro Arg Ala His Gln Phe Lys Ala Lys

275 280 285

Pro His Ala Lys Ser Val Ala Lys Phe Lys Lys Arg Met Lys Glu Leu

290 295 300

Thr Cys Arg Ser Trp Gly Val Ser Asn Ser Tyr Lys Val Glu Lys Leu
305 310 315 320
Asn Gln Leu Ile Arg Gly Trp Ile Asn Tyr Phe Lys Ile Gly Ser Met

325 330 335
Lys Thr Leu Cys Lys Glu Leu Asp Ser Arg Ile Arg Tyr Arg Leu Arg
340 345 350
Met Cys Ile Trp Lys Gln Trp Lys Thr Pro Gln Asn Gln Glu Lys Asn
355 360 365
Leu Val Lys Leu Gly Ile Asp Arg Asn Thr Ala Arg Arg Val Ala Tyr
370 375 380
Thr Gly Lys Arg Ile Ala Tyr Val Cys Asn Lys Gly Ala Val Asn Val

385 390 395 400
Ala Ile Ser Asn Lys Arg Leu Ala Ser Phe Gly Leu Ile Ser Met Leu
405 410 415
Asp Tyr Tyr Ile Glu Lys Cys Val Thr Cys
420 425

<210> 100
<211> 412
<212> PRT
<213> Geobacillus stearothermophilus
<400> 100
Ala Leu Leu Glu Arg Ile Leu Ala Arg Asp Asn Leu Ile Thr Ala Leu
1 5 10 15
Lys Arg Val Glu Ala Asn Gln Gly Ala Pro Gly Ile Asp Gly Val Ser

20 25 30
Thr Asp Gln Leu Arg Asp Tyr Ile Arg Ala His Trp Ser Thr Ile His
35 40 45
Ala Gln Leu Leu Ala Gly Thr Tyr Arg Pro Ala Pro Val Arg Arg Val
50 55 60
Glu Ile Pro Lys Pro Gly Gly Gly Thr Arg Gln Leu Gly Ile Pro Thr
65 70 75 80

Val Val Asp Arg Leu Ile Gln Gln Ala Ile Leu Gln Glu Leu Thr Pro

85 90 95

Ile Phe Asp Pro Asp Phe Ser Ser Ser Ser Phe Gly Phe Arg Pro Gly

100 105 110

Arg Asn Ala His Asp Ala Val Arg Gln Ala Gln Gly Tyr Ile Gln Glu

115 120 125

Gly Tyr Arg Tyr Val Val Asp Met Asp Leu Glu Lys Phe Phe Asp Arg

130 135 140

Val Asn His Asp Ile Leu Met Ser Arg Val Ala Arg Lys Val Lys Asp

145 150 155 160

Lys Arg Val Leu Lys Leu Ile Arg Ala Tyr Leu Gln Ala Gly Val Met

165 170 175

Ile Glu Gly Val Lys Val Gln Thr Glu Glu Gly Thr Pro Gln Gly Gly

180 185 190

Pro Leu Ser Pro Leu Leu Ala Asn Ile Leu Leu Asp Asp Leu Asp Lys

195 200 205

Glu Leu Glu Lys Arg Gly Leu Lys Phe Cys Arg Tyr Ala Asp Asp Cys

210 215 220

Asn Ile Tyr Val Lys Ser Leu Arg Ala Gly Gln Arg Val Lys Gln Ser

225 230 235 240

Ile Gln Arg Phe Leu Glu Lys Thr Leu Lys Leu Lys Val Asn Glu Glu

245 250 255

Lys Ser Ala Val Asp Arg Pro Trp Lys Arg Ala Phe Leu Gly Phe Ser

260 265 270

Phe Thr Pro Glu Arg Lys Ala Arg Ile Arg Leu Ala Pro Arg Ser Ile

275 280 285

Gln Arg Leu Lys Gln Arg Ile Arg Gln Leu Thr Asn Pro Asn Trp Ser

290 295 300

Ile Ser Met Pro Glu Arg Ile His Arg Val Asn Gln Tyr Val Met Gly

305 310 315 320

Trp Ile Gly Tyr Phe Arg Leu Val Glu Thr Pro Ser Val Leu Gln Thr

325 330 335

Ile Glu Gly Trp Ile Arg Arg Arg Leu Arg Leu Cys Gln Trp Leu Gln

340 345 350

Trp Lys Arg Val Arg Thr Arg Ile Arg Glu Leu Arg Ala Leu Gly Leu

355 360 365

Lys Glu Thr Ala Val Met Glu Ile Ala Asn Thr Arg Lys Gly Ala Trp

370 375 380

Arg Thr Thr Lys Thr Pro Gln Leu His Gln Ala Leu Gly Lys Thr Tyr

385 390 395 400

Trp Thr Ala Gln Gly Leu Lys Ser Leu Thr Gln Arg

405 410

<210> 101

<400> 101

000

<210> 102

<400> 102

000

<210> 103

<400> 103

000

<210> 104

<400> 104

000

<210> 105

<400> 105

000

<210> 106

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 106

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 107

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 107

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210

> 108

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 108

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 109

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 109

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Lys Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 110

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 110

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Arg Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 111

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 111

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Lys Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210

> 112

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 112

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Pro Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 113

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 113

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Gly Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 114

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 114

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Lys Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 115

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 115

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Lys Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210

> 116

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 116

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 117

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 117

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Gly Gly Ser Ser Leu

515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Gln Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asn Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro

675

<210> 118

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 118

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Arg Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 119

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 119

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Pro Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Lys Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210
> 120
<211> 677
<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 120

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Leu Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Lys Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Ala Leu Phe Asp Glu Ala Leu Arg Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Lys Phe Leu

290 295 300

Gly Thr Ala Gly Asn Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Gly Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Lys Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Gly Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575
Lys Leu Asn Val Tyr Thr Asn Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile Gln Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asn Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 121
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 121

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Leu Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Lys Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Ala Leu Phe Asn Glu Ala Leu Arg Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Lys Phe Leu

290 295 300

Gly Thr Ala Gly Asn Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Gly Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Lys Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Gly Gly Ser Ser Leu

515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asn Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590
Ile Gln Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asn Gln Ala Ala

645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 122
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 122
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670
Glu Asn Ser Ser Pro
675

<210> 123
<211> 2117
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 123
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15
Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn
20 25 30

Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys

35 40 45

Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn

50 55 60

Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr

65 70 75 80

Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg

85 90 95

Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp

100 105 110

Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp

115 120 125

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val

130 135 140

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu

145 150 155 160

Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu

165 170 175

Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu

180 185 190

Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln

195 200 205

Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val

210 215 220

Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu

225 230 235 240

Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe

245 250 255

Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser

260 265 270

Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr

275 280 285

Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr

290 295 300

Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu

305 310 315 320

Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser

325 330 335

Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu

340 345 350

Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile

355 360 365

Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly

370 375 380

Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys

385 390 395 400

Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu

405 410 415

Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile

420 425 430

His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr

435 440 445

Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe

450 455 460

Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe

465 470 475 480

Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe

485 490 495

Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg

500 505 510

Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys

515 520 525

His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys
530 535 540
Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly

545 550 555 560
Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys
565 570 575
Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys
580 585 590
Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser
595 600 605
Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe

610 615 620
Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr
625 630 635 640
Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr
645 650 655
Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg
660 665 670
Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile

675 680 685
Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp
690 695 700
Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu
705 710 715 720
Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp
725 730 735
Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys

740 745 750
Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val
755 760 765
Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu

770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys

785 790 795 800

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu

805 810 815

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr

820 825 830

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe

850 855 860

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys

865 870 875 880

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys

885 890 895

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln

900 905 910

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu

915 920 925

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln

930 935 940

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

945 950 955 960

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu

965 970 975

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

980 985 990

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

995 1000 1005

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1010 1015 1020

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1025 1030 1035

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1040 1045 1050

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1055 1060 1065

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1070 1075 1080

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1085 1090 1095

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1100 1105 1110

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1115 1120 1125

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1130 1135 1140

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala

1145 1150 1155

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1160 1165 1170

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1175 1180 1185

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1190 1195 1200

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1205 1210 1215

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1220 1225 1230

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1235 1240 1245

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1250 1255 1260

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1265 1270 1275

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1280 1285 1290

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1295 1300 1305

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1310 1315 1320

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1325 1330 1335

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1340 1345 1350

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1355 1360 1365

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1370 1375 1380

Gly Gly Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu

1385 1390 1395

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly

1400 1405 1410

Ser Ser Gly Gly Ser Ser Thr Leu Asn Ile Glu Asp Glu Tyr Arg

1415 1420 1425

Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr

1430 1435 1440

Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly Met

1445 1450 1455

Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala

1460 1465 1470

Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu

1475 1480 1485

Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln

1490 1495 1500

Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu

1505 1510 1515

Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp

1520 1525 1530

Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val

1535 1540 1545

Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

1550 1555 1560

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg

1565 1570 1575

Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp

1580 1585 1590

Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro

1595 1600 1605

Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His

1610 1615 1620

Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu

1625 1630 1635

Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu

1640 1645 1650

Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn

1655 1660 1665

Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys

1670 1675 1680

Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp

1685 1690 1695

Leu Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro

1700 1705 1710

Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly Thr Ala Gly Phe

1715 1720 1725

Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu

1730 1735 1740

Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp Gly Pro Asp

1745 1750 1755

Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala

1760 1765 1770

Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu Phe

1775 1780 1785

Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

1790 1795 1800

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu

1805 1810 1815

Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala

1820 1825 1830

Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly

1835 1840 1845

Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val

1850 1855 1860

Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His

1865 1870 1875

Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro

1880 1885 1890

Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu

1895 1900 1905

Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly

1910 1915 1920

Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp His

1925 1930 1935

Thr Trp Tyr Thr Asp Gly Ser Ser Leu Leu Gln Glu Gly Gln Arg

1940 1945 1950

Lys Ala Gly Ala Ala Val Thr Thr Glu Thr Glu Val Ile Trp Ala

1955 1960 1965

Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg Ala Glu Leu Ile

1970 1975 1980

Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys Leu Asn

1985 1990 1995

Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His Ile His

2000 2005 2010

Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly Lys

2015 2020 2025

Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

2030 2035 2040

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln

2045 2050 2055

Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln

2060 2065 2070

Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr

2075 2080 2085

Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys Arg Thr

2090 2095 2100

Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val

2105 2110 2115

<210> 124

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 124

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys

1 5 10 15

Arg Lys Val

<210> 125

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 125

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys

1 5 10 15

Val

<210> 126

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 126

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 127
<211> 33
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 127
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1 5 10 15
Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

Ser

<210> 128
<211> 677
<212> PRT
<213> Unknown
<220><223> moloney murine leukemia virus
<400> 128
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 129

<400> 129

000

<210> 130

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 130

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 131

<211> 76

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 131

gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60

ggcaccgagt cggtgc 76

<210> 132

<211> 677

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<400> 132

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210
> 133
<211> 21
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 133
Ser Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys
1 5 10 15
Lys Lys Arg Lys Val
20
<210> 134
<211> 2117
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 134
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn
20 25 30

Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys
35 40 45
Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn
50 55 60
Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr
65 70 75 80

Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg
85 90 95
Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp
100 105 110
Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp
115 120 125
Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val
130 135 140

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu
145 150 155 160
Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu
165 170 175
Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu
180 185 190
Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln
195 200 205

Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val
210 215 220
Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu
225 230 235 240
Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe
245 250 255
Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser
260 265 270

Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr

275 280 285

Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr

290 295 300

Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu

305 310 315 320

Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser

325 330 335

Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu

340 345 350

Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile

355 360 365

Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly

370 375 380

Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys

385 390 395 400

Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu

405 410 415

Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile

420 425 430

His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr

435 440 445

Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe

450 455 460

Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe

465 470 475 480

Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe

485 490 495

Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg

500 505 510

Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys

515 520 525

His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys
530 535 540
Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly
545 550 555 560
Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys
565 570 575
Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys
580 585 590

Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser
595 600 605
Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe
610 615 620
Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr
625 630 635 640
Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr
645 650 655

Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg
660 665 670
Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile
675 680 685
Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp
690 695 700
Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu
705 710 715 720

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp
725 730 735
Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys
740 745 750
Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val
755 760 765
Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu

770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
785 790 795 800
Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu
805 810 815
His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr
820 825 830
Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe
850 855 860
Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys
865 870 875 880
Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys
885 890 895
Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln
900 905 910

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
915 920 925
Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln
930 935 940
Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
945 950 955 960
Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
965 970 975

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
980 985 990
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
995 1000 1005
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1010 1015 1020

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1025 1030 1035

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1040 1045 1050
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1055 1060 1065
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1070 1075 1080
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1085 1090 1095

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1100 1105 1110
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1115 1120 1125
Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1130 1135 1140
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala
1145 1150 1155

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1160 1165 1170
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1175 1180 1185
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1190 1195 1200
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1205 1210 1215

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1220 1225 1230
Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1235 1240 1245
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1250 1255 1260

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1265 1270 1275

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1280 1285 1290

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1295 1300 1305

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1310 1315 1320

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1325 1330 1335

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1340 1345 1350

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1355 1360 1365

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1370 1375 1380

Gly Gly Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu

1385 1390 1395

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly

1400 1405 1410

Ser Ser Gly Gly Ser Ser Thr Leu Asn Ile Glu Asp Glu Tyr Arg

1415 1420 1425

Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr

1430 1435 1440

Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly Met

1445 1450 1455

Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala

1460 1465 1470

Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu

1475 1480 1485

Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln
1490 1495 1500
Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu
1505 1510 1515

Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp
1520 1525 1530
Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val
1535 1540 1545
Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
1550 1555 1560
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg
1565 1570 1575

Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp
1580 1585 1590
Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro
1595 1600 1605
Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His
1610 1615 1620
Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu
1625 1630 1635

Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu
1640 1645 1650
Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn
1655 1660 1665
Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys
1670 1675 1680
Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp
1685 1690 1695

Leu Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro
1700 1705 1710
Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly Lys Ala Gly Phe

1715 1720 1725

Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu

1730 1735 1740

Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp Gly Pro Asp

1745 1750 1755

Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala

1760 1765 1770

Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu Phe

1775 1780 1785

Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

1790 1795 1800

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu

1805 1810 1815

Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala

1820 1825 1830

Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly

1835 1840 1845

Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val

1850 1855 1860

Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His

1865 1870 1875

Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro

1880 1885 1890

Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu

1895 1900 1905

Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly

1910 1915 1920

Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp His

1925 1930 1935

Thr Trp Tyr Thr Asp Gly Ser Ser Leu Leu Gln Glu Gly Gln Arg

1940 1945 1950

Lys Ala Gly Ala Ala Val Thr Thr Glu Thr Glu Val Ile Trp Ala
1955 1960 1965
Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg Ala Glu Leu Ile
1970 1975 1980
Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys Leu Asn
1985 1990 1995

Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His Ile His
2000 2005 2010
Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly Lys
2015 2020 2025
Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
2030 2035 2040
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln
2045 2050 2055

Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln
2060 2065 2070
Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr
2075 2080 2085
Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys Arg Thr
2090 2095 2100
Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val
2105 2110 2115

<210> 135
<400> 135
000
<210> 136
<211> 19
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 136
Asn Leu Ser Lys Arg Pro Ala Ala Ile Lys Lys Ala Gly Gln Ala Lys

1 5 10 15

Lys Lys Lys

<210> 137

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 137

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365
<210> 138
<400> 138
000
<210> 139
<211> 677
<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 139

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 140

<400> 140

000

<210> 141

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 141

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365
<210> 142
<400> 142
000
<210> 143
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 143
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 144

<400> 144

000

<210> 145

<211> 2117

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 145

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15
Arg Lys Val Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr
20 25 30
Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe
35 40 45
Pro Gln Ala Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln
50 55 60
Ala Pro Leu Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile
65 70 75 80
Lys Gln Tyr Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His
85 90 95
Ile Gln Arg Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro
100 105 110
Trp Asn Thr Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr
115 120 125
Arg Pro Val Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile
130 135 140
His Pro Thr Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro
145 150 155 160
Ser His Gln Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys
165 170 175
Leu Arg Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg
180 185 190
Asp Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro
195 200 205
Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg
210 215 220
Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln
225 230 235 240
Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln

245 250 255

Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg

260 265 270

Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu

275 280 285

Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys

290 295 300

Glu Thr Val Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg

305 310 315 320

Glu Phe Leu Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe

325 330 335

Ala Glu Met Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu

340 345 350

Phe Asn Trp Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln

355 360 365

Ala Leu Leu Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro

370 375 380

Phe Glu Leu Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu

385 390 395 400

Thr Gln Lys Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys

405 410 415

Lys Leu Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val

420 425 430

Ala Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly

435 440 445

Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys

450 455 460

Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln

465 470 475 480

Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala

485 490 495

Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His
500 505 510
Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu
515 520 525
Thr Asp Gln Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly
530 535 540
Ser Ser Leu Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr

545 550 555 560
Thr Glu Thr Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser
565 570 575
Ala Gln Arg Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala
580 585 590
Glu Gly Lys Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala
595 600 605
Thr Ala His Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr

610 615 620
Ser Glu Gly Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu
625 630 635 640
Lys Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly
645 650 655
His Gln Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp
660 665 670
Gln Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr

675 680 685
Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Ser Gly Gly Ser
690 695 700
Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
705 710 715 720
Ser Gly Gly Ser Ser Gly Gly Ser Ser Asp Lys Lys Tyr Ser Ile Gly
725 730 735
Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu

740 745 750

Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg

755 760 765

His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly

770 775 780

Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr

785 790 795 800

Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn

805 810 815

Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser

820 825 830

Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly

835 840 845

Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr

850 855 860

His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg

865 870 875 880

Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe

885 890 895

Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu

900 905 910

Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro

915 920 925

Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu

930 935 940

Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu

945 950 955 960

Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu

965 970 975

Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu

980 985 990

Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala

995 1000 1005

Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn

1010 1015 1020

Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr

1025 1030 1035

Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr

1040 1045 1050

Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg

1055 1060 1065

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser

1070 1075 1080

Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu

1085 1090 1095

Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

1100 1105 1110

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

1115 1120 1125

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His

1130 1135 1140

Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr

1145 1150 1155

Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr

1160 1165 1170

Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

1175 1180 1185

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro

1190 1195 1200

Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

1205 1210 1215

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu

1220 1225 1230

Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val

1235 1240 1245

Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg

1250 1255 1260

Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp

1265 1270 1275

Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys

1280 1285 1290

Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile

1295 1300 1305

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His

1310 1315 1320

Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu

1325 1330 1335

Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

1340 1345 1350

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

1355 1360 1365

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg

1370 1375 1380

Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile

1385 1390 1395

Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser

1400 1405 1410

Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

1415 1420 1425

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly

1430 1435 1440

Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

1445 1450 1455

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
1460 1465 1470
Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val

1475 1480 1485
Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys
1490 1495 1500
Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu
1505 1510 1515
Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln
1520 1525 1530
Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg

1535 1540 1545
Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp
1550 1555 1560
Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp
1565 1570 1575
Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
1580 1585 1590
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

1595 1600 1605
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg
1610 1615 1620
Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
1625 1630 1635
Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
1640 1645 1650
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn

1655 1660 1665
Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val
1670 1675 1680
Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1685 1690 1695

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1700 1705 1710

Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys

1715 1720 1725

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1730 1735 1740

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1745 1750 1755

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1760 1765 1770

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1775 1780 1785

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1790 1795 1800

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1805 1810 1815

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1820 1825 1830

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1835 1840 1845

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1850 1855 1860

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1865 1870 1875

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1880 1885 1890

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1895 1900 1905

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1910 1915 1920

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1925 1930 1935

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1940 1945 1950

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1955 1960 1965

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1970 1975 1980

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile

1985 1990 1995

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

2000 2005 2010

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

2015 2020 2025

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

2030 2035 2040

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

2045 2050 2055

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

2060 2065 2070

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

2075 2080 2085

Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly Gly Ser Lys Arg Thr

2090 2095 2100

Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val

2105 2110 2115

<210> 146

<400> 146

000

<210> 147

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 147

```
Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1               5                  10                  15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
            20                  25                  30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
        35                  40                  45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
    50                  55                  60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65                  70                  75                  80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
                85                  90                  95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
            100                 105                 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
        115                 120                 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
    130                 135                 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145                 150                 155                 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
                165                 170                 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
            180                 185                 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
        195                 200                 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
    210                 215                 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
```

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 148

<400> 148

000

<210> 149

<211> 677

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<400> 149

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp
325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 150

<211> 2144

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 150

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys

1 5 10 15

Arg Lys Val Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr

20 25 30

Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe
35 40 45
Pro Gln Ala Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln
50 55 60
Ala Pro Leu Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile
65 70 75 80
Lys Gln Tyr Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His
85 90 95

Ile Gln Arg Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro
100 105 110
Trp Asn Thr Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr
115 120 125
Arg Pro Val Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile
130 135 140
His Pro Thr Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro
145 150 155 160

Ser His Gln Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys
165 170 175
Leu Arg Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg
180 185 190
Asp Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro
195 200 205
Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg
210 215 220

Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln
225 230 235 240
Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln
245 250 255
Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg
260 265 270

Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu
275 280 285

Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys
290 295 300
Glu Thr Val Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg
305 310 315 320
Glu Phe Leu Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe
325 330 335
Ala Glu Met Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu
340 345 350

Phe Asn Trp Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln
355 360 365
Ala Leu Leu Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro
370 375 380
Phe Glu Leu Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu
385 390 395 400
Thr Gln Lys Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys
405 410 415

Lys Leu Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val
420 425 430
Ala Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly
435 440 445
Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys
450 455 460
Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln
465 470 475 480

Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala
485 490 495
Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His
500 505 510
Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu

515 520 525

Thr Asp Gln Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly

530 535 540

Ser Ser Leu Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr

545 550 555 560

Thr Glu Thr Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser

565 570 575

Ala Gln Arg Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala

580 585 590

Glu Gly Lys Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala

595 600 605

Thr Ala His Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr

610 615 620

Ser Glu Gly Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu

625 630 635 640

Lys Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly

645 650 655

His Gln Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp

660 665 670

Gln Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr

675 680 685

Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Ser Gly Gly Ser

690 695 700

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

705 710 715 720

Ala Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly Ser Ala Ala

725 730 735

Pro Ala Ala Lys Lys Lys Lys Leu Asp Gly Ser Gly Ser Gly Gly Ser

740 745 750

Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr

755 760 765

Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser
770 775 780
Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys
785 790 795 800

Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala
805 810 815
Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn
820 825 830
Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val
835 840 845
Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu
850 855 860

Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu
865 870 875 880
Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys
885 890 895
Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala
900 905 910
Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp
915 920 925

Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val
930 935 940
Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly
945 950 955 960
Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg
965 970 975
Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu
980 985 990

Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys
995 1000 1005
Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

1010 1015 1020

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly

1025 1030 1035

Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

1040 1045 1050

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr

1055 1060 1065

Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His

1070 1075 1080

His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu

1085 1090 1095

Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly

1100 1105 1110

Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr

1115 1120 1125

Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu

1130 1135 1140

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg

1145 1150 1155

Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu

1160 1165 1170

Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

1175 1180 1185

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

1190 1195 1200

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala

1205 1210 1215

Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe

1220 1225 1230

Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu

1235 1240 1245

```
Arg Met  Thr Asn Phe Asp Lys  Asn Leu Pro Asn Glu  Lys Val Leu
    1250                 1255                 1260
Pro Lys  His Ser Leu Leu Tyr  Glu Tyr Phe Thr Val  Tyr Asn Glu
    1265                 1270                 1275
Leu Thr  Lys Val Lys Tyr Val  Thr Glu Gly Met Arg  Lys Pro Ala
    1280                 1285                 1290

Phe Leu  Ser Gly Glu Gln Lys  Lys Ala Ile Val Asp  Leu Leu Phe
    1295                 1300                 1305
Lys Thr  Asn Arg Lys Val Thr  Val Lys Gln Leu Lys  Glu Asp Tyr
    1310                 1315                 1320
Phe Lys  Lys Ile Glu Cys Phe  Asp Ser Val Glu Ile  Ser Gly Val
    1325                 1330                 1335
Glu Asp  Arg Phe Asn Ala Ser  Leu Gly Thr Tyr His  Asp Leu Leu
    1340                 1345                 1350

Lys Ile  Ile Lys Asp Lys Asp  Phe Leu Asp Asn Glu  Glu Asn Glu
    1355                 1360                 1365
Asp Ile  Leu Glu Asp Ile Val  Leu Thr Leu Thr Leu  Phe Glu Asp
    1370                 1375                 1380
Arg Glu  Met Ile Glu Glu Arg  Leu Lys Thr Tyr Ala  His Leu Phe
    1385                 1390                 1395
Asp Asp  Lys Val Met Lys Gln  Leu Lys Arg Arg Arg  Tyr Thr Gly
    1400                 1405                 1410

Trp Gly  Arg Leu Ser Arg Lys  Leu Ile Asn Gly Ile  Arg Asp Lys
    1415                 1420                 1425
Gln Ser  Gly Lys Thr Ile Leu  Asp Phe Leu Lys Ser  Asp Gly Phe
    1430                 1435                 1440
Ala Asn  Arg Asn Phe Met Gln  Leu Ile His Asp Asp  Ser Leu Thr
    1445                 1450                 1455
Phe Lys  Glu Asp Ile Gln Lys  Ala Gln Val Ser Gly  Gln Gly Asp
    1460                 1465                 1470

Ser Leu  His Glu His Ile Ala  Asn Leu Ala Gly Ser  Pro Ala Ile
```

1475 1480 1485

Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

1490 1495 1500

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met

1505 1510 1515

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

1520 1525 1530

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser

1535 1540 1545

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn

1550 1555 1560

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr

1565 1570 1575

Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val

1580 1585 1590

Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp

1595 1600 1605

Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp

1610 1615 1620

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp

1625 1630 1635

Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp

1640 1645 1650

Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

1655 1660 1665

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

1670 1675 1680

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr

1685 1690 1695

Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu

1700 1705 1710

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr
1715 1720 1725
Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr
1730 1735 1740
Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys
1745 1750 1755
Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val
1760 1765 1770

Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr
1775 1780 1785
Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr
1790 1795 1800
Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile
1805 1810 1815
Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg
1820 1825 1830

Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn
1835 1840 1845
Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu
1850 1855 1860
Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys
1865 1870 1875
Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr
1880 1885 1890

Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys
1895 1900 1905
Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
1910 1915 1920
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu
1925 1930 1935
Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu

1940 1945 1950

Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
1955 1960 1965
Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu
1970 1975 1980
Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu
1985 1990 1995
Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
2000 2005 2010

Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile
2015 2020 2025
Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp
2030 2035 2040
Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg
2045 2050 2055
Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu
2060 2065 2070

Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg
2075 2080 2085
Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile
2090 2095 2100
His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
2105 2110 2115
Gln Leu Gly Gly Asp Ser Gly Gly Ser Lys Arg Thr Ala Asp Gly
2120 2125 2130

Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val
2135 2140
<210> 151
<400> 151
000
<210> 152

<400> 152

000

<210> 153

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 153

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 154

<211> 2529

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 154

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys

1 5 10 15

Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn

20 25 30

Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys

35 40 45

Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn

50 55 60

Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr

65 70 75 80

Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg

85 90 95

Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp

100 105 110

Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp

115 120 125

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val

130 135 140

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu

145 150 155 160

Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu

165 170 175

Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu

180 185 190

Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln

195 200 205

Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val

210 215 220

Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu

225 230 235 240

Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe

245 250 255

Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser
260 265 270
Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr
275 280 285
Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr
290 295 300

Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu
305 310 315 320
Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser
325 330 335
Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu
340 345 350
Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile
355 360 365

Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly
370 375 380
Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys
385 390 395 400
Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu
405 410 415
Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile
420 425 430

His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr
435 440 445
Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe
450 455 460
Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe
465 470 475 480
Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe
485 490 495

Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg

500 505 510

Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys

515 520 525

His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys

530 535 540

Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly

545 550 555 560

Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys

565 570 575

Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys

580 585 590

Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser

595 600 605

Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe

610 615 620

Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr

625 630 635 640

Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr

645 650 655

Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg

660 665 670

Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile

675 680 685

Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp

690 695 700

Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu

705 710 715 720

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp

725 730 735

Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys

740 745 750

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val
755 760 765

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu
770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
785 790 795 800

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu
805 810 815

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr
820 825 830

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe
850 855 860

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys
865 870 875 880

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys
885 890 895

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln
900 905 910

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
915 920 925

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln
930 935 940

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
945 950 955 960

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
965 970 975

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
980 985 990

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

995 1000 1005

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1010 1015 1020
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1025 1030 1035
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1040 1045 1050
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1055 1060 1065

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1070 1075 1080
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1085 1090 1095
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1100 1105 1110
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1115 1120 1125

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1130 1135 1140
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala
1145 1150 1155
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1160 1165 1170
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1175 1180 1185

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1190 1195 1200
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1205 1210 1215
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1220 1225 1230

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1235 1240 1245

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1250 1255 1260
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1265 1270 1275
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1280 1285 1290
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1295 1300 1305

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1310 1315 1320
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1325 1330 1335
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1340 1345 1350
Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1355 1360 1365

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1370 1375 1380
Gly Gly Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu
1385 1390 1395
Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly
1400 1405 1410
Ser Ser Gly Gly Ser Ser Gly Ile Gln Gly Leu Ala Lys Leu Ile
1415 1420 1425

Ala Asp Val Ala Pro Ser Ala Ile Arg Glu Asn Asp Ile Lys Ser
1430 1435 1440
Tyr Phe Gly Arg Lys Val Ala Ile Asp Ala Ser Met Ser Ile Tyr
1445 1450 1455
Gln Phe Leu Ile Ala Val Arg Gln Gly Gly Asp Val Leu Gln Asn

1460 1465 1470

Glu Glu Gly Glu Thr Thr Ser His Leu Met Gly Met Phe Tyr Arg

1475 1480 1485

Thr Ile Arg Met Met Glu Asn Gly Ile Lys Pro Val Tyr Val Phe

1490 1495 1500

Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu Leu Ala Lys Arg

1505 1510 1515

Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln Gln Ala Gln

1520 1525 1530

Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys Arg Leu

1535 1540 1545

Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu Leu

1550 1555 1560

Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu

1565 1570 1575

Ala Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala

1580 1585 1590

Ala Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu

1595 1600 1605

Met Arg His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln

1610 1615 1620

Glu Phe His Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln

1625 1630 1635

Glu Gln Phe Val Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys

1640 1645 1650

Glu Ser Ile Arg Gly Ile Gly Pro Lys Arg Ala Val Asp Leu Ile

1655 1660 1665

Gln Lys His Lys Ser Ile Glu Glu Ile Val Arg Arg Leu Asp Pro

1670 1675 1680

Asn Lys Tyr Pro Val Pro Glu Asn Trp Leu His Lys Glu Ala His

1685 1690 1695

```
Gln Leu  Phe Leu Glu Pro Glu  Val Leu Asp Pro Glu  Ser Val Glu
    1700                1705                1710
Leu Lys  Trp Ser Glu Pro Asn  Glu Glu Glu Leu Ile  Lys Phe Met
    1715                1720                1725

Cys Gly  Glu Lys Gln Phe Ser  Glu Glu Arg Ile Arg  Ser Gly Val
    1730                1735                1740
Lys Arg  Leu Ser Lys Ser Arg  Gln Gly Ser Thr Gln  Gly Arg Leu
    1745                1750                1755
Asp Asp  Phe Phe Lys Val Thr  Gly Ser Leu Ser Ser  Ala Lys Arg
    1760                1765                1770
Lys Glu  Pro Glu Pro Lys Gly  Ser Thr Lys Lys Lys  Ala Lys Thr
    1775                1780                1785

Gly Ala  Ala Gly Lys Phe Lys  Arg Gly Lys Ser Gly  Gly Ser Ser
    1790                1795                1800
Gly Gly  Ser Ser Gly Ser Glu  Thr Pro Gly Thr Ser  Glu Ser Ala
    1805                1810                1815
Thr Pro  Glu Ser Ser Gly Gly  Ser Ser Gly Gly Ser  Ser Thr Leu
    1820                1825                1830
Asn Ile  Glu Asp Glu Tyr Arg  Leu His Glu Thr Ser  Lys Glu Pro
    1835                1840                1845

Asp Val  Ser Leu Gly Ser Thr  Trp Leu Ser Asp Phe  Pro Gln Ala
    1850                1855                1860
Trp Ala  Glu Thr Gly Gly Met  Gly Leu Ala Val Arg  Gln Ala Pro
    1865                1870                1875
Leu Ile  Ile Pro Leu Lys Ala  Thr Ser Thr Pro Val  Ser Ile Lys
    1880                1885                1890
Gln Tyr  Pro Met Ser Gln Glu  Ala Arg Leu Gly Ile  Lys Pro His
    1895                1900                1905

Ile Gln  Arg Leu Leu Asp Gln  Gly Ile Leu Val Pro  Cys Gln Ser
    1910                1915                1920
Pro Trp  Asn Thr Pro Leu Leu  Pro Val Lys Lys Pro  Gly Thr Asn
```

1925 1930 1935

Asp Tyr Arg Pro Val Gln Asp Leu Arg Glu Val Asn Lys Arg Val

1940 1945 1950

Glu Asp Ile His Pro Thr Val Pro Asn Pro Tyr Asn Leu Leu Ser

1955 1960 1965

Gly Leu Pro Pro Ser His Gln Trp Tyr Thr Val Leu Asp Leu Lys

1970 1975 1980

Asp Ala Phe Phe Cys Leu Arg Leu His Pro Thr Ser Gln Pro Leu

1985 1990 1995

Phe Ala Phe Glu Trp Arg Asp Pro Glu Met Gly Ile Ser Gly Gln

2000 2005 2010

Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe Lys Asn Ser Pro Thr

2015 2020 2025

Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala Asp Phe Arg Ile

2030 2035 2040

Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp Asp Leu Leu

2045 2050 2055

Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr Arg Ala

2060 2065 2070

Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala Lys

2075 2080 2085

Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

2090 2095 2100

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr

2105 2110 2115

Val Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu

2120 2125 2130

Phe Leu Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe

2135 2140 2145

Ala Glu Met Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr

2150 2155 2160

Leu Phe Asn Trp Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile
2165 2170 2175
Lys Gln Ala Leu Leu Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu
2180 2185 2190
Thr Lys Pro Phe Glu Leu Phe Val Asp Glu Lys Gln Gly Tyr Ala
2195 2200 2205

Lys Gly Val Leu Thr Gln Lys Leu Gly Pro Trp Arg Arg Pro Val
2210 2215 2220
Ala Tyr Leu Ser Lys Lys Leu Asp Pro Val Ala Ala Gly Trp Pro
2225 2230 2235
Pro Cys Leu Arg Met Val Ala Ala Ile Ala Val Leu Thr Lys Asp
2240 2245 2250
Ala Gly Lys Leu Thr Met Gly Gln Pro Leu Val Ile Leu Ala Pro
2255 2260 2265

His Ala Val Glu Ala Leu Val Lys Gln Pro Pro Asp Arg Trp Leu
2270 2275 2280
Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu Leu Asp Thr
2285 2290 2295
Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro Ala Thr
2300 2305 2310
Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu Asp
2315 2320 2325

Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
2330 2335 2340
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser
2345 2350 2355
Leu Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr
2360 2365 2370
Glu Thr Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser
2375 2380 2385

Ala Gln Arg Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met

2390 2395 2400

Ala Glu Gly Lys Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala

2405 2410 2415

Phe Ala Thr Ala His Ile His Gly Glu Ile Tyr Arg Arg Arg Gly

2420 2425 2430

Trp Leu Thr Ser Glu Gly Lys Glu Ile Lys Asn Lys Asp Glu Ile

2435 2440 2445

Leu Ala Leu Leu Lys Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile

2450 2455 2460

Ile His Cys Pro Gly His Gln Lys Gly His Ser Ala Glu Ala Arg

2465 2470 2475

Gly Asn Arg Met Ala Asp Gln Ala Ala Arg Lys Ala Ala Ile Thr

2480 2485 2490

Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile Glu Asn Ser Ser Pro

2495 2500 2505

Ser Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro

2510 2515 2520

Lys Lys Lys Arg Lys Val

2525

<210> 155

<400> 155

000

<210> 156

<400> 156

000

<210> 157

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 157

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

```
        755                 760                 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
    770                 775                 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785                 790                 795                 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
                805                 810                 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
            820                 825                 830
Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
        835                 840                 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

    850                 855                 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865                 870                 875                 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
                885                 890                 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
            900                 905                 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

        915                 920                 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
    930                 935                 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945                 950                 955                 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
                965                 970                 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

            980                 985                 990
Gly Thr Ala Leu Ile Lys Lys Tyr  Pro Lys Leu Glu Ser  Glu Phe Val
        995                 1000                1005
```

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 158

<400> 158

000

<210> 159

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 159

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670
Glu Asn Ser Ser Pro
675
<210> 160
<400> 160
000
<210> 161
<211> 379
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 161
Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser Ala
1 5 10 15
Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala Ile
20 25 30

Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln Gly

35 40 45

Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu Met

50 55 60

Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys Pro

65 70 75 80

Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu Leu

85 90 95

Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln Gln

100 105 110

Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys Arg

115 120 125

Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu Leu

130 135 140

Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu Ala

145 150 155 160

Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala Ala Thr

165 170 175

Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg His

180 185 190

Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His Leu

195 200 205

Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val Asp

210 215 220

Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly Ile

225 230 235 240

Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile Glu

245 250 255

Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu Asn

260 265 270

Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val Leu

275 280 285

Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu Glu
290 295 300

Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg Ile
305 310 315 320

Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr Gln
325 330 335

Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser Ala
340 345 350

Lys Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala Lys
355 360 365

Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys
370 375

<210> 162

<211> 1803

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 162

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Gly Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile
20 25 30

Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile
35 40 45

Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu
50 55 60

Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg
65 70 75 80

His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu
85 90 95

Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val
100 105 110

Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu
115 120 125
Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu
130 135 140
Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn

145 150 155 160
Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg
165 170 175
Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr
180 185 190
Ser Asp Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala
195 200 205
Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu

210 215 220
Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe
225 230 235 240
Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys
245 250 255
Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala
260 265 270
Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg

275 280 285
Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu
290 295 300
Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys
305 310 315 320
Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser
325 330 335
Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys

340 345 350
Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp

355 360 365

Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln

370 375 380

Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu

385 390 395 400

Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu

405 410 415

Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn

420 425 430

Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp

435 440 445

Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile

450 455 460

Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile

465 470 475 480

Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu

485 490 495

Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu

500 505 510

Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile

515 520 525

Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys

530 535 540

Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile

545 550 555 560

Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His

565 570 575

Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val

580 585 590

Leu Val Lys Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe

595 600 605

Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys
610 615 620
Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr
625 630 635 640
Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val
645 650 655
Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr

660 665 670
Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu
675 680 685
Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg
690 695 700
Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His
705 710 715 720
Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu

725 730 735
Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe
740 745 750
Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu
755 760 765
Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp
770 775 780
Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg

785 790 795 800
Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly
805 810 815
Asn Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn
820 825 830
Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met
835 840 845
Tyr His His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu

850 855 860

Gln Tyr Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr

865 870 875 880

Gly Asn Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile

885 890 895

Lys Lys Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile

900 905 910

Thr Asp Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu

915 920 925

Lys Pro Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe

930 935 940

Val Thr Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu

945 950 955 960

Val Asn Ser Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser

965 970 975

Asn Gln Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys

980 985 990

Ile Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu

995 1000 1005

Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr

1010 1015 1020

Leu Glu Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr

1025 1030 1035

Ile Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile

1040 1045 1050

Leu Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile

1055 1060 1065

Ile Lys Lys Gly Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser

1070 1075 1080

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1085 1090 1095

Gly Ser Ser Gly Gly Ser Ser Thr Leu Asn Ile Glu Asp Glu Tyr

1100 1105 1110

Arg Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser

1115 1120 1125

Thr Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly

1130 1135 1140

Met Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys

1145 1150 1155

Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln

1160 1165 1170

Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp

1175 1180 1185

Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu

1190 1195 1200

Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln

1205 1210 1215

Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

1220 1225 1230

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His

1235 1240 1245

Gln Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu

1250 1255 1260

Arg Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg

1265 1270 1275

Asp Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu

1280 1285 1290

Pro Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu

1295 1300 1305

His Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile

1310 1315 1320

Leu Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu

1325 1330 1335

Leu Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly

1340 1345 1350

Asn Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln

1355 1360 1365

Lys Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg

1370 1375 1380

Trp Leu Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr

1385 1390 1395

Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly Thr Ala Gly

1400 1405 1410

Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met Ala Ala Pro

1415 1420 1425

Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp Gly Pro

1430 1435 1440

Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr

1445 1450 1455

Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

1460 1465 1470

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln

1475 1480 1485

Lys Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys

1490 1495 1500

Leu Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val

1505 1510 1515

Ala Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met

1520 1525 1530

Gly Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu

1535 1540 1545

Val Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr

1550 1555 1560

His Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly

1565 1570 1575

Pro Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu

1580 1585 1590

Glu Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala His

1595 1600 1605

Gly Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp

1610 1615 1620

His Thr Trp Tyr Thr Asp Gly Ser Ser Leu Leu Gln Glu Gly Gln

1625 1630 1635

Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr Glu Val Ile Trp

1640 1645 1650

Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg Ala Glu Leu

1655 1660 1665

Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys Leu

1670 1675 1680

Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His Ile

1685 1690 1695

His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

1700 1705 1710

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala

1715 1720 1725

Leu Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His

1730 1735 1740

Gln Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp

1745 1750 1755

Gln Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser

1760 1765 1770

Thr Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys Arg

1775 1780 1785

Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val

1790 1795 1800

<210> 163

<211> 2117

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 163

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn
20 25 30

Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys
35 40 45

Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn
50 55 60

Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr
65 70 75 80

Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg
85 90 95

Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp
100 105 110

Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp
115 120 125

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val
130 135 140

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu
145 150 155 160

Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu
165 170 175

Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu
180 185 190

Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln

195 200 205

Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val
210 215 220
Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu
225 230 235 240
Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe
245 250 255
Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser
260 265 270

Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr
275 280 285
Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr
290 295 300
Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu
305 310 315 320
Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser
325 330 335

Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu
340 345 350
Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile
355 360 365
Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly
370 375 380
Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys
385 390 395 400

Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu
405 410 415
Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile
420 425 430
His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr
435 440 445

Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe
450 455 460

Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe
465 470 475 480
Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe
485 490 495
Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg
500 505 510
Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys
515 520 525

His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys
530 535 540
Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly
545 550 555 560
Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys
565 570 575
Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys
580 585 590

Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser
595 600 605
Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe
610 615 620
Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr
625 630 635 640
Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr
645 650 655

Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg
660 665 670
Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile
675 680 685
Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp

690 695 700

Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu

705 710 715 720

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp

725 730 735

Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys

740 745 750

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val

755 760 765

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu

770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys

785 790 795 800

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu

805 810 815

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr

820 825 830

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe

850 855 860

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys

865 870 875 880

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys

885 890 895

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln

900 905 910

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu

915 920 925

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln

930 935 940

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
945 950 955 960
Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
965 970 975

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
980 985 990
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
995 1000 1005
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1010 1015 1020
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1025 1030 1035

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1040 1045 1050
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1055 1060 1065
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1070 1075 1080
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1085 1090 1095

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1100 1105 1110
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1115 1120 1125
Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1130 1135 1140
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala
1145 1150 1155

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1160 1165 1170
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1175 1180 1185

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1190 1195 1200

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1205 1210 1215

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1220 1225 1230

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1235 1240 1245

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1250 1255 1260

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1265 1270 1275

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1280 1285 1290

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1295 1300 1305

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1310 1315 1320

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1325 1330 1335

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln

1340 1345 1350

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1355 1360 1365

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1370 1375 1380

Gly Gly Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu

1385 1390 1395

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly

1400 1405 1410

Ser Ser Gly Gly Ser Ser Thr Leu Asn Ile Glu Asp Glu Tyr Arg
1415 1420 1425
Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr
1430 1435 1440
Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly Met
1445 1450 1455

Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala
1460 1465 1470
Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu
1475 1480 1485
Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln
1490 1495 1500
Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu
1505 1510 1515

Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp
1520 1525 1530
Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val
1535 1540 1545
Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
1550 1555 1560
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg
1565 1570 1575

Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp
1580 1585 1590
Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro
1595 1600 1605
Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His
1610 1615 1620
Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu
1625 1630 1635

Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu

1640 1645 1650

Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn

1655 1660 1665

Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys

1670 1675 1680

Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp

1685 1690 1695

Leu Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro

1700 1705 1710

Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly Lys Ala Gly Phe

1715 1720 1725

Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu

1730 1735 1740

Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp Gly Pro Asp

1745 1750 1755

Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala

1760 1765 1770

Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu Phe

1775 1780 1785

Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

1790 1795 1800

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu

1805 1810 1815

Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala

1820 1825 1830

Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly

1835 1840 1845

Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val

1850 1855 1860

Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His

1865 1870 1875

Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro
1880 1885 1890
Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu
1895 1900 1905
Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly
1910 1915 1920
Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp His
1925 1930 1935

Thr Trp Tyr Thr Asp Gly Ser Ser Leu Leu Gln Glu Gly Gln Arg
1940 1945 1950
Lys Ala Gly Ala Ala Val Thr Thr Glu Thr Glu Val Ile Trp Ala
1955 1960 1965
Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg Ala Glu Leu Ile
1970 1975 1980
Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys Leu Asn
1985 1990 1995

Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His Ile His
2000 2005 2010
Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly Lys
2015 2020 2025
Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
2030 2035 2040
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln
2045 2050 2055

Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln
2060 2065 2070
Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr
2075 2080 2085
Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys Arg Thr
2090 2095 2100
Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val

2105 2110 2115

<210> 164

<211> 2117

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<400> 164

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn
20 25 30

Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys
35 40 45

Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn
50 55 60

Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr
65 70 75 80

Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg
85 90 95

Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp
100 105 110

Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp
115 120 125

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val
130 135 140

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu
145 150 155 160

Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu
165 170 175

Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu
180 185 190

Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln
195 200 205

Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val
210 215 220

Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu
225 230 235 240

Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe
245 250 255

Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser
260 265 270

Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr
275 280 285

Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr
290 295 300

Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu
305 310 315 320

Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser
325 330 335

Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu
340 345 350

Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile
355 360 365

Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly
370 375 380

Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys
385 390 395 400

Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu
405 410 415

Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile
420 425 430

His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr

435 440 445

Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe

450 455 460

Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe

465 470 475 480

Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe

485 490 495

Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg

500 505 510

Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys

515 520 525

His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys

530 535 540

Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly

545 550 555 560

Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys

565 570 575

Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys

580 585 590

Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser

595 600 605

Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe

610 615 620

Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr

625 630 635 640

Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr

645 650 655

Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg

660 665 670

Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile

675 680 685

Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp
690 695 700

Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu
705 710 715 720

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp
725 730 735

Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys
740 745 750

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val
755 760 765

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu
770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
785 790 795 800

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu
805 810 815

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr
820 825 830

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe
850 855 860

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys
865 870 875 880

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys
885 890 895

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln
900 905 910

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
915 920 925

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln

930 935 940

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
945 950 955 960
Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
965 970 975
Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
980 985 990
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
995 1000 1005

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1010 1015 1020
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1025 1030 1035
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1040 1045 1050
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1055 1060 1065

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1070 1075 1080
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1085 1090 1095
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1100 1105 1110
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1115 1120 1125

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1130 1135 1140
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala
1145 1150 1155
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1160 1165 1170

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1175 1180 1185

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1190 1195 1200
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1205 1210 1215
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1220 1225 1230
Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1235 1240 1245

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1250 1255 1260
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1265 1270 1275
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1280 1285 1290
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1295 1300 1305

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1310 1315 1320
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1325 1330 1335
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Glu
1340 1345 1350
Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1355 1360 1365

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1370 1375 1380
Gly Gly Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu
1385 1390 1395
Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly

1400 1405 1410

Ser Ser Gly Gly Ser Ser Thr Leu Asn Ile Glu Asp Glu Tyr Arg

1415 1420 1425

Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr

1430 1435 1440

Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly Met

1445 1450 1455

Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala

1460 1465 1470

Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu

1475 1480 1485

Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln

1490 1495 1500

Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu

1505 1510 1515

Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp

1520 1525 1530

Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val

1535 1540 1545

Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

1550 1555 1560

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg

1565 1570 1575

Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp

1580 1585 1590

Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro

1595 1600 1605

Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His

1610 1615 1620

Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu

1625 1630 1635

Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu
1640 1645 1650
Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn
1655 1660 1665

Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys
1670 1675 1680
Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp
1685 1690 1695
Leu Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro
1700 1705 1710
Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly Lys Ala Gly Phe
1715 1720 1725

Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu
1730 1735 1740
Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp Gly Pro Asp
1745 1750 1755
Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala
1760 1765 1770
Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu Phe
1775 1780 1785

Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
1790 1795 1800
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu
1805 1810 1815
Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala
1820 1825 1830
Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly
1835 1840 1845

Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val
1850 1855 1860
Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His

1865 1870 1875

Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro

1880 1885 1890

Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu

1895 1900 1905

Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala His Gly

1910 1915 1920

Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp His

1925 1930 1935

Thr Trp Tyr Thr Asp Gly Ser Ser Leu Leu Gln Glu Gly Gln Arg

1940 1945 1950

Lys Ala Gly Ala Ala Val Thr Thr Glu Thr Glu Val Ile Trp Ala

1955 1960 1965

Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg Ala Glu Leu Ile

1970 1975 1980

Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys Leu Asn

1985 1990 1995

Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His Ile His

2000 2005 2010

Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly Lys

2015 2020 2025

Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

2030 2035 2040

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln

2045 2050 2055

Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln

2060 2065 2070

Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr

2075 2080 2085

Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys Arg Thr

2090 2095 2100

Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val
2105 2110 2115

<210> 165

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (6)..(150)

<223> may be absent

<400> 165

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
50 55 60

Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
65 70 75 80

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
85 90 95

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
100 105 110

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
115 120 125

Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
130 135 140

Ser Gly Gly Gly Gly Ser
145 150

<210> 166

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (2)..(30)

<223> may be absent

<400> 166

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
1 5 10 15

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
20 25 30

<210> 167

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (6)..(150)

<223> may be absent

<400> 167

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu
1 5 10 15

Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala
20 25 30

Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala
35 40 45

Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala
50 55 60

Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys
65 70 75 80

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu
85 90 95

Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala

100 105 110

Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala

115 120 125

Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala

130 135 140

Lys Glu Ala Ala Ala Lys

145 150

<210> 168

<211> 90

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (4)..(90)

<223> may be absent

<400> 168

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

1 5 10 15

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

20 25 30

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

35 40 45

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

50 55 60

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

65 70 75 80

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

85 90

<210> 169

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (5)..(120)

<223> may be absent

<400> 169

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
1 5 10 15

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
35 40 45

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
50 55 60

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
65 70 75 80

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
85 90 95

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
100 105 110

Ser Gly Gly Ser Ser Gly Gly Ser
115 120

<210> 170

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(1)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (3)..(3)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature
<222> (3)..(60)
<223> may be absent
<220><221> misc_feature
<222> (5)..(5)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (7)..(7)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (9)..(9)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (11)..(11)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (13)..(13)

<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (15)..(15)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (17)..(17)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (19)..(19)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (21)..(21)
<223> Xaa can be any naturally occurring amino acid
<220><221> misc_feature
<222> (23)..(23)

<223

> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (25)..(25)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (27)..(27)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (29)..(29)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (31)..(31)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (33)..(33)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (35)..(35)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (37)..(37)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (39)..(39)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (41)..(41)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (43)..(43)

<223> Xaa can be any naturally occurring amino acid

<220

><221> misc_feature

<222> (45)..(45)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (47)..(47)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (49)..(49)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (51)..(51)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (53)..(53)

<223> Xaa can be any naturally occurring amino acid

<220><221

> misc_feature

<222> (55)..(55)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (57)..(57)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (59)..(59)

<223> Xaa can be any naturally occurring amino acid

<400> 170

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

1 5 10 15

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

20 25 30

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

35 40 45

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

50 55 60

<210> 171

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 171

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1 5 10 15

<210> 172

<211> 32

<212> PRT

<213

> Artificial Sequence

<220><223> Synthetic

<400> 172

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser

20 25 30

<210> 173

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 173

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

1 5 10

<210> 174

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 174

Ser Gly Gly Ser

1

<210> 175

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 175

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ala Gly Ser Tyr Pro Tyr Asp Val

20 25 30

Pro Asp Tyr Ala Gly Ser Ala Ala Pro Ala Ala Lys Lys Lys Lys Leu

35 40 45

Asp Gly Ser Gly Ser Gly Gly Ser Ser Gly Gly Ser

50 55 60

<210> 176

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 176

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

1 5 10 15

Gly Ser Gly Gly Ser

20

<210> 177

<400> 177

000

<210> 178

<400> 178

000

<210> 179

<400> 179

000

<210> 180

<400> 180

000

<210> 181

<400> 181

000

<210> 182

<400> 182

000

<210> 183

<400> 183

000

<210> 184

<400> 184

000

<210> 185

<400> 185

000

<210> 186

<400> 186

000

<210> 187

<400> 187

000

<210> 188

<400> 188

000

<210> 189

<400> 189

000

<210> 190

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 190

Ala Val Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys

1 5 10 15

Lys Lys Leu Asp

20

<210> 191

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 191

Met Ser Arg Arg Arg Lys Ala Asn Pro Thr Lys Leu Ser Glu Asn Ala

1 5 10 15

Lys Lys Leu Ala Lys Glu Val Glu Asn

20 25

<210> 192

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400>

192

Pro Ala Ala Lys Arg Val Lys Leu Asp

1 5

<210> 193

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 193

Lys Leu Lys Ile Lys Arg Pro Val Lys

1 5

<210> 194

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 194

Val Ser Arg Lys Arg Pro Arg Pro

1 5

<210> 195

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 195

Glu Gly Ala Pro Pro Ala Lys Arg Ala Arg

1 5 10

<210> 196

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 196

Pro Pro Gln Pro Lys Lys Lys Pro Leu Asp Gly Glu

1 5 10

<210> 197

<400> 197

000

<210> 198

<211> 380

<212> PRT

<213> Homo sapiens

<400> 198

Met Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser

1 5 10 15

Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala

20 25 30

Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln
35 40 45
Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu
50 55 60
Met Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys
65 70 75 80

Pro Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu
85 90 95
Leu Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln
100 105 110
Gln Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys
115 120 125
Arg Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu
130 135 140

Leu Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu
145 150 155 160
Ala Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala Ala
165 170 175
Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg
180 185 190
His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His
195 200 205

Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val
210 215 220
Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly
225 230 235 240
Ile Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile
245 250 255
Glu Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu
260 265 270

Asn Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val

275 280 285

Leu Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu

290 295 300

Glu Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg

305 310 315 320

Ile Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr

325 330 335

Gln Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser

340 345 350

Ala Lys Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala

355 360 365

Lys Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys

370 375 380

<210> 199

<211> 380

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 199

Met Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser

1 5 10 15

Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala

20 25 30

Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln

35 40 45

Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu

50 55 60

Met Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys

65 70 75 80

Pro Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu

85 90 95

Leu Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln

100 105 110

Gln Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys

115 120 125

Arg Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu

130 135 140

Leu Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu

145 150 155 160

Ala Ser Cys Ala Ala Leu Val Arg Ala Gly Lys Val Tyr Ala Ala Ala

165 170 175

Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg

180 185 190

His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His

195 200 205

Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val

210 215 220

Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly

225 230 235 240

Ile Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile

245 250 255

Glu Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu

260 265 270

Asn Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val

275 280 285

Leu Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu

290 295 300

Glu Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg

305 310 315 320

Ile Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr

325 330 335

Gln Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser

340 345 350

Ala Lys Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala
355 360 365

Lys Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys
370 375 380

<210> 200

<211> 380

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 200

Met Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser
1 5 10 15

Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala
20 25 30

Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln
35 40 45

Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu
50 55 60

Met Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys
65 70 75 80

Pro Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu
85 90 95

Leu Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln
100 105 110

Gln Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys
115 120 125

Arg Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu
130 135 140

Leu Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu
145 150 155 160

Ala Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala Ala
165 170 175

Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ala Pro Val Leu Met Arg
180 185 190

His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His
195 200 205
Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val
210 215 220
Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly
225 230 235 240
Ile Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile
245 250 255

Glu Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu
260 265 270
Asn Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val
275 280 285
Leu Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu
290 295 300
Glu Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg
305 310 315 320

Ile Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr
325 330 335
Gln Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser
340 345 350
Ala Lys Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala
355 360 365
Lys Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys
370 375 380

<210> 201

<

211> 380

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 201

Met Gly Ile Gln Gly Leu Ala Lys Leu Ile Ala Asp Val Ala Pro Ser
1 5 10 15
Ala Ile Arg Glu Asn Asp Ile Lys Ser Tyr Phe Gly Arg Lys Val Ala
20 25 30
Ile Asp Ala Ser Met Ser Ile Tyr Gln Phe Leu Ile Ala Val Arg Gln
35 40 45
Gly Gly Asp Val Leu Gln Asn Glu Glu Gly Glu Thr Thr Ser His Leu

50 55 60
Met Gly Met Phe Tyr Arg Thr Ile Arg Met Met Glu Asn Gly Ile Lys
65 70 75 80
Pro Val Tyr Val Phe Asp Gly Lys Pro Pro Gln Leu Lys Ser Gly Glu
85 90 95
Leu Ala Lys Arg Ser Glu Arg Arg Ala Glu Ala Glu Lys Gln Leu Gln
100 105 110
Gln Ala Gln Ala Ala Gly Ala Glu Gln Glu Val Glu Lys Phe Thr Lys

115 120 125
Arg Leu Val Lys Val Thr Lys Gln His Asn Asp Glu Cys Lys His Leu
130 135 140
Leu Ser Leu Met Gly Ile Pro Tyr Leu Asp Ala Pro Ser Glu Ala Glu
145 150 155 160
Ala Ser Cys Ala Ala Leu Val Lys Ala Gly Lys Val Tyr Ala Ala Ala
165 170 175
Thr Glu Asp Met Asp Cys Leu Thr Phe Gly Ser Pro Val Leu Met Arg

180 185 190
His Leu Thr Ala Ser Glu Ala Lys Lys Leu Pro Ile Gln Glu Phe His
195 200 205
Leu Ser Arg Ile Leu Gln Glu Leu Gly Leu Asn Gln Glu Gln Phe Val
210 215 220
Asp Leu Cys Ile Leu Leu Gly Ser Asp Tyr Cys Glu Ser Ile Arg Gly
225 230 235 240

Ile Gly Pro Lys Arg Ala Val Asp Leu Ile Gln Lys His Lys Ser Ile

245 250 255

Glu Glu Ile Val Arg Arg Leu Asp Pro Asn Lys Tyr Pro Val Pro Glu

260 265 270

Asn Trp Leu His Lys Glu Ala His Gln Leu Phe Leu Glu Pro Glu Val

275 280 285

Leu Asp Pro Glu Ser Val Glu Leu Lys Trp Ser Glu Pro Asn Glu Glu

290 295 300

Glu Leu Ile Lys Phe Met Cys Gly Glu Lys Gln Phe Ser Glu Glu Arg

305 310 315 320

Ile Arg Ser Gly Val Lys Arg Leu Ser Lys Ser Arg Gln Gly Ser Thr

325 330 335

Gln Gly Arg Leu Asp Asp Phe Phe Lys Val Thr Gly Ser Leu Ser Ser

340 345 350

Ala Arg Arg Lys Glu Pro Glu Pro Lys Gly Ser Thr Lys Lys Lys Ala

355 360 365

Lys Thr Gly Ala Ala Gly Lys Phe Lys Arg Gly Lys

370 375 380

<210> 202

<211> 908

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 202

Met Gly Val Asn Asp Leu Trp Gln Ile Leu Glu Pro Val Lys Gln His

1 5 10 15

Ile Pro Leu Arg Asn Leu Gly Gly Lys Thr Ile Ala Val Asp Leu Ser

20 25 30

Leu Trp Val Cys Glu Ala Gln Thr Val Lys Lys Met Met Gly Ser Val

35 40 45

Met Lys Pro His Leu Arg Asn Leu Phe Phe Arg Ile Ser Tyr Leu Thr

50 55 60

Gln Met Asp Val Lys Leu Val Phe Val Met Glu Gly Glu Pro Pro Lys
65 70 75 80

Leu Lys Ala Asp Val Ile Ser Lys Arg Asn Gln Ser Arg Tyr Gly Ser
85 90 95

Ser Gly Lys Ser Trp Ser Gln Lys Thr Gly Arg Ser His Phe Lys Ser
100 105 110

Val Leu Arg Glu Cys Leu His Met Leu Glu Cys Leu Gly Ile Pro Trp
115 120 125

Val Gln Ala Ala Gly Glu Ala Glu Ala Met Cys Ala Tyr Leu Asn Ala
130 135 140

Gly Gly His Val Asp Gly Cys Leu Thr Asn Asp Gly Asp Thr Phe Leu
145 150 155 160

Tyr Gly Ala Gln Thr Val Tyr Arg Asn Phe Thr Met Asn Thr Lys Asp
165 170 175

Pro His Val Asp Cys Tyr Thr Met Ser Ser Ile Lys Ser Lys Leu Gly
180 185 190

Leu Asp Arg Asp Ala Leu Val Gly Leu Ala Ile Leu Leu Gly Cys Asp
195 200 205

Tyr Leu Pro Lys Gly Val Pro Gly Val Gly Lys Glu Gln Ala Leu Lys
210 215 220

Leu Ile Gln Ile Leu Lys Gly Gln Ser Leu Leu Gln Arg Phe Asn Arg
225 230 235 240

Trp Asn Glu Thr Ser Cys Asn Ser Ser Pro Gln Leu Leu Val Thr Lys
245 250 255

Lys Leu Ala His Cys Ser Val Cys Ser His Pro Gly Ser Pro Lys Asp
260 265 270

His Glu Arg Asn Gly Cys Arg Leu Cys Lys Ser Asp Lys Tyr Cys Glu
275 280 285

Pro His Asp Tyr Glu Tyr Cys Cys Pro Cys Glu Trp His Arg Thr Glu
290 295 300

His Asp Arg Gln Leu Ser Glu Val Glu Asn Asn Ile Lys Lys Lys Ala
305 310 315 320
Cys Cys Cys Glu Gly Phe Pro Phe His Glu Val Ile Gln Glu Phe Leu
325 330 335
Leu Asn Lys Asp Lys Leu Val Lys Val Ile Arg Tyr Gln Arg Pro Asp
340 345 350
Leu Leu Leu Phe Gln Arg Phe Thr Leu Glu Lys Met Glu Trp Pro Asn
355 360 365

His Tyr Ala Cys Glu Lys Leu Leu Val Leu Leu Thr His Tyr Asp Met
370 375 380
Ile Glu Arg Lys Leu Gly Ser Arg Asn Ser Asn Gln Leu Gln Pro Ile
385 390 395 400
Arg Ile Val Lys Thr Arg Ile Arg Asn Gly Val His Cys Phe Glu Ile
405 410 415
Glu Trp Glu Lys Pro Glu His Tyr Ala Met Glu Asp Lys Gln His Gly
420 425 430

Glu Phe Ala Leu Leu Thr Ile Glu Glu Glu Ser Leu Phe Glu Ala Ala
435 440 445
Tyr Pro Glu Ile Val Ala Val Tyr Gln Lys Gln Lys Leu Glu Ile Lys
450 455 460
Gly Lys Lys Gln Lys Arg Ile Lys Pro Lys Glu Asn Asn Leu Pro Glu
465 470 475 480
Pro Asp Glu Val Met Ser Phe Gln Ser His Met Thr Leu Lys Pro Thr
485 490 495

Cys Glu Ile Phe His Lys Gln Asn Ser Lys Leu Asn Ser Gly Ile Ser
500 505 510
Pro Asp Pro Thr Leu Pro Gln Glu Ser Ile Ser Ala Ser Leu Asn Ser
515 520 525
Leu Leu Leu Pro Lys Asn Thr Pro Cys Leu Asn Ala Gln Glu Gln Phe
530 535 540
Met Ser Ser Leu Arg Pro Leu Ala Ile Gln Gln Ile Lys Ala Val Ser

545 550 555 560

Lys Ser Leu Ile Ser Glu Ser Ser Gln Pro Asn Thr Ser Ser His Asn
565 570 575
Ile Ser Val Ile Ala Asp Leu His Leu Ser Thr Ile Asp Trp Glu Gly
580 585 590
Thr Ser Phe Ser Asn Ser Pro Ala Ile Gln Arg Asn Thr Phe Ser His
595 600 605
Asp Leu Lys Ser Glu Val Glu Ser Glu Leu Ser Ala Ile Pro Asp Gly
610 615 620

Phe Glu Asn Ile Pro Glu Gln Leu Ser Cys Glu Ser Glu Arg Tyr Thr
625 630 635 640
Ala Asn Ile Lys Lys Val Leu Asp Glu Asp Ser Asp Gly Ile Ser Pro
645 650 655
Glu Glu His Leu Leu Ser Gly Ile Thr Asp Leu Cys Leu Gln Asp Leu
660 665 670
Pro Leu Lys Glu Arg Ile Phe Thr Lys Leu Ser Tyr Pro Gln Asp Asn
675 680 685

Leu Gln Pro Asp Val Asn Leu Lys Thr Leu Ser Ile Leu Ser Val Lys
690 695 700
Glu Ser Cys Ile Ala Asn Ser Gly Ser Asp Cys Thr Ser His Leu Ser
705 710 715 720
Lys Asp Leu Pro Gly Ile Pro Leu Gln Asn Glu Ser Arg Asp Ser Lys
725 730 735
Ile Leu Lys Gly Asp Gln Leu Leu Gln Glu Asp Tyr Lys Val Asn Thr
740 745 750

Ser Val Pro Tyr Ser Val Ser Asn Thr Val Val Lys Thr Cys Asn Val
755 760 765
Arg Pro Pro Asn Thr Ala Leu Asp His Ser Arg Lys Val Asp Met Gln
770 775 780
Thr Thr Arg Lys Ile Leu Met Lys Lys Ser Val Cys Leu Asp Arg His
785 790 795 800

Ser Ser Asp Glu Gln Ser Ala Pro Val Phe Gly Lys Ala Lys Tyr Thr
805 810 815

Thr Gln Arg Met Lys His Ser Ser Gln Lys His Asn Ser Ser His Phe
820 825 830

Lys Glu Ser Gly His Asn Lys Leu Ser Ser Pro Lys Ile His Ile Lys
835 840 845

Glu Thr Glu Gln Cys Val Arg Ser Tyr Glu Thr Ala Glu Asn Glu Glu
850 855 860

Ser Cys Phe Pro Asp Ser Thr Lys Ser Ser Leu Ser Ser Leu Gln Cys
865 870 875 880

His Lys Lys Glu Asn Asn Ser Gly Thr Cys Leu Asp Ser Pro Leu Pro
885 890 895

Leu Arg Gln Arg Leu Lys Leu Arg Phe Gln Ser Thr
900 905

<210> 203

<211> 1186

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 203

Met Gly Val Gln Gly Leu Trp Lys Leu Leu Glu Cys Ser Gly Arg Gln
1 5 10 15

Val Ser Pro Glu Ala Leu Glu Gly Lys Ile Leu Ala Val Asp Ile Ser
20 25 30

Ile Trp Leu Asn Gln Ala Leu Lys Gly Val Arg Asp Arg His Gly Asn
35 40 45

Ser Ile Glu Asn Pro His Leu Leu Thr Leu Phe His Arg Leu Cys Lys
50 55 60

Leu Leu Phe Phe Arg Ile Arg Pro Ile Phe Val Phe Asp Gly Asp Ala
65 70 75 80

Pro Leu Leu Lys Lys Gln Thr Leu Val Lys Arg Arg Gln Arg Lys Asp

85 90 95

Leu Ala Ser Ser Asp Ser Arg Lys Thr Thr Glu Lys Leu Leu Lys Thr

100 105 110

Phe Leu Lys Arg Gln Ala Ile Lys Thr Ala Phe Arg Ser Lys Arg Asp

115 120 125

Glu Ala Leu Pro Ser Leu Thr Gln Val Arg Arg Glu Asn Asp Leu Tyr

130 135 140

Val Leu Pro Pro Leu Gln Glu Glu Glu Lys His Ser Ser Glu Glu Glu

145 150 155 160

Asp Glu Lys Glu Trp Gln Glu Arg Met Asn Gln Lys Gln Ala Leu Gln

165 170 175

Glu Glu Phe Phe His Asn Pro Gln Ala Ile Asp Ile Glu Ser Glu Asp

180 185 190

Phe Ser Ser Leu Pro Pro Glu Val Lys His Glu Ile Leu Thr Asp Met

195 200 205

Lys Glu Phe Thr Lys Arg Arg Arg Thr Leu Phe Glu Ala Met Pro Glu

210 215 220

Glu Ser Asp Asp Phe Ser Gln Tyr Gln Leu Lys Gly Leu Leu Lys Lys

225 230 235 240

Asn Tyr Leu Asn Gln His Ile Glu His Val Gln Lys Glu Met Asn Gln

245 250 255

Gln His Ser Gly His Ile Arg Arg Gln Tyr Glu Asp Glu Gly Gly Phe

260 265 270

Leu Lys Glu Val Glu Ser Arg Arg Val Val Ser Glu Asp Thr Ser His

275 280 285

Tyr Ile Leu Ile Lys Gly Ile Gln Ala Lys Thr Val Ala Glu Val Asp

290 295 300

Ser Glu Ser Leu Pro Ser Ser Ser Lys Met His Gly Met Ser Phe Asp

305 310 315 320

Val Lys Ser Ser Pro Cys Glu Lys Leu Lys Thr Glu Lys Glu Pro Asp

325 330 335

Ala Thr Pro Pro Ser Pro Arg Thr Leu Leu Ala Met Gln Ala Ala Leu

340 345 350

Leu Gly Ser Ser Ser Glu Glu Glu Leu Glu Ser Glu Asn Arg Arg Gln

355 360 365

Ala Arg Gly Arg Asn Ala Pro Ala Ala Val Asp Glu Gly Ser Ile Ser

370 375 380

Pro Arg Thr Leu Ser Ala Ile Lys Arg Ala Leu Asp Asp Asp Glu Asp

385 390 395 400

Val Lys Val Cys Ala Gly Asp Asp Val Gln Thr Gly Gly Pro Gly Ala

405 410 415

Glu Glu Met Arg Ile Asn Ser Ser Thr Glu Asn Ser Asp Glu Gly Leu

420 425 430

Lys Val Arg Asp Gly Lys Gly Ile Pro Phe Thr Ala Thr Leu Ala Ser

435 440 445

Ser Ser Val Asn Ser Ala Glu Glu His Val Ala Ser Thr Asn Glu Gly

450 455 460

Arg Glu Pro Thr Asp Ser Val Pro Lys Glu Gln Met Ser Leu Val His

465 470 475 480

Val Gly Thr Glu Ala Phe Pro Ile Ser Asp Glu Ser Met Ile Lys Asp

485 490 495

Arg Lys Asp Arg Leu Pro Leu Glu Ser Ala Val Val Arg His Ser Asp

500 505 510

Ala Pro Gly Leu Pro Asn Gly Arg Glu Leu Thr Pro Ala Ser Pro Thr

515 520 525

Cys Thr Asn Ser Val Ser Lys Asn Glu Thr His Ala Glu Val Leu Glu

530 535 540

Gln Gln Asn Glu Leu Cys Pro Tyr Glu Ser Lys Phe Asp Ser Ser Leu

545 550 555 560

Leu Ser Ser Asp Asp Glu Thr Lys Cys Lys Pro Asn Ser Ala Ser Glu

565 570 575

Val Ile Gly Pro Val Ser Leu Gln Glu Thr Ser Ser Ile Val Ser Val

580 585 590

Pro Ser Glu Ala Val Asp Asn Val Glu Asn Val Val Ser Phe Asn Ala

595 600 605

Lys Glu His Glu Asn Phe Leu Glu Thr Ile Gln Glu Gln Gln Thr Thr

610 615 620

Glu Ser Ala Gly Gln Asp Leu Ile Ser Ile Pro Lys Ala Val Glu Pro

625 630 635 640

Met Glu Ile Asp Ser Glu Glu Ser Glu Ser Asp Gly Ser Phe Ile Glu

645 650 655

Val Gln Ser Val Ile Ser Asp Glu Glu Leu Gln Ala Glu Phe Pro Glu

660 665 670

Thr Ser Lys Pro Pro Ser Glu Gln Gly Glu Glu Glu Leu Val Gly Thr

675 680 685

Arg Glu Gly Glu Ala Pro Ala Glu Ser Glu Ser Leu Leu Arg Asp Asn

690 695 700

Ser Glu Arg Asp Asp Val Asp Gly Glu Pro Gln Glu Ala Glu Lys Asp

705 710 715 720

Ala Glu Asp Ser Leu His Glu Trp Gln Asp Ile Asn Leu Glu Glu Leu

725 730 735

Glu Thr Leu Glu Ser Asn Leu Leu Ala Gln Gln Asn Ser Leu Lys Ala

740 745 750

Gln Lys Gln Gln Gln Glu Arg Ile Ala Ala Thr Val Thr Gly Gln Met

755 760 765

Phe Leu Glu Ser Gln Glu Leu Leu Arg Leu Phe Gly Ile Pro Tyr Ile

770 775 780

Gln Ala Pro Met Glu Ala Glu Ala Gln Cys Ala Ile Leu Asp Leu Thr

785 790 795 800

Asp Gln Thr Ser Gly Thr Ile Thr Asp Asp Ser Asp Ile Trp Leu Phe

805 810 815

Gly Ala Arg His Val Tyr Arg Asn Phe Phe Asn Lys Asn Lys Phe Val

820 825 830

Glu Tyr Tyr Gln Tyr Val Asp Phe His Asn Gln Leu Gly Leu Asp Arg
835 840 845
Asn Lys Leu Ile Asn Leu Ala Tyr Leu Leu Gly Ser Asp Tyr Thr Glu

850 855 860
Gly Ile Pro Thr Val Gly Cys Val Thr Ala Met Glu Ile Leu Asn Glu
865 870 875 880
Phe Pro Gly His Gly Leu Glu Pro Leu Leu Lys Phe Ser Glu Trp Trp
885 890 895
His Glu Ala Gln Lys Asn Pro Lys Ile Arg Pro Asn Pro His Asp Thr
900 905 910
Lys Val Lys Lys Lys Leu Arg Thr Leu Gln Leu Thr Pro Gly Phe Pro

915 920 925
Asn Pro Ala Val Ala Glu Ala Tyr Leu Lys Pro Val Val Asp Asp Ser
930 935 940
Lys Gly Ser Phe Leu Trp Gly Lys Pro Asp Leu Asp Lys Ile Arg Glu
945 950 955 960
Phe Cys Gln Arg Tyr Phe Gly Trp Asn Arg Thr Lys Thr Asp Glu Ser
965 970 975
Leu Phe Pro Val Leu Lys Gln Leu Asp Ala Gln Gln Thr Gln Leu Arg

980 985 990
Ile Asp Ser Phe Phe Arg Leu Ala Gln Gln Glu Lys Glu Asp Ala Lys
995 1000 1005
Arg Ile Lys Ser Gln Arg Leu Asn Arg Ala Val Thr Cys Met Leu
1010 1015 1020
Arg Lys Glu Lys Glu Ala Ala Ala Ser Glu Ile Glu Ala Val Ser
1025 1030 1035
Val Ala Met Glu Lys Glu Phe Glu Leu Leu Asp Lys Ala Lys Arg

1040 1045 1050
Lys Thr Gln Lys Arg Gly Ile Thr Asn Thr Leu Glu Glu Ser Ser
1055 1060 1065
Ser Leu Lys Arg Lys Arg Leu Ser Asp Ser Lys Arg Lys Asn Thr

1070 1075 1080

Cys Gly Gly Phe Leu Gly Glu Thr Cys Leu Ser Glu Ser Ser Asp

1085 1090 1095

Gly Ser Ser Ser Glu Asp Ala Glu Ser Ser Ser Leu Met Asn Val

1100 1105 1110

Gln Arg Arg Thr Ala Ala Lys Glu Pro Lys Thr Ser Ala Ser Asp

1115 1120 1125

Ser Gln Asn Ser Val Lys Glu Ala Pro Val Lys Asn Gly Gly Ala

1130 1135 1140

Thr Thr Ser Ser Ser Ser Asp Ser Asp Asp Asp Gly Gly Lys Glu

1145 1150 1155

Lys Met Val Leu Val Thr Ala Arg Ser Val Phe Gly Lys Lys Arg

1160 1165 1170

Arg Lys Leu Arg Arg Ala Arg Gly Arg Lys Arg Lys Thr

1175 1180 1185

<210> 204

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(8)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (9)..(20)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (21)..(21)

<223> n is a, c, g, or t

<400> 204

nnnnnnnnnn nnnnnnnnnn ngg 23

<210> 205

<211> 15

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(12)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (13)..(13)

<223> n is a, c, g, or t

<400> 205

nnnnnnnnnn nnngg 15

<210> 206

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(9)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (10)..(20)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (21)..(22)

<223> n is a, c, g, or t

<400> 206

nnnnnnnnnn nnnnnnnnnn ngg 23

<210> 207

<211> 14

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(11)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (12)..(12)

<223> n is a, c, g, or t

<400> 207

nnnnnnnnnn nngg 14

<210> 208

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(9)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (10)..(20)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (21)..(22)

<223> n is a, c, g, or t

<400> 208

nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> 209

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(12)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (13)..(14)

<223> n is a, c, g, or t

<400> 209

nnnnnnnnnn nnnnagaaw 19

<210> 210

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(9)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (10)..(20)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (21)..(22)

<223> n is a, c, g, or t

<400> 210

nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> 211

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(11)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (12)..(13)

<223> n is a, c, g, or t

<400> 211

nnnnnnnnnn nnnagaaw 18

<210> 212

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220

><221> misc_feature

<222> (1)..(8)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (9)..(20)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (21)..(21)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (24)..(24)

<223> n is a, c, g, or t

<400> 212

nnnnnnnnnn nnnnnnnnnn nggng 25

<210> 213

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(12)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (13)..(14)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (16)..(16)

<223> n is a, c, g, or t

<400> 213

nnnnnnnnnn nnnggng 17

<210> 214

<400> 214

000

<210> 215

<400> 215

000

<210> 216

<211> 125

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(8)

<223> n is a, c, g, or t

<400> 216

nnnnnnnngt ttttgtactc tcaagattta gaaataaatc ttgcagaagc tacaaagata 60

aggcttcatg ccgaaatcaa caccctgtca ttttatggca gggtgttttc gttatttaat 120

ttttt 125

<210> 217

<211> 121

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(18)

<223> n is a, c, g, or t

<400> 217

nnnnnnnnnn nnnnnnnngt ttttgtactc tcagaaatgc agaagctaca aagataaggc 60

ttcatgccga aatcaacacc ctgtcatttt atggcagggt gttttcgtta tttaattttt 120

t 121

<210> 218

<211> 109

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(20)

<223> n is a, c, g, or t

<400> 218

nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttttt 109

<210> 219

<211> 102

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(20)

<223> n is a, c, g, or t

<400> 219

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tt 102

<210> 220

<211> 88

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(20)

<223> n is a, c, g, or t

<400> 220

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gttttttt 88

<210> 221

<211> 76

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(20)

<223> n is a, c, g, or t

<400> 221

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcatt tttttt 76

<210> 222

<211> 82

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 222

guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu ccguuaucaa cuugaaaaag 60

uggcaccgag ucggugcuuu uu 82

<210> 223

<211> 905

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic

<400> 223

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcg ttcactgccg 600

tataggcagg gcccagactg agcacgtgag ttttagagct agaaatagca agttaaaata 660

aggctagtcc gttatcaact tgaaaaagtg ggaccgagtc ggtcctctgc catcaaagcg 720

tgctcagtct gttttagggt catgaaggtt tttcttttcc tgagaaaaca acacgtattg 780

ttttctcagg ttttgctttt tggccttttt ctagcttaaa aaaaaaaaaa gcaaaagatg 840

ctggtggttg gcactcctgg tttccaggac ggggttcaaa tccctgcggc gtctttgctt 900

tgact 905

<210> 224

<211> 942

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 224

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcg ttcactgccg 600

tataggcagg gcccagactg agcacgtgag ttttagagct agaaatagca agttaaaata 660

aggctagtcc gttatcaact tgaaaaagtg ggaccgagtc ggtcctctgc catcaaagcg 720

tgctcagtct gttttgtttt ggctgggttt ttccttgttc gcaccggaca cctccagtga 780

ccagacggca aggtttttat cccagtgtat attggaaaaa catgttatac ttttgacaat 840

ttaacgtgcc tagagctcaa attaaactaa taccataacg taatgcaact tacaacataa 900

ataaaggtca atgtttaatc cataaaaaaa aaaaaaaaaa aa 942

<210> 225

<211> 1370

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 225

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcg ttcactgccg 600

tataggcagg gcccagactg agcacgtgag ttttagagct agaaatagca agttaaaata 660

aggctagtcc gttatcaact tgaaaaagtg ggaccgagtc ggtcctctgc catcaaagcg 720

tgctcagtct gttttgtttt ggctgggttt ttccttgttc gcaccggaca cctccagtga 780

ccagacggca aggtttttat cccagtgtat attggaaaaa catgttatac ttttgacaat 840

ttaacgtgcc tagagctcaa attaaactaa taccataacg taatgcaact tacaacataa 900

ataaaggtca atgtttaatc cataaaaaaa aaaaaaaaaa aaacacactg ttttggctgg 960

gtttttcctt gttcgcaccg gacacctcca gtgaccagac ggcaaggttt ttatcccagt 1020

gtatattgga aaaacatgtt atacttttga caatttaacg tgcctagagc tcaaattaaa 1080

ctaataccat aacgtaatgc aacttacaac ataaataaag gtcaatgttt aatccataaa 1140

aaaaaaaaaa aaaaaatctc tctgttttgg ctgggttttt ccttgttcgc accggacacc 1200

tccagtgacc agacggcaag gtttttatcc cagtgtatat tggaaaaaca tgttatactt 1260

ttgacaattt aacgtgccta gagctcaaat taaactaata ccataacgta atgcaactta 1320

caacataaat aaaggtcaat gtttaatcca taaaaaaaaa aaaaaaaaaa 1370

<210> 226

<211> 807

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 226

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcg ttcactgccg 600

tataggcagg gcccagactg agcacgtgag ttttagagct agaaatagca agttaaaata 660

aggctagtcc gttatcaact tgaaaaagtg ggaccgagtc ggtcctctgc catcaaagcg 720

tgctcagtct gtttgtttca aaagtagact gtacgctaag ggtcatatct ttttttgttt 780

ggtttgtgtc ttggttggcg tcttaaa 807

<210> 227

<211> 743

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 227

ctaaggacca gcttctttgg gagagaacag acgcaggggc gggagggaaa aagggagagg 60

cagacgtcac ttccccttgg cggctctggc agcagattgg tcggttgagt ggcagaaagg 120

cagacgggga ctgggcaagg cactgtcggt gacatcacgg acagggcgac ttctatgtag 180

atgaggcagc gcagaggctg ctgcttcgcc acttgctgct tcaccacgaa ggagttcccg 240

tgccctggga gcgggttcag gaccgctgat cggaagtgag aatcccagct gtgtgtcagg 300

gctggaaagg gctcgggagt gcgcggggca agtgaccgtg tgtgtaaaga gtgaggcgta 360

tgaggctgtg tcggggcaga ggcccaagat ctcagttcac tgccgtatag gcagggccca 420

gactgagcac gtgagtttta gagctagaaa tagcaagtta aaataaggct agtccgttat 480

caacttgaaa aagtgggacc gagtcggtcc tctgccatca aagcgtgctc agtctgtttc 540

agcaagttca gagaaatctg aacttgctgg atttttggag cagggagatg gaataggagc 600

ttgctccgtc cactccacgc atcgacctgg tattgcagta cctccaggaa cggtgcaccc 660

actttctgga gtttcaaaag tagactgtac gctaagggtc atatcttttt ttgtttggtt 720

tgtgtcttgg ttggcgtctt aaa 743

<210> 228

<211> 141

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 228

ggcccagact gagcacgtga gttttagagc tagctcatga aaatgagcta gcaagttaaa 60

ataaggctag tccgttatca acttgaaaaa gtgggaccga gtcggtcctc tgccatcaaa 120

gcgtgctcag tctgtttttt t 141

<210> 229

<211> 129

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 229

ggcccagact gagcacgtga gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtcctctg ccatcaaagc gtgctcagtc 120

tgttttttt 129

<210> 230

<211> 197

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 230

ggcccagact gagcacgtga gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtcctctg ccatcaaagc gtgctcagtc 120

tgggccggca tggtcccagc ctcctcgctg gcgccggctg ggcaacatgc ttcggcatgg 180

cgaatgggac ttttttt 197

<210> 231

<211> 171

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 231

ggtgggagac gtcccaccgg cccagactga gcacgtgagt tttagagcta gaaatagcaa 60

gttaaaataa ggctagtccg ttatcaactt gaaaaagtgg gaccgagtcg gtcctctgcc 120

atcaaagctt cgaccgtgct cagtctggtg ggagacgtcc cacctttttt t 171

<210> 232

<211> 180

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 232

gagcagcatg gcgtcgctgc tcacggccca gactgagcac gtgagtttta gagctagaaa 60

tagcaagtta aaataaggct agtccgttat caacttgaaa aagtgggacc gagtcggtcc 120

tctgccatca aagcttcgac cgtgctcagt ctccatcagt tgacaccctg aggttttttt 180

<210> 233

<211> 188

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 233

gcagacctaa gtggugacat atggtctggg cccagactga gcacgtgagt tttagagcta 60

uacgtagcaa gttaaaataa ggctagtccg ttatcaactt uacgaagtgg gaccgagtcg 120

gtcctctgcc atcaaagctt cgaccgtgct cagtctgcat gcgattagaa ataatcgcat 180

gttttttt 188

<210> 234

<211> 125

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 234

tctgccatca aagctgcgac cgtgctcagt ctggtgggag acgtcccacc ggccggcatg 60

gtcccagcct cctcgctggc gccggctggg caacatgctt cggcatggcg aatgggactt 120

ttttt 125

<210> 235

<211> 328

<212> PRT

<213> Unknown

<220><223> Bordetella phage

<400> 235

Met Gly Lys Arg His Arg Asn Leu Ile Asp Gln Ile Thr Thr Trp Glu
1 5 10 15

Asn Leu Leu Asp Ala Tyr Arg Lys Thr Ser His Gly Lys Arg Arg Thr
20 25 30

Trp Gly Tyr Leu Glu Phe Lys Glu Tyr Asp Leu Ala Asn Leu Leu Ala
35 40 45

Leu Gln Ala Glu Leu Lys Ala Gly Asn Tyr Glu Arg Gly Pro Tyr Arg
50 55 60

Glu Phe Leu Val Tyr Glu Pro Lys Pro Arg Leu Ile Ser Ala Leu Glu
65 70 75 80

Phe Lys Asp Arg Leu Val Gln His Ala Leu Cys Asn Ile Val Ala Pro
85 90 95

Ile Phe Glu Ala Gly Leu Leu Pro Tyr Thr Tyr Ala Cys Arg Pro Asp
100 105 110

Lys Gly Thr His Ala Gly Val Cys His Val Gln Ala Glu Leu Arg Arg
115 120 125

Thr Arg Ala Thr His Phe Leu Lys Ser Asp Phe Ser Lys Phe Phe Pro
130 135 140

Ser Ile Asp Arg Ala Ala Leu Tyr Ala Met Ile Asp Lys Lys Ile His
145 150 155 160

Cys Ala Ala Thr Arg Arg Leu Leu Arg Val Val Leu Pro Asp Glu Gly
165 170 175

Val Gly Ile Pro Ile Gly Ser Leu Thr Ser Gln Leu Phe Ala Asn Val
180 185 190

Tyr Gly Gly Ala Val Asp Arg Leu Leu His Asp Glu Leu Lys Gln Arg
195 200 205
His Trp Ala Arg Tyr Met Asp Asp Ile Val Val Leu Gly Asp Asp Pro
210 215 220
Glu Glu Leu Arg Ala Val Phe Tyr Arg Leu Arg Asp Phe Ala Ser Glu
225 230 235 240
Arg Leu Gly Leu Lys Ile Ser His Trp Gln Val Ala Pro Val Ser Arg

245 250 255
Gly Ile Asn Phe Leu Gly Tyr Arg Ile Trp Pro Thr His Lys Leu Leu
260 265 270
Arg Lys Ser Ser Val Lys Arg Ala Lys Arg Lys Val Ala Asn Phe Ile
275 280 285
Lys His Gly Glu Asp Glu Ser Leu Gln Arg Phe Leu Ala Ser Trp Ser
290 295 300
Gly His Ala Gln Trp Ala Asp Thr His Asn Leu Phe Thr Trp Met Glu

305 310 315 320
Glu Gln Tyr Gly Ile Ala Cys His
325

<210> 236
<211> 128
<212> PRT
<213> Unknown
<220><223> Bordetella phage
<400> 236
Met Glu Pro Ile Glu Glu Ala Thr Lys Cys Tyr Asp Gln Met Leu Ile
1 5 10 15
Val Glu Arg Tyr Glu Arg Val Ile Ser Tyr Leu Tyr Pro Ile Ala Gln
20 25 30
Ser Ile Pro Arg Lys His Gly Val Ala Arg Glu Met Phe Leu Lys Cys

35 40 45
Leu Leu Gly Gln Val Glu Leu Phe Ile Val Ala Gly Lys Ser Asn Gln
50 55 60

Val Ser Lys Leu Tyr Ala Ala Asp Ala Gly Leu Ala Met Leu Arg Phe

65 70 75 80

Trp Leu Arg Phe Leu Ala Gly Ile Gln Lys Pro His Ala Met Thr Pro

85 90 95

His Gln Val Glu Thr Ala Gln Val Leu Ile Ala Glu Val Gly Arg Ile

100 105 110

Leu Gly Ser Trp Ile Ala Arg Val Asn Arg Lys Gly Gln Ala Gly Lys

115 120 125

<210> 237

<211> 169

<212> RNA

<213> Unknown

<220><223> Bordetella phage

<400> 237

accuucuugc auggcucugc caacgcuacg gcuuggcggg cuggccuuuc cucaauaggu 60

ggucagccgg uucuguccug cuucggcgaa cacguuacac gguucggcaa aacgucgauu 120

acugaaaaug gaaaggcggg gccgacuuca agggcaggcu gggaaauaa 169

<210> 238

<211> 149

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 238

accuucuugc auggcucugc caacgcuacg gcuuggcggg cuggccuuuc cucaauaggu 60

ggucagccgg uucuguccug cuucggcgaa cacguuacac gguucggcaa aacgucgauu 120

acugaaaaug gaaaggcggg gccgacuuc 149

<210> 239

<211> 85

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 239

accuucuugc auggcucugc caacgcuacg gcuuggcggg cuggccuuuc cucaauaggu 60

ggucaaaggg caggcuggga aauaa 85

<210> 240

<211> 160

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 240

cauggcucug ccaacgcuac ggcuuggcgg gcuggccuuu ccucaauagg uggucagccg 60

guucuguccu gcuucggcga acacguuaca cgguucggca aaacgucgau uacugaaaau 120

ggaaaggcgg ggccgacuuc aagggcaggc ugggaaauaa 160

<210> 241

<211> 140

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 241

cauggcucug ccaacgcuac ggcuuggcgg gcuggccuuu ccucaauagg uggucagccg 60

guucuguccu gcuucggcga acacguuaca cgguucggca aaacgucgau uacugaaaau 120

ggaaaggcgg ggccgacuuc 140

<210> 242

<211> 76

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 242

cauggcucug ccaacgcuac ggcuuggcgg gcuggccuuu ccucaauagg uggucaaagg 60

gcaggcuggg aaauaa 76

<210> 243

<211> 56

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 243

cauggcucug ccaacgcuac ggcuuggcgg gcuggccuuu ccucaauagg ugguca 56

<210> 244

<211> 169

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 244

accuucuugc auggcucugc caacgcuacg gcuuggcggg cuggccuuuc cucaauagau 60

gagccgccgg uucuguccug cuucggcgaa cacguuacac gguucggcaa aacgucgauu 120

acugaaaaug gaaaggcggg gccgacuuca agggcaggcu gggaaauaa 169

<210> 245

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 245

Gly Leu Asn Asp Ile Phe Glu Ala Gln Lys Ile Glu Trp His Glu

1 5 10 15

<210> 246

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 246

Glu Pro Glu Ala

1

<210> 247

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 247

Lys Arg Arg Trp Lys Lys Asn Phe Ile Ala Val Ser Ala Ala Asn Arg

1 5 10 15

Phe Lys Lys Ile Ser Ser Ser Gly Ala Leu

20 25

<210> 248

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 248

Glu Glu Glu Glu Glu Glu

1 5

<210> 249

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 249

Gly Ala Pro Val Pro Tyr Pro Asp Pro Leu Glu Pro Arg

1 5 10

<210> 250

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 250

Asp Tyr Lys Asp Asp Asp Asp Lys

1 5

<210> 251

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 251

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1 5

<210> 252

<211> 1

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 252

His

1

<210> 253

<211> 2

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 253

His His

1

<210> 254

<211> 3

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 254

His His His

1

<210> 255

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 255

His His His His

1

<210> 256

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 256

His His His His His

1 5

<210> 257

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 257

His His His His His His

1 5

<210> 258

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 258

His His His His His His His

1 5

<210> 259

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 259

His His His His His His His His

1 5

<210> 260

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 260

His His His His His His His His His

1 5

<210> 261

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 261

His His His His His His His His His His

1 5 10

<210> 262

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (2)..(25)

<223> may be absent

<400> 262

His His His His His His His His His His His His His His His His His

1 5 10 15

His His His His His His His His His

20 25

<210> 263

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 263

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu

1 5 10

<210> 264

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 264

Thr Lys Glu Asn Pro Arg Ser Asn Gln Glu Glu Ser Tyr Asp Asp Asn

1 5 10 15

Glu Ser

<210> 265

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 265

Thr Glu Thr Ser Gln Val Ala Pro Ala

1 5

<210> 266

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 266

Lys Glu Thr Ala Ala Ala Lys Phe Glu Arg Gln His Met Asp Ser

1 5 10 15

<210> 267

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 267

Met Asp Glu Lys Thr Thr Gly Trp Arg Gly Gly His Val Val Glu Gly
1 5 10 15

Leu Ala Gly Glu Leu Glu Gln Leu Arg Ala Arg Leu Glu His His Pro
20 25 30

Gln Gly Gln Arg Glu Pro
35

<210> 268

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 268

Ser Leu Ala Glu Leu Leu Asn Ala Gly Leu Gly Gly Ser
1 5 10

<210> 269

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 269

Thr Gln Asp Pro Ser Arg Val Gly
1 5

<210> 270

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 270

Pro Asp Arg Val Arg Ala Val Ser His Trp Ser Ser
1 5 10

<210> 271

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 271

Trp Ser His Pro Gln Phe Glu Lys

1 5

<210> 272

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 272

Cys Cys Pro Gly Cys Cys

1 5

<210> 273

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 273

Glu Val His Thr Asn Gln Asp Pro Leu Asp

1 5 10

<210> 274

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 274

Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr

1 5 10

<210> 275

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 275

Tyr Thr Asp Ile Glu Met Asn Arg Leu Gly Lys

1 5 10

<210> 276

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 276

Asp Leu Tyr Asp Asp Asp Asp Lys

1 5

<210> 277

<211> 65

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 277

accuucuugc auggcucugc caacgcuacg gcuuggcggg cuggccuuuc cucaauaggu 60

gguca 65

<210> 278

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 278

Asp Thr Tyr Arg Tyr Ile

1 5

<210> 279

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 279

Thr Asp Phe Tyr Leu Lys

1 5

<210> 280

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 280

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly

1 5 10

<210> 281

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 281

Gln Tyr Pro Ala Leu Thr

1 5

<210> 282

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 282

Ser Ser Thr Ser Ser Asp Phe Arg Asp Arg

1 5 10

<210> 283

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 283

Lys Asp His Leu Ile His Asn Val His Lys Glu Phe His Ala His Ala

1 5 10 15

His Asn Lys

<210> 284

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 284

Gln Pro Glu Leu Ala Pro Glu Asp

1 5

<210> 285

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 285

Arg Arg Arg Arg Arg

1 5

<210> 286

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 286

Cys Cys Cys Cys

1

<210> 287

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 287

Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe Phe

1 5 10

<210> 288

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 288

Asn Ala Asn Asn Pro Asp Trp Asp Phe

1 5

<210> 289

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 289

Lys Glu Thr Ala Ala Ala Lys Phe Glu Arg Gln His Met Asp Ser

1 5 10 15

<210> 290

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 290

Tyr Thr Asp Ile Glu Met Asn Arg Leu Gly Lys

1 5 10

<210> 291

<211> 253

<212> PRT

<213> Homo sapiens

<400> 291

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

Ile Leu Leu Ile Ser Phe Leu Ile Phe Leu Ile Val Gly

245 250

<210> 292

<211> 2365

<212> DNA

<213> Homo sapiens

<400> 292

gcgaaccttg gctgctggat gctggttctc tttgtggcca catggagtga cctgggcctc 60

tgcaagaagc gcccgaagcc tggaggatgg aacactgggg gcagccgata cccggggcag 120

ggcagccctg gaggcaaccg ctacccacct cagggcggtg gtggctgggg gcagcctcat 180

ggtggtggct gggggcagcc tcatggtggt ggctgggggc agccccatgg tggtggctgg 240

ggacagcctc atggtggtgg ctggggtcaa ggaggtggca cccacagtca gtggaacaag 300

ccgagtaagc caaaaaccaa catgaagcac atggctggtg ctgcagcagc tggggcagtg 360

gtgggggggcc ttggcggcta catgctggga agtgccatga gcaggcccat catacatttc 420

ggcagtgact atgaggaccg ttactatcgt gaaaacatgc accgttaccc caaccaagtg 480

tactacaggc ccatggatga gtacagcaac cagaacaact ttgtgcacga ctgcgtcaat 540

atcacaatca agcagcacac ggtcaccaca accaccaagg gggagaactt caccgagacc 600

gacgttaaga tgatggagcg cgtggttgag cagatgtgta tcacccagta cgagagggaa 660

tctcaggcct attaccagag aggatcgagc atggtcctct tctcctctcc acctgtgatc 720

ctcctgatct ctttcctcat cttcctgata gtgggatgag gaaggtcttc ctgttttcac 780

catctttcta atctttttcc agcttgaggg aggcggtatc cacctgcagc ccttttagtg 840

gtggtgtctc actctttctt ctctctttgt cccggatagg ctaatcaata cccttggcac 900

tgatgggcac tggaaaacat agagtagacc tgagatgctg gtcaagcccc ctttgattga 960

gttcatcatg agccgttgct aatgccaggc cagtaaaagt ataacagcaa ataaccattg 1020

gttaatctgg acttatttttt ggacttagtg caacaggttg aggctaaaac aaatctcaga 1080

acagtctgaa ataccttttgc ctggatacct ctggctcctt cagcagctag agctcagtat 1140

actaatgccc tatcttagta gagatttcat agctatttag agatattttc cattttaaga 1200

aaacccgaca acatttctgc caggtttgtt aggaggccac atgatactta ttcaaaaaaa 1260

tcctagagat tcttagctct tgggatgcag gctcagcccg ctggagcatg agctctgtgt 1320

gtaccgagaa ctggggtgat gttttacttt tcacagtatg ggctacacag cagctgttca 1380

acaagagtaa atattgtcac aacactgaac ctctggctag aggacatatt cacagtgaac 1440

ataactgtaa catatatgaa aggcttctgg gacttgaaat caaatgtttg ggaatggtgc 1500

ccttggaggc aacctcccat tttagatgtt taaaggaccc tatatgtggc attcctttct 1560

ttaaactata ggtaattaag gcagctgaaa agtaaattgc cttctagaca ctgaaggcaa 1620

atctcctttg tccatttacc tggaaaccag aatgattttg acatacagga gagctgcagt 1680

tgtgaaagca ccatcatcat agaggatgat gtaattaaaa aatggtcagt gtgcaaagaa 1740

aagaactgct tgcatttctt tatttctgtc tcataattgt caaaaaccag aattaggtca 1800

agttcatagt ttctgtaatt ggcttttgaa tcaaagaata gggagacaat ctaaaaaata 1860

tcttaggttg gagatgacag aaatatgatt gatttgaagt ggaaaaagaa attctgttaa 1920

tgttaattaa agtaaaatta ttccctgaat tgtttgatat tgtcacctag cagatatgta 1980

ttacttttct gcaatgttat tattggcttg cactttgtga gtattctatg taaaaatata 2040

tatgtatata aaatatatat tgcataggac agacttagga gttttgttta gagcagttaa 2100

catctgaagt gtctaatgca ttaacttttg taaggtactg aatacttaat atgtgggaaa 2160

ccctttgcg tggtccttag gcttacaatg tgcactgaat cgtttcatgt aagaatccaa 2220

agtggacacc attaacaggt ctttgaaata tgcatgtact ttatattttc tatatttgta 2280

actttgcatg ttcttgtttt gttatataaa aaaattgtaa atgtttaata tctgactgaa 2340

attaaacgag cgaagatgag cacca 2365

<210> 293

<400> 293

000

<210> 294

<400> 294

000

<210> 295

<400> 295

000

<210> 296

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 296

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Ala Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 297

<400> 297

000

<210> 298

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 298

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Gly Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 299

<400> 299

000

<210> 300

<400> 300

000

<210> 301

<400> 301

000

<210> 302

<400> 302

000

<210> 303

<400> 303

000

<210> 304

<400> 304

000

<210> 305

<400> 305

000

<210> 306

<400> 306

000

<210> 307

<400> 307

000

<210> 308

<400> 308

000

<210> 309

<400> 309

000

<210> 310

<400> 310

000

<210> 311

<400> 311

000

<210> 312

<400> 312

000

<210> 313

<400> 313

000

<210> 314

<400> 314

000

<

210> 315

<400> 315

000

<210> 316

<400> 316

000

<210> 317

<400> 317

000

<210> 318

<400> 318

000

<210> 319

<400> 319

000

<210> 320

<400> 320

000

<210> 321

<400> 321

000

<210> 322

<400> 322

000

<210> 323

<400> 323

000

<210> 324

<400> 324

000

<210> 325

<400> 325

000

<210> 326

<400> 326

000

<210> 327

<400> 327

000

<210> 328

<400> 328

000

<210> 329

<400> 329

000

<210> 330

<400> 330

000

<210

> 331

<211> 122

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 331

aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 60

aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 120

ta 122

<210> 332

<211> 252

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 332

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tctaagatac 120

attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa 180

atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca agttctatgg 240

ccctgactgg aa 252

<210> 333

<211> 146

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 333

ccatggccca acttgtttat tgcagcttat aatggttaca aataaagcaa tagcatcaca 60

aatttcacaa ataaagcatt ttttcactg cattctagtt gtggtttgtc caaactcatc 120

aatgtatctt atcatgtctg gatctc 146

<210> 334

<211> 276

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 334

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgagatcca 120

gacatgataa gatacattga tgagtttgga caaaccacaa ctagaatgca gtgaaaaaaa 180

tgctttattt gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat 240

aaacaagttg ggccatggct atggccctga ctggaa 276

<210> 335

<211> 244

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 335

tgatcataat caagccatat cacatctgta gaggtttact tgctttaaaa aacctccaca 60

cctccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact tgtttattgc 120

agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata aagcattttt 180

ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttatc atgtctggat 240

ctgc 244

<210> 336

<211> 374

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 336

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgcagatcc 120

agacatgata agatacattg atgagtttgg acaaaccaca actagaatgc agtgaaaaaa 180

atgctttatt tgtgaaattt gtgatgctat tgctttattt gtaaccatta taagctgcaa 240

taaacaagtt aacaacaaca attgcattca ttttatgttt caggttcagg gggaggtgtg 300

gaggtttttt aaagcaagta aacctctaca gatgtgatat ggcttgatta tgatcactat 360

ggccctgact ggaa 374

<210> 337

<211> 479

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 337

acgggtggca tccctgtgac ccctccccag tgcctctcct ggccctggaa gttgccactc 60

cagtgcccac cagccttgtc ctaataaaat taagttgcat cattttgtct gactaggtgt 120

ccttctataa tattatgggg tggaggggggg tggtatggag caaggggcaa gttgggaaga 180

caacctgtag ggcctgcggg gtctattggg aaccaagctg gagtgcagtg gcacaatctt 240

ggctcactgc aatctccgcc tcctgggttc aagcgattct cctgcctcag cctcccgagt 300

tgttgggatt ccaggcatgc atgaccaggc tcagctaatt tttgtttttt tggtagagac 360

ggggtttcac catattggcc aggctggtct ccaactccta atctcaggtg atctacccac 420

cttggcctcc caaattgctg ggattacagg cgtgaaccac tgctcccttc cctgtcctt 479

<210> 338

<211> 609

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 338

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcaaggacag 120

ggaagggagc agtggttcac gcctgtaatc ccagcaattt gggaggccaa ggtgggtaga 180

tcacctgaga ttaggagttg gagaccagcc tggccaatat ggtgaaaccc cgtctctacc 240

aaaaaaacaa aaattagctg agcctggtca tgcatgcctg gaatcccaac aactcgggag 300

gctgaggcag gagaatcgct tgaacccagg aggcggagat tgcagtgagc caagattgtg 360

ccactgcact ccagcttggt tcccaataga ccccgcaggc cctacaggtt gtcttcccaa 420

cttgcccctt gctccatacc acccccctcc accccataat attatagaag gacacctagt 480

cagacaaaat gatgcaactt aattttatta ggacaaggct ggtgggcact ggagtggcaa 540

cttccagggc caggagaggc actggggagg ggtcacaggg atgccacccg tctatggccc 600

tgactggaa 609

<210> 339

<211> 228

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 339

cgactgtgcc ttctagttgc cagccatctg ttgtttgccc ctcccccgtg ccttccttga 60

ccctggaagg tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgcatt 120

gtctgagtag gtgtcattct attctggggg gtggggtggg gcaggacagc aaggggggagg 180

attgggaaga caatagcagg catgctgggg atgcggtggg ctctatgg 228

<210> 340

<211> 358

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 340

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcccatagag 120

cccaccgcat ccccagcatg cctgctattg tcttcccaat cctccccctt gctgtcctgc 180

cccaccccac cccccagaat agaatgacac ctactcagac aatgcgatgc aatttcctca 240

ttttattagg aaaggacagt gggagtggca ccttccaggg tcaaggaagg cacgggggag 300

gggcaaacaa cagatggctg gcaactagaa ggcacagtcg ctatggccct gactggaa 358

<210> 341

<211> 527

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 341

ttcactcctc aggtgcaggc tgcctatcag aaggtggtgg ctggtgtggc caatgccctg 60

gctcacaaat accactgaga tctttttccc tctgccaaaa attatgggga catcatgaag 120

ccccttgagc atctgacttc tggctaataa aggaaattta ttttcattgc aatagtgtgt 180

tggaattttt tgtgtctctc actcggaagg acatatggga gggcaaatca tttaaaacat 240

cagaatgagt atttggttta gagtttggca acatatgccc atatgctggc tgccatgaac 300

aaaggttggc tataaagagg tcatcagtat atgaaacagc cccctgctgt ccattcctta 360

ttccatagaa aagccttgac ttgaggttag atttttttta tattttgttt tgtgttattt 420

ttttctttaa catccctaaa attttcctta catgttttac tagccagatt tttcctcctc 480

tcctgactac tcccagtcat agctgtccct cttctcttat ggagatc 527

<210> 342

<211> 657

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 342

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgatctcca 120

taagagaaga gggacagcta tgactgggag tagtcaggag aggaggaaaa atctggctag 180

taaaacatgt aaggaaaatt ttagggatgt taaagaaaaa aataacacaa aacaaaatat 240

aaaaaaaatc taacctcaag tcaaggcttt tctatggaat aaggaatgga cagcaggggg 300

ctgtttcata tactgatgac ctctttatag ccaacctttg ttcatggcag ccagcatatg 360

ggcatatgtt gccaaactct aaaccaaata ctcattctga tgttttaaat gatttgccct 420

cccatatgtc cttccgagtg agagacacaa aaaattccaa cacactattg caatgaaaat 480

aaatttcctt tattagccag aagtcagatg ctcaaggggc ttcatgatgt ccccataatt 540

tttggcagag ggaaaaagat ctcagtggta tttgtgagcc agggcattgg ccacaccagc 600

caccaccttc tgataggcag cctgcacctg aggagtgaac tatggccctg actggaa 657

<210> 343

<211> 271

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 343

gggggaggct aactgaaaca cggaaggaga caataccgga aggaacccgc gctatgacgg 60

caataaaaag acagaataaa acgcacgggt gttgggtcgt ttgttcataa acgcggggtt 120

cggtcccagg gctggcactc tgtcgatacc ccaccgagac cccattgggg ccaatacgcc 180

cgcgtttctt ccttttcccc accccacccc ccaagttcgg gtgaaggccc agggctcgca 240

gccaacgtcg gggcggcagg ccctgccata g 271

<210> 344

<211> 401

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 344

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcctatggca 120

gggcctgccg ccccgacgtt ggctgcgagc cctgggcctt cacccgaact tgggggggtgg 180

ggtggggaaa aggaagaaac gcgggcgtat tggccccaat ggggtctcgg tggggtatcg 240

acagagtgcc agccctggga ccgaaccccg cgtttatgaa caaacgaccc aacacccgtg 300

cgttttattc tgtcttttta ttgccgtcat agcgcgggtt ccttccggta ttgtctcctt 360

ccgtgtttca gttagcctcc cccctatggc cctgactgga a 401

<210> 345

<211> 171

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 345

tagggtcatg aaggtttttc ttttcctgag aaaacaacac gtattgtttt ctcaggtttt 60

gctttttggc ctttttctag cttaaaaaaa aaaaaagcaa aagatgctgg tggttggcac 120

tcctggtttc caggacgggg ttcaaatccc tgcggcgtct ttgctttgac t 171

<210> 346

<211> 301

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 346

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcagtcaaag 120

caaagacgcc gcagggattt gaaccccgtc ctggaaacca ggagtgccaa ccaccagcat 180

cttttgcttt ttttttttt aagctagaaa aaggccaaaa agcaaaacct gagaaaacaa 240

tacgtgttgt tttctcagga aaagaaaaac cttcatgacc ctactatggc cctgactgga 300

a 301

<210> 347

<211> 208

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 347

tgttttggct gggtttttcc ttgttcgcac cggacacctc cagtgaccag acggcaaggt 60

ttttatccca gtgtatattg gaaaaacatg ttatactttt gacaatttaa cgtgcctaga 120

gctcaaatta aactaatacc ataacgtaat gcaacttaca acataaataa aggtcaatgt 180

ttaatccata aaaaaaaaaa aaaaaaaa 208

<210> 348

<211> 338

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 348

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tctttttttt 120

ttttttttt tatggattaa acattgacct ttatttatgt tgtaagttgc attacgttat 180

ggtattagtt taatttgagc tctaggcacg ttaaattgtc aaaagtataa catgtttttc 240

caatatacac tgggataaaa accttgccgt ctggtcactg gaggtgtccg gtgcgaacaa 300

ggaaaaaccc agccaaaaca ctatggccct gactggaa 338

<210> 349

<211> 636

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 349

tgttttggct gggtttttcc ttgttcgcac cggacacctc cagtgaccag acggcaaggt 60

ttttatccca gtgtatattg gaaaaacatg ttatactttt gacaatttaa cgtgcctaga 120

gctcaaatta aactaatacc ataacgtaat gcaacttaca acataaataa aggtcaatgt 180

ttaatccata aaaaaaaaaa aaaaaaaaac acactgtttt ggctgggttt ttccttgttc 240

gcaccggaca cctccagtga ccagacggca aggtttttat cccagtgtat attggaaaaa 300

catgttatac ttttgacaat ttaacgtgcc tagagctcaa attaaactaa taccataacg 360

taatgcaact tacaacataa ataaaggtca atgtttaatc cataaaaaaa aaaaaaaaaa 420

aatctctctg ttttggctgg gtttttcctt gttcgcaccg gacacctcca gtgaccagac 480

ggcaaggttt ttatcccagt gtatattgga aaaacatgtt atacttttga caatttaacg 540

tgcctagagc tcaaattaaa ctaataccat aacgtaatgc aacttacaac ataaataaag 600

gtcaatgttt aatccataaa aaaaaaaaaa aaaaaa 636

<210> 350

<211> 766

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 350

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tctttttttt 120

ttttttttt tatggattaa acattgacct ttatttatgt tgtaagttgc attacgttat 180

ggtattagtt taatttgagc tctaggcacg ttaaattgtc aaaagtataa catgtttttc 240

caatatacac tgggataaaa accttgccgt ctggtcactg gaggtgtccg gtgcgaacaa 300

ggaaaaaccc agccaaaaca gagagatttt ttttttttt tttttatgga ttaaacattg 360

acctttattt atgttgtaag ttgcattacg ttatggtatt agtttaattt gagctctagg 420

cacgttaaat tgtcaaaagt ataacatgtt tttccaatat acactgggat aaaaaccttg 480

ccgtctggtc actggaggtg tccggtgcga acaaggaaaa acccagccaa aacagtgtgt 540

ttttttttt ttttttttta tggattaaac attgaccttt atttatgttg taagttgcat 600

tacgttatgg tattagttta atttgagctc taggcacgtt aaattgtcaa aagtataaca 660

tgtttttcca atatacactg ggataaaaac cttgccgtct ggtcactgga ggtgtccggt 720

gcgaacaagg aaaaacccag ccaaaacact atggccctga ctggaa 766

<210> 351

<211> 204

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 351

cagcaagttc agagaaatct gaacttgctg gatttttgga gcagggagat ggaataggag 60

cttgctccgt ccactccacg catcgacctg gtattgcagt acctccagga acggtgcacc 120

cactttctgg agtttcaaaa gtagactgta cgctaagggt catatctttt tttgtttggt 180

ttgtgtcttg gttggcgtct taaa 204

<210> 352

<211> 334

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 352

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tctttaagac 120

gccaaccaag acacaaacca aacaaaaaaa gatatgaccc ttagcgtaca gtctactttt 180

gaaactccag aaagtgggtg caccgttcct ggaggtactg caataccagg tcgatgcgtg 240

gagtggacgg agcaagctcc tattccatct ccctgctcca aaaatccagc aagttcagat 300

ttctctgaac ttgctgctat ggccctgact ggaa 334

<210> 353

<211> 73

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 353

gtttcaaaag tagactgtac gctaagggtc atatcttttt ttgtttggtt tgtgtcttgg 60

ttggcgtctt aaa 73

<210> 354

<211> 203

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 354

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tctttaagac 120

gccaaccaag acacaaacca aacaaaaaaa gatatgaccc ttagcgtaca gtctactttt 180

gaaacctatg gccctgactg gaa 203

<210> 355

<211> 73

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 355

gcccggctag ctcagtcggt agagcatgag actcttaatc tcagggtcgt gggttcgagc 60

cccacgttgg gcg 73

<210> 356

<211> 203

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 356

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tccgcccaac 120

gtggggctcg aacccacgac cctgagatta agagtctcat gctctaccga ctgagctagc 180

cgggcctatg gccctgactg gaa 203

<210> 357

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 357

gagacggtcg ggtccagata ttcgtatctg tcgagtagag tgtgggctc 49

<210> 358

<211> 179

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 358

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgagcccac 120

actctactcg acagatacga atatctggac ccgaccgtct cctatggccc tgactggaa 179

<210> 359

<211> 98

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 359

gacgcaactg aatgaaatgg tgaaggacgg gtccaggtgt ggctgcttcg gcagtgcagc 60

ttgttgagta gagtgtgagc tccgtaacta gtcgcgtc 98

<210> 360

<211> 228

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 360

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgacgcgac 120

tagttacgga gctcacactc tactcaacaa gctgcactgc cgaagcagcc acacctggac 180

ccgtccttca ccatttcatt cagttgcgtc ctatggccct gactggaa 228

<210> 361

<211> 95

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 361

gatgtaactg aatgaaatgg tgaaggacgg gtccagtagg ctgcttcggc agcctacttg 60

ttgagtagag tgtgagctcc gtaactagtt acatc 95

<210> 362

<211> 225

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 362

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgatgtaac 120

tagttacgga gctcacactc tactcaacaa gtaggctgcc gaagcagcct actggacccg 180

tccttcacca tttcattcag ttacatccta tggccctgac tggaa 225

<210> 363

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 363

ggcacgtacg aagggacggt gcggagagga gagtacgtgc 40

<210> 364

<211> 170

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 364

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgcacgtac 120

tctcctctcc gcaccgtccc ttcgtacgtg ccctatggcc ctgactggaa 170

<210> 365

<211> 68

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 365

ggccggcatg gtcccagcct cctcgctggc gccggctggg caacatgctt cggcatggcg 60

aatgggac 68

<210> 366

<211> 198

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 366

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcgtcccatt 120

cgccatgccg aagcatgttg cccagccggc gccagcgagg aggctgggac catgccggcc 180

ctatggccct gactggaa 198

<210> 367

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 367

ggactctagg actggacttt ggact 25

<210> 368

<211> 155

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 368

atccttccag tcagggccat gtttgagagc tagaaatagc aagtttaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gggaccgagt cggtccacct gaaccgtata tcagtccaaa 120

gtccagtcct agagtcccta tggccctgac tggaa 155

<210> 369

<211> 1719

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 369

gttcgttgca acaaattgat gagcaatgct tttttataat gccaactttg tacaaaaaag 60

ttggcatgac aagttccagg ctttggtttt cgctgctgct ggcggcagcg ttcgcaggac 120

gggcgacggc cctctggccc tggcctcaga acttccaaac ctccgaccag cgctacgtcc 180

tttacccgaa caactttcaa ttccagtacg atgtcagctc ggccgcgcag cccggctgct 240

cagtcctcga cgaggccttc cagcgctatc gtgacctgct tttcggttcc gggtcttggc 300

cccgtcctta cctcacaggg aaacggcata cactggagaa gaatgtgttg gttgtctctg 360

tagtcacacc tggatgtaac cagcttccta ctttggagtc agtggagaat tataccctga 420

ccataaatga tgaccagtgt ttactcctct ctgagactgt ctggggagct ctccgaggtc 480

tggagacttt tagccagctt gtttggaaat ctgctgaggg cacattcttt atcaacaaga 540

ctgagattga ggactttccc cgctttcctc accggggctt gctgttggat acatctcgcc 600

attacctgcc actctctagc atcctggaca ctctggatgt catggcgtac aataaattga 660

acgtgttcca ctggcatctg gtagatgatc cttccttccc atatgagagc ttcacttttc 720

cagagctcat gagaaagggg tcctacaacc ctgtcaccca catctacaca gcacaggatg 780

tgaaggaggt cattgaatac gcacggctcc ggggtatccg tgtgcttgca gagtttgaca 840

ctcctggcca cactttgtcc tggggaccag gtatccctgg attactgact ccttgctacc 900

ctgggtctga gccctctggc acctttggac cagtgaatcc cagtctcaat aatacctatg 960

agttcatgag cacattcttc ttagaagtca gctctgtctt cccagatttt tatcttcatc 1020

ttggaggaga tgaggttgat ttcacctgct ggaagtccaa cccagagatc caggacttta 1080

tgaggaagaa aggcttcggt gaggacttca agcagctgga gtccttctac atccagacgc 1140

tgctggacat cgtctcttct tatggcaagg gctatgtggt gtggcaggag gtgtttgata 1200

ataaagtaaa gattcagcca gacacaatca tacaggtgtg gcgagaggat attccagtga 1260

actatatgaa ggagctggaa ctggtcacca aggccggctt ccgggccctt ctctctgccc 1320

cctggtacct gaaccgtata tcctatggcc ctgactggaa ggatttctac gtagtggaac 1380

ccctggcatt tgaaggtacc cctgagcaga aggctctggt gattggtgga gaggcttgta 1440

tgtggggaga atatgtggac aacacaaacc tggtccccag gctctggccc agagcagggg 1500

ctgttgccga aaggctgtgg agcaacaagt tgacatctga cctgacattt gcctatgaac 1560

gtttgtcaca cttccgctgt gagttgctga ggcgaggtgt ccaggcccaa cccctcaatg 1620

taggcttctg tgagcaggag tttgaacaga cctgcccaac tttcttgtac aaagttggca 1680

ttataagaaa gcattgctta tcaatttgtt gcaacgaac 1719

<210> 370

<211> 529

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 370

Met Thr Ser Ser Arg Leu Trp Phe Ser Leu Leu Leu Ala Ala Ala Phe

1 5 10 15

Ala Gly Arg Ala Thr Ala Leu Trp Pro Trp Pro Gln Asn Phe Gln Thr

20 25 30

Ser Asp Gln Arg Tyr Val Leu Tyr Pro Asn Asn Phe Gln Phe Gln Tyr

35 40 45

Asp Val Ser Ser Ala Ala Gln Pro Gly Cys Ser Val Leu Asp Glu Ala

50 55 60

Phe Gln Arg Tyr Arg Asp Leu Leu Phe Gly Ser Gly Ser Trp Pro Arg

65 70 75 80

Pro Tyr Leu Thr Gly Lys Arg His Thr Leu Glu Lys Asn Val Leu Val

85 90 95

Val Ser Val Val Thr Pro Gly Cys Asn Gln Leu Pro Thr Leu Glu Ser

100 105 110

Val Glu Asn Tyr Thr Leu Thr Ile Asn Asp Asp Gln Cys Leu Leu Leu

115 120 125

Ser Glu Thr Val Trp Gly Ala Leu Arg Gly Leu Glu Thr Phe Ser Gln

130 135 140

Leu Val Trp Lys Ser Ala Glu Gly Thr Phe Phe Ile Asn Lys Thr Glu

145 150 155 160

Ile Glu Asp Phe Pro Arg Phe Pro His Arg Gly Leu Leu Leu Asp Thr

165 170 175

Ser Arg His Tyr Leu Pro Leu Ser Ser Ile Leu Asp Thr Leu Asp Val

180 185 190

Met Ala Tyr Asn Lys Leu Asn Val Phe His Trp His Leu Val Asp Asp

195 200 205

Pro Ser Phe Pro Tyr Glu Ser Phe Thr Phe Pro Glu Leu Met Arg Lys

210 215 220

Gly Ser Tyr Asn Pro Val Thr His Ile Tyr Thr Ala Gln Asp Val Lys

225 230 235 240

Glu Val Ile Glu Tyr Ala Arg Leu Arg Gly Ile Arg Val Leu Ala Glu

245 250 255

Phe Asp Thr Pro Gly His Thr Leu Ser Trp Gly Pro Gly Ile Pro Gly

260 265 270

Leu Leu Thr Pro Cys Tyr Pro Gly Ser Glu Pro Ser Gly Thr Phe Gly

275 280 285

Pro Val Asn Pro Ser Leu Asn Asn Thr Tyr Glu Phe Met Ser Thr Phe

290 295 300

Phe Leu Glu Val Ser Ser Val Phe Pro Asp Phe Tyr Leu His Leu Gly

305 310 315 320

Gly Asp Glu Val Asp Phe Thr Cys Trp Lys Ser Asn Pro Glu Ile Gln

325 330 335

Asp Phe Met Arg Lys Lys Gly Phe Gly Glu Asp Phe Lys Gln Leu Glu

340 345 350

Ser Phe Tyr Ile Gln Thr Leu Leu Asp Ile Val Ser Ser Tyr Gly Lys

355 360 365

Gly Tyr Val Val Trp Gln Glu Val Phe Asp Asn Lys Val Lys Ile Gln

370 375 380

Pro Asp Thr Ile Ile Gln Val Trp Arg Glu Asp Ile Pro Val Asn Tyr

385 390 395 400

Met Lys Glu Leu Glu Leu Val Thr Lys Ala Gly Phe Arg Ala Leu Leu

405 410 415

Ser Ala Pro Trp Tyr Leu Asn Arg Ile Ser Tyr Gly Pro Asp Trp Lys

420 425 430

Asp Phe Tyr Val Val Glu Pro Leu Ala Phe Glu Gly Thr Pro Glu Gln

435 440 445

Lys Ala Leu Val Ile Gly Gly Glu Ala Cys Met Trp Gly Glu Tyr Val

450 455 460

Asp Asn Thr Asn Leu Val Pro Arg Leu Trp Pro Arg Ala Gly Ala Val

465 470 475 480

Ala Glu Arg Leu Trp Ser Asn Lys Leu Thr Ser Asp Leu Thr Phe Ala

485 490 495

Tyr Glu Arg Leu Ser His Phe Arg Cys Glu Leu Leu Arg Arg Gly Val

500 505 510

Gln Ala Gln Pro Leu Asn Val Gly Phe Cys Glu Gln Glu Phe Glu Gln

515 520 525

Thr

<210> 371

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(5)

<223> n is a, c, g, or t

<400> 371

nnnnntcatc ctgataaact gcaaa 25

<210> 372

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(5)

<223> n is a, c, g, or t

<400> 372

nnnnncagtt tgtacggatg actgatgagt cccaaatagg acgaaacgcg cttcggtgcg 60

tctcatcctg ataaactgca aa 82

<210> 373

<211> 333

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 373

acctgaaccg tatatcgacg caccgaagcg cgtttcgtcc tatttgggac tcatcaggat 60

atacggttca ggtgatatac ggttcaggtg acgcaccgaa gcgcgtttcg tcctatttgg 120

gactcatcag acctgaaccg tatatcatcc ttccagtcag ggccatgttt gagagctaga 180

aatagcaagt ttaaataagg ctagtccgtt atcaacttga aaaagtggga ccgagtcggt 240

ccacctgaac cgtatatcga cgcaccgaag cgcgtttcgt cctatttggg actcatcagg 300

atatacggtt caggtctatg gccctgactg gaa 333

<210> 374

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 374

cctgggtcaa tccttggggc ccagactgag cacg 34

<210> 375

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 375

cctgggtcaa tccttggggc ccagactgag cacgtgatgg cagaggaaag g 51

<210> 376

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223

> Synthetic

<400> 376

cctttcctct gccatcacgt gctcagtctg ggccccaagg attgacccag g 51

<210> 377

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 377

tgatggcaga ggaaagg 17

<210> 378

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 378

gcaggcttta aaggaaccaa ttc 23

<210> 379

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 379

gcaggcttta aaggaaccaa ttccctgggt caatccttgg ggc 43

<210> 380

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 380

ctctggagga tctagcggag 20

<210> 381

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 381

ctctggagga tctagcggag ttttttttt ttttttttt 40

<210> 382

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 382

ctctggagga tctagcggag ccccccccc cccc 34

<210> 383

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 383

ggcccagact gagcacgtga 20

<210> 384

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 384

ggcccagact gagcacgtga 20

<210> 385

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 385

ggcccagact gagcacgtga 20

<210> 386

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 386

ggcccagact gagcacgtga 20

<210> 387

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 387

ggcccagact gagcacgtga 20

<210> 388

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 388

ggcccagact gagcacgtga 20

<210> 389

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 389

ggcccagact gagcacgtga 20

<210> 390

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 390

ggcccagact gagcacgtga 20

<210> 391

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 391

ggcccagact gagcacgtga 20

<210> 392

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 392

ggcccagact gagcacgtga 20

<210> 393

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 393

ggcccagact gagcacgtga 20

<210> 394

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 394

tctgctatca cgtgctcagt ctg 23

<210> 395

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 395

gcacatacta gcccctgtct 20

<210> 396

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 396

Gln Tyr Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu

1 5 10 15

<210> 397

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 397

catgatataa aagcaaattc taaatttata ggtataactg aacta 45

<210> 398

<211> 535

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 398

Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn

1 5 10 15

Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln

20 25 30

Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp

35 40 45

Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly

50 55 60

Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val

65 70 75 80

Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val

85 90 95

Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu

100 105 110

Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly

115 120 125

Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser

130 135 140

Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser

145 150 155 160

Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp

165 170 175

Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg

180 185 190

Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val

195 200 205

Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly

210 215 220

Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu

225 230 235 240

Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu

245 250 255

His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val

260 265 270

Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val

275 280 285

Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu

290 295 300

Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala

305 310 315 320

Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser

325 330 335

Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp

340 345 350

Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe

355 360 365

Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His

370 375 380

Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr

385 390 395 400

Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His

405 410 415

Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val

420 425 430

Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr

435 440 445

His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile

450 455 460

Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly

465 470 475 480

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser

485 490 495

Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu

500 505 510

Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser

515 520 525

Leu Phe Phe Glu Ile Lys Ser

530 535

<210> 399

<211> 1605

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 399

ggcgccgacg acgtggtgga cagcagcaag agcttcgtga tggagaactt cagcagctac 60

cacggcacca agcccggcta cgtggacagc atccagaagg gcatccagaa gcccaagagc 120

ggcacccagg gcaactacga cgacgactgg aaggagttct acagcaccga caacaagtac 180

gacgccgccg gctacagcgt ggacaacgag aacccccctga gcggcaaggc cggcggcgtg 240

gtgaaggtga cctaccccgg cctgaccaag gtgctggccc tgaaggtgga caacgccgag 300

accatcaaga aggagctggg cctgagcctg accgagcccc tgatggagca ggtgggcacc 360

gaggagttca tcaagaggtt cggcgacggc gccagcaggg tggtgctgag cctgcccttc 420

gccgagggca gcagcagcgt ggagtacatc aacaactggg agcaggccaa ggccctgagc 480

gtggagctgg agatcaactt cgagaccagg ggcaagaggg gccaggacgc catgtacgag 540

tacatggccc aggcctgcgc cggcaacagg gtgaggagga gcgtgggcag cagcctgagc 600

tgcatcaacc tggactggga cgtgatcagg gacaagacca agaccaagat cgagagcctg 660

aaggagcacg gccccatcaa gaacaagatg agcgagagcc ccaacaagac cgtgagcgag 720

gagaaggcca agcagtacct ggaggagttc caccagaccg ccctggagca ccccgagctg 780

agcgagctga agaccgtgac cggcaccaac cccgtgttcg ccggcgccaa ctacgccgcc 840

tgggccgtga acgtggccca ggtgatcgac agcgagaccg ccgacaacct ggagaagacc 900

accgccgccc tgagcatcct gcccggcatc ggcagcgtga tgggcatcgc cgacggcgcc 960

gtgcaccaca acaccgagga gatcgtggcc cagagcatcg ccctgagcag cctgatggtg 1020

gcccaggcca tccccctggt gggcgagctg gtggacatcg gcttcgccgc ctacaacttc 1080

gtggagagca tcatcaacct gttccaggtg gtgcacaaca gctacaacag gcccgcctac 1140

agccccggcc acaagaccca gcccttcctg cacgacggct acgccgtgag ctggaacacc 1200

gtggaggaca gcatcatcag gaccggcttc cagggcgaga gcggccacga catcaagatc 1260

accgccgaga acaccccccct gcccatcgcc ggcgtgctgc tgcccaccat ccccggcaag 1320

ctggacgtga acaagagcaa gacccacatc agcgtgaacg gcaggaagat caggatgagg 1380

tgcagggcca tcgacggcga cgtgaccttc tgcaggccca agagccccgt gtacgtgggc 1440

aacggcgtgc acgccaacct gcacgtggcc ttccacagga gcagcagcga gaagatccac 1500

agcaacgaga tcagcagcga cagcatcggc gtgctgggct accagaagac cgtggaccac 1560

accaaggtga acagcaagct gagcctgttc ttcgagatca agagc 1605

<210> 400

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 400

Gly Thr Tyr Arg Leu Ile Pro Asn Ala Arg Ala Asn Leu Thr Ala

1 5 10 15

<210> 401

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 401

ggcacctaca ggctgatccc caacgccagg gccaacctga ccgcc 45

<210> 402

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 402

Pro Ser Lys Phe Phe Thr Ile Ser Asp Ser Ala Thr Phe Ala Pro Gly
1 5 10 15

Pro Val Ser Asn Ala
20

<210> 403

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 403

ccgagcaaat tttttaccat tagcgatagc gcgacctttg cgccgggccc ggtgagcaac 60

gcg 63

<210> 404

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223>

Synthetic

<400> 404

Pro Ser Lys Leu Phe Ile Met Leu Asp Asn Ala Thr Phe Ala Pro Gly
1 5 10 15

Pro Val Val Asn Ala
20

<210> 405

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 405

ccgagcaaac tgtttattat gctggataac gcgacctttg cgccgggccc ggtggtgaac 60

gcg 63

<210> 406

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 406

Thr Pro Pro Pro Tyr Gln Val Ser Cys Gly Gly Glu Ser Asp Arg Ala

1 5 10 15

Ser Ala Arg Val Ile Asp Pro Ala Ala Gln Ser

20 25

<210> 407

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 407

acccccccc cctaccaggt gagctgcggc ggcgagagcg acagggccag cgccagggtg 60

atcgaccccg ccgcccagag c 81

<210> 408

<211> 134

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 408

Pro Glu Tyr Ala Tyr Lys Ile Val Lys Asn Lys Lys Met Glu Asp Gly

1 5 10 15

Phe Leu Gln Gly Met Val Asp Gly Trp Tyr Gly His His Ser Asn Glu
20 25 30
Gln Gly Ser Gly Leu Met Glu Asn Glu Arg Thr Leu Asp Lys Ala Asn
35 40 45
Pro Asn Asn Asp Leu Cys Ser Trp Ser Asp His Glu Ala Ser Ser Asn

50 55 60
Asn Thr Asn Gln Glu Asp Leu Leu Gln Arg Glu Ser Arg Arg Lys Lys
65 70 75 80
Arg Ile Gly Thr Ser Thr Leu Asn Gln Arg Gly Asn Cys Asn Thr Lys
85 90 95
Cys Gln Thr Glu Glu Ala Arg Leu Lys Arg Glu Glu Val Ser Leu Val
100 105 110
Lys Ser Asp Gln Cys Ser Asn Gly Ser Leu Gln Cys Arg Ala Asn Asn

115 120 125
Ser Thr Glu Gln Val Asp
130

<210> 409
<211> 402
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 409
cccgagtacg cctacaagat cgtgaagaac aagaagatgg aggacggctt cctgcagggc 60
atggtggacg gctggtacgg ccaccacagc aacgagcagg gcagcggcct gatggagaac 120
gagaggaccc tggacaaggc caaccccaac aacgacctgt gcagctggag cgaccacgag 180
gccagcagca acaacaccaa ccaggaggac ctgctgcaga gggagagcag gaggaagaag 240

aggatcggca ccagcaccct gaaccagagg ggcaactgca acaccaagtg ccagaccgag 300
gaggccaggc tgaagaggga ggaggtgagc ctggtgaaga gcgaccagtg cagcaacggc 360
agcctgcagt gcagggccaa caacagcacc gagcaggtgg ac 402
<210> 410
<211> 154
<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 410

Thr Lys Ser Thr Asn Ser Arg Ser Gly Gly Ile Ser Gly Pro Asp Asn
1 5 10 15

Glu Ala Pro Val Gly Glu Ala Pro Ser Pro Tyr Gly Asp Asn Pro Arg

20 25 30

Pro Asn Asp Gly Asn Asn Ile Arg Ile Gly Ser Lys Gly Tyr Asn Gly
35 40 45

Ile Ile Thr Asp Thr Ile Glu Glu Ser Cys Ser Cys Tyr Pro Asp Ala
50 55 60

Lys Val Val Lys Ser Val Glu Leu Asp Ser Thr Ile Trp Thr Ser Gly
65 70 75 80

Ser Ser Pro Asn Gln Lys Ile Ile Thr Ile Gly Trp Asp Pro Asn Gly

85 90 95

Trp Thr Gly Thr Pro Met Ser Pro Asn Gly Ala Tyr Gly Thr Asp Gly
100 105 110

Pro Ser Asn Gly Gln Ala Asn Gln His Gln Ala Glu Ser Ile Ser Ala
115 120 125

Gly Asn Ser Ser Leu Cys Pro Ile Arg Asp Asn Trp His Gly Ser Asn
130 135 140

Arg Ser Trp Ser Trp Pro Asp Gly Ala Glu
145 150

<210> 411

<211> 462

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 411

accaagagca ccaacagcag gagcggcggc atcagcggcc ccgacaacga ggcccccgtg 60

ggcgaggccc ccagccccta cggcgacaac cccaggccca acgacggcaa caacatcagg 120

atcggcagca agggctacaa cggcatcatc accgacacca tcgaggagag ctgcagctgc 180

taccccgacg ccaaggtggt gaagagcgtg gagctggaca gcaccatctg gaccagcggc 240

agcagcccca accagaagat catcaccatc ggctgggacc ccaacggctg gaccggcacc 300

cccatgagcc ccaacggcgc ctacggcacc gacggcccca gcaacggcca ggccaaccag 360

caccaggccg agagcatcag cgccggcaac agcagcctgt gccccatcag ggacaactgg 420

cacggcagca acaggagctg gagctggccc gacggcgccg ag 462

<210> 412

<211> 45

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 412

Glu Lys Ile Val Leu Leu Leu Ala Met Met Glu Lys Ile Val Leu Leu
1 5 10 15

Leu Ala Lys Cys Gln Thr Pro Met Gly Ala Ile Lys Ala Val Asp Gly
20 25 30

Val Thr Asn Lys Cys Pro Tyr Leu Gly Ser Pro Ser Phe
35 40 45

<210> 413

<211> 135

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 413

gagaagatcg tgctgctgct ggccatgatg gagaagatcg tgctgctgct ggccaagtgc 60

cagaccccca tgggcgccat caaggccgtg gacggcgtga ccaacaagtg cccctacctg 120

ggcagcccca gcttc 135

<210> 414

<211> 36

<212> PRT

<213> Artificial Sequence

<

220><223> Synthetic

<400> 414

Ile Arg Pro Cys Phe Trp Val Glu Leu Asn Pro Asn Gln Lys Ile Ile
1 5 10 15
Thr Ile Arg Pro Cys Phe Trp Val Glu Leu Ile Cys Tyr Pro Asp Ala
20 25 30
Gly Glu Ile Thr
35

<210> 415
<211> 108
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 415
atcaggccct gcttctgggt ggagctgaac cccaaccaga agatcatcac catcaggccc 60

tgcttctggg tggagctgat ctgctacccc gacgccggcg agatcacc 108

<210> 416
<211> 45
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 416
Met Glu Lys Ile Val Leu Leu Leu Ala Glu Lys Ile Val Leu Leu Leu
1 5 10 15
Ala Met Cys Pro Tyr Leu Gly Ser Pro Ser Phe Lys Cys Gln Thr Pro
20 25 30
Met Gly Ala Ile Lys Ala Val Asp Gly Val Thr Asn Lys
35 40 45

<210> 417
<211> 135
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 417
atggagaaga tcgtgctgct gctggccgag aagatcgtgc tgctgctggc catgtgcccc 60

tacctgggca gccccagctt caagtgccag acccccatgg gcgccatcaa ggccgtggac 120

ggcgtgacca acaag 135

<210> 418

<211> 36

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 418

Asn Pro Asn Gln Lys Ile Ile Thr Ile Cys Tyr Pro Asp Ala Gly Glu

1 5 10 15

Ile Thr Ile Arg Pro Cys Phe Trp Val Glu Leu Arg Pro Cys Phe Trp

20 25 30

Val Glu Leu Ile

35

<210> 419

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 419

aaccccaacc agaagatcat caccatctgc taccccgacg ccggcgagat caccatcagg 60

ccctgcttct gggtggagct gaggccctgc ttctgggtgg agctgatc 108

<210> 420

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 420

Met Val Ser Leu Val Lys Ser Asp Gln Ile Gly Thr Ser Thr Leu Asn

1 5 10 15

Gln Arg

<210> 421

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 421

atggtgagcc tggtgaagag cgaccagatc ggcaccagca ccctgaacca gagg 54

<210> 422

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 422

Tyr Asn Gly Ile Ile Thr Asp Thr Ile

1 5

<210> 423

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 423

tacaacggca tcatcaccga caccatc 27

<210> 424

<211> 63

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 424

Met Glu Lys Ile Val Leu Leu Leu Ala Glu Lys Ile Val Leu Leu Leu

1 5 10 15

Ala Met Met Val Ser Leu Val Lys Ser Asp Gln Cys Pro Tyr Leu Gly

20 25 30

Ser Pro Ser Phe Ile Gly Thr Ser Thr Leu Asn Gln Arg Lys Cys Gln

35 40 45

Thr Pro Met Gly Ala Ile Lys Ala Val Asp Gly Val Thr Asn Lys

50 55 60

<210> 425

<211> 135

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 425

atggagaaga tcgtgctgct gctggccgag aagatcgtgc tgctgctggc catgatggtg 60

agcctggtga agagcgacca gtgcccctac ctgggcagcc ccagcttcat cggcaccagc 120

accctgaacc agagg 135

<210> 426

<211> 45

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 426

Asn Pro Asn Gln Lys Ile Ile Thr Ile Tyr Asn Gly Ile Ile Thr Asp

1 5 10 15

Thr Ile Cys Tyr Pro Asp Ala Gly Glu Ile Thr Ile Arg Pro Cys Phe

20 25 30

Trp Val Glu Leu Arg Pro Cys Phe Trp Val Glu Leu Ile

35 40 45

<210> 427

<211> 135

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 427

aaccccaacc agaagatcat caccatctac aacggcatca tcaccgacac catctgctac 60

cccgacgccg gcgagatcac catcaggccc tgcttctggg tggagctgag gccctgcttc 120

tgggtggagc tgatc 135

<210> 428

<211> 535

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 428

Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15

Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln
20 25 30

Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45

Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50 55 60

Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65 70 75 80

Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
85 90 95

Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100 105 110

Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115 120 125

Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130 135 140

Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145 150 155 160

Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165 170 175

Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg
180 185 190

Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val
195 200 205

Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly
210 215 220
Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu
225 230 235 240
Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
245 250 255
His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val
260 265 270

Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val
275 280 285
Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu
290 295 300
Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala
305 310 315 320
Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser
325 330 335

Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp
340 345 350
Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe
355 360 365
Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His
370 375 380
Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr
385 390 395 400

Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His
405 410 415
Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val
420 425 430
Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr
435 440 445
His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile

450 455 460

Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly
465 470 475 480
Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser
485 490 495
Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu
500 505 510
Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser
515 520 525

Leu Phe Phe Glu Ile Lys Ser
530 535

<210> 429
<211> 27
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 429
ggctaaccgt gccatttgat caggtca 27

<210> 430
<211> 27
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 430
ggctaaccgt gcaaattaac aaactaa 27

<210> 431
<211> 28
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic

<400> 431
ggccatctcg tgcaaattaa caaactaa 28

<210> 432

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 432

ggtcctctgc catctcgtgc aaattaacaa actaa 35

<210> 433

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 433

ggcttccttt cctctgccat ctcgtgcaaa ttaacaaact aa 42

<210> 434

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 434

ggctaaccgt gccatttgat caggtca 27

<210> 435

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 435

ggctaaccgt gcaaattaac aaactaa 27

<210> 436

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 436

ggccatctcg tgcaaattaa caaactaa 28

<210> 437

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 437

ggtcctctgc catctcgtgc aaattaacaa actaa 35

<210> 438

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 438

ggcttccttt cctctgccat ctcgtgcaaa ttaacaaact aa 42

<210> 439

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 439

tctgccatct cgtgctc 17

<210> 440

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 440

tctgccatct cgtgctc 17

<210> 441

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 441

tctgccatca tcgtgctc 18

<210> 442

<211> 16

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 442

tctgccatcc gtgctc 16

<210> 443

<211> 1517

<212> PRT

<213> Streptococcus pyogenes

<400> 443

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn
20 25 30

Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys
35 40 45

Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn
50 55 60

Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr
65 70 75 80

Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg
85 90 95

Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp
100 105 110

Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp
115 120 125

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val

130 135 140

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu

145 150 155 160

Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu

165 170 175

Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu

180 185 190

Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln

195 200 205

Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val

210 215 220

Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu

225 230 235 240

Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe

245 250 255

Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser

260 265 270

Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr

275 280 285

Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr

290 295 300

Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu

305 310 315 320

Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser

325 330 335

Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu

340 345 350

Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile

355 360 365

Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly

370 375 380

Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys
385 390 395 400
Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu
405 410 415
Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile
420 425 430
His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr

435 440 445
Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe
450 455 460
Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe
465 470 475 480
Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe
485 490 495
Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg

500 505 510
Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys
515 520 525
His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys
530 535 540
Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly
545 550 555 560
Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys

565 570 575
Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys
580 585 590
Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser
595 600 605
Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe
610 615 620
Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr

625 630 635 640

Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr

645 650 655

Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg

660 665 670

Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile

675 680 685

Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp

690 695 700

Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu

705 710 715 720

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp

725 730 735

Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys

740 745 750

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val

755 760 765

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu

770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys

785 790 795 800

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu

805 810 815

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr

820 825 830

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe

850 855 860

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys

865 870 875 880

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys

885 890 895

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln

900 905 910

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu

915 920 925

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln

930 935 940

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

945 950 955 960

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu

965 970 975

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

980 985 990

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

995 1000 1005

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1010 1015 1020

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1025 1030 1035

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1040 1045 1050

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1055 1060 1065

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1070 1075 1080

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1085 1090 1095

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1100 1105 1110

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1115 1120 1125

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1130 1135 1140

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala

1145 1150 1155

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1160 1165 1170

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1175 1180 1185

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1190 1195 1200

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1205 1210 1215

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1220 1225 1230

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1235 1240 1245

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1250 1255 1260

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1265 1270 1275

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1280 1285 1290

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1295 1300 1305

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1310 1315 1320

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1325 1330 1335

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1340 1345 1350

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1355 1360 1365

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1370 1375 1380

Gly Gly Asp Ser Gly Gly Ser Ser Gly Gly Ser Cys Leu Ser Tyr

1385 1390 1395

Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu Pro Ile Gly

1400 1405 1410

Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser Val Asp

1415 1420 1425

Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His Asp

1430 1435 1440

Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser

1445 1450 1455

Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly

1460 1465 1470

Gln Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu

1475 1480 1485

Met Arg Val Asp Asn Leu Pro Asn Ser Gly Gly Ser Lys Arg Thr

1490 1495 1500

Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val

1505 1510 1515

<210> 444

<400> 444

000

<210> 445

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 445

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

```
        755                 760                 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
    770                 775                 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785                 790                 795                 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
                805                 810                 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
            820                 825                 830
Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
        835                 840                 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
    850                 855                 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865                 870                 875                 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
                885                 890                 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
            900                 905                 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
        915                 920                 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
    930                 935                 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945                 950                 955                 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
                965                 970                 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
            980                 985                 990

Gly Thr Ala Leu Ile Lys Lys Tyr  Pro Lys Leu Glu Ser  Glu Phe Val
        995                 1000                1005
```

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 446

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 446

Ser Gly Gly Ser Ser Gly Gly Ser

1 5

<210> 447

<211> 106

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 447

Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu

1 5 10 15

Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser
20 25 30
Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His
35 40 45
Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
50 55 60
Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln
65 70 75 80

Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg
85 90 95
Val Asp Asn Leu Pro Asn Ser Gly Gly Ser
100 105

<210> 448
<400> 448
000
<210> 449
<400> 449
000
<210> 450
<211> 780
<212> PRT
<213> Staphylococcus pyogenes
<400> 450
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15
Arg Lys Val Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn

20 25 30
Val Tyr Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn
35 40 45
Gly Phe Ile Ala Ser Asn Cys Phe Asn Ser Gly Ser Glu Thr Pro Gly
50 55 60

Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly
65 70 75 80
Ser Ser Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser

85 90 95
Lys Glu Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro
100 105 110
Gln Ala Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala
115 120 125
Pro Leu Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys
130 135 140
Gln Tyr Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile

145 150 155 160
Gln Arg Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp
165 170 175
Asn Thr Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg
180 185 190
Pro Val Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His
195 200 205
Pro Thr Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser

210 215 220
His Gln Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu
225 230 235 240
Arg Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp
245 250 255
Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln
260 265 270
Gly Phe Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp

275 280 285
Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr
290 295 300
Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln

305 310 315 320

Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala

325 330 335

Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly

340 345 350

Tyr Leu Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu

355 360 365

Thr Val Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu

370 375 380

Phe Leu Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala

385 390 395 400

Glu Met Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe

405 410 415

Asn Trp Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala

420 425 430

Leu Leu Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe

435 440 445

Glu Leu Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr

450 455 460

Gln Lys Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys

465 470 475 480

Leu Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala

485 490 495

Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln

500 505 510

Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln

515 520 525

Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala

530 535 540

Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu

545 550 555 560

Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn
565 570 575
Cys Leu Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr
580 585 590
Asp Gln Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser

595 600 605
Ser Leu Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr
610 615 620
Glu Thr Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala
625 630 635 640
Gln Arg Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu
645 650 655
Gly Lys Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr

660 665 670
Ala His Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser
675 680 685
Glu Gly Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys
690 695 700
Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His
705 710 715 720
Gln Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln

725 730 735
Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu
740 745 750
Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys Arg Thr Ala Asp
755 760 765
Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val
770 775 780
<210> 451
<400> 451
000
<210> 452

<211> 35

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 452

Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn Val Tyr Asp
1 5 10 15

Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn Gly Phe Ile
20 25 30

Ala Ser Asn
35

<210> 453

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 453

Cys Phe Asn Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
1 5 10 15

Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser
20 25

<210> 454

<211> 677

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<400> 454

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210

> 455

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 455

gtgccatcac gtgctcagtc t 21

<210> 456

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 456

gagcgatcca ggtgctgcag a 21

<210> 457

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 457

ggagcccttc ttcttctgct 20

<210> 458

<211> 881

<212> PRT

<213> Staphylococcus aureus

<400> 458

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Gly Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile
20 25 30

Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile
35 40 45

Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu

50 55 60

Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg
65 70 75 80

His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu
85 90 95

Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val
100 105 110

Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu

115 120 125

Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu
130 135 140

Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn

145 150 155 160

Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg

165 170 175

Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr

180 185 190

Ser Asp Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala

195 200 205

Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu

210 215 220

Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe

225 230 235 240

Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys

245 250 255

Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala

260 265 270

Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg

275 280 285

Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu

290 295 300

Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys

305 310 315 320

Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser

325 330 335

Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys

340 345 350

Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp

355 360 365

Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln

370 375 380

Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu

385 390 395 400

Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu
405 410 415
Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn
420 425 430
Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp

435 440 445
Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile
450 455 460
Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile
465 470 475 480
Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu
485 490 495
Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu

500 505 510
Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile
515 520 525
Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys
530 535 540
Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile
545 550 555 560
Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His

565 570 575
Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val
580 585 590
Leu Val Lys Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe
595 600 605
Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys
610 615 620
Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr

625 630 635 640
Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val

645 650 655

Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr

660 665 670

Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu

675 680 685

Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg

690 695 700

Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His

705 710 715 720

Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu

725 730 735

Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe

740 745 750

Glu Glu Lys Gln Ala Glu Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr

755 760 765

Val Glu Tyr Gly Leu Leu Pro Ile Gly Lys Ile Val Glu Lys Arg Ile

770 775 780

Glu Cys Thr Val Tyr Ser Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln

785 790 795 800

Pro Val Ala Gln Trp His Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr

805 810 815

Cys Leu Glu Asp Gly Ser Leu Ile Arg Ala Thr Lys Asp His Lys Phe

820 825 830

Met Thr Val Asp Gly Gln Met Leu Pro Ile Asp Glu Ile Phe Glu Arg

835 840 845

Glu Leu Asp Leu Met Arg Val Asp Asn Leu Pro Asn Ser Gly Gly Ser

850 855 860

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys

865 870 875 880

Val

<210> 459

<400> 459

000

<210

> 460

<211> 739

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 460

Gly Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val

1 5 10 15

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly

20 25 30

Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg

35 40 45

Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile

50 55 60

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His

65 70 75 80

Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu

85 90 95

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu

100 105 110

Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr

115 120 125

Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala

130 135 140

Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys

145 150 155 160

Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr

165 170 175

Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln

180 185 190

Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg

195 200 205

Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys

210 215 220

Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe

225 230 235 240

Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr

245 250 255

Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn

260 265 270

Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe

275 280 285

Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu

290 295 300

Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys

305 310 315 320

Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr

325 330 335

Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala

340 345 350

Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu

355 360 365

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser

370 375 380

Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile

385 390 395 400

Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala

405 410 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln

420 425 430

Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro

435 440 445

Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile

450 455 460

Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg

465 470 475 480

Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys

485 490 495

Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr

500 505 510

Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp

515 520 525

Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu

530 535 540

Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro

545 550 555 560

Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys

565 570 575

Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu

580 585 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile

595 600 605

Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu

610 615 620

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp

625 630 635 640

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu

645 650 655

Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys

660 665 670

Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp

675 680 685

Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp

690 695 700

Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys

705 710 715 720

Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys

725 730 735

Gln Ala Glu

<210> 461

<400> 461

000

<210> 462

<211> 102

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 462

Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu

1 5 10 15

Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser

20 25 30

Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His

35 40 45

Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser

50 55 60

Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln

65 70 75 80

Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg

85 90 95

Val Asp Asn Leu Pro Asn

100

<210> 463

<400> 463

000

<210> 464

<400> 464

000

<210> 465

<211> 1099

<212> PRT

<213> Staphylococcus aureus

<400> 465

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn
20 25 30

Val Tyr Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn
35 40 45

Gly Phe Ile Ala Ser Asn Cys Phe Asn Glu Ile Glu Thr Glu Gln Glu
50 55 60

Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp
65 70 75 80

Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
85 90 95

Lys Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly
100 105 110

Asn Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn
115 120 125

Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met
130 135 140

Tyr His His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu
145 150 155 160

Gln Tyr Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr

165 170 175

Gly Asn Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile

180 185 190

Lys Lys Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile

195 200 205

Thr Asp Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu

210 215 220

Lys Pro Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe

225 230 235 240

Val Thr Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu

245 250 255

Val Asn Ser Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser

260 265 270

Asn Gln Ala Glu Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys

275 280 285

Ile Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu

290 295 300

Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu

305 310 315 320

Glu Asn Met Asn Asp Lys Arg Pro Pro His Ile Ile Lys Thr Ile Ala

325 330 335

Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn

340 345 350

Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly

355 360 365

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

370 375 380

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser

385 390 395 400

Ser Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys

405 410 415

Glu Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln
420 425 430

Ala Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro
435 440 445

Leu Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln
450 455 460

Tyr Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln
465 470 475 480

Arg Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn
485 490 495

Thr Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro
500 505 510

Val Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro
515 520 525

Thr Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His
530 535 540

Gln Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg
545 550 555 560

Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro
565 570 575

Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly
580 585 590

Phe Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu
595 600 605

Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val
610 615 620

Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly
625 630 635 640

Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser
645 650 655

Ala Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr

660 665 670

Leu Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr

675 680 685

Val Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe

690 695 700

Leu Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu

705 710 715 720

Met Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn

725 730 735

Trp Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu

740 745 750

Leu Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu

755 760 765

Leu Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln

770 775 780

Lys Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu

785 790 795 800

Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala

805 810 815

Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro

820 825 830

Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro

835 840 845

Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu

850 855 860

Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn

865 870 875 880

Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys

885 890 895

Leu Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp

900 905 910

Gln Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser
915 920 925

Leu Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu
930 935 940

Thr Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln
945 950 955 960

Arg Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly
965 970 975

Lys Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala
980 985 990

His Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu
995 1000 1005

Gly Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys
1010 1015 1020

Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly
1025 1030 1035

His Gln Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala
1040 1045 1050

Asp Gln Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr
1055 1060 1065

Ser Thr Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys
1070 1075 1080

Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys
1085 1090 1095

Val

<210> 466

<400> 466

000

<210> 467

<211> 314

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 467

Cys Phe Asn Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile Phe Ile
1 5 10 15

Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr Lys Tyr
20 25 30

Ser His Arg Val Asp Lys Lys Pro Asn Arg Lys Leu Ile Asn Asp Thr
35 40 45

Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile Val Asn
50 55 60

Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu
65 70 75 80

Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp Pro Gln
85 90 95

Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp Glu Lys
100 105 110

Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu Thr Lys
115 120 125

Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys Tyr Tyr
130 135 140

Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr Pro Asn
145 150 155 160

Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg Phe Asp
165 170 175

Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys Asn Leu
180 185 190

Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys Cys Tyr
195 200 205

Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu Phe Ile
210 215 220

Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile Asn Gly Glu Leu Tyr

225 230 235 240

Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu Val Asn
245 250 255

Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn Asp Lys
260 265 270

Arg Pro Pro His Ile Ile Lys Thr Ile Ala Ser Lys Thr Gln Ser Ile
275 280 285

Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr Glu Val Lys Ser
290 295 300

Lys Lys His Pro Gln Ile Ile Lys Lys Gly
305 310

<210> 468

<400> 468

000

<210> 469

<400> 469

000

<210> 470

<400> 470

000

<210> 471

<211> 677

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<400> 471

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670

Glu Asn Ser Ser Pro
675

<210> 472
<211> 413
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 472
Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15
His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val
20 25 30
Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp
35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60

Ala Ile Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80
Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95
Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met
100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr
115 120 125
Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140
Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160
Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190
Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205
Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220
Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu
225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255
Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270
His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285
Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu

325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 473

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 473

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr

1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Val Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60

Ala Ile Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly

65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr

115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr

130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg

145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser

180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp

195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met

210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile

260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu

290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320

His Leu Tyr Ser Met Lys Tyr Thr Asn Val Val Pro Leu Tyr Asp Leu

325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys
405 410

<210> 474

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 474

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val
20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp
35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met
100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Pro Tyr Ser Glu Tyr
115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu
225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly
340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365

His Asp Met Leu Ala Glu Gly Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 475

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 475

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr

1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly

65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr

115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr

130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg

145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser

180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp

195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met

210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile

260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu

290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu

325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 476

<211> 413

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 476

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val
20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp
35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60

Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met
100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Pro Tyr Ser Glu Tyr
115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu
165 170 175

Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220
Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu
225 230 235 240
Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr
245 250 255
Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile
260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala
275 280 285
Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300
Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320
His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly
340 345 350
Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365
His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380
Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu
385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys
405 410

<210> 477
<211> 413
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 477

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15
His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val
20 25 30
Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45
Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60
Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80
Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro
85 90 95
Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110
Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr
115 120 125
Asn Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr
130 135 140
Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg
145 150 155 160
Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175
Glu Arg Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser
180 185 190
Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp
195 200 205
Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met
210 215 220
Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240
Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile

260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu

290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu

325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Gly Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 478

<211> 413

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 478

Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr

1 5 10 15

His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp

35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly

50 55 60

Ala Ile Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly

65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr

115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr

130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg

145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175

Glu Arg Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser

180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp

195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met

210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile

260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu
290 295 300
Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu
305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu
325 330 335
Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly
340 345 350
Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu
355 360 365
His Asp Met Leu Ala Glu Gly Leu Arg Tyr Ser Val Ile Arg Glu Val
370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu
385 390 395 400
His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys
405 410

<210> 479
<211> 413
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 479
Cys Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1 5 10 15
His Arg Ile Glu Asp Val Val Asp Gly Arg Lys Pro Ile His Val Val

20 25 30
Ala Val Ala Lys Asp Gly Thr Leu Leu Ala Arg Pro Val Val Ser Trp
35 40 45
Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50 55 60
Ala Thr Val Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gly Pro

85 90 95

Gly Gly Ser Gly Asn Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met

100 105 110

Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr

115 120 125

Asp Pro Thr Ser Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr

130 135 140

Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp Ala Lys Arg

145 150 155 160

Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Ala His Leu Leu

165 170 175

Glu Arg Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser

180 185 190

Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp

195 200 205

Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met

210 215 220

Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu

225 230 235 240

Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr

245 250 255

Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile

260 265 270

His Arg Ala Leu Asp Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala

275 280 285

Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu

290 295 300

Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu

305 310 315 320

His Leu Tyr Ser Met Lys Tyr Lys Asn Val Val Pro Leu Tyr Asp Leu

325 330 335

Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Gly Gly Ser Gly

340 345 350

Ala Ser Arg Val Gln Ala Phe Ala Asp Ala Leu Asp Asp Lys Phe Leu

355 360 365

His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val

370 375 380

Leu Pro Thr Arg Arg Ala Arg Thr Phe Asp Leu Glu Val Glu Glu Leu

385 390 395 400

His Thr Leu Val Ala Glu Gly Val Val Val His Asn Cys

405 410

<210> 480

<400> 480

000

<210> 481

<400> 481

000

<210> 482

<400> 482

000

<210> 483

<400> 483

000

<210> 484

<400> 484

000

<210> 485

<400> 485

000

<210> 486

<400> 486

000

<210> 487

<211> 240

<212> PRT

<213> Homo sapiens

<400> 487

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 488

<211> 108

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 488

Met Gly Val Gln Val Glu Thr Ile Ser Pro Gly Asp Gly Arg Thr Phe

1 5 10 15

Pro Lys Arg Gly Gln Thr Cys Val Val His Tyr Thr Gly Met Leu Glu

20 25 30

Asp Gly Lys Lys Phe Asp Ser Ser Arg Asp Arg Asn Lys Pro Phe Lys

35 40 45

Phe Met Leu Gly Lys Gln Glu Val Ile Arg Gly Trp Glu Glu Gly Val

50 55 60

Ala Gln Met Ser Val Gly Gln Arg Ala Lys Leu Thr Ile Ser Pro Asp

65 70 75 80

Tyr Ala Tyr Gly Ala Thr Gly His Pro Gly Ile Ile Pro Pro His Ala

85 90 95

Thr Leu Val Phe Asp Val Glu Leu Leu Lys Leu Glu

100 105

<210> 489

<211> 108

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 489

Met Gly Val Gln Val Glu Thr Ile Ser Pro Gly Asp Gly Arg Thr Phe

1 5 10 15

Pro Lys Arg Gly Gln Thr Cys Val Val His Tyr Thr Gly Met Leu Glu

20 25 30

Asp Gly Lys Lys Val Asp Ser Ser Arg Asp Arg Asn Lys Pro Phe Lys

35 40 45

Phe Met Leu Gly Lys Gln Glu Val Ile Arg Gly Trp Glu Glu Gly Val

50 55 60

Ala Gln Met Ser Val Gly Gln Arg Ala Lys Leu Thr Ile Ser Pro Asp

65 70 75 80

Tyr Ala Tyr Gly Ala Thr Gly His Pro Gly Ile Ile Pro Pro His Ala

85 90 95

Thr Leu Val Phe Asp Val Glu Leu Leu Lys Leu Glu

100 105

<210> 490

<211> 165

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 490

Met Val Asn Pro Thr Val Phe Phe Asp Ile Ala Val Asp Gly Glu Pro

1 5 10 15

Leu Gly Arg Val Ser Phe Glu Leu Phe Ala Asp Lys Val Pro Lys Thr

20 25 30

Ala Glu Asn Phe Arg Ala Leu Ser Thr Gly Glu Lys Gly Phe Gly Tyr

35 40 45

Lys Gly Ser Cys Phe His Arg Ile Ile Pro Gly Phe Met Cys Gln Gly

50 55 60

Gly Asp Phe Thr Arg His Asn Gly Thr Gly Gly Lys Ser Ile Tyr Gly

65 70 75 80

Glu Lys Phe Glu Asp Glu Asn Phe Ile Leu Lys His Thr Gly Pro Gly

85 90 95

Ile Leu Ser Met Ala Asn Ala Gly Pro Asn Thr Asn Gly Ser Gln Phe

100 105 110

Phe Ile Cys Thr Ala Lys Thr Glu Trp Leu Asp Gly Lys His Val Val

115 120 125

Phe Gly Lys Val Lys Glu Gly Met Asn Ile Val Glu Ala Met Glu Arg

130 135 140

Phe Gly Ser Arg Asn Gly Lys Thr Ser Lys Lys Ile Thr Ile Ala Asp

145 150 155 160

Cys Gly Gln Leu Glu

165

<210> 491

<211> 108

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 491

Met Gly Val Gln Val Glu Thr Ile Ser Pro Gly Asp Gly Arg Thr Phe

1 5 10 15

Pro Lys Arg Gly Gln Thr Cys Val Val His Tyr Thr Gly Met Leu Glu

20 25 30

Asp Gly Lys Lys Phe Asp Ser Ser Arg Asp Arg Asn Lys Pro Phe Lys

35 40 45

Phe Met Leu Gly Lys Gln Glu Val Ile Arg Gly Trp Glu Glu Gly Val

50 55 60

Ala Gln Met Ser Val Gly Gln Arg Ala Lys Leu Thr Ile Ser Pro Asp

65 70 75 80

Tyr Ala Tyr Gly Ala Thr Gly His Pro Gly Ile Ile Pro Pro His Ala

85 90 95

Thr Leu Val Phe Asp Val Glu Leu Leu Lys Leu Glu

100 105

<210> 492

<400> 492

000

<210> 493

<211> 216

<212> PRT

<213> Homo sapiens

<400> 493

Met Leu Arg Leu Ser Glu Arg Asn Met Lys Val Leu Leu Ala Ala Ala

1 5 10 15

Leu Ile Ala Gly Ser Val Phe Phe Leu Leu Leu Pro Gly Pro Ser Ala

20 25 30

Ala Asp Glu Lys Lys Lys Gly Pro Lys Val Thr Val Lys Val Tyr Phe

35 40 45

Asp Leu Arg Ile Gly Asp Glu Asp Val Gly Arg Val Ile Phe Gly Leu

50 55 60

Phe Gly Lys Thr Val Pro Lys Thr Val Asp Asn Phe Val Ala Leu Ala

65 70 75 80

Thr Gly Glu Lys Gly Phe Gly Tyr Lys Asn Ser Lys Phe His Arg Val

85 90 95

Ile Lys Asp Phe Met Ile Gln Gly Gly Asp Phe Thr Arg Gly Asp Gly

100 105 110

Thr Gly Gly Lys Ser Ile Tyr Gly Glu Arg Phe Pro Asp Glu Asn Phe

115 120 125

Lys Leu Lys His Tyr Gly Pro Gly Trp Val Ser Met Ala Asn Ala Gly

130 135 140

Lys Asp Thr Asn Gly Ser Gln Phe Phe Ile Thr Thr Val Lys Thr Ala

145 150 155 160

Trp Leu Asp Gly Lys His Val Val Phe Gly Lys Val Leu Glu Gly Met

165 170 175

Glu Val Val Arg Lys Val Glu Ser Thr Lys Thr Asp Ser Arg Asp Lys

180 185 190

Pro Leu Lys Asp Val Ile Ile Ala Asp Cys Gly Lys Ile Glu Val Glu

195 200 205

Lys Pro Phe Ala Ile Ala Lys Glu

210 215

<210> 494

<211> 212

<212> PRT

<213> Mus musculus

<400> 494

Met Ser Pro Gly Pro Arg Leu Leu Leu Pro Ala Val Leu Cys Leu Gly
1 5 10 15

Leu Gly Ala Leu Val Ser Ser Ser Gly Ser Ser Gly Val Arg Lys Arg
20 25 30

Gly Pro Ser Val Thr Asp Lys Val Phe Phe Asp Val Arg Ile Gly Asp
35 40 45

Lys Asp Val Gly Arg Ile Val Ile Gly Leu Phe Gly Asn Val Val Pro
50 55 60

Lys Thr Val Glu Asn Phe Val Ala Leu Ala Thr Gly Glu Lys Gly Tyr
65 70 75 80

Gly Tyr Lys Gly Ser Ile Phe His Arg Val Ile Lys Asp Phe Met Ile
85 90 95

Gln Gly Gly Asp Phe Thr Ala Arg Asp Gly Thr Gly Gly Met Ser Ile
100 105 110

Tyr Gly Glu Thr Phe Pro Asp Glu Asn Phe Lys Leu Lys His Tyr Gly
115 120 125

Ile Gly Trp Val Ser Met Ala Asn Ala Gly Pro Asp Thr Asn Gly Ser
130 135 140

Gln Phe Phe Ile Thr Leu Thr Lys Pro Thr Trp Leu Asp Gly Lys His
145 150 155 160

Val Val Phe Gly Lys Val Leu Asp Gly Met Thr Val Val His Ser Ile
165 170 175

Glu Leu Gln Ala Thr Asp Gly His Asp Arg Pro Leu Thr Asp Cys Thr
180 185 190

Ile Val Asn Ser Gly Lys Ile Asp Val Lys Thr Pro Phe Val Val Glu
195 200 205

Val Pro Asp Trp
210

<210> 495

<211> 108

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 495

Met Gly Val Gln Val Glu Thr Ile Ser Pro Gly Asp Gly Arg Thr Phe

1 5 10 15

Pro Lys Arg Gly Gln Thr Cys Val Val His Tyr Thr Gly Met Leu Glu

20 25 30

Asp Gly Lys Lys Val Asp Ser Ser Arg Asp Arg Asn Lys Pro Phe Lys

35 40 45

Phe Met Leu Gly Lys Gln Glu Val Ile Arg Gly Trp Glu Glu Gly Val

50 55 60

Ala Gln Met Ser Val Gly Gln Arg Ala Lys Leu Thr Ile Ser Pro Asp

65 70 75 80

Tyr Ala Tyr Gly Ala Thr Gly His Pro Gly Ile Ile Pro Pro His Ala

85 90 95

Thr Leu Val Phe Asp Val Glu Leu Leu Lys Leu Glu

100 105

<210> 496

<211> 187

<212> PRT

<213> Homo sapiens

<400> 496

Met Val Gly Ser Leu Asn Cys Ile Val Ala Val Ser Gln Asn Met Gly

1 5 10 15

Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe

20 25 30

Arg Tyr Phe Gln Arg Met Thr Thr Thr Ser Ser Val Glu Gly Lys Gln

35 40 45

Asn Leu Val Ile Met Gly Lys Lys Thr Trp Phe Ser Ile Pro Glu Lys

50 55 60

Asn Arg Pro Leu Lys Gly Arg Ile Asn Leu Val Leu Ser Arg Glu Leu

65 70 75 80

Lys Glu Pro Pro Gln Gly Ala His Phe Leu Ser Arg Ser Leu Asp Asp

85 90 95

Ala Leu Lys Leu Thr Glu Gln Pro Glu Leu Ala Asn Lys Val Asp Met

100 105 110

Val Trp Ile Val Gly Gly Ser Ser Val Tyr Lys Glu Ala Met Asn His

115 120 125

Pro Gly His Leu Lys Leu Phe Val Thr Arg Ile Met Gln Asp Phe Glu

130 135 140

Ser Asp Thr Phe Phe Pro Glu Ile Asp Leu Glu Lys Tyr Lys Leu Leu

145 150 155 160

Pro Glu Tyr Pro Gly Val Leu Ser Asp Val Gln Glu Glu Lys Gly Ile

165 170 175

Lys Tyr Lys Phe Glu Val Tyr Glu Lys Asn Asp

180 185

<210> 497

<211> 159

<212> PRT

<213> Escherichia coli

<400> 497

Met Ile Ser Leu Ile Ala Ala Leu Ala Val Asp Arg Val Ile Gly Met

1 5 10 15

Glu Asn Ala Met Pro Trp Asn Leu Pro Ala Asp Leu Ala Trp Phe Lys

20 25 30

Arg Asn Thr Leu Asn Lys Pro Val Ile Met Gly Arg His Thr Trp Glu

35 40 45

Ser Ile Gly Arg Pro Leu Pro Gly Arg Lys Asn Ile Ile Leu Ser Ser

50 55 60

Gln Pro Gly Thr Asp Asp Arg Val Thr Trp Val Lys Ser Val Asp Glu

65 70 75 80

Ala Ile Ala Ala Cys Gly Asp Val Pro Glu Ile Met Val Ile Gly Gly

85 90 95

Gly Arg Val Tyr Glu Gln Phe Leu Pro Lys Ala Gln Lys Leu Tyr Leu

100 105 110

Thr His Ile Asp Ala Glu Val Glu Gly Asp Thr His Phe Pro Asp Tyr

115 120 125

Glu Pro Asp Asp Trp Glu Ser Val Phe Ser Glu Phe His Asp Ala Asp

130 135 140

Ala Gln Asn Ser His Ser Tyr Cys Phe Glu Ile Leu Glu Arg Arg

145 150 155

<210> 498

<211> 777

<212> PRT

<213> Homo sapiens

<400> 498

Met Asp Ser Lys Glu Ser Leu Thr Pro Gly Arg Glu Glu Asn Pro Ser

1 5 10 15

Ser Val Leu Ala Gln Glu Arg Gly Asp Val Met Asp Phe Tyr Lys Thr

20 25 30

Leu Arg Gly Gly Ala Thr Val Lys Val Ser Ala Ser Ser Pro Ser Leu

35 40 45

Ala Val Ala Ser Gln Ser Asp Ser Lys Gln Arg Arg Leu Leu Val Asp

50 55 60

Phe Pro Lys Gly Ser Val Ser Asn Ala Gln Gln Pro Asp Leu Ser Lys

65 70 75 80

Ala Val Ser Leu Ser Met Gly Leu Tyr Met Gly Glu Thr Glu Thr Lys

85 90 95

Val Met Gly Asn Asp Leu Gly Phe Pro Gln Gln Gly Gln Ile Ser Leu

100 105 110

Ser Ser Gly Glu Thr Asp Leu Lys Leu Leu Glu Glu Ser Ile Ala Asn

115 120 125

Leu Asn Arg Ser Thr Ser Val Pro Glu Asn Pro Lys Ser Ser Ala Ser
130 135 140
Thr Ala Val Ser Ala Ala Pro Thr Glu Lys Glu Phe Pro Lys Thr His
145 150 155 160
Ser Asp Val Ser Ser Glu Gln Gln His Leu Lys Gly Gln Thr Gly Thr
165 170 175

Asn Gly Gly Asn Val Lys Leu Tyr Thr Thr Asp Gln Ser Thr Phe Asp
180 185 190
Ile Leu Gln Asp Leu Glu Phe Ser Ser Gly Ser Pro Gly Lys Glu Thr
195 200 205
Asn Glu Ser Pro Trp Arg Ser Asp Leu Leu Ile Asp Glu Asn Cys Leu
210 215 220
Leu Ser Pro Leu Ala Gly Glu Asp Asp Ser Phe Leu Leu Glu Gly Asn
225 230 235 240

Ser Asn Glu Asp Cys Lys Pro Leu Ile Leu Pro Asp Thr Lys Pro Lys
245 250 255
Ile Lys Asp Asn Gly Asp Leu Val Leu Ser Ser Pro Ser Asn Val Thr
260 265 270
Leu Pro Gln Val Lys Thr Glu Lys Glu Asp Phe Ile Glu Leu Cys Thr
275 280 285
Pro Gly Val Ile Lys Gln Glu Lys Leu Gly Thr Val Tyr Cys Gln Ala
290 295 300

Ser Phe Pro Gly Ala Asn Ile Ile Gly Asn Lys Met Ser Ala Ile Ser
305 310 315 320
Val His Gly Val Ser Thr Ser Gly Gly Gln Met Tyr His Tyr Asp Met
325 330 335
Asn Thr Ala Ser Leu Ser Gln Gln Gln Asp Gln Lys Pro Ile Phe Asn
340 345 350
Val Ile Pro Pro Ile Pro Val Gly Ser Glu Asn Trp Asn Arg Cys Gln
355 360 365

Gly Ser Gly Asp Asp Asn Leu Thr Ser Leu Gly Thr Leu Asn Phe Pro

370 375 380

Gly Arg Thr Val Phe Ser Asn Gly Tyr Ser Ser Pro Ser Met Arg Pro

385 390 395 400

Asp Val Ser Ser Pro Pro Ser Ser Ser Ser Thr Ala Thr Thr Gly Pro

405 410 415

Pro Pro Lys Leu Cys Leu Val Cys Ser Asp Glu Ala Ser Gly Cys His

420 425 430

Tyr Gly Val Leu Thr Cys Gly Ser Cys Lys Val Phe Phe Lys Arg Ala

435 440 445

Val Glu Gly Gln His Asn Tyr Leu Cys Ala Gly Arg Asn Asp Cys Ile

450 455 460

Ile Asp Lys Ile Arg Arg Lys Asn Cys Pro Ala Cys Arg Tyr Arg Lys

465 470 475 480

Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys Lys Lys

485 490 495

Ile Lys Gly Ile Gln Gln Ala Thr Thr Gly Val Ser Gln Glu Thr Ser

500 505 510

Glu Asn Pro Gly Asn Lys Thr Ile Val Pro Ala Thr Leu Pro Gln Leu

515 520 525

Thr Pro Thr Leu Val Ser Leu Leu Glu Val Ile Glu Pro Glu Val Leu

530 535 540

Tyr Ala Gly Tyr Asp Ser Ser Val Pro Asp Ser Thr Trp Arg Ile Met

545 550 555 560

Thr Thr Leu Asn Met Leu Gly Gly Arg Gln Val Ile Ala Ala Val Lys

565 570 575

Trp Ala Lys Ala Ile Pro Gly Phe Arg Asn Leu His Leu Asp Asp Gln

580 585 590

Met Thr Leu Leu Gln Tyr Ser Trp Met Phe Leu Met Ala Phe Ala Leu

595 600 605

Gly Trp Arg Ser Tyr Arg Gln Ser Ser Ala Asn Leu Leu Cys Phe Ala

610 615 620

Pro Asp Leu Ile Ile Asn Glu Gln Arg Met Thr Leu Pro Cys Met Tyr
625 630 635 640
Asp Gln Cys Lys His Met Leu Tyr Val Ser Ser Glu Leu His Arg Leu
645 650 655
Gln Val Ser Tyr Glu Glu Tyr Leu Cys Met Lys Thr Leu Leu Leu Leu
660 665 670
Ser Ser Val Pro Lys Asp Gly Leu Lys Ser Gln Glu Leu Phe Asp Glu
675 680 685

Ile Arg Met Thr Tyr Ile Lys Glu Leu Gly Lys Ala Ile Val Lys Arg
690 695 700
Glu Gly Asn Ser Ser Gln Asn Trp Gln Arg Phe Tyr Gln Leu Thr Lys
705 710 715 720
Leu Leu Asp Ser Met His Glu Val Val Glu Asn Leu Leu Asn Tyr Cys
725 730 735
Phe Gln Thr Phe Leu Asp Lys Thr Met Ser Ile Glu Phe Pro Glu Met
740 745 750

Leu Ala Glu Ile Ile Thr Asn Gln Ile Pro Lys Tyr Ser Asn Gly Asn
755 760 765
Ile Lys Lys Leu Leu Phe His Gln Lys
770 775

<210> 499
<211> 452
<212> RNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 499

cacgguaggc acuguuagga guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuugc cucgguccaa uccuuccgga 120
gugcaggugg aaaccaucuc cccaggagac gggcgcaccu uccccaagcg cggccagacc 180

ugcguggugc acuacaccgg gaugcuugaa gauggaaaga aauuugauuc cucccgggac 240
agaaacaagc ccuuuaaguu uaugcuaggc aagcaggagg ugauccgagg cugggaagaa 300
gggguugccc agaugagugu gggucagaga gccaaacuga cuauaucucc agauuaugcc 360

uauggugcca cugggcaccc aggcaucauc ccaccacaug ccacucucgu cuucgaugug 420

gagcuucuaa aacuggaaua acagugccua cc 452

<210> 500

<211> 464

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 500

ggcccagact gagcacgtga gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctgga ggaagcaggg cttcctttcc 120

tctgccatca ttccagtttt agaagctcca catcgaagac gagagtggca tgtggtggga 180

tgatgcctgg gtgcccagtg gcaccatagg cataatctgg agatatagtc agtttggctc 240

tctgacccac actcatctgg gcaacccctt cttcccagcc tcggatcacc tcctgcttgc 300

ctagcataaa cttaaagggc ttgtttctgt cccgggagga atcaaatttc tttccatctt 360

caagcatccc ggtgtagtgc accacgcagg tctggccgcg cttggggaag gtgcgcccgt 420

ctcctgggga gatggtttcc acctgcactc ccgtgctcag tctg 464

<210> 501

<400> 501

000

<210> 502

<400> 502

000

<210> 503

<400> 503

000

<210> 504

<400> 504

000

<210> 505

<400> 505

000

<210> 506

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 506

ggcccagact gagcacgtga 20

<210> 507

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 507

ggcccagact gagcacgtga 20

<210> 508

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 508

ggcccagact gagcacgtga 20

<210> 509

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 509

ggcccagact gagcacgtga 20

<210> 510

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 510

ggcccagact gagcacgtga 20

<210> 511

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 511

ggcactgcgg ctggaggtgg 20

<210> 512

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 512

gagtccgagc agaagaagaa 20

<210> 513

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 513

ggaatccctt ctgcagcacc 20

<210> 514

<400> 514

000

<210> 515

<400> 515

000

<210> 516

<400> 516

000

<210> 517

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 517

ataacttcgt atagcataca ttatacgaag ttat 34

<210> 518

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 518

taactttaaa taatgccaat tatttaaagt ta 32

<210> 519

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 519

tcaatttctg agaactgtca ttctcggaaa ttga 34

<210> 520

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 520

ctcgtgtccg ataactgtaa ttatcggaca tgat 34

<210> 521

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 521

gaagttccta ttctctagaa agtataggaa cttc 34

<210> 522

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 522

gagtttcatt aaggaataac taattcccta atgaaactc 39

<210> 523

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 523

ggttgcttaa gaataagtaa ttcttaagca acc 33

<210> 524

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 524

acgaaaaatg gtaaggaata gaccattcct taccattttt ggt 43

<210> 525

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 525

ttgatgaaag aataacgtat tctttcatca a 31

<210> 526

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 526

gtgcgtcaaa taataacgta ttatttgaca ctt 33

<210> 527

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 527

aataggtctg agaacgccca ttctcagacg tatt 34

<210> 528

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 528

tgaatgtcct ataattacac ttataggaca ttca 34

<210> 529

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 529

gaaactttaa ataataagtc ttatttaaag tttc 34

<210> 530

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 530

aaacgatatc agacatttgt ctgataatgc ttcattatca gacaaatgtc tgatatcgtt 60

t 61

<210> 531

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 531

atatatacgt atatagacat atatacgtat atat 34

<210> 532

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 532

ataactctat ataatgtatg ctatatagag ttat 34

<210> 533

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223

> Synthetic

<400> 533

acaacatcct attacaccct atatgccaac atgg 34

<210> 534

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 534

aacccactgc ttaagcctca ataaagcttg cctt 34

<210> 535

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 535

gatacaacgt atataccttt ctatacgttg ttta 34

<210> 536

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 536

ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg tacaaacc 48

<210> 537

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 537

ggcttgtcga cgacggcggt ctccgtcgtc aggatcat 38

<210> 538

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 538

gtgccccaac tggggtaacc tttgagttct ctcagttggg gg 42

<210> 539

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 539

tgcgggtgcc agggcgtgcc cttgggctcc ccgggcgcgt actcc 45

<210> 540

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 540

tgttccccaa agcgatacca cttgaagcag tggtactgct tgtgggtaca 50

<210> 541

<211> 64

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 541

gcatgttccc caaagcgata ccacttgaag cagtggtact gcttgtgggt acactctgcg 60

ggtg 64

<210> 542

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 542

ggtgctgggt tgttgtctct ggacagtgat ccatgggaaa ctactcagca cc 52

<210> 543

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 543

caggtttttg acgaaagtga tccagatgat ccag 34

<210> 544

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 544

attttaggta tatgattttg tttattagtg taaataacac tatgtaccta aaat 54

<210> 545

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 545

caaaggatca ctgaatcaaa agtattgctc atccacgcga aa 42

<210> 546

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 546

tttgtgcgga actacgaaca gttcattaat acgaagtgta caaacttcca tacaa 55

<210> 547

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 547

cgaaaatgta tggaggcact tgtatcaata taggatgtat accttcgaag acactt 56

<210> 548

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 548

gagtttttat ttcgtttatt tcaattaagg taactaaaaa actcctttta agg 53

<210> 549

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 549

atgccaacac aattaacatc tcaatcaagg taaatgcttt ttgctttttt tgc 53

<210> 550

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 550

ttcctcgttt tctctcgttg gaagaagaag aaacgagaaa 40

<210> 551

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 551

tttcggatca agctatgaag gacgcaaaga gggaactaaa 40

<210> 552

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 552

ttcctcgttt tctctcgttg gacggaaacg aatcgagaaa 40

<210> 553

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 553

tttcggatca agctatgaag gacgcaaaga gggaactaaa 40

<210> 554

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 554

gtagtgtatc tcacaggtcc acggttggcc gtggactgct gaagaacatt cc 52

<210> 555

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 555

gaaggtgttg gtgcggggtt ggccgtggtc gaggtggggt 40

<210> 556

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 556

aaaaaaatac agcgtttttc atgtacaact atactagttg tagtgcctaa aa 52

<210> 557

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 557

ttgtaaagga gactgataat ggcatgtaca actatactcg tcggtaaaaa ggca 54

<210> 558

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 558

tccagcccaa cagtgttagt ctttgctctt acccagttgg gcggga 46

<210> 559

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 559

gatcagctcc gcgggcaaga cctttctcct tcacggggtg gaaggtc 47

<210> 560

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 560

ctagttttaa agttggttat tagttactgt gatatttatc acggtaccca ataaccaatg 60

aat 63

<210> 561

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 561

cggaaggtag cgtcaacgat aggtgtaact gtcgtgtttg taacggtact tccaacagct 60

ggcgccgcca c 71

<210> 562

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 562

caatgaaaaa ctaggcatgt agaagttgtt tgt 33

<210> 563

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 563

tttccacaga caactcacgt ggaggtagtc ac 32

<210> 564

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 564

tgtagtaagt atcttaatat acagctttat ctgtttttta agatacttac tacttt 56

<210> 565

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 565

cacccattgt gttcacagga tatacagctt tatctgtact gatattaatg acatgctg 58

<210> 566

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 566

agttgtggcc atgtgtccat ctgggggcag atggagacgg ggtcaca 47

<210> 567

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 567

atctggatgt gggtgtccat ctgcgggcag acgccgcagt cgaagcacgg 50

<210> 568

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 568

accttgatct cggtgtccat cgccgggcag acgccgcagt cgaagcacgg 50

<210> 569

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 569

ccctaatacg caagtcgata actctcctgg gagcgttgac aacttgcgca ccctgatctg 60

<210> 570

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 570

gagcgccgga tcagggagtg gacggcctgg gagcgctaca cgctgtggct gcggtcggtg 60

c 61

<210> 571

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 571

gctcatgtat gtgtctacgc gagattctcg cccgagaact tctgcaaggc actgctcttg 60

gct 63

<210> 572

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 572

ggacggcgca gaaggggagt agctcttcgc cggaccgtcg acatactgct cagctcgtc 59

<210> 573

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 573

atggataaaa aaatacagcg tttttcatgt acaactatac tagttgtagt gcctaaataa 60

tgctt 65

<210> 574

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 574

gtttgtaaag gagactgata atggcatgta caactatact cgtcggtaaa aaggcatctt 60

at 62

<210> 575

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 575

aaaaattaca aagttttcaa cccttgattt gaattagcgg tcaaataatt tgtaattcgt 60

tt 62

<210> 576

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 576

ttccaaagag cgcccaacgc gacctgaaat ttgaataaga ctgctgcttg tgtaaaggcg 60

atgatt 66

<210> 577

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 577

gtgttataaa cctgtgtgag agttaagttt acatgcctaa ccttaacttt tacgcaggtt 60

cagctt 66

<210> 578

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 578

agacgagaaa cgttccgtcc gtctgggtca gttgggcaaa gttgatgacc gggtcgtccg 60

tt 62

<210> 579

<211> 64

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 579

ttaataaact atggaagtat gtacagtctt gcaatgttga gtgaacaaac ttccataata 60

aaat 64

<210> 580

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 580

caatcatcag ataactatgg cggcacgtgc attaaccacg gttgtatccc gtctaaagta 60

ctcgt 65

<210> 581

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 581

gtggttgttt ttgttggaag tgtgtatcag gtatctgcat agttattccg aacttccaat 60

ta 62

<210> 582

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 582

tttatattgc gaaaaataat tggcgaacga ggtaactgga tacctcatcc gccaattaaa 60

atttg 65

<210> 583

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 583

ggaaaatata aataatttta gtaacctaca tctcaatcaa ggatagtaaa actctcactc 60

tt 62

<210> 584

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 584

agcacgctga taatcagcaa gaccaccaac atttccacca atgtaaaagc tttaacctta 60

gc 62

<210> 585

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 585

gtttatatgt ttactaataa gacgctctca acccataaag tcttattagt aaacatattt 60

caact 65

<210> 586

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 586

atggattttg cagattccca gatgccccta cagaaagagg tacaaaacat ttattggaat 60

taatt 65

<210> 587

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 587

tttttgtatg ttagttgtgt cactgggtag acctaaatag tgacacaact gctattaaaa 60

tttaa 65

<210> 588

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 588

gttcgtggta actatgggtg gtacaggtgc cacattagtt gtaccattta tgtttatgtg 60

gttaac 66

<210> 589

<211> 64

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 589

caataacggt tgtatttgta gaacttgacc agttgtttta gtaacataaa tacaactccg 60

aata 64

<210> 590

<211> 64

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 590

gcatacattg ttgttgtttt tccagatcca gttggtcctg taaatataag caatccatgt 60

gagt 64

<210> 591

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 591

gtttagtatc tcgttatctc tcgttggagg gagaagaaac gggataccaa aa 52

<210> 592

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 592

taacttttc ggatcgagtt atgatggacg taaagaggga acaaagcatc ta 52

<210> 593

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 593

tagtttccaa tgttacagga actgctggca gaatccaaca cattggaagt cg 52

<210> 594

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 594

cggtctccat cgggatctgc tgatcgagca gcatgccgac ca 42

<210> 595

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 595

taaccgcaag tgtacatccc tcggctggcc gagacaagta cagttgcgac ag 52

<210> 596

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 596

cggtctccat cgggatctgc tgatcgagca gcatgccgac ca 42

<210> 597

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 597

atgtggtcct ttagatccac tgacgtgggt cagtgtctct aaaggactcg cg 52

<210> 598

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 598

catcagggcg gtcaggccgt agatgtggaa gaaacggcag cacggcgagg acg 53

<210> 599

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 599

cgtttcctgt aaaccgaggt tttggataaa ca 32

<210> 600

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 600

cgtttcctgt aaaccgaggt tttggataat gg 32

<210> 601

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 601

gagttctctt aaaccaaggt ttaggattga aa 32

<210> 602

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 602

gagttctctt aaaccaaggt attggataac ag 32

<210> 603

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 603

tggttcttga aaaccaaggt ttttgataaa gc 32

<210> 604

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 604

aaattttcct ttttggaagg tttttgataa cca 33

<210> 605

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 605

gccttcccct aaaccaacgt ttttatgccg cc 32

<210> 606

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 606

gccttccccc aaaccaaggt aatcaagaac gc 32

<210> 607

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 607

ttgtgaaatt tgggtacacc ctaatcatac aa 32

<210> 608

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 608

cgtatgatta gggtgtatat taattt 26

<210> 609

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 609

caaatgattt tattttgact aataatgacc tacttacatt aatttactga taattaaaga 60

gattttaaat atacaactta ttcacctaaa ggatgacaaa a 101

<210> 610

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 610

gcactttagg tgaaaaaggt t 21

<210> 611

<211> 126

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 611

taacattaat cacttaaaaa tcatcgcatt acactaatct gtggttaaat gatagactac 60

ataatgcgac aaaacgcaac atatccagtc actatgaatc aactacttag atagtattag 120

tgacct 126

<210> 612

<211> 259

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 612

ctaagtggtt tgggacaaaa atgggacata caaatctttg catcggtttg caaggctttg 60

catgtctttc gaagatggga cgtgtgagcg caggtatgac gtggtatgtg ttgacttaaa 120

aggtagttct tataattcgt aatgcgaagg tcgtaggttc gactcctatt atcggcacca 180

gttaaatcaa atacttacgt attattcgtg ccttccttat ttttactgtg ggacatattt 240

gggacagaag taccaaaaa 259

<210> 613

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 613

gcagcgcatt cgtaatgcga aggtcgt 27

<210> 614

<211> 412

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 614

gcgatcccca tccgcgacgt gccaactagg tctcctctcg tcgtgaacaa ggctaccggg 60

ttgcaactcc tgtgcaactc tcaggcttca acgcgcttct acgacctgca atttctttcc 120

acttagagga tgcagccgag aggggtaaaa acctatcttg accggcccat atgtggtcgg 180

cagacaccca ttcttccaaa ctagctacgc gggttcgatt cccgtcgccc gctccgctgg 240

tcagagggtg ttttcgccct ctggccattt ttctttccag gggtctgcaa ctcttgtgcg 300

actcttctga cctgggcata cgcggttgca acgcatccct gatctggcta ctttcgatgc 360

tgacaaacga atagagcccc ccgcctgcgc gaacagacga ggggcattca ca 412

<210> 615

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 615

gagcgggcga cgggaatcga acccgcgtag ctagtttgga aga 43

<210> 616

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 616

gtcaaccagt atcccggtgc 20

<210> 617

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 617

ggggtcccag gtgctgacgt 20

<210> 618

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 618

gatgtacaga gagcccaggg c 21

<210> 619

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 619

gtcaaccagt atcccggtgc 20

<210> 620

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 620

gatgtacaga gagcccaggg c 21

<210> 621

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 621

ggggtcccag gtgctgacgt 20

<210> 622

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 622

Phe Cys Pro Phe

1

<210> 623

<400> 623

000

<210> 624

<400> 624

000

<210> 625

<211> 209

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 625

ttcattcagc aagttcagag aaatctgaac ttgctggatt tttggagcag ggagatggaa 60

taggagcttg ctccgtccac tccacgcatc gacctggtat tgcagtacct ccaggaacgg 120

tgcacccact ttctggagtt tcaaaagtag actgtacgct aagggtcata tctttttttg 180

tttggtttgt gtcttggttg gcgtcttaa 209

<210> 626

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 626

gccttgatac caacctgccc a 21

<210> 627

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 627

gctggaactg gtcaccaagg c 21

<210> 628

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 628

gtacctgaac cgtatatcct a 21

<210> 629

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 629

gtacctgaac cgtatatctt a 21

<210> 630

<211> 535

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 630

Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn

1 5 10 15

Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln

20 25 30

Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp Asp
35 40 45
Asp Trp Lys Glu Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly
50 55 60

Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val
65 70 75 80
Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val
85 90 95
Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu
100 105 110
Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg Phe Gly
115 120 125

Asp Gly Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser
130 135 140
Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser
145 150 155 160
Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
165 170 175
Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg
180 185 190

Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp Asp Val
195 200 205
Ile Arg Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly
210 215 220
Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr Val Ser Glu
225 230 235 240
Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
245 250 255

His Pro Glu Leu Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val
260 265 270
Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala Gln Val

275 280 285

Ile Asp Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu

290 295 300

Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile Ala Asp Gly Ala

305 310 315 320

Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser

325 330 335

Ser Leu Met Val Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp

340 345 350

Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe

355 360 365

Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His

370 375 380

Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val Ser Trp Asn Thr

385 390 395 400

Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His

405 410 415

Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val

420 425 430

Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr

435 440 445

His Ile Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile

450 455 460

Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly

465 470 475 480

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser

485 490 495

Glu Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu

500 505 510

Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser

515 520 525

Leu Phe Phe Glu Ile Lys Ser

530 535

<210> 631

<400> 631

000

<210> 632

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 632

ggcccagact gagcacgtga 20

<210> 633

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 633

ggcccagact gagcacgtga 20

<210> 634

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 634

ggcccagact gagcacgtga 20

<210> 635

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 635

ggcccagact gagcacgtga 20

<210> 636

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 636

ggcccagact gagcacgtga 20

<210> 637

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 637

gtcatcttag tcattacctg 20

<210> 638

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 638

gtcatcttag tcattacctg 20

<210> 639

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 639

gtcatcttag tcattacctg 20

<210> 640

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 640

gcatggtgca cctgactcct g 21

<210> 641

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 641

tctgccatcc cgtgctcagt ctg 23

<210> 642

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 642

tctgccagca cgtgctcagt ctg 23

<210> 643

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 643

tctgccaaca cgtgctcagt ctg 23

<210> 644

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 644

tggaggaagc agggcttcct ttcctctgaa aacacgtgct cagtctg 47

<210> 645

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 645

tctgccatca aagcgtgctc agtctg 26

<210> 646

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 646

aacgaacacc tcatgtaatg actaagatg 29

<210> 647

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 647

aacgaacacc tcacgtaatg actaagatg 29

<210> 648

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 648

aacgaacacc tcagtacgta atgactaaga tg 32

<210> 649

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 649

agacttctcc acaggagtca ggtgcac 27

<210> 650

<400> 650

000

<210> 651

<400> 651

000

<210> 652

<400> 652

000

<210> 653

<400> 653

000

<210> 654

<400> 654

000

<210> 655

<400> 655

000

<210> 656

<400> 656

000

<210> 657

<400> 657

000

<210> 658

<400> 658

000

<210> 659

<400> 659

000

<210> 660

<400> 660

000

<210> 661

<400> 661

000

<210> 662

<400> 662

000

<210> 663

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 663

ggcccagact gagcacgtga 20

<210> 664

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 664

ggcccagact gagcacgtga 20

<210> 665

<211> 20

<212> DNA

<213> Artificial Sequence

<

220><223> Synthetic

<400> 665

ggcccagact gagcacgtga 20

<210> 666

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 666

ggcccagact gagcacgtga 20

<210> 667

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 667

ggcactgcgg ctggaggtgg 20

<210> 668

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 668

ggcactgcgg ctggaggtgg 20

<210> 669

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 669

ggcactgcgg ctggaggtgg 20

<210> 670

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 670

ggcactgcgg ctggaggtgg 20

<210> 671

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 671

gagtccgagc agaagaagaa 20

<210> 672

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 672

gagtccgagc agaagaagaa 20

<210> 673

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 673

gagtccgagc agaagaagaa 20

<210> 674

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 674

gagtccgagc agaagaagaa 20

<210> 675

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 675

ggaatccctt ctgcagcacc 20

<210> 676

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 676

ggaatccctt ctgcagcacc 20

<210> 677

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 677

ggaatccctt ctgcagcacc 20

<210> 678

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 678

tctgccatct cgtgctcagt ctg 23

<210> 679

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 679

tctgccatca aagcgtgctc agtctg 26

<210> 680

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 680

tctgccatcc gtgctcagtc tg 22

<210> 681

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 681

tctgcgatca cgtgctcagt ctg 23

<210> 682

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 682

ttaacgccca cctccagcc 19

<210> 683

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 683

ttaacccccc cctccagcc 19

<210> 684

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 684

ttaacccctt acacctccag cc 22

<210> 685

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 685

ttaacccccc ctccagcc 18

<210> 686

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 686

gtgatgggag cacttcttct tctgctcgga 30

<210> 687

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 687

gtgatgggag ccctgcttct tctgctcgga 30

<210> 688

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 688

gtgatgggag cccttcgcat tcttctgctc gga 33

<210> 689

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 689

gtgatgggag ttcttcttct gctcgga 27

<210> 690

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 690

ggaaaagcga tgcaggtgct gcagaaggga t 31

<210> 691

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 691

ggaaaagcga tccaggcgct gcagaaggga t 31

<210> 692

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 692

ggaaaagcga tccaatcggt gctgcagaag ggat 34

<210> 693

<400> 693

000

<210> 694

<400> 694

000

<210> 695

<400> 695

000

<210> 696

<400> 696

000

<210> 697

<400> 697

000

<210> 698

<400> 698

000

<210> 699

<400> 699

000

<210> 700

<211> 677

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<400> 700

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp
325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 701
<211> 677
<212> PRT
<213> Artificial Sequence

<220><223> Synthetic

<400> 701

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675
<210> 702
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 702

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670

Glu Asn Ser Ser Pro
675

<210> 703

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 703

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Lys Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670

Glu Asn Ser Ser Pro
675

<210> 704

<211> 677

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 704

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Arg Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 705
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 705
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Lys Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 706
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 706
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Pro Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670

Glu Asn Ser Ser Pro
675

<210> 707
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 707
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Gly Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 708
<211> 677
<212> PRT
<213> Artificial Sequence
<220
><223> Synthetic
<400> 708
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Lys Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 709
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 709
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Lys Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 710
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 710
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670

Glu Asn Ser Ser Pro
675

<210> 711
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 711

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Gly Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Gln Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asn Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 712

<211> 677

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 712

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

```
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
                165                 170                 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
            180                 185                 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
        195                 200                 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
    210                 215                 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225                 230                 235                 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
                245                 250                 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
            260                 265                 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
        275                 280                 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Arg Phe Leu
    290                 295                 300
Gly Thr Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305                 310                 315                 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
                325                 330                 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
            340                 345                 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
        355                 360                 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
    370                 375                 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385                 390                 395                 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
```

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670

Glu Asn Ser Ser Pro
675

<210> 713

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 713

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Pro Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Lys Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 714

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 714

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Leu Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Lys Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Ala Leu Phe Asp Glu Ala Leu Arg Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Lys Phe Leu
290 295 300

Gly Thr Ala Gly Asn Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430

Val Ile Gly Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Lys Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Gly Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asn Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile Gln Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asn Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 715

<211> 677

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 715

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Leu Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Lys Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Ala Leu Phe Asn Glu Ala Leu Arg Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Lys Phe Leu
290 295 300

Gly Thr Ala Gly Asn Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430

Val Ile Gly Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Lys Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Gly Gly Ser Ser Leu
515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575

Lys Leu Asn Val Tyr Thr Asn Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590

Ile Gln Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asn Gln Ala Ala
645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670

Glu Asn Ser Ser Pro
675

<210> 716

<211> 677

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 716

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540

Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670

Glu Asn Ser Ser Pro

675

<210> 717

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 717

gcttctccag ccctggcctg ggtcaatcct tggggcccag actgagcacg tgttggcaga 60

ggaaaggaag ccctgcttcc tccagagggc gtcgcaggac 100

<210> 718

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 718

gcttctccag ccctggcctg ggtcaatcct tggggcccag actgagcacg tgttttcaga 60

ggaaaggaag ccctgcttcc tccagagggc gtcgcaggac 100

<210> 719

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 719

gcttctccag ccctggcctg ggtcaatcct tggggcccag actgagcacg ggatggcaga 60

ggaaaggaag ccctgcttcc tccagagggc gtcgcaggac 100

<210> 720

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 720

gcttctccag ccctggcctg ggtcaatcct tggggcccag actgagcacg tgctggcaga 60

ggaaaggaag ccctgcttcc tccagagggc gtcgcaggac 100

<210> 721

<211> 103

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 721

gcttctccag ccctggcctg ggtcaatcct tggggcccag actgagcacg ctttgatggc 60

agaggaaagg aagccctgct tcctccagag ggcgtcgcag gac 103

<210> 722

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 722

cccagtttac acgtctcata tgccccttgg cagtcatctt agtcattaca tgaggtgttc 60

gttgtaactc atataaactg agttcccatg ttttgcttaa 100

<210> 723

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 723

cccagtttac acgtctcata tgccccttgg cagtcatctt agtcattacg tgaggtgttc 60

gttgtaactc atataaactg agttcccatg ttttgcttaa 100

<210> 724

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 724

cagtttacac gtctcatatg ccccttggca gtcatcttag tcattacgta ctgaggtgtt 60

cgttgtaact catataaact gagttcccat gttttgctta 100

<210> 725

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 725

acttcatcca cgttcacctt gccccacagg gcagtaacgg cagacttctc cacaggagtc 60

agatgcacca tggtgtctgt ttgaggttgc tagtgaacac 100

<210> 726

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 726

acttcatcca cgttcacctt gccccacagg gcagtaacgg cagacttctc ctcaggagtc 60

aggtgcacca tggtgtctgt ttgaggttgc tagtgaacac 100

<210> 727

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 727

gtgttcacta gcaacctcaa acagacacca tggtgcacct gactcctgag gagaagtctg 60

ccgttactgc cctgtggggc aaggtgaacg tggatgaagt 100

<210> 728

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 728

gtgttcacta gcaacctcaa acagacacca tggtgcacct gactcctgat gagaagtctg 60

ccgttactgc cctgtggggc aaggtgaacg tggatgaagt 100

<210> 729

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 729

cacatggctg gtgctgcagc agctggggca gtggtggggg gccttggcgt ctacatgctg 60

ggaagtgcca tgagcaggcc catcatacat ttcggcagtg 100

<210> 730

<211> 9224

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 730

atgtctaaag gtgaagaatt attcactggt gttgtcccaa ttttggttga attagatggt 60

gatgttaatg gtcacaaatt ttctgtctcc ggtgaaggtg aaggtgatgc tacttacggt 120

aaattgacct taaaatttat ttgtactact ggtaaattgc cagttccatg gccaacctta 180

gtcactactt tcggttatgg tgttcaatgt tttgctagat acccagatca tatgaaacaa 240

catgactttt tcaagtctgc catgccagaa ggttatgttc aagaaagaac tatttttttc 300

aaagatgacg gtaactacaa gaccagagct gaagtcaagt ttgaaggtga taccttagtt 360

aatagaatcg aattaaaagg tattgatttt aaagaagatg gtaacatttt aggtcacaaa 420

ttggaataca actataactc tcacaatgtt tacatcatgg ctgacaaaca aaagaatggt 480

atcaaagtta acttcaaaat tagacacaac attgaagatg gttctgttca attagctgac 540

cattatcaac aaaatactcc aattggtgat ggtccagtct tgttaccaga caaccattac 600

ttatccactc aatctgcctt atccaaagat ccaaacgaaa agagagacca catggtcttg 660

ttagaatttg ttactgctgc tggtattacc catggtatgg atgaattgta caaagctagc 720

aacctgggtc aatccttggg gcccagactg agcacgtgat ggcagagcac aggagacgtc 780

atggtttcaa aaggtgaaga agataatatg gctattatta aagaatttat gagatttaaa 840

gttcatatgg aaggttcagt taatggtcat gaatttgaaa ttgaaggtga aggtgaaggt 900

agaccatatg aaggtactca aactgctaaa ttgaaagtta ctaaaggtgg tccattacca 960

tttgcttggg atattttgtc accacaattt atgtatggtt caaaagctta tgttaaacat 1020

ccagctgata ttccagatta tttaaaattg tcatttccag aaggttttaa atgggaaaga 1080

gttatgaatt ttgaagatgg tggtgttgtt actgttactc aagattcatc attacaagat 1140

ggtgaattta tttataaagt taaattgaga ggtactaatt ttccatcaga tggtccagtt 1200

atgcaaaaaa aaactatggg ttgggaagct tcatcagaaa gaatgtatcc agaagatggt 1260

gctttaaaag gtgaaattaa acaaagattg aaattaaaag atggtggtca ttatgatgct 1320

gaagttaaaa ctacttataa agctaaaaaa ccagttcaat taccaggtgc ttataatgtt 1380

aatattaaat tggatattac ttcacataat gaagattata ctattgttga acaatatgaa 1440

agagctgaag gtagacattc aactggtggt atggatgaat tatataaagg taccgctcga 1500

gcagctgtga ttgattgagt cgacttggtt gaacacgttg ccaaggctta agtgaattta 1560

ctttaaatct tgcatttaaa taaattttct ttttatagct ttatgactta gtttcaattt 1620

atatactatt ttaatgacat tttcgattcg gatccactag ttctagagcg gccgccaccg 1680

cggtggagct ccagcttttg ttccctttag tgagggttaa ttgcgcgctt ggcgtaatca 1740

tggtcatagc tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatagga 1800

gccggaagca taaagtgtaa agcctggggt gcctaatgag tgaggtaact cacattaatt 1860

gcgttgcgct cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga 1920

atcggccaac gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc 1980

actgactcgc tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg 2040

gtaatacggt tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc 2100

cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc 2160

ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga 2220

ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc 2280

ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat 2340

agctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg 2400

cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc 2460

aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 2520

gcgaggtatg taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact 2580

agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt 2640

ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag 2700

cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg 2760

tctgacgctc agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa 2820

aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata 2880

tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg 2940

atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata 3000

cgggagggct taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg 3060

gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct 3120

gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt 3180

tcgccagtta atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc 3240

tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga 3300

tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt 3360
aagttggccg cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc 3420
atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa 3480
tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca 3540
catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca 3600
aggatcttac cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct 3660
tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc 3720

gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa 3780
tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt 3840
tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgaacga 3900
agcatctgtg cttcattttg tagaacaaaa atgcaacgcg agagcgctaa tttttcaaac 3960
aaagaatctg agctgcattt ttacagaaca gaaatgcaac gcgaaagcgc tattttacca 4020
acgaagaatc tgtgcttcat ttttgtaaaa caaaaatgca acgcgagagc gctaattttt 4080
caaacaaaga atctgagctg cattttttaca gaacagaaat gcaacgcgag agcgctattt 4140

taccaacaaa gaatctatac ttctttttttg ttctacaaaa atgcatcccg agagcgctat 4200
ttttctaaca aagcatctta gattactttt tttctccttt gtgcgctcta taatgcagtc 4260
tcttgataac tttttgcact gtaggtccgt taaggttaga agaaggctac tttggtgtct 4320
attttctctt ccataaaaaa agcctgactc cacttcccgc gtttactgat tactagcgaa 4380
gctgcgggtg catttttttca agataaaggc atccccgatt atattctata ccgatgtgga 4440
ttgcgcatac tttgtgaaca gaaagtgata gcgttgatga ttcttcattg gtcagaaaat 4500
tatgaacggt ttcttctatt ttgtctctat atactacgta taggaaatgt ttacattttc 4560

gtattgtttt cgattcactc tatgaatagt tcttactaca atttttttgt ctaaagagta 4620
atactagaga taaacataaa aaatgtagag gtcgagttta gatgcaagtt caaggagcga 4680
aaggtggatg ggtaggttat atagggatat agcacagaga tatatagcaa agagatactt 4740
ttgagcaatg tttgtggaag cggtattcgc aatattttag tagctcgtta cagtccggtg 4800
cgtttttggt tttttgaaag tgcgtcttca gagcgctttt ggttttcaaa agcgctctga 4860
agttcctata ctttctagag aataggaact tcggaatagg aacttcaaag cgtttccgaa 4920
aacgagcgct tccgaaaatg caacgcgagc tgcgcacata cagctcactg ttcacgtcgc 4980

acctatatct gcgtgttgcc tgtatatata tatacatgag aagaacggca tagtgcgtgt 5040

ttatgcttaa atgcgtactt atatgcgtct atttatgtag gatgaaaggt agtctagtac 5100

ctcctgtgat attatcccat tccatgcggg gtatcgtatg cttccttcag cactaccctt 5160

tagctgttct atatgctgcc actcctcaat tggattagtc tcatccttca atgctatcat 5220

ttcctttgat attggatcat actaagaaac cattattatc atgacattaa cctataaaaa 5280

taggcgtatc acgaggccct ttcgtctcgc gcgtttcggt gatgacggtg aaaacctctg 5340

acacatgcag ctcccggaga cggtcacagc ttgtctgtaa gcggatgccg ggagcagaca 5400

agcccgtcag ggcgcgtcag cgggtgttgg cgggtgtcgg ggctggctta actatgcggc 5460

atcagagcag attgtactga gagtgcacca tatcgactac gtcgtaaggc cgtttctgac 5520

agagtaaaat tcttgaggga actttcacca ttatgggaaa tgcttcaaga aggtattgac 5580

ttaaactcca tcaaatggtc aggtcattga gtgtttttta tttgttgtat ttttttttt 5640

ttagagaaaa tcctccaata tcaaattagg aatcgtagtt tcatgatttt ctgttacacc 5700

taacttttg tgtggtgccc tcctccttgt caatattaat gttaaagtgc aattcttttt 5760

ccttatcacg ttgagccatt agtatcaatt tgcttacctg tattccttta ctatcctcct 5820

ttttctcctt cttgataaat gtatgtagat tgcgtatata gtttcgtcta ccctatgaac 5880

atattccatt ttgtaatttc gtgtcgtttc tattatgaat ttcatttata aagtttatgt 5940

acaaatatca taaaaaaaga gaatcttttt aagcaaggat tttcttaact tcttcggcga 6000

cagcatcacc gacttcggtg gtactgttgg aaccacctaa atcaccagtt ctgatacctg 6060

catccaaaac ctttttaact gcatcttcaa tggccttacc ttcttcaggc aagttcaatg 6120

acaatttcaa catcattgca gcagacaaga tagtggcgat agggtcaacc ttattctttg 6180

gcaaatctgg agcagaaccg tggcatggtt cgtacaaacc aaatgcggtg ttcttgtctg 6240

gcaaagaggc caaggacgca gatggcaaca aacccaagga acctgggata acggaggctt 6300

catcggagat gatatcacca aacatgttgc tggtgattat aataccattt aggtgggttg 6360

ggttcttaac taggatcatg gcggcagaat caatcaattg atgttgaacc ttcaatgtag 6420

ggaattcgtt cttgatggtt tcctccacag tttttctcca taatcttgaa gaggccaaaa 6480

gattagcttt atccaaggac caaataggca atggtggctc atgttgtagg gccatgaaag 6540

cggccattct tgtgattctt tgcacttctg gaacggtgta ttgttcacta tcccaagcga 6600

caccatcacc atcgtcttcc tttctcttac caaagtaaat acctcccact aattctctga 6660

caacaacgaa gtcagtacct ttagcaaatt gtggcttgat tggagataag tctaaaagag 6720

agtcggatgc aaagttacat ggtcttaagt tggcgtacaa ttgaagttct ttacggattt 6780

ttagtaaacc ttgttcaggt ctaacactac cggtacccca tttaggacca gccacagcac 6840

ctaacaaaac ggcatcaacc ttcttggagg cttccagcgc ctcatctgga agtgggacac 6900

ctgtagcatc gatagcagca ccaccaatta aatgattttc gaaatcgaac ttgacattgg 6960

aacgaacatc agaaatagct ttaagaacct taatggcttc ggctgtgatt tcttgaccaa 7020

cgtggtcacc tggcaaaacg acgatcttct taggggcaga catagggca gacattagaa 7080

tggtatatcc ttgaaatata tatatatatt gctgaaatgt aaaaggtaag aaaagttaga 7140

aagtaagacg attgctaacc acctattgga aaaaacaata ggtccttaaa taatattgtc 7200

aacttcaagt attgtgatgc aagcatttag tcatgaacgc ttctctattc tatatgaaaa 7260

gccggttccg gcctctcacc tttccttttt ctcccaattt ttcagttgaa aaaggtatat 7320

gcgtcaggcg acctctgaaa ttaacaaaaa atttccagtc atcgaatttg attctgtgcg 7380

atagcgcccc tgtgtgttct cgttatgttg aggaaaaaaa taatggttgc taagagattc 7440

gaactcttgc atcttacgat acctgagtat tcccacagtt aactgcggtc aagatatttc 7500

ttgaatcagg cgccttagac cgctcggcca aacaaccaat tacttgttga gaaatagagt 7560

ataattatcc tataaatata acgtttttga acacacatga acaaggaagt acaggacaat 7620

tgattttgaa gagaatgtgg attttgatgt aattgttggg attccatttt taataaggca 7680

ataatattag gtatgtggat atactagaag ttctcctcga ccgtcgatat gcggtgtgaa 7740

ataccgcaca gatgcgtaag gagaaaatac cgcatcagga aattgtaaac gttaatattt 7800

tgttaaaatt cgcgttaaat ttttgttaaa tcagctcatt ttttaaccaa taggccgaaa 7860

tcggcaaaat cccttataaa tcaaaagaat agaccgagat agggttgagt gttgttccag 7920

tttggaacaa gagtccacta ttaaagaacg tggactccaa cgtcaaaggg cgaaaaaccg 7980

tctatcaggg cgatggccca ctacgtgaac catcacccta atcaagtttt ttggggtcga 8040

ggtgccgtaa agcactaaat cggaacccta aagggagccc ccgatttaga gcttgacggg 8100

gaaagccggc gaacgtggcg agaaaggaag ggaagaaagc gaaaggagcg ggcgctaggg 8160

cgctggcaag tgtagcggtc acgctgcgcg taaccaccac acccgccgcg cttaatgcgc 8220

cgctacaggg cgcgtcgcgc cattcgccat tcaggctgcg caactgttgg gaagggcgat 8280

cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat 8340

taagttgggt aacgccaggg ttttcccagt cacgacgttg taaaacgacg gccagtgagc 8400

gcgcgtaata cgactcacta tagggcgaat tgggtaccgg gccccccctc gaggtcgacg 8460

gtatcgataa gcttgatatc gaattcctgc agcccggggg atccgttaga atcattttga 8520

ataaaaaaca cgctttttca gttcgagttt atcattatca atactgccat ttcaaagaat 8580

acgtaaataa ttaatagtag tgattttcct aactttattt agtcaaaaaa ttagcctttt 8640

aattctgctg taacccgtac atgcccaaaa taggggggcgg gttacacaga atatataaca 8700

tcgtaggtgt ctgggtgaac agtttattcc tggcatccac taaatataat ggagcccgct 8760

ttttaagctg gcatccagaa aaaaaaagaa tcccagcacc aaaatattgt tttcttcacc 8820

aaccatcagt tcataggtcc attctcttag cgcaactaca gagaacaggg gcacaaacag 8880

gcaaaaaacg ggcacaacct caatggagtg atgcaacctg cctggagtaa atgatgacac 8940

aaggcaattg acccacgcat gtatctatct cattttctta caccttctat taccttctgc 9000

tctctctgat ttggaaaaag ctgaaaaaaa aggttgaaac cagttccctg aaattattcc 9060

cctacttgac taataagtat ataaagacgg taggtattga ttgtaattct gtaaatctat 9120

ttcttaaact tcttaaattc tacttttata gttagtcttt tttttagttt taaaacacca 9180

agaacttagt ttcgaataaa cacacataaa caaacaaaga attc 9224

<210> 731

<211> 9225

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 731

atgtctaaag gtgaagaatt attcactggt gttgtcccaa ttttggttga attagatggt 60

gatgttaatg gtcacaaatt ttctgtctcc ggtgaaggtg aaggtgatgc tacttacggt 120

aaattgacct taaaatttat ttgtactact ggtaaattgc cagttccatg gccaacctta 180

gtcactactt tcggttatgg tgttcaatgt tttgctagat acccagatca tatgaaacaa 240

catgactttt tcaagtctgc catgccagaa ggttatgttc aagaaagaac tatttttttc 300

aaagatgacg gtaactacaa gaccagagct gaagtcaagt ttgaaggtga taccttagtt 360

aatagaatcg aattaaaagg tattgatttt aaagaagatg gtaacatttt aggtcacaaa 420

ttggaataca actataactc tcacaatgtt tacatcatgg ctgacaaaca aaagaatggt 480

atcaaagtta acttcaaaat tagacacaac attgaagatg gttctgttca attagctgac 540

cattatcaac aaaatactcc aattggtgat ggtccagtct tgttaccaga caaccattac 600

ttatccactc aatctgcctt atccaaagat ccaaacgaaa agagagacca catggtcttg 660

ttagaatttg ttactgctgc tggtattacc catggtatgg atgaattgta caaagctagc 720

aaacctgggt caatccttgg ggcccagact gagcacgtga tggcagagca caggagacgt 780

catggtttca aaaggtgaag aagataatat ggctattatt aaagaattta tgagatttaa 840

agttcatatg gaaggttcag ttaatggtca tgaatttgaa attgaaggtg aaggtgaagg 900

tagaccatat gaaggtactc aaactgctaa attgaaagtt actaaaggtg gtccattacc 960

atttgcttgg gatattttgt caccacaatt tatgtatggt tcaaaagctt atgttaaaca 1020

tccagctgat attccagatt atttaaaatt gtcatttcca gaaggtttta aatgggaaag 1080

agttatgaat tttgaagatg gtggtgttgt tactgttact caagattcat cattacaaga 1140

tggtgaattt atttataaag ttaaattgag aggtactaat tttccatcag atggtccagt 1200

tatgcaaaaa aaaactatgg gttgggaagc ttcatcagaa agaatgtatc cagaagatgg 1260

tgctttaaaa ggtgaaatta aacaaagatt gaaattaaaa gatggtggtc attatgatgc 1320

tgaagttaaa actacttata aagctaaaaa accagttcaa ttaccaggtg cttataatgt 1380

taatattaaa ttggatatta cttcacataa tgaagattat actattgttg aacaatatga 1440

aagagctgaa ggtagacatt caactggtgg tatggatgaa ttatataaag gtaccgctcg 1500

agcagctgtg attgattgag tcgacttggt tgaacacgtt gccaaggctt aagtgaattt 1560

actttaaatc ttgcatttaa ataaattttc tttttatagc tttatgactt agtttcaatt 1620

tatatactat tttaatgaca ttttcgattc ggatccacta gttctagagc ggccgccacc 1680

gcggtggagc tccagctttt gttcccttta gtgagggtta attgcgcgct tggcgtaatc 1740

atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatagg 1800

agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgaggtaac tcacattaat 1860

tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg 1920

aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg cttcctcgct 1980

cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc 2040

ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg 2100

ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg 2160

ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg 2220

actataaaga taccaggcgt ttcccccctgg aagctccctc gtgcgctctc ctgttccgac 2280

cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca 2340

tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt 2400

gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc 2460

caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag 2520

agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac 2580

tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg gaaaaagagt 2640

tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa 2700

gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg 2760

gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga gattatcaaa 2820

aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat 2880
atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac ctatctcagc 2940
gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga taactacgat 3000
acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgctcacc 3060
ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtggtcc 3120
tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag 3180
ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg tggtgtcacg 3240

ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg 3300
atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg ttgtcagaag 3360
taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt 3420
catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt cattctgaga 3480
atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata ataccgcgcc 3540
acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc 3600
aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac ccaactgatc 3660

ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc 3720
cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct tccttttca 3780
atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat 3840
ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc cacctgaacg 3900
aagcatctgt gcttcatttt gtagaacaaa aatgcaacgc gagagcgcta atttttcaaa 3960
caaagaatct gagctgcatt tttacagaac agaaatgcaa cgcgaaagcg ctattttacc 4020
aacgaagaat ctgtgcttca tttttgtaaa acaaaaatgc aacgcgagag cgctaatttt 4080

tcaaacaaag aatctgagct gcatttttac agaacagaaa tgcaacgcga gagcgctatt 4140
ttaccaacaa agaatctata cttctttttt gttctacaaa aatgcatccc gagagcgcta 4200
tttttctaac aaagcatctt agattacttt ttttctcctt tgtgcgctct ataatgcagt 4260
ctcttgataa ctttttgcac tgtaggtccg ttaaggttag aagaaggcta ctttggtgtc 4320
tattttctct tccataaaaa aagcctgact ccacttcccg cgtttactga ttactagcga 4380
agctgcgggt gcatttttc aagataaagg catccccgat tatattctat accgatgtgg 4440
attgcgcata ctttgtgaac agaaagtgat agcgttgatg attcttcatt ggtcagaaaa 4500

ttatgaacgg tttcttctat tttgtctcta tatactacgt ataggaaatg tttacatttt 4560

cgtattgttt tcgattcact ctatgaatag ttcttactac aatttttttg tctaaagagt 4620

aatactagag ataaacataa aaaatgtaga ggtcgagttt agatgcaagt tcaaggagcg 4680

aaaggtggat gggtaggtta tatagggata tagcacagag atatatagca aagagatact 4740

tttgagcaat gtttgtggaa gcggtattcg caatatttta gtagctcgtt acagtccggt 4800

gcgtttttgg ttttttgaaa gtgcgtcttc agagcgcttt tggttttcaa aagcgctctg 4860

aagttcctat actttctaga gaataggaac ttcggaatag gaacttcaaa gcgtttccga 4920

aaacgagcgc ttccgaaaat gcaacgcgag ctgcgcacat acagctcact gttcacgtcg 4980

cacctatatc tgcgtgttgc ctgtatatat atatacatga gaagaacggc atagtgcgtg 5040

tttatgctta aatgcgtact tatatgcgtc tatttatgta ggatgaaagg tagtctagta 5100

cctcctgtga tattatccca ttccatgcgg ggtatcgtat gcttccttca gcactaccct 5160

ttagctgttc tatatgctgc cactcctcaa ttggattagt ctcatccttc aatgctatca 5220

tttcctttga tattggatca tactaagaaa ccattattat catgacatta acctataaaa 5280

ataggcgtat cacgaggccc tttcgtctcg cgcgtttcgg tgatgacggt gaaaacctct 5340

gacacatgca gctcccggag acggtcacag cttgtctgta agcggatgcc gggagcagac 5400

aagcccgtca gggcgcgtca gcgggtgttg gcgggtgtcg gggctggctt aactatgcgg 5460

catcagagca gattgtactg agagtgcacc atatcgacta cgtcgtaagg ccgtttctga 5520

cagagtaaaa ttcttgaggg aactttcacc attatgggaa atgcttcaag aaggtattga 5580

cttaaactcc atcaaatggt caggtcattg agtgtttttt atttgttgta ttttttttt 5640

tttagagaaa atcctccaat atcaaattag gaatcgtagt ttcatgattt tctgttacac 5700

ctaacttttt gtgtggtgcc ctcctccttg tcaatattaa tgttaaagtg caattctttt 5760

tccttatcac gttgagccat tagtatcaat ttgcttacct gtattccttt actatcctcc 5820

tttttctcct tcttgataaa tgtatgtaga ttgcgtatat agtttcgtct accctatgaa 5880

catattccat tttgtaattt cgtgtcgttt ctattatgaa tttcatttat aaagtttatg 5940

tacaaatatc ataaaaaaag agaatctttt taagcaagga ttttcttaac ttcttcggcg 6000

acagcatcac cgacttcggt ggtactgttg gaaccaccta aatcaccagt tctgatacct 6060

gcatccaaaa ccttttaac tgcatcttca atggccttac cttcttcagg caagttcaat 6120

gacaatttca acatcattgc agcagacaag atagtggcga tagggtcaac cttattcttt 6180

ggcaaatctg gagcagaacc gtggcatggt tcgtacaaac caaatgcggt gttcttgtct 6240

ggcaaagagg ccaaggacgc agatggcaac aaacccaagg aacctgggat aacggaggct 6300

tcatcggaga tgatatcacc aaacatgttg ctggtgatta taataccatt taggtgggtt 6360

gggttcttaa ctaggatcat ggcggcagaa tcaatcaatt gatgttgaac cttcaatgta 6420

gggaattcgt tcttgatggt ttcctccaca gtttttctcc ataatcttga agaggccaaa 6480

agattagctt tatccaagga ccaaataggc aatggtggct catgttgtag ggccatgaaa 6540

gcggccattc ttgtgattct ttgcacttct ggaacggtgt attgttcact atcccaagcg 6600

acaccatcac catcgtcttc ctttctctta ccaaagtaaa tacctcccac taattctctg 6660

acaacaacga agtcagtacc tttagcaaat tgtggcttga ttggagataa gtctaaaaga 6720

gagtcggatg caaagttaca tggtcttaag ttggcgtaca attgaagttc tttacggatt 6780

tttagtaaac cttgttcagg tctaacacta ccggtacccc atttaggacc agccacagca 6840

cctaacaaaa cggcatcaac cttcttggag gcttccagcg cctcatctgg aagtgggaca 6900

cctgtagcat cgatagcagc accaccaatt aaatgatttt cgaaatcgaa cttgacattg 6960

gaacgaacat cagaaatagc tttaagaacc ttaatggctt cggctgtgat ttcttgacca 7020

acgtggtcac ctggcaaaac gacgatcttc ttaggggcag acataggggc agacattaga 7080

atggtatatc cttgaaatat atatatatat tgctgaaatg taaaaggtaa gaaaagttag 7140

aaagtaagac gattgctaac cacctattgg aaaaaacaat aggtccttaa ataatattgt 7200

caacttcaag tattgtgatg caagcattta gtcatgaacg cttctctatt ctatatgaaa 7260

agccggttcc ggcctctcac ctttcctttt tctcccaatt tttcagttga aaaaggtata 7320

tgcgtcaggc gacctctgaa attaacaaaa aatttccagt catcgaattt gattctgtgc 7380

gatagcgccc ctgtgtgttc tcgttatgtt gaggaaaaaa ataatggttg ctaagagatt 7440

cgaactcttg catcttacga tacctgagta ttcccacagt taactgcggt caagatattt 7500

cttgaatcag gcgccttaga ccgctcggcc aaacaaccaa ttacttgttg agaaatagag 7560

tataattatc ctataaatat aacgtttttg aacacacatg aacaaggaag tacaggacaa 7620

ttgattttga agagaatgtg gattttgatg taattgttgg gattccattt ttaataaggc 7680

aataatatta ggtatgtgga tatactagaa gttctcctcg accgtcgata tgcggtgtga 7740

aataccgcac agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgttaatatt 7800

ttgttaaaat tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa 7860

atcggcaaaa tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca 7920

gtttggaaca agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc 7980

gtctatcagg gcgatggccc actacgtgaa ccatcaccct aatcaagttt tttggggtcg 8040

aggtgccgta aagcactaaa tcggaaccct aaagggagcc cccgatttag agcttgacgg 8100

ggaaagccgg cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg 8160

gcgctggcaa gtgtagcggt cacgctgcgc gtaaccacca cacccgccgc gcttaatgcg 8220

ccgctacagg gcgcgtcgcg ccattcgcca ttcaggctgc gcaactgttg ggaagggcga 8280

tcggtgcggg cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga 8340
ttaagttggg taacgccagg gttttcccag tcacgacgtt gtaaaacgac ggccagtgag 8400
cgcgcgtaat acgactcact atagggcgaa ttgggtaccg ggcccccccct cgaggtcgac 8460
ggtatcgata agcttgatat cgaattcctg cagcccgggg gatccgttag aatcattttg 8520
aataaaaaac acgctttttc agttcgagtt tatcattatc aatactgcca tttcaaagaa 8580
tacgtaaata attaatagta gtgattttcc taactttatt tagtcaaaaa attagccttt 8640
taattctgct gtaacccgta catgcccaaa atagggggcg ggttacacag aatatataac 8700

atcgtaggtg tctgggtgaa cagtttattc ctggcatcca ctaaatataa tggagcccgc 8760
tttttaagct ggcatccaga aaaaaaaaga atcccagcac caaaatattg ttttcttcac 8820
caaccatcag ttcataggtc cattctctta gcgcaactac agagaacagg ggcacaaaca 8880
ggcaaaaaac gggcacaacc tcaatggagt gatgcaacct gcctggagta aatgatgaca 8940
caaggcaatt gacccacgca tgtatctatc tcattttctt acaccttcta ttaccttctg 9000
ctctctctga tttggaaaaa gctgaaaaaa aaggttgaaa ccagttccct gaaattattc 9060
ccctacttga ctaataagta tataaagacg gtaggtattg attgtaattc tgtaaatcta 9120

tttcttaaac ttcttaaatt ctacttttat agttagtctt ttttttagtt ttaaaacacc 9180
aagaacttag tttcgaataa acacacataa acaaacaaag aattc 9225

<210> 732
<211> 9223
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 732

atgtctaaag gtgaagaatt attcactggt gttgtcccaa ttttggttga attagatggt 60
gatgttaatg gtcacaaatt ttctgtctcc ggtgaaggtg aaggtgatgc tacttacggt 120
aaattgacct taaaatttat ttgtactact ggtaaattgc cagttccatg gccaacctta 180
gtcactactt tcggttatgg tgttcaatgt tttgctagat acccagatca tatgaaacaa 240

catgactttt tcaagtctgc catgccagaa ggttatgttc aagaaagaac tatttttttc 300
aaagatgacg gtaactacaa gaccagagct gaagtcaagt ttgaaggtga taccttagtt 360
aatagaatcg aattaaaagg tattgatttt aaagaagatg gtaacatttt aggtcacaaa 420
ttggaataca actataactc tcacaatgtt tacatcatgg ctgacaaaca aaagaatggt 480

atcaaagtta acttcaaaat tagacacaac attgaagatg gttctgttca attagctgac 540

cattatcaac aaaatactcc aattggtgat ggtccagtct tgttaccaga caaccattac 600

ttatccactc aatctgcctt atccaaagat ccaaacgaaa agagagacca catggtcttg 660

ttagaatttg ttactgctgc tggtattacc catggtatgg atgaattgta caaagctagc 720

aaacctgggt caatccttgg ggcccagact gagcacgtga tggcagagca caggacgtca 780

tggtttcaaa aggtgaagaa gataatatgg ctattattaa agaatttatg agatttaaag 840

ttcatatgga aggttcagtt aatggtcatg aatttgaaat tgaaggtgaa ggtgaaggta 900

gaccatatga aggtactcaa actgctaaat tgaaagttac taaaggtggt ccattaccat 960

ttgcttggga tattttgtca ccacaattta tgtatggttc aaaagcttat gttaaacatc 1020

cagctgatat tccagattat ttaaaattgt catttccaga aggttttaaa tgggaaagag 1080

ttatgaattt tgaagatggt ggtgttgtta ctgttactca agattcatca ttacaagatg 1140

gtgaatttat ttataaagtt aaattgagag gtactaattt tccatcagat ggtccagtta 1200

tgcaaaaaaa aactatgggt tgggaagctt catcagaaag aatgtatcca gaagatggtg 1260

ctttaaaagg tgaaattaaa caaagattga aattaaaaga tggtggtcat tatgatgctg 1320

aagttaaaac tacttataaa gctaaaaaac cagttcaatt accaggtgct tataatgtta 1380

atattaaatt ggatattact tcacataatg aagattatac tattgttgaa caatatgaaa 1440

gagctgaagg tagacattca actggtggta tggatgaatt atataaaggt accgctcgag 1500

cagctgtgat tgattgagtc gacttggttg aacacgttgc caaggcttaa gtgaatttac 1560

tttaaatctt gcatttaaat aaattttctt tttatagctt tatgacttag tttcaattta 1620

tatactattt taatgacatt ttcgattcgg atccactagt tctagagcgg ccgccaccgc 1680

ggtggagctc cagcttttgt tccctttagt gagggttaat tgcgcgcttg gcgtaatcat 1740

ggtcatagct gtttcctgtg tgaaattgtt atccgctcac aattccacac aacataggag 1800

ccggaagcat aaagtgtaaa gcctggggtg cctaatgagt gaggtaactc acattaattg 1860

cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa 1920

tcggccaacg cgcggggaga ggcggtttgc gtattgggcg ctcttccgct tcctcgctca 1980

ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg 2040

taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc 2100

agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc 2160

cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac 2220

tataaagata ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc 2280

tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata 2340

gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc 2400
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca 2460
acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag 2520
cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta 2580
gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg 2640
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt gtttgcaagc 2700
agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt 2760

ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa 2820
ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat 2880
atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga 2940
tctgtctatt tcgttcatcc atagttgcct gactccccgt cgtgtagata actacgatac 3000
gggagggctt accatctggc cccagtgctg caatgatacc gcgagaccca cgctcaccgg 3060
ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggtcctg 3120
caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt 3180

cgccagttaa tagtttgcgc aacgttgttg ccattgctac aggcatcgtg gtgtcacgct 3240
cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga gttacatgat 3300
cccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt gtcagaagta 3360
agttggccgc agtgttatca ctcatggtta tggcagcact gcataattct cttactgtca 3420
tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat 3480
agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat accgcgccac 3540
atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga aaactctcaa 3600

ggatcttacc gctgttgaga tccagttcga tgtaacccac tcgtgcaccc aactgatctt 3660
cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg 3720
caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc ctttttcaat 3780
attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt 3840
agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcca cctgaacgaa 3900
gcatctgtgc ttcattttgt agaacaaaaa tgcaacgcga gagcgctaat ttttcaaaca 3960
aagaatctga gctgcatttt tacagaacag aaatgcaacg cgaaagcgct attttaccaa 4020

cgaagaatct gtgcttcatt tttgtaaaac aaaaatgcaa cgcgagagcg ctaatttttc 4080

aaacaaagaa tctgagctgc atttttacag aacagaaatg caacgcgaga gcgctatttt 4140

accaacaaag aatctatact tctttttgt tctacaaaaa tgcatcccga gagcgctatt 4200

tttctaacaa agcatcttag attacttttt ttctcctttg tgcgctctat aatgcagtct 4260

cttgataact ttttgcactg taggtccgtt aaggttagaa gaaggctact ttggtgtcta 4320

ttttctcttc cataaaaaaa gcctgactcc acttcccgcg tttactgatt actagcgaag 4380

ctgcgggtgc atttttttcaa gataaaggca tccccgatta tattctatac cgatgtggat 4440

tgcgcatact ttgtgaacag aaagtgatag cgttgatgat tcttcattgg tcagaaaatt 4500

atgaacggtt tcttctattt tgtctctata tactacgtat aggaaatgtt tacattttcg 4560

tattgttttc gattcactct atgaatagtt cttactacaa ttttttttgtc taaagagtaa 4620

tactagagat aaacataaaa aatgtagagg tcgagtttag atgcaagttc aaggagcgaa 4680

aggtggatgg gtaggttata tagggatata gcacagagat atatagcaaa gagatacttt 4740

tgagcaatgt ttgtggaagc ggtattcgca atattttagt agctcgttac agtccggtgc 4800

gtttttggtt ttttgaaagt gcgtcttcag agcgcttttg gttttcaaaa gcgctctgaa 4860

gttcctatac tttctagaga ataggaactt cggaatagga acttcaaagc gtttccgaaa 4920

acgagcgctt ccgaaaatgc aacgcgagct gcgcacatac agctcactgt tcacgtcgca 4980

cctatatctg cgtgttgcct gtatatatat atacatgaga agaacggcat agtgcgtgtt 5040

tatgcttaaa tgcgtactta tatgcgtcta tttatgtagg atgaaaggta gtctagtacc 5100

tcctgtgata ttatcccatt ccatgcgggg tatcgtatgc ttccttcagc actacccttt 5160

agctgttcta tatgctgcca ctcctcaatt ggattagtct catccttcaa tgctatcatt 5220

tcctttgata ttggatcata ctaagaaacc attattatca tgacattaac ctataaaaat 5280

aggcgtatca cgaggccctt tcgtctcgcg cgtttcggtg atgacggtga aaacctctga 5340

cacatgcagc tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa 5400

gcccgtcagg gcgcgtcagc gggtgttggc gggtgtcggg gctggcttaa ctatgcggca 5460

tcagagcaga ttgtactgag agtgcaccat atcgactacg tcgtaaggcc gtttctgaca 5520

gagtaaaatt cttgagggaa ctttcaccat tatgggaaat gcttcaagaa ggtattgact 5580

taaactccat caaatggtca ggtcattgag tgtttttttat ttgttgtatt ttttttttttt 5640

tagagaaaat cctccaatat caaattagga atcgtagttt catgattttc tgttacacct 5700

aacttttttgt gtggtgccct cctccttgtc aatattaatg ttaaagtgca attcttttttc 5760

cttatcacgt tgagccatta gtatcaattt gcttacctgt attcctttac tatcctcctt 5820

tttctccttc ttgataaatg tatgtagatt gcgtatatag tttcgtctac cctatgaaca 5880

tattccattt tgtaatttcg tgtcgtttct attatgaatt tcatttataa agtttatgta 5940

caaatatcat aaaaaaagag aatcttttta agcaaggatt ttcttaactt cttcggcgac 6000

agcatcaccg acttcggtgg tactgttgga accacctaaa tcaccagttc tgatacctgc 6060

atccaaaacc tttttaactg catcttcaat ggccttacct tcttcaggca agttcaatga 6120

caatttcaac atcattgcag cagacaagat agtggcgata gggtcaacct tattctttgg 6180

caaatctgga gcagaaccgt ggcatggttc gtacaaacca aatgcggtgt tcttgtctgg 6240

caaagaggcc aaggacgcag atggcaacaa acccaaggaa cctgggataa cggaggcttc 6300

atcggagatg atatcaccaa acatgttgct ggtgattata ataccattta ggtgggttgg 6360

gttcttaact aggatcatgg cggcagaatc aatcaattga tgttgaacct tcaatgtagg 6420

gaattcgttc ttgatggttt cctccacagt ttttctccat aatcttgaag aggccaaaag 6480

attagcttta tccaaggacc aaataggcaa tggtggctca tgttgtaggg ccatgaaagc 6540

ggccattctt gtgattcttt gcacttctgg aacggtgtat tgttcactat cccaagcgac 6600

accatcacca tcgtcttcct ttctcttacc aaagtaaata cctcccacta attctctgac 6660

aacaacgaag tcagtacctt tagcaaattg tggcttgatt ggagataagt ctaaaagaga 6720

gtcggatgca aagttacatg gtcttaagtt ggcgtacaat tgaagttctt tacggatttt 6780

tagtaaacct tgttcaggtc taacactacc ggtaccccat ttaggaccag ccacagcacc 6840

taacaaaacg gcatcaacct tcttggaggc ttccagcgcc tcatctggaa gtgggacacc 6900

tgtagcatcg atagcagcac caccaattaa atgattttcg aaatcgaact tgacattgga 6960

acgaacatca gaaatagctt taagaacctt aatggcttcg gctgtgattt cttgaccaac 7020

gtggtcacct ggcaaaacga cgatcttctt aggggcagac ataggggcag acattagaat 7080

ggtatatcct tgaaatatat atatatattg ctgaaatgta aaaggtaaga aaagttagaa 7140

agtaagacga ttgctaacca cctattggaa aaaacaatag gtccttaaat aatattgtca 7200

acttcaagta ttgtgatgca agcatttagt catgaacgct tctctattct atatgaaaag 7260

ccggttccgg cctctcacct ttccttttc tcccaatttt tcagttgaaa aaggtatatg 7320

cgtcaggcga cctctgaaat taacaaaaaa tttccagtca tcgaatttga ttctgtgcga 7380

tagcgcccct gtgtgttctc gttatgttga ggaaaaaaat aatggttgct aagagattcg 7440

aactcttgca tcttacgata cctgagtatt cccacagtta actgcggtca agatatttct 7500

tgaatcaggc gccttagacc gctcggccaa acaaccaatt acttgttgag aaatagagta 7560

taattatcct ataaatataa cgtttttgaa cacacatgaa caaggaagta caggacaatt 7620

gattttgaag agaatgtgga ttttgatgta attgttggga ttccattttt aataaggcaa 7680

taatattagg tatgtggata tactagaagt tctcctcgac cgtcgatatg cggtgtgaaa 7740

taccgcacag atgcgtaagg agaaaatacc gcatcaggaa attgtaaacg ttaatatttt 7800

gttaaaattc gcgttaaatt tttgttaaat cagctcattt tttaaccaat aggccgaaat 7860
cggcaaaatc ccttataaat caaaagaata gaccgagata gggttgagtg ttgttccagt 7920
ttggaacaag agtccactat taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt 7980
ctatcagggc gatggcccac tacgtgaacc atcaccctaa tcaagttttt tggggtcgag 8040
gtgccgtaaa gcactaaatc ggaaccctaa agggagcccc cgatttagag cttgacgggg 8100
aaagccggcg aacgtggcga gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc 8160
gctggcaagt gtagcggtca cgctgcgcgt aaccaccaca cccgccgcgc ttaatgcgcc 8220

gctacagggc gcgtcgcgcc attcgccatt caggctgcgc aactgttggg aagggcgatc 8280
ggtgcgggcc tcttcgctat tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt 8340
aagttgggta acgccagggt tttcccagtc acgacgttgt aaaacgacgg ccagtgagcg 8400
cgcgtaatac gactcactat agggcgaatt gggtaccggg ccccccctcg aggtcgacgg 8460
tatcgataag cttgatatcg aattcctgca gcccgggggа tccgttagaa tcattttgaa 8520
taaaaaacac gctttttcag ttcgagttta tcattatcaa tactgccatt tcaaagaata 8580
cgtaaataat taatagtagt gattttccta actttattta gtcaaaaaat tagcctttta 8640

attctgctgt aacccgtaca tgcccaaaat agggggcggg ttacacagaa tatataacat 8700
cgtaggtgtc tgggtgaaca gtttattcct ggcatccact aaatataatg gagcccgctt 8760
tttaagctgg catccagaaa aaaaaagaat cccagcacca aaatattgtt ttcttcacca 8820
accatcagtt cataggtcca ttctcttagc gcaactacag agaacagggg cacaaacagg 8880
caaaaaacgg gcacaacctc aatggagtga tgcaacctgc ctggagtaaa tgatgacaca 8940
aggcaattga cccacgcatg tatctatctc attttcttac accttctatt accttctgct 9000
ctctctgatt tggaaaaagc tgaaaaaaaa ggttgaaacc agttccctga aattattccc 9060

ctacttgact aataagtata taaagacggt aggtattgat tgtaattctg taaatctatt 9120
tcttaaactt cttaaattct acttttatag ttagtctttt ttttagtttt aaaacaccaa 9180
gaacttagtt tcgaataaac acacataaac aaacaaagaa ttc 9223

<210> 733
<211> 9753
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 733

atgaaacgga cagccgacgg aagcgagttc gagtcaccaa agaagaagcg gaaagtcgac 60

aagaagtaca gcatcggcct ggacatcggc accaactctg tgggctgggc cgtgatcacc 120

gacgagtaca aggtgcccag caagaaattc aaggtgctgg gcaacaccga ccggcacagc 180

atcaagaaga acctgatcgg agccctgctg ttcgacagcg gcgaaacagc cgaggccacc 240

cggctgaaga gaaccgccag aagaagatac accagacgga agaaccggat ctgctatctg 300

caagagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca cagactggaa 360

gagtccttcc tggtggaaga ggataagaag cacgagcggc accccatctt cggcaacatc 420

gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgag aaagaaactg 480

gtggacagca ccgacaaggc cgacctgcgg ctgatctatc tggccctggc ccacatgatc 540

aagttccggg gccacttcct gatcgagggc gacctgaacc ccgacaacag cgacgtggac 600

aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggaaaa ccccatcaac 660

gccagcggcg tggacgccaa ggccatcctg tctgccagac tgagcaagag cagacggctg 720

gaaaatctga tcgcccagct gcccggcgag aagaagaatg gcctgttcgg aaacctgatt 780

gccctgagcc tgggcctgac ccccaacttc aagagcaact tcgacctggc cgaggatgcc 840

aaactgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc 900

ggcgaccagt acgccgacct gtttctggcc gccaagaacc tgtccgacgc catcctgctg 960

agcgacatcc tgagagtgaa caccgagatc accaaggccc ccctgagcgc ctctatgatc 1020

aagagatacg acgagcacca ccaggacctg accctgctga aagctctcgt gcggcagcag 1080

ctgcctgaga agtacaaaga gattttcttc gaccagagca agaacggcta cgccggctac 1140

attgacggcg gagccagcca ggaagagttc tacaagttca tcaagcccat cctggaaaag 1200

atggacggca ccgaggaact gctcgtgaag ctgaacagag aggacctgct gcggaagcag 1260

cggaccttcg acaacggcag catcccccac cagatccacc tgggagagct gcacgccatt 1320

ctgcggcggc aggaagattt ttacccattc ctgaaggaca accgggaaaa gatcgagaag 1380

atcctgacct tccgcatccc ctactacgtg ggccctctgg ccaggggaaa cagcagattc 1440

gcctggatga ccagaaagag cgaggaaacc atcacccccct ggaacttcga ggaagtggtg 1500

gacaagggcg cttccgccca gagcttcatc gagcggatga ccaacttcga taagaacctg 1560

cccaacgaga aggtgctgcc caagcacagc ctgctgtacg agtacttcac cgtgtataac 1620

gagctgacca aagtgaaata cgtgaccgag ggaatgagaa agcccgcctt cctgagcggc 1680

gagcagaaaa aggccatcgt ggacctgctg ttcaagacca accggaaagt gaccgtgaag 1740

cagctgaaag aggactactt caagaaaatc gagtgcttcg actccgtgga aatctccggc 1800

gtggaagatc ggttcaacgc ctccctgggc acataccacg atctgctgaa aattatcaag 1860

gacaaggact tcctggacaa tgaggaaaac gaggacattc tggaagatat cgtgctgacc 1920
ctgacactgt ttgaggacag agagatgatc gaggaacggc tgaaaaccta tgcccacctg 1980
ttcgacgaca aagtgatgaa gcagctgaag cggcggagat acaccggctg gggcaggctg 2040
agccggaagc tgatcaacgg catccgggac aagcagtccg gcaagacaat cctggatttc 2100
ctgaagtccg acggcttcgc caacagaaac ttcatgcagc tgatccacga cgacagcctg 2160
acctttaaag aggacatcca gaaagcccag gtgtccggcc agggcgatag cctgcacgag 2220
cacattgcca atctggccgg cagccccgcc attaagaagg gcatcctgca gacagtgaag 2280

gtggtggacg agctcgtgaa agtgatgggc cggcacaagc ccgagaacat cgtgatcgaa 2340
atggccagag agaaccagac cacccagaag ggacagaaga acagccgcga gagaatgaag 2400
cggatcgaag agggcatcaa agagctgggc agccagatcc tgaaagaaca ccccgtggaa 2460
aacacccagc tgcagaacga gaagctgtac ctgtactacc tgcagaatgg gcgggatatg 2520
tacgtggacc aggaactgga catcaaccgg ctgtccgact acgatgtgga cgctatcgtg 2580
cctcagagct ttctgaagga cgactccatc gacaacaagg tgctgaccag aagcgacaag 2640
aaccggggca agagcgacaa cgtgccctcc gaagaggtcg tgaagaagat gaagaactac 2700

tggcggcagc tgctgaacgc caagctgatt acccagagaa agttcgacaa tctgaccaag 2760
gccgagagag gcggcctgag cgaactggat aaggccggct tcatcaagag acagctggtg 2820
gaaacccggc agatcacaaa gcacgtggca cagatcctgg actcccggat gaacactaag 2880
tacgacgaga atgacaagct gatccgggaa gtgaaagtga tcaccctgaa gtccaagctg 2940
gtgtccgatt tccggaagga tttccagttt tacaaagtgc gcgagatcaa caactaccac 3000
cacgcccacg acgcctacct gaacgccgtc gtgggaaccg ccctgatcaa aaagtaccct 3060
aagctggaaa gcgagttcgt gtacggcgac tacaaggtgt acgacgtgcg gaagatgatc 3120

gccaagagcg agcaggaaat cggcaaggct accgccaagt acttcttcta cagcaacatc 3180
atgaactttt tcaagaccga gattaccctg gccaacggcg agatccggaa gcggcctctg 3240
atcgagacaa acggcgaaac cggggagatc gtgtgggata agggccggga ttttgccacc 3300
gtgcggaaag tgctgagcat gccccaagtg aatatcgtga aaaagaccga ggtgcagaca 3360
ggcggcttca gcaaagagtc tatcctgccc aagaggaaca gcgataagct gatcgccaga 3420
aagaaggact gggaccctaa gaagtacggc ggcttcgaca gccccaccgt ggcctattct 3480
gtgctggtgg tggccaaagt ggaaaagggc aagtccaaga aactgaagag tgtgaaagag 3540

ctgctgggga tcaccatcat ggaaagaagc agcttcgaga agaatcccat cgactttctg 3600

gaagccaagg gctacaaaga agtgaaaaag gacctgatca tcaagctgcc taagtactcc 3660

ctgttcgagc tggaaaacgg ccggaagaga atgctggcct ctgccggcga actgcagaag 3720

ggaaacgaac tggccctgcc ctccaaatat gtgaacttcc tgtacctggc cagccactat 3780

gagaagctga agggctcccc cgaggataat gagcagaaac agctgtttgt ggaacagcac 3840

aagcactacc tggacgagat catcgagcag atcagcgagt tctccaagag agtgatcctg 3900

gccgacgcta atctggacaa agtgctgtcc gcctacaaca agcaccggga taagcccatc 3960

agagagcagg ccgagaatat catccacctg tttaccctga ccaatctggg agcccctgcc 4020

gccttcaagt actttgacac caccatcgac cggaagaggt acaccagcac caaagaggtg 4080

ctggacgcca ccctgatcca ccagagcatc accggcctgt acgagacacg gatcgacctg 4140

tctcagctgg gaggtgactc tggaggatct agcggaggat cctctggcag cgagacacca 4200

ggaacaagcg agtcagcaac accagagagc agtggcggca gcagcggcgg cagcagcacc 4260

ctaaatatag aagatgagta tcggctacat gagacctcaa aagagccaga tgtttctcta 4320

gggtccacat ggctgtctga ttttcctcag gcctgggcgg aaaccggggg catgggactg 4380

gcagttcgcc aagctcctct gatcatacct ctgaaagcaa cctctacccc cgtgtccata 4440

aaacaatacc ccatgtcaca agaagccaga ctggggatca agccccacat acagagactg 4500

ttggaccagg gaatactggt accctgccag tccccctgga acacgcccct gctacccgtt 4560

aagaaaccag ggactaatga ttataggcct gtccaggatc tgagagaagt caacaagcgg 4620

gtggaagaca tccaccccac cgtgcccaac ccttacaacc tcttgagcgg gctcccaccg 4680

tcccaccagt ggtacactgt gcttgattta aaggatgcct ttttctgcct gagactccac 4740

cccaccagtc agcctctctt cgcctttgag tggagagatc cagagatggg aatctcagga 4800

caattgacct ggaccagact cccacagggt ttcaaaaaca gtcccaccct gtttgatgag 4860

gcactgcaca gagacctagc agacttccgg atccagcacc cagacttgat cctgctacag 4920

tacgtggatg acttactgct ggccgccact tctgagctag actgccaaca aggtactcgg 4980

gccctgttac aaaccctagg gaacctcggg tatcgggcct cggccaagaa agcccaaatt 5040

tgccagaaac aggtcaagta tctggggtat cttctaaaag agggtcagag atggctgact 5100

gaggccagaa aagagactgt gatggggcag cctactccga agacccctcg acaactaagg 5160

gagttcctag ggacggcagg cttctgtcgc ctctggatcc ctgggtttgc agaaatggca 5220

gccccccctgt accctctcac caaaacgggg actctgttta attggggccc agaccaacaa 5280

aaggcctatc aagaaatcaa gcaagctctt ctaactgccc cagccctggg gttgccagat 5340

ttgactaagc cctttgaact ctttgtcgac gagaagcagg gctacgccaa aggtgtccta 5400

acgcaaaaac tgggaccttg gcgtcggccg gtggcctacc tgtccaaaaa gctagaccca 5460

gtagcagctg ggtggccccc ttgcctacgg atggtagcag ccattgccgt actgacaaag 5520

gatgcaggca agctaaccat gggacagcca ctagtcattc tggccccccа tgcagtagag 5580

gcactagtca aacaaccccc cgaccgctgg ctttccaacg cccggatgac tcactatcag 5640

gccttgcttt tggacacgga ccgggtccag ttcggaccgg tggtagccct gaacccggct 5700

acgctgctcc cactgcctga ggaagggctg caacacaact gccttgatat cctggccgaa 5760

gcccacggaa cccgacccga cctaacggac cagccgctcc cagacgccga ccacacctgg 5820

tacacggatg gaagcagtct cttacaagag ggacagcgta aggcgggagc tgcggtgacc 5880

accgagaccg aggtaatctg ggctaaagcc ctgccagccg ggacatccgc tcagcgggct 5940

gaactgatag cactcaccca ggccctaaag atggcagaag gtaagaagct aaatgtttat 6000

actgatagcc gttatgcttt tgctactgcc catatccatg gagaaatata cagaaggcgt 6060

gggttgctca catcagaagg caaagagatc aaaaataaag acgagatctt ggccctacta 6120

aaagccctct ttctgcccaa aagacttagc ataatccatt gtccaggaca tcaaaaggga 6180

cacagcgccg aggctagagg caaccggatg gctgaccaag cggcccgaaa ggcagccatc 6240

acagagactc cagacacctc taccctcctc atagaaaatt catcaccctc tggcggctca 6300

aaaagaaccg ccgacggcag cgaattcgag cccaagaaga agaggaaagt ctaaccggtc 6360

atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 6420

gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 6480

ccactgtcct ttcctaataa aatgagaaaa ttgcatcgca ttgtctgagt aggtgtcatt 6540

ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 6600

ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 6660

cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 6720

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 6780

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6840

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6900

gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6960

ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 7020

caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 7080

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 7140

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 7200

cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 7260

ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 7320

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg 7380
accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 7440
cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 7500
cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 7560
gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 7620
aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 7680
aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 7740

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 7800
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7860
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7920
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7980
ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 8040
accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 8100
agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 8160

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 8220
tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 8280
cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 8340
tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 8400
ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 8460
gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 8520
tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 8580

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 8640
gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 8700
cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 8760
gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 8820
ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8880
ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8940
atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 9000

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 9060

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 9120

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 9180

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 9240

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 9300

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 9360

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 9420

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 9480

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 9540

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 9600

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 9660

aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 9720

ctaatacgac tcactatagg gagagccgcc acc 9753

<210> 734

<211> 9753

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 734

atgaaacgga cagccgacgg aagcgagttc gagtcaccaa agaagaagcg gaaagtcgac 60

aagaagtaca gcatcggcct ggacatcggc accaactctg tgggctgggc cgtgatcacc 120

gacgagtaca aggtgcccag caagaaattc aaggtgctgg gcaacaccga ccggcacagc 180

atcaagaaga acctgatcgg agccctgctg ttcgacagcg gcgaaacagc cgaggccacc 240

cggctgaaga gaaccgccag aagaagatac accagacgga agaaccggat ctgctatctg 300

caagagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca cagactggaa 360

gagtccttcc tggtggaaga ggataagaag cacgagcggc accccatctt cggcaacatc 420

gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgag aaagaaactg 480

gtggacagca ccgacaaggc cgacctgcgg ctgatctatc tggccctggc ccacatgatc 540

aagttccggg gccacttcct gatcgagggc gacctgaacc ccgacaacag cgacgtggac 600

aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggaaaa ccccatcaac 660

gccagcggcg tggacgccaa ggccatcctg tctgccagac tgagcaagag cagacggctg 720

gaaaatctga tcgcccagct gcccggcgag aagaagaatg gcctgttcgg aaacctgatt 780

gccctgagcc tgggcctgac ccccaacttc aagagcaact tcgacctggc cgaggatgcc 840

aaactgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc 900

ggcgaccagt acgccgacct gtttctggcc gccaagaacc tgtccgacgc catcctgctg 960

agcgacatcc tgagagtgaa caccgagatc accaaggccc ccctgagcgc ctctatgatc 1020

aagagatacg acgagcacca ccaggacctg accctgctga aagctctcgt gcggcagcag 1080

ctgcctgaga agtacaaaga gattttcttc gaccagagca agaacggcta cgccggctac 1140

attgacggcg gagccagcca ggaagagttc tacaagttca tcaagcccat cctggaaaag 1200

atggacggca ccgaggaact gctcgtgaag ctgaacagag aggacctgct gcggaagcag 1260

cggaccttcg acaacggcag catcccccac cagatccacc tgggagagct gcacgccatt 1320

ctgcggcggc aggaagattt ttacccattc ctgaaggaca accgggaaaa gatcgagaag 1380

atcctgacct tccgcatccc ctactacgtg ggccctctgg ccaggggaaa cagcagattc 1440

gcctggatga ccagaaagag cgaggaaacc atcacccccc ggaacttcga ggaagtggtg 1500

gacaagggcg cttccgccca gagcttcatc gagcggatga ccaacttcga taagaacctg 1560

cccaacgaga aggtgctgcc caagcacagc ctgctgtacg agtacttcac cgtgtataac 1620

gagctgacca aagtgaaata cgtgaccgag ggaatgagaa agcccgcctt cctgagcggc 1680

gagcagaaaa aggccatcgt ggacctgctg ttcaagacca accggaaagt gaccgtgaag 1740

cagctgaaag aggactactt caagaaaatc gagtgcttcg actccgtgga aatctccggc 1800

gtggaagatc ggttcaacgc ctccctgggc acataccacg atctgctgaa aattatcaag 1860

gacaaggact tcctggacaa tgaggaaaac gaggacattc tggaagatat cgtgctgacc 1920

ctgacactgt ttgaggacag agagatgatc gaggaacggc tgaaaaccta tgcccacctg 1980

ttcgacgaca aagtgatgaa gcagctgaag cggcggagat acaccggctg gggcaggctg 2040

agccggaagc tgatcaacgg catccgggac aagcagtccg gcaagacaat cctggatttc 2100

ctgaagtccg acggcttcgc caacagaaac ttcatgcagc tgatccacga cgacagcctg 2160

acctttaaag aggacatcca gaaagcccag gtgtccggcc agggcgatag cctgcacgag 2220

cacattgcca atctggccgg cagccccgcc attaagaagg gcatcctgca gacagtgaag 2280

gtggtggacg agctcgtgaa agtgatgggc cggcacaagc ccgagaacat cgtgatcgaa 2340

atggccagag agaaccagac cacccagaag ggacagaaga acagccgcga gagaatgaag 2400

cggatcgaag agggcatcaa agagctgggc agccagatcc tgaaagaaca ccccgtggaa 2460

aacacccagc tgcagaacga gaagctgtac ctgtactacc tgcagaatgg gcgggatatg 2520

tacgtggacc aggaactgga catcaaccgg ctgtccgact acgatgtgga cgctatcgtg 2580

cctcagagct ttctgaagga cgactccatc gacaacaagg tgctgaccag aagcgacaag 2640

aaccggggca agagcgacaa cgtgccctcc gaagaggtcg tgaagaagat gaagaactac 2700

tggcggcagc tgctgaacgc caagctgatt acccagagaa agttcgacaa tctgaccaag 2760

gccgagagag gcggcctgag cgaactggat aaggccggct tcatcaagag acagctggtg 2820

gaaacccggc agatcacaaa gcacgtggca cagatcctgg actcccggat gaacactaag 2880

tacgacgaga atgacaagct gatccgggaa gtgaaagtga tcaccctgaa gtccaagctg 2940

gtgtccgatt tccggaagga tttccagttt tacaaagtgc gcgagatcaa caactaccac 3000

cacgcccacg acgcctacct gaacgccgtc gtgggaaccg ccctgatcaa aaagtaccct 3060

aagctggaaa gcgagttcgt gtacggcgac tacaaggtgt acgacgtgcg gaagatgatc 3120

gccaagagcg agcaggaaat cggcaaggct accgccaagt acttcttcta cagcaacatc 3180

atgaactttt tcaagaccga gattaccctg gccaacggcg agatccggaa gcggcctctg 3240

atcgagacaa acggcgaaac cggggagatc gtgtgggata agggccggga ttttgccacc 3300

gtgcggaaag tgctgagcat gccccaagtg aatatcgtga aaaagaccga ggtgcagaca 3360

ggcggcttca gcaaagagtc tatcctgccc aagaggaaca gcgataagct gatcgccaga 3420

aagaaggact gggaccctaa gaagtacggc ggcttcgaca gccccaccgt ggcctattct 3480

gtgctggtgg tggccaaagt ggaaaagggc aagtccaaga aactgaagag tgtgaaagag 3540

ctgctgggga tcaccatcat ggaaagaagc agcttcgaga agaatcccat cgactttctg 3600

gaagccaagg gctacaaaga agtgaaaaag gacctgatca tcaagctgcc taagtactcc 3660

ctgttcgagc tggaaaacgg ccggaagaga atgctggcct ctgccggcga actgcagaag 3720

ggaaacgaac tggccctgcc ctccaaatat gtgaacttcc tgtacctggc cagccactat 3780

gagaagctga agggctcccc cgaggataat gagcagaaac agctgtttgt ggaacagcac 3840

aagcactacc tggacgagat catcgagcag atcagcgagt tctccaagag agtgatcctg 3900

gccgacgcta atctggacaa agtgctgtcc gcctacaaca agcaccggga taagcccatc 3960

agagagcagg ccgagaatat catccacctg tttaccctga ccaatctggg agcccctgcc 4020

gccttcaagt actttgacac caccatcgac cggaagaggt acaccagcac caaagaggtg 4080

ctggacgcca ccctgatcca ccagagcatc accggcctgt acgagacacg gatcgacctg 4140

tctcagctgg gaggtgactc tggaggatct agcggaggat cctctggcag cgagacacca 4200

ggaacaagcg agtcagcaac accagagagc agtggcggca gcagcggcgg cagcagcacc 4260

ctaaatatag aagatgagta tcggctacat gagacctcaa aagagccaga tgtttctcta 4320

gggtccacat ggctgtctga ttttcctcag gcctgggcgg aaaccggggg catgggactg 4380

gcagttcgcc aagctcctct gatcatacct ctgaaagcaa cctctacccc cgtgtccata 4440

aaacaatacc ccatgtcaca agaagccaga ctggggatca agccccacat acagagactg 4500

ttggaccagg gaatactggt accctgccag tccccctgga acacgcccct gctacccgtt 4560

aagaaaccag ggactaatga ttataggcct gtccaggatc tgagagaagt caacaagcgg 4620

gtggaagaca tccaccccac cgtgcccaac ccttacaacc tcttgagcgg gctcccaccg 4680

tcccaccagt ggtacactgt gcttgattta aaggatgcct ttttctgcct gagactccac 4740

cccaccagtc agcctctctt cgcctttgag tggagagatc cagagatggg aatctcagga 4800

caattgacct ggaccagact cccacagggt ttcaaaaaca gtcccaccct gtttaatgag 4860

gcactgcaca gagacctagc agacttccgg atccagcacc cagacttgat cctgctacag 4920

tacgtggatg acttactgct ggccgccact tctgagctag actgccaaca aggtactcgg 4980

gccctgttac aaaccctagg gaacctcggg tatcgggcct cggccaagaa agcccaaatt 5040

tgccagaaac aggtcaagta tctggggtat cttctaaaag agggtcagag atggctgact 5100

gaggccagaa aagagactgt gatggggcag cctactccga agacccctcg acaactaagg 5160

gagttcctag ggaaggcagg cttctgtcgc ctcttcatcc ctgggtttgc agaaatggca 5220

gccccccctgt accctctcac caaaccgggg actctgttta attggggccc agaccaacaa 5280

aaggcctatc aagaaatcaa gcaagctctt ctaactgccc cagccctggg gttgccagat 5340

ttgactaagc cctttgaact ctttgtcgac gagaagcagg gctacgccaa aggtgtccta 5400

acgcaaaaac tgggaccttg gcgtcggccg gtggcctacc tgtccaaaaa gctagaccca 5460

gtagcagctg ggtggccccc ttgcctacgg atggtagcag ccattgccgt actgacaaag 5520

gatgcaggca agctaaccat gggacagcca ctagtcattc tggccccca tgcagtagag 5580

gcactagtca aacaaccccc cgaccgctgg ctttccaacg cccggatgac tcactatcag 5640

gccttgcttt tggacacgga ccgggtccag ttcggaccgg tggtagccct gaacccggct 5700

acgctgctcc cactgcctga ggaagggctg caacacaact gccttgatat cctggccgaa 5760

gcccacggaa cccgacccga cctaacggac cagccgctcc cagacgccga ccacacctgg 5820

tacacggatg gaagcagtct cttacaagag ggacagcgta aggcgggagc tgcggtgacc 5880

accgagaccg aggtaatctg ggctaaagcc ctgccagccg ggacatccgc tcagcgggct 5940

gaactgatag cactcaccca ggccctaaag atggcagaag gtaagaagct aaatgtttat 6000

actgatagcc gttatgcttt tgctactgcc catatccatg gagaaatata cagaaggcgt 6060

gggtggctca catcagaagg caaagagatc aaaaataaag acgagatctt ggccctacta 6120

aaagccctct ttctgcccaa aagacttagc ataatccatt gtccaggaca tcaaaaggga 6180

cacagcgccg aggctagagg caaccggatg gctgaccaag cggcccgaaa ggcagccatc 6240

acagagactc cagacacctc taccctcctc atagaaaatt catcaccctc tggcggctca 6300

aaaagaaccg ccgacggcag cgaattcgag cccaagaaga agaggaaagt ctaaccggtc 6360
atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 6420
gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 6480
ccactgtcct ttcctaataa aatgagaaaa ttgcatcgca ttgtctgagt aggtgtcatt 6540
ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 6600
ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 6660
cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 6720

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 6780
cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6840
tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6900
gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6960
ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 7020
caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 7080
aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 7140

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 7200
cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 7260
ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 7320
gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg 7380
accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 7440
cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 7500
cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 7560

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 7620
aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 7680
aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 7740
actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 7800
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7860
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7920
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7980

ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 8040

accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 8100

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 8160

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 8220

tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 8280

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 8340

tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 8400

ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 8460

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 8520

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 8580

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 8640

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 8700

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 8760

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 8820

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8880

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8940

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 9000

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 9060

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 9120

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 9180

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 9240

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 9300

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 9360

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 9420

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 9480

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 9540

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 9600

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 9660

aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 9720

ctaatacgac tcactatagg gagagccgcc acc 9753

<210> 735

<211> 2305

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 735

gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60

ataattagaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240

cgaaacaccg gcccagactg agcacgtgag ttttagagct agaaatagca agttaaaata 300

aggctagtcc gttatcaact tgaaaaagtg ggaccgagtc ggtcctctgc catcaaagcg 360

tgctcagtct gtttttttaa gcttgggccg ctcgaggtac ctctctacat atgacatgtg 420

agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca 480

taggctccgc ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa 540

cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc 600

tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc 660

gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct 720

gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg 780

tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag 840

gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt ggcctaacta 900

cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg 960

aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt 1020

tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt 1080

ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt tggtcatgag 1140

attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat 1200

ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc 1260

tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgtagat 1320

aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac cgcgagaccc 1380

acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag 1440

aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag 1500

agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta caggcatcgt 1560

ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg 1620

agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt 1680

tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac tgcataattc 1740

tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc 1800

attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa 1860

taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg 1920

aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca ctcgtgcacc 1980

caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag 2040

gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt 2100

cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt 2160

tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc 2220

acctgacgtc gctagctgta caaaaaagca ggctttaaag gaaccaattc agtcgactgg 2280

atccggtacc aaggtcgggc aggaa 2305

<210> 736

<211> 11591

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 736

gtcgacggat cgggagatct cccgatcccc tatggtgcac tctcagtaca atctgctctg 60

atgccgcata gttaagccag tatctgctcc ctgcttgtgt gttggaggtc gctgagtagt 120

gcgcgagcaa aatttaagct acaacaaggc aaggcttgac cgacaattgc atgaagaatc 180

tgcttagggt taggcgtttt gcgctgcttc gcgatgtacg ggccagatat acgcgttgac 240

attgattatt gactagttat taatagtaat caattacggg gtcattagtt catagcccat 300

atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga ccgcccaacg 360

acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca atagggactt 420

tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca gtacatcaag 480

tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg cccgcctggc 540

attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc tacgtattag 600

tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt ggatagcggt 660

ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt ttgttttggc 720

accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg 780

gcggtaggcg tgtacggtgg gaggtctata taagcagcgc gttttgcctg tactgggtct 840

ctctggttag accagatctg agcctgggag ctctctggct aactagggaa cccactgctt 900

aagcctcaat aaagcttgcc ttgagtgctt caagtagtgt gtgcccgtct gttgtgtgac 960

tctggtaact agagatccct cagacccttt tagtcagtgt ggaaaatctc tagcagtggc 1020

gcccgaacag ggacttgaaa gcgaaaggga aaccagagga gctctctcga cgcaggactc 1080

ggcttgctga agcgcgcacg gcaagaggcg aggggcggcg actggtgagt acgccaaaaa 1140

ttttgactag cggaggctag aaggagagag atgggtgcga gagcgtcagt attaagcggg 1200

ggagaattag atcgcgatgg gaaaaaattc ggttaaggcc agggggaaag aaaaaatata 1260

aattaaaaca tatagtatgg gcaagcaggg agctagaacg attcgcagtt aatcctggcc 1320

tgttagaaac atcagaaggc tgtagacaaa tactgggaca gctacaacca tcccttcaga 1380

caggatcaga agaacttaga tcattatata atacagtagc aaccctctat tgtgtgcatc 1440

aaaggataga gataaaagac accaaggaag ctttagacaa gatagaggaa gagcaaaaca 1500

aaagtaagac caccgcacag caagcggccg ctgatcttca gacctggagg aggagatatg 1560

agggacaatt ggagaagtga attatataaa tataaagtag taaaaattga accattagga 1620

gtagcaccca ccaaggcaaa gagaagagtg gtgcagagag aaaaaagagc agtgggaata 1680

ggagctttgt tccttgggtt cttgggagca gcaggaagca ctatgggcgc agcgtcaatg 1740

acgctgacgg tacaggccag acaattattg tctggtatag tgcagcagca gaacaatttg 1800

ctgagggcta ttgaggcgca acagcatctg ttgcaactca cagtctgggg catcaagcag 1860

ctccaggcaa gaatcctggc tgtggaaaga tacctaaagg atcaacagct cctggggatt 1920

tggggttgct ctggaaaact catttgcacc actgctgtgc cttggaatgc tagttggagt 1980

aataaatctc tggaacagat ttggaatcac acgacctgga tggagtggga cagagaaatt 2040

aacaattaca caagcttaat acactcctta attgaagaat cgcaaaacca gcaagaaaag 2100

aatgaacaag aattattgga attagataaa tgggcaagtt tgtggaattg gtttaacata 2160

acaaattggc tgtggtatat aaaattattc ataatgatag taggaggctt ggtaggttta 2220

agaatagttt ttgctgtact ttctatagtg aatagagtta ggcagggata ttcaccatta 2280

tcgtttcaga cccacctccc aaccccgagg ggacccgaca ggcccgaagg aatagaagaa 2340

gaaggtggag agagagacag agacagatcc attcgattag tgaacggatc ggcactgcgt 2400

gcgccaattc tgcagacaaa tggcagtatt catccacaat tttaaaagaa aaggggggat 2460

tgggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca tacaaactaa 2520

agaattacaa aaacaaatta caaaaattca aaattttcgg gtttattaca gggacagcag 2580

agatccagtt tggttaatta aggtaccaag gtcgggcagg aagagggcct atttcccatg 2640

attccttcat atttgcatat acgatacaag gctgttagag agataattag aattaatttg 2700

actgtaaaca caaagatatt agtacaaaat acgtgacgta gaaagtaata atttcttggg 2760

tagtttgcag ttttaaaatt atgttttaaa atggactatc atatgcttac cgtaacttga 2820

aagtatttcg atttcttggc tttatatatc ttgtggaaag gacgaaacac cgcgggctgg 2880

agctgttcgc gcgttttaga gctagaaata gcaagttaaa ataaggctag tccgttatca 2940

acttgaaaaa gtggcaccga gtcggtgcaa gatgcaagcg cgaacagctc cagttttttt 3000

gaattcagtg caagtgggtt ttaggaccag gatgaggcgg ggtgggggtg cctacctgac 3060

gaccgacccc gacccactgg acaagcaccc aacccccatt ccccaaattg cgcatcccct 3120

atcagagagg gggaggggaa acaggatgcg gcgaggcgcg tgcgcactgc cagcttcagc 3180

accgcggaca gtgccttcgc ccccgcctgg cggcgcgcgc caccgccgcc tcagcactga 3240

aggcgcgctg acgtcactcg ccggtccccc gcaaactccc cttcccggcc accttggtcg 3300

cgtccgcgcc gccgccggcc cagccggacc gcaccacgcg aggcgcgaga taggggggca 3360

cgggcgcgac catctgcgct gcggcgccgg cgactcagcg ctgcctcagt ctgcggtggg 3420

cagcggagga gtcgtgtcgt gcctgagagc gcagtcgaga atctagagcg ctgccaccat 3480

gaaacggaca gccgacggaa gcgagttcga gtcaccaaag aagaagcgga aagtcgacaa 3540

gaagtacagc atcggcctgg acatcggcac caactctgtg ggctgggccg tgatcaccga 3600

cgagtacaag gtgcccagca agaaattcaa ggtgctgggc aacaccgacc ggcacagcat 3660

caagaagaac ctgatcggag ccctgctgtt cgacagcggc gaaacagccg aggccacccg 3720

gctgaagaga accgccagaa gaagatacac cagacggaag aaccggatct gctatctgca 3780

agagatcttc agcaacgaga tggccaaggt ggacgacagc ttcttccaca gactggaaga 3840

gtccttcctg gtggaagagg ataagaagca cgagcggcac cccatcttcg gcaacatcgt 3900

ggacgaggtg gcctaccacg agaagtaccc caccatctac cacctgagaa agaaactggt 3960

ggacagcacc gacaaggccg acctgcggct gatctatctg gccctggccc acatgatcaa 4020

gttccggggc cacttcctga tcgagggcga cctgaacccc gacaacagcg acgtggacaa 4080

gctgttcatc cagctggtgc agacctacaa ccagctgttc gaggaaaacc ccatcaacgc 4140

cagcggcgtg gacgccaagg ccatcctgtc tgccagactg agcaagagca gacggctgga 4200

aaatctgatc gcccagctgc ccggcgagaa gaagaatggc ctgttcggaa acctgattgc 4260

cctgagcctg ggcctgaccc ccaacttcaa gagcaacttc gacctggccg aggatgccaa 4320

actgcagctg agcaaggaca cctacgacga cgacctggac aacctgctgg cccagatcgg 4380

cgaccagtac gccgacctgt ttctggccgc caagaacctg tccgacgcca tcctgctgag 4440

cgacatcctg agagtgaaca ccgagatcac caaggccccc ctgagcgcct ctatgatcaa 4500

gagatacgac gagcaccacc aggacctgac cctgctgaaa gctctcgtgc ggcagcagct 4560

gcctgagaag tacaaagaga ttttcttcga ccagagcaag aacggctacg ccggctacat 4620

tgacggcgga gccagccagg aagagttcta caagttcatc aagcccatcc tggaaaagat 4680

ggacggcacc gaggaactgc tcgtgaagct gaacagagag gacctgctgc ggaagcagcg 4740

gaccttcgac aacggcagca tcccccacca gatccacctg ggagagctgc acgccattct 4800

gcggcggcag gaagattttt acccattcct gaaggacaac cgggaaaaga tcgagaagat 4860

cctgaccttc cgcatccccc actacgtggg ccctctggcc aggggaaaca gcagattcgc 4920

ctggatgacc agaaagagcg aggaaaccat caccccctgg aacttcgagg aagtggtgga 4980

caagggcgct tccgcccaga gcttcatcga gcggatgacc aacttcgata agaacctgcc 5040

caacgagaag gtgctgccca agcacagcct gctgtacgag tacttcaccg tgtataacga 5100

gctgaccaaa gtgaaatacg tgaccgaggg aatgagaaag cccgccttcc tgagcggcga 5160

gcagaaaaag gccatcgtgg acctgctgtt caagaccaac cggaaagtga ccgtgaagca 5220

gctgaaagag gactacttca agaaaatcga gtgcctgtcc tacgagacag agatcctgac 5280

agtggagtat ggcctgctgc caatcggcaa gatcgtggag aagaggatcg agtgtaccgt 5340

gtactctgtg gataacaatg gcaacatcta tacacagccc gtggcacagt ggcacgatag 5400

gggagagcag gaggtgttcg agtattgcct ggaggacggc agcctgatca gggcaaccaa 5460

ggaccacaag ttcatgacag tggatggcca gatgctgccc atcgacgaga ttttcgagcg 5520

ggagctggac ctgatgagag tggataacct gcctaatagc ggaggcagta aaagaacagc 5580

agacgggagt gagtttgagc ccaagaaaaa gagaaaggtg ggatccggcg caacaaactt 5640

ctctctgctg aaacaagccg gagatgtcga agagaatcct ggaccggtga gcaagggcga 5700

ggagctgttc accggggtgg tgcccatcct ggtcgagctg gacggcgacg taaacggcca 5760

caagttcagc gtgtccggcg agggcgaggg cgatgccacc tacggcaagc tgaccctgaa 5820

gttcatctgc accaccggca agctgcccgt gccctggccc accctcgtga ccaccctgac 5880

ctacggcgtg cagtgcttca gccgctaccc cgaccacatg aagcagcacg acttcttcaa 5940

gtccgccatg cccgaaggct acgtccagga gcgcaccatc ttcttcaagg acgacggcaa 6000

ctacaagacc cgcgccgagg tgaagttcga gggcgacacc ctggtgaacc gcatcgagct 6060

gaagggcatc gacttcaagg aggacggcaa catcctgggg cacaagctgg agtacaacta 6120

caacagccac aacgtctata tcatggccga caagcagaag aacggcatca aggtgaactt 6180

caagatccgc cacaacatcg aggacggcag cgtgcagctc gccgaccact accagcagaa 6240

caccccatc ggcgacggcc ccgtgctgct gcccgacaac cactacctga gcacccagtc 6300
cgccctgagc aaagacccca acgagaagcg cgatcacatg gtcctgctgg agttcgtgac 6360
cgccgccggg atcactctcg gcatggacga gctgtacaag tccggactca gatctcgaga 6420
ggaggaggag gagacagaca gcaggatgcc ccacctcgac agccccggca gctcccagcc 6480
gagacgctcc ttcctctcaa gggtgatcag ggcagcgcta ccgttgcagc tgcttctgct 6540
gctgctgctg ctcctggcct gcctgctacc tgcctctgaa gatgactaca gctgcaccca 6600
ggccaacaac tttgcccgat ccttctaccc catgctgcgg tacaccaacg ggccacctcc 6660

cacctaaacg cgttaagtcg acaatcaacc tctggattac aaaatttgtg aaagattgac 6720
tggtattctt aactatgttg ctccttttac gctatgtgga tacgctgctt taatgccttt 6780
gtatcatgct attgcttccc gtatggcttt cattttctcc tccttgtata aatcctggtt 6840
gctgtctctt tatgaggagt tgtggcccgt tgtcaggcaa cgtggcgtgg tgtgcactgt 6900
gtttgctgac gcaaccccca ctggttgggg cattgccacc acctgtcagc tcctttccgg 6960
gactttcgct ttccccctcc ctattgccac ggcggaactc atcgccgcct gccttgcccg 7020
ctgctggaca ggggctcggc tgttgggcac tgacaattcc gtggtgttgt cggggaaatc 7080

atcgtccttt ccttggctgc tcgcctgtgt tgccacctgg attctgcgcg ggacgtcctt 7140
ctgctacgtc ccttcggccc tcaatccagc ggaccttcct tcccgcggcc tgctgccggc 7200
tctgcggcct cttccgcgtc ttcgccttcg ccctcagacg agtcggatct ccctttgggc 7260
cgcctccccg cgtcgacttt aagaccaatg acttacaagg cagctgtaga tcttagccac 7320
tttttaaaag aaaaggggag actggaaggg ctaattcact cccaacgaag acaagatctg 7380
ctttttgctt gtactgggtc tctctggtta gaccagatct gagcctggga gctctctggc 7440
taactaggga acccactgct taagcctcaa taaagcttgc cttgagtgct tcaagtagtg 7500

tgtgcccgtc tgttgtgtga ctctggtaac tagagatccc tcagaccctt ttagtcagtg 7560
tggaaaatct ctagcagggc ccgtttaaac ccgctgatca gcctcgactg tgccttctag 7620
ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac 7680
tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca 7740
ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag 7800
caggcatgct ggggatgcgg tgggctctat ggcttctgag gcggaaagaa ccagctgggg 7860
ctctaggggg tatccccacg cgccctgtag cggcgcatta agcgcggcgg gtgtggtggt 7920

tacgcgcagc gtgaccgcta cacttgccag cgccctagcg cccgctcctt tcgctttctt 7980

cccttccttt ctcgccacgt tcgccggctt tccccgtcaa gctctaaatc gggggctccc 8040

tttagggttc cgatttagtg ctttacggca cctcgacccc aaaaaacttg attagggtga 8100

tggttcacgt agtgggccat cgccctgata gacggttttt cgccctttga cgttggagtc 8160

cacgttcttt aatagtggac tcttgttcca aactggaaca acactcaacc ctatctcggt 8220

ctattctttt gatttataag ggattttgcc gatttcggcc tattggttaa aaaatgagct 8280

gatttaacaa aaatttaacg cgaattaatt ctgtggaatg tgtgtcagtt agggtgtgga 8340

aagtccccag gctccccagc aggcagaagt atgcaaagca tgcatctcaa ttagtcagca 8400

accaggtgtg gaaagtcccc aggctcccca gcaggcagaa gtatgcaaag catgcatctc 8460

aattagtcag caaccatagt cccgccccta actccgccca tcccgcccct aactccgccc 8520

agttccgccc attctccgcc ccatggctga ctaatttttt ttatttatgc agaggccgag 8580

gccgcctctg cctctgagct attccagaag tagtgaggag gcttttttgg aggcctaggc 8640

ttttgcaaaa agctcccggg agcttgtata tccattttcg gatctgatca gcacgtgttg 8700

acaattaatc atcggcatag tatatcggca tagtataata cgacaaggtg aggaactaaa 8760

ccatggccaa gttgaccagt gccgttccgg tgctcaccgc gcgcgacgtc gccggagcgg 8820

tcgagttctg gaccgaccgg ctcgggttct cccgggactt cgtggaggac gacttcgccg 8880

gtgtggtccg ggacgacgtg accctgttca tcagcgcggt ccaggaccag gtggtgccgg 8940

acaacaccct ggcctgggtg tgggtgcgcg gcctggacga gctgtacgcc gagtggtcgg 9000

aggtcgtgtc cacgaacttc cgggacgcct ccgggccggc catgaccgag atcggcgagc 9060

agccgtgggg gcgggagttc gccctgcgcg acccggccgg caactgcgtg cacttcgtgg 9120

ccgaggagca ggactgacac gtgctacgag atttcgattc caccgccgcc ttctatgaaa 9180

ggttgggctt cggaatcgtt ttccgggacg ccggctggat gatcctccag cgcggggatc 9240

tcatgctgga gttcttcgcc caccccaact tgtttattgc agcttataat ggttacaaat 9300

aaagcaatag catcacaaat ttcacaaata aagcattttt ttcactgcat tctagttgtg 9360

gtttgtccaa actcatcaat gtatcttatc atgtctgtat accgtcgacc tctagctaga 9420

gcttggcgta atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc 9480

cacacaacat acgagccgga agcataaagt gtaaagcctg gggtgcctaa tgagtgagct 9540

aactcacatt aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc 9600

agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt 9660

ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag 9720

ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca 9780

tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt 9840

tccataggct ccgccccccct gacgagcatc acaaaaatcg acgctcaagt cagaggtggc 9900

gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc ctcgtgcgct 9960

ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg 10020

tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca 10080

agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta tccggtaact 10140

atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta 10200

acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tggtggccta 10260

actacggcta cactagaaga acagtatttg gtatctgcgc tctgctgaag ccagttacct 10320

tcggaaaaag agttggtagc tcttgatccg gcaaacaaac caccgctggt agcggtggtt 10380

tttttgtttg caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga 10440

tcttttctac ggggtctgac gctcagtgga acgaaaactc acgttaaggg attttggtca 10500

tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga agttttaaat 10560

caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg 10620

cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc cccgtcgtgt 10680

agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 10740

acccacgctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 10800

gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 10860

ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca 10920

tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa 10980

ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctccttc ggtcctccga 11040

tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca gcactgcata 11100

attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 11160

agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 11220

ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa cgttcttcgg 11280

ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 11340

cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 11400

gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 11460

tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 11520

tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 11580

tgccacctga c 11591

<210> 737

<400> 737

000

<210> 738

<211> 8323

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 738

taacaaattg gctgtggtat ataaaattat tcataatgat agtaggaggc ttggtaggtt 60

taagaatagt ttttgctgta ctttctatag tgaatagagt taggcaggga tattcaccat 120

tatcgtttca gacccacctc ccaaccccga ggggacccag agagggccta tttcccatga 180

ttccttcata tttgcatata cgatacaagg ctgttagaga gataattaga attaatttga 240

ctgtaaacac aaagatatta gtacaaaata cgtgacgtag aaagtaataa tttcttgggt 300

agtttgcagt tttaaaatta tgttttaaaa tggactatca tatgcttacc gtaacttgaa 360

agtatttcga tttcttggct ttatatatct tgtggaaagg acgaaacacc gccgcgcgcg 420

cgaaaaagcc ggttttagag ctagaaatag caagttaaaa taaggctagt ccgttatcaa 480

cttgaaaaag tggcaccgag tcggtgcttt tttaagcttg gcgtaactag atcttgagac 540

aaatggcagt attcatccac aattttaaaa gaaaaggggg gattgggggg tacagtgcag 600

gggaaagaat agtagacata atagcaacag acatacaaac taaagaatta caaaaacaaa 660

ttacaaaaat tcaaaatttt cgggtttatt acagggacag cagagatcca ctttggcgcc 720

ggctcgaggg ggcccgggtg caaagatgga taaagtttta aacagagagg aatctttgca 780

gctaatggac cttctaggtc ttgaaaggag tgggaattgg ctccggtgcc cgtcagtggg 840

cagagcgcac atcgcccaca gtccccgaga agttgggggg aggggtcggc aattgatccg 900

gtgcctagag aaggtggcgc ggggtaaact gggaaagtga tgtcgtgtac tggctccgcc 960

ttttttcccga gggtggggga gaaccgtata taagtgcagt agtcgccgtg aacgttcttt 1020

ttcgcaacgg gtttgccgcc agaacacagg taagtgccgt gtgtggttcc cgcgggcctg 1080

gcctctttac gggttatggc ccttgcgtgc cttgaattac ttccacctgg ctgcagtacg 1140

tgattcttga tcccgagctt cgggttggaa gtgggtggga gagttcgagg ccttgcgctt 1200

aaggagcccc ttcgcctcgt gcttgagttg aggcctggcc tgggcgctgg ggccgccgcg 1260

tgcgaatctg gtggcacctt cgcgcctgtc tcgctgcttt cgataagtct ctagccattt 1320

aaaatttttg atgacctgct gcgacgcttt ttttctggca agatagtctt gtaaatgcgg 1380

gccaagatct gcacactggt atttcggttt ttggggccgc gggcggcgac ggggcccgtg 1440

cgtcccagcg cacatgttcg gcgaggcggg gcctgcgagc gcggccaccg agaatcggac 1500

gggggtagtc tcaagctggc cggcctgctc tggtgcctgg cctcgcgccg ccgtgtatcg 1560

ccccgccctg ggcggcaagg ctggcccggt cggcaccagt tgcgtgagcg gaaagatggc 1620

cgcttcccgg ccctgctgca gggagctcaa aatggaggac gcggcgctcg ggagagcggg 1680

cgggtgagtc acccacacaa aggaaaaggg cctttccgtc ctcagccgtc gcttcatgtg 1740

actccacgga gtaccgggcg ccgtccaggc acctcgatta gttctcgagc ttttggagta 1800

cgtcgtcttt aggttggggg gaggggtttt atgcgatgga gtttccccac actgagtggg 1860

tggagactga agttaggcca gcttggcact tgatgtaatt ctccttggaa tttgcccttt 1920

ttgagtttgg atcttggttc attctcaagc ctcagacagt ggttcaaagt ttttttcttc 1980

catttcaggt gtcgtgacgt acggccacca tgaccgagta caagcccacg gtgcgcctcg 2040

ccacccgcga cgacgtcccc agggccgtac gcaccctcgc cgccgcgttc gccgactacc 2100

ccgccacgcg ccacaccgtc gatccggacc gccacatcga gcgggtcacc gagctgcaag 2160

aactcttcct cacgcgcgtc gggctcgaca tcggcaaggt gtgggtcgcg gacgacggcg 2220

ccgccgtggc ggtctggacc acgccggaga gcgtcgaagc gggggcggtg ttcgccgaga 2280

tcggcccgcg catggccgag ttgagcggtt cccggctggc cgcgcagcaa cagatggaag 2340

gcctcctggc gccgcaccgg cccaaggagc ccgcgtggtt cctggccacc gtcggagtct 2400

cgcccgacca ccagggcaag ggtctgggca gcgccgtcgt gctccccgga gtggaggcgg 2460

ccgagcgcgc cggggtgccc gccttcctgg agacctccgc gccccgcaac ctccccttct 2520

acgagcggct cggcttcacc gtcaccgccg acgtcgaggt gcccgaagga ccgcgcacct 2580

ggtgcatgac ccgcaagccc ggtgcctgaa cgcgttaagt cgacaatcaa cctctggatt 2640

acaaaatttg tgaaagattg actggtattc ttaactatgt tgctcctttt acgctatgtg 2700

gatacgctgc tttaatgcct ttgtatcatg ctattgcttc ccgtatggct ttcattttct 2760

cctccttgta taaatcctgg ttgctgtctc tttatgagga gttgtggccc gttgtcaggc 2820

aacgtggcgt ggtgtgcact gtgtttgctg acgcaacccc cactggttgg ggcattgcca 2880

ccacctgtca gctcctttcc gggactttcg ctttcccccc ccctattgcc acggcggaac 2940

tcatcgccgc ctgccttgcc cgctgctgga caggggctcg gctgttgggc actgacaatt 3000

ccgtggtgtt gtcggggaaa tcatcgtcct ttccttggct gctcgcctgt gttgccacct 3060

ggattctgcg cgggacgtcc ttctgctacg tcccttcggc cctcaatcca gcggaccttc 3120

cttcccgcgg cctgctgccg gctctgcggc ctcttccgcg tcttcgcctt cgccctcaga 3180

cgagtcggat ctccctttgg gccgcctccc cgcgtcgact ttaagaccaa tgacttacaa 3240

ggcagctgta gatcttagcc actttttaaa agaaaagggg ggactggaag ggctaattca 3300

ctcccaacga agacaagatc tgctttttgc ttgtactggg tctctctggt tagaccagat 3360

ctgagcctgg gagctctctg gctaactagg gaacccactg cttaagcctc aataaagctt 3420

gccttgagtg cttcaagtag tgtgtgcccg tctgttgtgt gactctggta actagagatc 3480

cctcagaccc ttttagtcag tgtggaaaat ctctagcagt acgtatagta gttcatgtca 3540

tcttattatt cagtatttat aacttgcaaa gaaatgaata tcagagagtg agaggaactt 3600

gtttattgca gcttataatg gttacaaata aagcaatagc atcacaaatt tcacaaataa 3660

agcatttttt tcactgcatt ctagttgtgg tttgtccaaa ctcatcaatg tatcttatca 3720

tgtctggctc tagctatccc gcccctaact ccgcccatcc cgcccctaac tccgcccagt 3780

tccgcccatt ctccgcccca tggctgacta atttttttta tttatgcaga ggccgaggcc 3840

gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctagggacg 3900

tacccaattc gccctatagt gagtcgtatt acgcgcgctc actggccgtc gttttacaac 3960

gtcgtgactg ggaaaaccct ggcgttaccc aacttaatcg ccttgcagca catccccctt 4020

tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa cagttgcgca 4080

gcctgaatgg cgaatgggac gcgccctgta gcggcgcatt aagcgcggcg ggtgtggtgg 4140

ttacgcgcag cgtgaccgct acacttgcca gcgccctagc gcccgctcct ttcgctttct 4200

tcccttcctt tctcgccacg ttcgccggct ttccccgtca agctctaaat cgggggctcc 4260

ctttagggtt ccgatttagt gctttacggc acctcgaccc caaaaaactt gattagggtg 4320

atggttcacg tagtgggcca tcgccctgat agacggtttt tcgccctttg acgttggagt 4380

ccacgttctt taatagtgga ctcttgttcc aaactggaac aacactcaac cctatctcgg 4440

tctattcttt tgatttataa gggattttgc cgatttcggc ctattggtta aaaaatgagc 4500

tgatttaaca aaaatttaac gcgaatttta acaaaatatt aacgcttaca atttaggtgg 4560

cacttttcgg ggaaatgtgc gcggaacccc tatttgttta tttttctaaa tacattcaaa 4620

tatgtatccg ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa 4680

gagtatgagt attcaacatt tccgtgtcgc ccttattccc ttttttgcgg cattttgcct 4740

tcctgttttt gctcacccag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg 4800

tgcacgagtg ggttacatcg aactggatct caacagcggt aagatccttg agagttttcg 4860

ccccgaagaa cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt 4920

atcccgtatt gacgccgggc aagagcaact cggtcgccgc atacactatt ctcagaatga 4980

cttggttgag tactcaccag tcacagaaaa gcatcttacg gatggcatga cagtaagaga 5040
attatgcagt gctgccataa ccatgagtga taacactgcg gccaacttac ttctgacaac 5100
gatcggagga ccgaaggagc taaccgcttt tttgcacaac atgggggatc atgtaactcg 5160
ccttgatcgt tgggaaccgg agctgaatga agccatacca aacgacgagc gtgacaccac 5220
gatgcctgta gcaatggcaa caacgttgcg caaactatta actggcgaac tacttactct 5280
agcttcccgg caacaattaa tagactggat ggaggcggat aaagttgcag gaccacttct 5340
gcgctcggcc cttccggctg gctggtttat tgctgataaa tctggagccg gtgagcgtgg 5400

gtctcgcggt atcattgcag cactggggcc agatggtaag ccctcccgta tcgtagttat 5460
ctacacgacg gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg 5520
tgcctcactg attaagcatt ggtaactgtc agaccaagtt tactcatata tactttagat 5580
tgatttaaaa cttcattttt aatttaaaag gatctaggtg aagatccttt ttgataatct 5640
catgaccaaa atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa 5700
gatcaaagga tcttcttgag atccttttt tctgcgcgta atctgctgct tgcaaacaaa 5760
aaaaccaccg ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc 5820

gaaggtaact ggcttcagca gagcgcagat accaaatact gttcttctag tgtagccgta 5880
gttaggccac cacttcaaga actctgtagc accgcctaca tacctcgctc tgctaatcct 5940
gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg 6000
atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag 6060
cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat gagaaagcgc 6120
cacgcttccc gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg 6180
agagcgcacg agggagcttc cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt 6240

tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg 6300
gaaaaacgcc agcaacgcgg cctttttacg gttcctggcc ttttgctggc cttttgctca 6360
catgttcttt cctgcgttat cccctgattc tgtggataac cgtattaccg cctttgagtg 6420
agctgatacc gctcgccgca gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc 6480
ggaagagcgc ccaatacgca aaccgcctct ccccgcgcgt tggccgattc attaatgcag 6540
ctggcacgac aggtttcccg actggaaagc gggcagtgag cgcaacgcaa ttaatgtgag 6600
ttagctcact cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg 6660

tggaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa 6720

gcgcgcaatt aaccctcact aaagggaaca aaagctggag ctgcaagctt aatgtagtct 6780

tatgcaatac tcttgtagtc ttgcaacatg gtaacgatga gttagcaaca tgccttacaa 6840

ggagagaaaa agcaccgtgc atgccgattg gtggaagtaa ggtggtacga tcgtgcctta 6900

ttaggaaggc aacagacggg tctgacatgg attggacgaa ccactgaatt gccgcattgc 6960

agagatattg tatttaagtg cctagctcga tacataaacg ggtctctctg gttagaccag 7020

atctgagcct gggagctctc tggctaacta gggaacccac tgcttaagcc tcaataaagc 7080

ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt gtgactctgg taactagaga 7140

tccctcagac ccttttagtc agtgtggaaa atctctagca gtggcgcccg aacagggact 7200

tgaaagcgaa agggaaacca gaggagctct ctcgacgcag gactcggctt gctgaagcgc 7260

gcacggcaag aggcgagggg cggcgactgg tgagtacgcc aaaaattttg actagcggag 7320

gctagaagga gagagatggg tgcgagagcg tcagtattaa gcgggggaga attagatcgc 7380

gatgggaaaa aattcggtta aggccagggg gaaagaaaaa atataaatta aaacatatag 7440

tatgggcaag cagggagcta gaacgattcg cagttaatcc tggcctgtta gaaacatcag 7500

aaggctgtag acaaatactg ggacagctac aaccatccct tcagacagga tcagaagaac 7560

ttagatcatt atataataca gtagcaaccc tctattgtgt gcatcaaagg atagagataa 7620

aagacaccaa ggaagcttta gacaagatag aggaagagca aaacaaaagt aagaccaccg 7680

cacagcaagc ggccgctgat cttcagacct ggaggaggag atatgaggga caattggaga 7740

agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc acccaccaag 7800

gcaaagagaa gagtggtgca gagagaaaaa agagcagtgg gaataggagc tttgttcctt 7860

gggttcttgg gagcagcagg aagcactatg ggcgcagcgt caatgacgct gacggtacag 7920

gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag ggctattgag 7980

gcgcaacagc atctgttgca actcacagtc tggggcatca agcagctcca ggcaagaatc 8040

ctggctgtgg aaagatacct aaaggatcaa cagctcctgg ggatttgggg ttgctctgga 8100

aaactcattt gcaccactgc tgtgccttgg aatgctagtt ggagtaataa atctctggaa 8160

cagatttgga atcacacgac ctggatggag tgggacagag aaattaacaa ttacacaagc 8220

ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaagaatga acaagaatta 8280

ttggaattag ataaatgggc aagtttgtgg aattggttta aca 8323

<210> 739

<211> 677

<212> PRT

<213> Unknown

<220><223> moloney murine leukemia virus

<400> 739

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190

Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys

565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys

625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 740
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 740

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu

1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln
500 505 510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala

645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210> 741
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 741
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala

20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu
515 520 525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr

530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg
545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly

595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile

660 665 670
Glu Asn Ser Ser Pro
675

<210> 742
<211> 677
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 742
Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu

35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr

50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg

65 70 75 80

Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr

85 90 95

Pro Leu Leu Pro Val Lys Leu Pro Gly Thr Asn Asp Tyr Ser Pro Val

100 105 110

Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr

115 120 125

Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln

130 135 140

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu

145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu

165 170 175

Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe

180 185 190

Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala

195 200 205

Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp

210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr

225 230 235 240

Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala

245 250 255

Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu

260 265 270

Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val

275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305 310 315 320

Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp

325 330 335

Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu

340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu

355 360 365

Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

370 375 380

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp

385 390 395 400

Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile

405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu

420 425 430

Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro

435 440 445

Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu

450 455 460

Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro

465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu

485 490 495

Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln

500 505 510

Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu

515 520 525

Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr
530 535 540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg

545 550 555 560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu

610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
660 665 670
Glu Asn Ser Ser Pro
675

<210
> 743
<400> 743
000
<210> 744
<400> 744
000
<210> 745
<400> 745
000
<210> 746
<400> 746
000

<210> 747

<400> 747

000

<210> 748

<400> 748

000

<210> 749

<400> 749

000

<210> 750

<400> 750

000

<210> 751

<400> 751

000

<210> 752

<400> 752

000

<210> 753

<400> 753

000

<210> 754

<400> 754

000

<210> 755

<400> 755

000

<210> 756

<400> 756

000

<210> 757

<400> 757

000

<210> 758

<400> 758

000

<210>

759

<400> 759

000

<210> 760

<400> 760

000

<210> 761

<400> 761

000

<210> 762

<400> 762

000

<210> 763

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 763

gccaacatga ggatcaccca tgtctgcagg gcc 33

<210> 764

<211> 131

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 764

Gly Ser Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly
1 5 10 15

Thr Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Val Ala
20 25 30

Glu Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys
35 40 45

Ser Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val
50 55 60

Glu Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Glu Glu Leu Pro
65 70 75 80

Val Ala Gly Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile
85 90 95
Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly
100 105 110
Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser
115 120 125
Gly Ile Tyr
130

<210> 765
<211> 1943
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic

<400> 765
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15
Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn
20 25 30
Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys
35 40 45
Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn
50 55 60

Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr
65 70 75 80
Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg
85 90 95
Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp
100 105 110
Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp

115 120 125

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val
130 135 140
Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu
145 150 155 160
Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu
165 170 175
Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu
180 185 190

Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln
195 200 205
Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val
210 215 220
Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu
225 230 235 240
Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe
245 250 255

Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser
260 265 270
Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr
275 280 285
Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr
290 295 300
Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu
305 310 315 320

Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser
325 330 335
Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu
340 345 350
Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile
355 360 365

Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly
370 375 380

Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys
385 390 395 400
Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu
405 410 415
Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile
420 425 430
His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr
435 440 445

Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe
450 455 460
Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe
465 470 475 480
Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe
485 490 495
Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg
500 505 510

Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys
515 520 525
His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys
530 535 540
Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly
545 550 555 560
Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys
565 570 575

Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys
580 585 590
Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser
595 600 605
Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe

610 615 620

Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr
625 630 635 640

Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr
645 650 655

Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg
660 665 670

Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile
675 680 685

Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp
690 695 700

Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu
705 710 715 720

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp
725 730 735

Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys
740 745 750

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val
755 760 765

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu
770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
785 790 795 800

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu
805 810 815

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr
820 825 830

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe
850 855 860

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys
865 870 875 880
Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys
885 890 895

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln
900 905 910
Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
915 920 925
Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln
930 935 940
Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
945 950 955 960

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
965 970 975
Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
980 985 990
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
995 1000 1005
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1010 1015 1020

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1025 1030 1035
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1040 1045 1050
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1055 1060 1065
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1070 1075 1080

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1085 1090 1095
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1100 1105 1110

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1115 1120 1125

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1130 1135 1140

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala

1145 1150 1155

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1160 1165 1170

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1175 1180 1185

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1190 1195 1200

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1205 1210 1215

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1220 1225 1230

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1235 1240 1245

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1250 1255 1260

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1265 1270 1275

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1280 1285 1290

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1295 1300 1305

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1310 1315 1320

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1325 1330 1335

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1340 1345 1350
Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1355 1360 1365
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1370 1375 1380

Gly Gly Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu
1385 1390 1395
Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly
1400 1405 1410
Ser Ser Gly Gly Ser Ser Thr Leu Asn Ile Glu Asp Glu Tyr Arg
1415 1420 1425
Leu His Glu Thr Ser Lys Glu Pro Asp Val Ser Leu Gly Ser Thr
1430 1435 1440

Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly Met
1445 1450 1455
Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala
1460 1465 1470
Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu
1475 1480 1485
Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln
1490 1495 1500

Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu
1505 1510 1515
Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln Asp
1520 1525 1530
Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val
1535 1540 1545
Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
1550 1555 1560

Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg

1565 1570 1575

Leu His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp

1580 1585 1590

Pro Glu Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro

1595 1600 1605

Gln Gly Phe Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His

1610 1615 1620

Arg Asp Leu Ala Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu

1625 1630 1635

Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu

1640 1645 1650

Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr Leu Gly Asn

1655 1660 1665

Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys Gln Lys

1670 1675 1680

Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg Trp

1685 1690 1695

Leu Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro

1700 1705 1710

Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly Lys Ala Gly Phe

1715 1720 1725

Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu

1730 1735 1740

Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp Gly Pro Asp

1745 1750 1755

Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala

1760 1765 1770

Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu Phe

1775 1780 1785

Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys

1790 1795 1800

Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu
1805 1810 1815

Asp Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala
1820 1825 1830

Ala Ile Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly
1835 1840 1845

Gln Pro Leu Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val
1850 1855 1860

Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His
1865 1870 1875

Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val Gln Phe Gly Pro
1880 1885 1890

Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu Pro Glu Glu
1895 1900 1905

Gly Leu Gln His Asn Cys Leu Asp Asn Ser Arg Leu Ile Asn Ser
1910 1915 1920

Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys
1925 1930 1935

Lys Lys Arg Lys Val
1940

<210> 766
<211> 503
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 766

Thr Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15

Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30

Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45

Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
85 90 95

Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu
145 150 155 160

His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
180 185 190
Lys Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala
195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220

Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
275 280 285

Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu

290 295 300

Gly Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met
305 310 315 320
Ala Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp
325 330 335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu
340 345 350

Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu
355 360 365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys
370 375 380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp
385 390 395 400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile
405 410 415

Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu
420 425 430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro
435 440 445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu
450 455 460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro
465 470 475 480

Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu
485 490 495
Asp Asn Ser Arg Leu Ile Asn
500

<210> 767
<211> 3
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic

<400> 767

Gly Gly Ser

1

<210> 768

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 768

Gly Gly Ser Gly Gly Ser

1 5

<210> 769

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 769

Gly Gly Ser Gly Gly Ser Gly Gly Ser

1 5

<210> 770

<400> 770

000

<210> 771

<400> 771

000

<210> 772

<400> 772

000

<210> 773

<400> 773

000

<210> 774

<400> 774

000

<210> 775

<400> 775

000

<210> 776

<400> 776

000

<210> 777

<400> 777

000

<210> 778

<400> 778

000

<210> 779

<400> 779

000

<210> 780

<400> 780

000

<210> 781

<211>

48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 781

aggaaccccg ggtctgactc gtgcactacc gtctcctttc ctctgcag 48

<210> 782

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 782

Ser Leu Gly Pro Arg Leu Ser Thr

1 5

<210> 783

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 783

aaccccgggt ctgactcgtg cactaccgtc tcgtgtcctc tgcag 45

<210> 784

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 784

Leu Gly Pro Arg Leu Ser Thr

1 5

<210> 785

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 785

tccttggggc ccagactgag cacgtgatgg cagaggaaag gagacgtc 48

<210> 786

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 786

tccttggggc ccagactgag cacgagatgg cagaggaaag gagacgtc 48

<210> 787

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 787

ttggggccca gactgagcac gtgatggcag agcacaggag acgtc 45

<210> 788

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 788

ttggggccca gactgagcac gagatggcag agcacaggag acgtc 45

<210> 789

<211> 297

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 789

atgtgggctg cctagaaagg catggatgag agaagcctgg agacagggat cccagggaaa 60

cgcccatgca attagtctat ttctgctgca agtaagcatg catttgtagg cttgatgctt 120

tttttctgct tctccagccc tggcctgggt caatccttgg ggcccagact gagcacgtga 180

tggcagagga aaggaagccc tgcttcctcc agagggcgtc gcaggacagc ttttcctaga 240

caggggctag tatgtgcagc tcctgcaccg ggatactggt tgacaagttt ggctggg 297

<210> 790

<211> 297

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 790

cccagccaaa cttgtcaacc agtatcccgg tgcaggagct gcacatacta gcccctgtct 60

aggaaaagct gtcctgcgac gccctctgga ggaagcaggg cttcctttcc tctgccatca 120

cgtgctcagt ctgggcccca aggattgacc caggccaggg ctggagaagc agaaaaaaag 180

catcaagcct acaaatgcat gcttacttgc agcagaaata gactaattgc atgggcgttt 240

ccctgggatc cctgtctcca ggcttctctc atccatgcct ttctaggcag cccacat 297

<210> 791

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 791

tggggcccag actgagcacg tgatggcaga ggaaaggaag 40

<210> 792

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 792

tggggcccag actgagcacg tgatggcaaa ggaaaggaag 40

<210> 793

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 793

tggggcccag actgagcacg agatggcaga ggaaaggaag 40

<210> 794

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 794

tggggcccag actgagcacg agatggcaaa ggaaaggaag 40

<210> 795

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223

> Synthetic

<400> 795

ctgagcacgt gatggcagag gaaaggaag 29

<210> 796

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 796

ctgagcacgg gaaaggaag 19

<210> 797

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 797

ctgagcacgt gatggcaaag gaaaggaag 29

<210> 798

<211> 9

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 798

ctgagcacg 9

<210> 799

<211> 16

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 799

ctgagcacgg gaaagg 16

<210> 800

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 800

ctgagcacgt gatggcaggg gaaaggaag 29

<210> 801

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 801

gactgagcac gtgatggcag aggaaaggaa g 31

<210> 802

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 802

gactgagcac gcaccatcat caccatcatt g 31

<210> 803

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 803

gactgagcac gtgatggcaa aggaaaggaa g 31

<210> 804

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 804

gactgagcac ggattacaag gatgacgacg a 31

<210> 805

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 805

Gly Leu Gly Gly Thr Met Leu

1 5

<210> 806

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 806

ggccttggcg gctacatgct g 21

<210> 807

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 807

ggccttggcg tctacatgct g 21

<210> 808

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 808

Gly Leu Gly Val Thr Met Leu

1 5

<210> 809

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 809

ggggcagtgg tgggggcct tggcggctac atgctg 36

<210> 810

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 810

gcagtggtgg ggggccttgg atgtagacgc caaggccccc cac 43

<210> 811

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 811

Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala

1 5 10 15

Asp His Tyr Gln Gln Asn Thr Pro Ile

20 25

<210> 812

<211>

24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 812

Asn Phe Lys Ile His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp

1 5 10 15

His Tyr Gln Gln Asn Thr Pro Ile

20

<210> 813

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 813

Asn Phe Lys Ile Arg Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp

1 5 10 15

His Tyr Gln Gln Asn Thr Pro Ile
20

<210> 814

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 814

Asn Phe Lys Ile Arg His Asn Glu Asp Gly Ser Val Gln Leu Ala Asp
1 5 10 15

His Tyr Gln Gln Asn Thr Pro Ile
20

<210> 815

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 815

Asn Phe Lys Ile Arg His Asn Gln Leu Ala Asp His Tyr Gln Gln Asn

1 5 10 15

Thr Pro Ile

<210> 816

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 816

His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
1 5 10 15

Asn Thr Pro Ile
20

<210> 817

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 817

Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln

1 5 10

<210> 818

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 818

actgcaagtg ca 12

<210> 819

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 819

tgcagtgttc ac 12

<210> 820

<211> 122

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 820

ggcccagacu gagcacguga guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcucug ccaucucgug cucagucugu 120

uu 122

<210> 821

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 821

ggcccagact gagcacgtga tggcagag 28

<210> 822

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 822

gagatggcag agcaggggg gggggggg 28

<210> 823

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 823

gagatggcag agcaaaaaaa aaaaaaaaaa a 31

<210> 824

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 824

ggcccagact gagcacgtga tggcagagca ca 32

<210> 825

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 825

ggcccagact gagcacgaga tggcagagca ca 32

<210> 826

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 826

tccagatggc acattgtcag aggg 24

<210> 827

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 827

tgacgggagg gcagaactag tcct 24

<210> 828

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 828

ttcagacagc atatttgagt catt 24

<210> 829

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 829

tggggcccag actgagcacg tgatggcagc ggaaaggaag 40

<210> 830

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 830

tggggcccag actgagcacg ctttgatggc agaggaaagg 40

<210> 831

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 831

tggggcccag actgagcacg cttttgatgg cagaggaaag 40

<210> 832

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 832

tggggcccag actgagcact gatggcagag gaaaggaag 39

<210> 833

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 833

tggggcccag actgagcatg atggcagagg aaaggaag 38

<210> 834

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 834

tggggcccag actgagcacg atggcagagg aaaggaag 38

<210> 835

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 835

tggggcccag actgagcacg gtgatggcag aggaaaggaa 40

<210> 836

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 836

tggggcccag actgagcacg gatggcagag gaaaggaag 39

<210> 837

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 837

tggggcccag actgagcacg ttgatggcag aggaaaggaa 40

<210> 838

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 838

tggggcccag actgatggca gaggaaagga ag 32

<210> 839

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 839

tggggcccag actgagcacg atgatggcag aggaaaggaa 40

<210> 840

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 840

tggggcccag actgagcacg cagaggaaag gaag 34

<210> 841

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 841

tggggctgat ggcagaggaa aggaag 26

<210> 842

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 842

tggggcccag actgagcaga ggaaaggaag 30

<210> 843

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 843

tggggcccag actgagcacg ctgatggcag aggaaaggaa 40

<210> 844

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 844

tggggcctga tggcagagga aaggaag 27

<210> 845

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 845

tggggtgatg gcagaggaaa ggaag 25

<210> 846

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 846

tggggcccag actgagcacg aggaaaggaa g 31

<210> 847

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 847

tggggcccag actgagcacg agaggaaagg aag 33

<210> 848

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 848

tggggcccag actgagcacg ggcagaggaa aggaag 36

<210> 849

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 849

tggggcccag actgaggaaa ggaag 25

<210> 850

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 850

tggggcagag gaaaggaag 19

<210> 851

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 851

tggggcccag actgagcacg gaaaggaag 29

<210> 852

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 852

tggggcccag atggcagagg aaaggaag 28

<210> 853

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 853

tggggcccag actgagcacg aaaggaag 28

<210> 854

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 854

tggatggcag aggaaaggaa g 21

<210> 855

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 855

tggggcccag actgagcttt gatggcagag gaaaggaag 39

<210> 856

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 856

tggggcccag actgagcacg aag 23

<210> 857

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 857

tggggcccag actgagcacg gaggaaagga ag 32

<210> 858

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 858

tggggcccag actgagcacg gaag 24

<210> 859

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 859

tggggcccag aggaaaggaa g 21

<210> 860

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 860

tggggcccag tgatggcaga ggaaaggaag 30

<210> 861

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 861

tggtgatggc agaggaaagg aag 23

<210> 862

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 862

tggggcccag actgagcacg tgtgatggca gaggaaagga 40

<210> 863

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 863

atggcagagg aaaggaag 18

<210> 864

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 864

tggggcccag actgtgatgg cagaggaaag gaag 34

<210> 865

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 865

tggggcccag actgagtgat ggcagaggaa aggaag 36

<210> 866

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 866

tggggccctg atggcagagg aaaggaag 28

<210> 867

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 867

tggggcccag actgagcacg gcagaggaaa ggaag 35

<210> 868

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 868

cctgaaccgt atatcctatg gccctgactg gaaggatttc 40

<210> 869

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 869

aaccgtatat ctatcctatg gccctgactg gaaggatttc 40

<210> 870

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 870

aaccgtatat ctatcctatt gccctgactg gaaggatttc 40

<210> 871

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 871

aaccgtatat ctatcctatg tccctgactg gaaggatttc 40

<210> 872

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 872

aaccgtatat ctatcctatg gccctgactg gaaggctttc 40

<210> 873

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 873

aaccgtatat ctatcctctg gccctgactg gaaggatttc 40

<210> 874

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 874

aaccgtatat ctatcctctg gccctgactg gaaggctttc 40

<210> 875

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 875

caccatggtg cacctgactc ctgaggagaa gtctgccgtt 40

<210> 876

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 876

caccatggtg cacctgactc ctgtggagaa gtctgccgtt 40

<210> 877

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 877

caccatggcg cacctgactc ctgtggagaa gtctgccgtt 40

<210> 878

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 878

caccatggtg cacctgactc ctgtggagaa gtctgtcgtt 40

<210> 879

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (5)..(24)

<223> n is a, c, g, or t

<400> 879

caccnnnnnn nnnnnnnnnn nnnngtttt 29

<210> 880

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (10)..(29)

<223> n is a, c, g, or t

<400> 880

ctctaaaacn nnnnnnnnnn nnnnnnnnn 29

<210

> 881

<211> 67

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 881

agagctagaa atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac 60

cgagtcg 67

<210> 882

<211> 67

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 882

gcaccgactc ggtgccactt tttcaagttg ataacggact agccttattt taacttgcta 60

tttctag 67

<

210> 883

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (5)..(34)

<223> n is a, c, g, or t

<400> 883

gtgcnnnnnn nnnnnnnnnn nnnnnnnnnn nnnn 34

<210> 884

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (5)..(34)

<223> n is a, c, g, or t

<400> 884

aaaannnnnn nnnnnnnnnn nnnnnnnnnn nnnn 34

<210

> 885

<211> 104

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 885

cagaggaccg actcggtccc acttttttcaa gttgataacg gactagcctt attttaactt 60

gctatttcta gctctaaaac tcacgtgctc agtctgggcc ggtg 104

<210> 886

<211> 105

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 886

atcacgcacc gactcggtgc cacttttca agttgataac ggactagcct tattttaact 60

tgctatttct agctctaaaa cttcttcttc tgctcggact cggtg 105

<210> 887

<211> 105

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 887

tttccgcacc gactcggtgc cacttttca agttgataac ggactagcct tattttaact 60

tgctatttct agctctaaaa cggtgctgca gaagggattc cggtg 105

<210> 888

<211> 105

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 888

tcgttgcacc gactcggtgc cacttttca agttgataac ggactagcct tattttaact 60

tgctatttct agctctaaaa ccaggtaatg actaagatga cggtg 105

<210> 889

<211> 75

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 889

tccttggggc ccagactgag cacgtgatgg cagaggaaag gaagccctgc ttcctccaga 60

gggcgtcgca ggaca 75

<210> 890

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 890

Ser Leu Gly Pro Arg Leu Ser Thr Trp Gln Arg Lys Gly Ser Pro Ala

1 5 10 15

Ser Ser Arg Gly Arg Arg Arg Thr

20

<210> 891

<211> 75

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 891

tgtcctgcga cgccctctgg aggaagcagg gcttcctttc ctctgccatc acgtgctcag 60

tctgggcccc aagga 75

<210> 892

<211> 201

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 892

tattattact ctatgttcta tttaagtttt catgttctaa aaatgtatcc cagtttacac 60

gtctcatatg ccccttggca gtcatcttag tcattacctg aggtgttcgt tgtaactcat 120

ataaactgag ttcccatgtt ttgcttaatg gttgagttcc gtttgtctgc acagcctgag 180

acattgctgg aaataaagaa g 201

<210> 893

<211> 201

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 893

cttctttatt tccagcaatg tctcaggctg tgcagacaaa cggaactcaa ccattaagca 60

aaacatggga actcagttta tatgagttac aacgaacacc tcaggtaatg actaagatga 120

ctgccaaggg gcatatgaga cgtgtaaact gggatacatt tttagaacat gaaaacttaa 180

atagaacata gagtaataat a 201

<210> 894

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 894

gtcatcttag tcattacctg 20

<210> 895

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 895

aacgaacacc gcaggtaatg actaagatg 29

<210> 896

<211> 76

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 896

gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60

ggcaccgagt cggtgc 76

<210> 897

<211> 126

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (99)..(123)

<223> n is a, c, g, or u

<400> 897

gaaggcaaua augcuaauua caguuuuaga gcuagaaaua gcaaguuaaa auaaggcuag 60

uccguuauca acuugaaaaa guggcaccga guaggugcnn nnnnnnnnnn nnnnnnnnnn 120

nnnuuu 126

<210> 898

<211> 135

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (99)..(123)

<223> n is a, c, g, or u

<220><221> misc_feature

<222> (128)..(132)

<223> n is a, c, g, or u

<400> 898

gaaggcaaua augcuaauua caguuuuaga gcuagaaaua gcaaguuaaa auaaggcuag 60

uccguuauca acuugaaaaa guggcaccga guaggugcnn nnnnnnnnnn nnnnnnnnnn 120

nnngaaannn nnuuu 135

<210> 899

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 899

Ala Ala Cys Ala Gly Asn Arg Val Arg Arg Ser Val Gly Ser Ser Leu

1 5 10 15

Lys Cys

<210> 900

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 900

Ala Ala Asp His Cys Pro Val Val Glu Val Asn Gly Val Thr Ile

1 5 10 15

<210> 901

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400

> 901

Ala Ala His Leu Pro Thr Gly Thr Pro Leu Asp Ile Asp

1 5 10

<210> 902

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 902

Ala Ala Leu Ala Val Trp Ala Gly Leu Ala Val Gln

1 5 10

<210> 903

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 903

Ala Ala Leu Gly Val Ala Thr Ala Ala Gln Ile Thr Ala Gly Ile

1 5 10 15

<210> 904

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 904

Ala Ala Leu Leu Asn Thr Pro Pro Pro Tyr Gln Val Ser Cys Gly Gly

1 5 10 15

Glu Ser Asp Arg Ala Thr Ala Arg

20

<210> 905

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 905

Ala Ala Gln Ser Phe Thr Gly Val Val Tyr Gly Thr His Thr Thr

1 5 10 15

<210> 906

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 906

Ala Cys Glu Val Glu Pro Ala Phe Gly His Ser Asp Ala Ala Cys

1 5 10 15

<210> 907

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 907

Ala Cys Thr Phe Trp Ala Val Asn Ala Tyr Ser Ser Gly Gly Tyr

1 5 10 15

<210> 908

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 908

Ala Asp Asp Pro Leu Leu Arg

1 5

<210> 909

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 909

Ala Asp Asp Pro Leu Leu Arg Thr

1 5

<210> 910

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 910

Ala Glu Met Ile Cys Asp Ile Asp Thr Tyr Ile Val Glu Ala Gly

1 5 10 15

<210> 911

<

211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 911

Ala Glu Met Ile Cys Asp Ile Asp Thr Tyr Ile Val Glu Ala Gly Leu

1 5 10 15

Ala Ser Phe Ile

20

<210> 912

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 912

Ala Glu Pro Leu Leu Ser Cys

1 5

<210> 913

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 913

Ala Phe Val Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val Tyr

1 5 10 15

<210> 914

<211> 31

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 914

Ala Gly Phe Ile Tyr Arg Glu Thr Phe Cys Ile Thr Thr Ile Tyr Lys

1 5 10 15

Thr Gly Gln Pro Ala Ala Asp His Tyr Tyr Ser Lys Val Thr Ala

20 25 30

<210> 915

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 915

Ala Gly Leu Leu Ala Cys Cys Ala Lys Cys Leu Tyr Tyr Leu Arg

1 5 10 15

<210> 916

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 916

Ala His Thr Thr Ser Asp Pro Trp His Pro Pro Gly

1 5 10

<210> 917

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<

400> 917

Ala Ile Ala Lys Leu Glu Asp Ala Lys Glu Leu Leu Glu Ser Ser

1 5 10 15

<210> 918

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 918

Ala Ile Asp Asn Leu Arg Ala Ser Leu Glu Thr Thr Asn Gln Ala

1 5 10 15

<210> 919

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 919

Ala Lys Gly Val Glu Phe Arg

1 5

<210> 920

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 920

Ala Lys Trp Ala Val Pro Thr Thr Arg Thr Asp Asp Lys Leu Arg

1 5 10 15

<210> 921

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 921

Ala Leu Ala Glu Val Leu Lys Lys Pro Val

1 5 10

<210> 922

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400

> 922

Ala Leu Gly Val Ile Asn Thr Leu Glu Trp Ile Pro Arg Phe Lys

1 5 10 15

<210> 923

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 923

Ala Leu His Gln Ser Met Leu Asn Ser Gln Ala Ile Asp Asn Leu

1 5 10 15

<210> 924

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 924

Ala Leu Ile Gly Ile Leu Ser Leu Phe Val

1 5 10

<210> 925

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 925

Ala Leu Leu Asn Thr Pro Pro Pro Tyr Gln Val Ser Cys Gly Gly Glu

1 5 10 15

Ser Asp Arg Ala

20

<210> 926

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 926

Ala Leu Leu Asn Thr Pro Pro Pro Tyr Gln Val Ser Cys Gly Gly Glu

1 5 10 15

Ser Asp Arg Ala Ser Ala Arg Val

20

<210> 927

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 927

Ala Leu Val Glu Gly Leu Ala Pro Gly Gly Gly Asn Cys His Leu

1 5 10 15

<210> 928

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 928

Ala Met Ala Ala Trp Ser Glu Arg Ala Gly Glu Ala

1 5 10

<210> 929

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 929

Ala Asn Cys Ala Ser Ile Leu Cys Lys Cys Tyr Thr Thr Gly Thr

1 5 10 15

<210> 930

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 930

Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val

1 5 10

<210> 931

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 931

Ala Pro Gly Pro Gly Glu Val Trp

1 5

<210> 932

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 932

Ala Pro Leu Pro Pro His Thr Thr Glu Arg Ile Glu Thr Arg Ser Ala

1 5 10 15

Arg His Pro Trp Arg Ile Arg

20

<210> 933

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 933

Ala Pro Pro Met Pro Pro Gln Pro Pro Arg Ala His Gly Gln His Tyr

1 5 10 15

Gly His His His His Gln Leu Pro Phe Leu Gly

20 25

<210> 934

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 934

Ala Pro Pro Thr Leu Pro Gln Pro Pro Cys Ala His Gly Gln His Tyr

1 5 10 15

Gly His His His His Gln Leu Pro Phe Leu Gly

20 25

<210> 935

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 935

Ala Pro Pro Thr Leu Pro Gln Pro Pro Arg Ala His Gly Gln His Tyr
1 5 10 15

Gly His His His His Gln Leu Pro Phe Leu Gly
20 25

<210> 936

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 936

Ala Pro Gln Pro Gly Pro Gln Pro Pro Gln Pro Pro Gln Pro Gln Pro
1 5 10 15

Glu Ala Pro Ala Pro Gln
20

<210> 937

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 937

Ala Gln Leu Ala Ser Tyr Phe Asn Pro Gly Gly Ser Tyr Tyr Lys
1 5 10 15

<210> 938

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 938

Ala Arg Ala Ala His Leu Pro Thr Gly Thr Pro Leu Asp
1 5 10

<210> 939

<211> 15

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 939

Ala Arg Leu Val Ser Glu Ile Ala Met His Thr Thr Glu Asp Lys

1 5 10 15

<210> 940

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 940

Ala Arg Leu Val Ser Glu Ile Ala Met His Thr Thr Glu Asp Lys Ile

1 5 10 15

Ser Arg Ala Val

20

<210> 941

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 941

Ala Ser Asp Val Glu Thr Ala Glu Gly Gly Glu Ile His Glu Leu Leu

1 5 10 15

Arg

<210> 942

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 942

Ala Ser Asp Val Glu Thr Ala Glu Gly Gly Glu Ile His Glu Leu Leu

1 5 10 15

Arg Leu Gln

<210> 943

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 943

Ala Ser Asp Val Glu Thr Ala Glu Gly Gly Glu Ile His Glu Leu Leu

1 5 10 15

Arg Leu Gln Ser Arg

20

<210> 944

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 944

Ala Ser Asp Val Glu Thr Ala Glu Gly Gly Glu Ile His Lys Leu Leu

1 5 10 15

Arg Leu Gln

<210> 945

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 945

Ala Ser Asp Val Leu Pro Gly His Trp Leu Gln Gly

1 5 10

<210> 946

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 946

Ala Ser Glu Leu Gly Ile Thr Ala Glu Asp Ala Arg Leu Val Ser

1 5 10 15

<210> 947

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 947

Ala Ser Glu Leu Gly Ile Thr Ala Glu Asp Ala Arg Leu Val Ser Glu

1 5 10 15

Ile Ala Met His

20

<210> 948

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 948

Ala Ser Glu Ser Ser Gln Asp Pro Gln Asp Ser Arg Arg

1 5 10

<210> 949

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 949

Ala Ser Ile Leu Cys Lys Cys Tyr Thr Thr Gly Thr Ile Ile Asn

1 5 10 15

<210> 950

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 950

Ala Ser Pro Val Cys Gln Arg His Ser Pro Asp Cys Ser Arg Leu

1 5 10 15

<210> 951

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 951

Ala Ser Gln Ala Arg Trp Thr Gly Ala Thr Arg Ala

1 5 10

<210> 952

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 952

Ala Ser Tyr Phe Asn Pro Gly Gly Ser Tyr Tyr Lys Gln Tyr His Pro

1 5 10 15

Thr Ala Cys Glu Val Glu Pro Ala Phe Gly His Ser

20 25

<210> 953

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 953

Ala Ser Tyr Lys Val Met Thr Arg Ser Ser His Gln Ser Leu Val

1 5 10 15

<210> 954

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 954

Ala Ser Tyr Val Lys Lys Pro Lys Glu Asp Val Asp

1 5 10

<210> 955

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 955

Ala Thr Ala Ala Gln Ile Thr Ala Gly Ile Ala Leu His Gln Ser

1 5 10 15

<210> 956

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 956

Ala Thr Pro Glu Arg Pro Arg Leu

1 5

<210> 957

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 957

Ala Val Cys Leu Gly Gly Leu Ile Gly Ile Pro Ala Leu Ile Cys

1 5 10 15

<210> 958

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 958

Ala Val Gly Pro Arg Gln Ala Gln Val Ser Phe

1 5 10

<210> 959

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 959

Ala Val Asn Ala Tyr Ser Ser Gly Gly Tyr Ala Gln Leu Ala Ser

1 5 10 15

<210> 960

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 960

Ala Val Ser Glu Thr Arg Gln Thr Trp Ala Glu Trp Ala Ala Ala

1 5 10 15

<210> 961

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 961

Ala Val Thr Ala Pro Asp Thr Ala Ala Asp Ser Glu Leu Arg Arg

1 5 10 15

<210> 962

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 962

Ala Val Thr Ala Pro Asp Thr Ala Ala Asp Ser Glu Leu Arg Arg Trp
1 5 10 15

Ile Lys Tyr Thr
20

<210> 963

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 963

Ala Tyr Gly Gly Ile Ile Lys Asp Ala Pro Pro Gly Ala Gly Phe Ile
1 5 10 15

Tyr Arg Glu Thr Phe Cys
20

<210> 964

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 964

Cys Ala Leu Pro Leu Ala Gly Leu Leu Ala Cys Cys Ala Lys Cys
1 5 10 15

<210> 965

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 965

Cys Ala Arg Ile Trp Asn Gly Thr Gln Arg Ala Cys Thr Phe Trp
1 5 10 15

<210

> 966

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 966

Cys Ala Arg Thr Leu Val Ser Gly Ser Phe Gly Asn Arg Phe Ile

1 5 10 15

<210> 967

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 967

Cys Ala Ser Pro Tyr Glu Gly Arg Tyr Arg Asp Met Tyr Asp Ala Leu

1 5 10 15

Arg Asx Arg Leu Leu Tyr

20

<210> 968

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 968

Cys Ala Val Phe Val Arg Ser Gly Gln Pro Val Ile Gly Ala

1 5 10

<210> 969

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 969

Cys Cys Ala Lys Cys Leu Tyr Tyr Leu Arg Gly Ala Ile Ala Pro Arg

1 5 10 15

<210> 970

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 970

Cys Cys Arg Gly Arg Cys Asn Lys Lys Gly Glu Gln Val Gly Met

1 5 10 15

<210> 971

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 971

Cys Glu Ile Pro Thr Asp Val Ser Cys Glu Gly Leu Gly Ala Trp

1 5 10 15

<210> 972

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 972

Cys Phe Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser Pro Met Glu Val

1 5 10 15

<210> 973

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 973

Cys Phe Gln Gln Ala Cys Lys Gly Lys Ile Gln Ala Leu Cys Glu

1 5 10 15

<210> 974

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 974

Cys Phe Gln Gln Ala Cys Lys Gly Lys Ile Gln Ala Leu Cys Glu Asn

1 5 10 15

Pro Glu Trp Ala Pro Leu Lys Asp Asn Arg Ile Pro Ser

20 25

<210> 975

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 975

Cys Gly Gly Glu Ser Asp Arg Ala Ser Ala Arg Val Ile Asp Pro

1 5 10 15

<210> 976

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 976

Cys Ile Thr Thr Ile Tyr Lys Thr Gly Gln Pro Ala Ala Asp His Tyr

1 5 10 15

Tyr Ser Lys Val Thr Ala

20

<210> 977

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 977

Cys Lys Gly Lys Ile Gln Ala Leu Cys Glu Asn Pro Glu Trp Ala

1 5 10 15

<210> 978

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 978

Cys Lys Pro Trp Gln Glu Ser Arg Lys Asn Lys Ala Gln

1 5 10

<210> 979

<211>

15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 979

Cys Asn Lys Lys Gly Glu Gln Val Gly Met Ser Arg Pro Gly Leu

1 5 10 15

<210> 980

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 980

Cys Asn Val Thr Thr Glu His Pro Phe Cys Asn Thr Pro His Gly

1 5 10 15

<210> 981

<211> 10

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 981

Cys Gln Val Gly Ser Ser Asn Ser Ala Phe

1 5 10

<210> 982

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 982

Cys Ser Gly Pro Thr Thr Ile Arg Gly Gln Phe Ser

1 5 10

<210> 983

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 983

Cys Thr Ser Pro Tyr Asp Gly Lys Tyr Trp Ser Met Tyr Ser Arg Leu

1 5 10 15

<210> 984

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 984

Cys Val Leu Ala Asp Ser Glu Ser Gly Gly His Ile Thr His Ser

1 5 10 15

<210> 985

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 985

Cys Tyr Thr Thr Gly Thr Ile Ile Asn Gln Asp Pro Asp Lys Ile Leu

1 5 10 15

Thr Tyr Ile Ala Ala Asp His Cys

20

<210> 986

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 986

Asp Ala Asp Asp Pro Leu Leu Arg

1 5

<210> 987

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 987

Asp Ala Arg Ala Ala His Leu Pro Thr Gly Thr Pro Leu Asp Ile

1 5 10 15

<210> 988

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 988

Asp Ala Arg Ala Ala His Leu Pro Thr Gly Thr Pro Leu Asp Ile Asp

1 5 10 15

Thr Ala Ser Glu

20

<210> 989

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 989

Asp Cys His Ala Pro Thr Tyr Leu Pro Ala Glu Val Asp Gly Asp

1 5 10 15

<210> 990

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 990

Asp Cys Ser Arg Leu Val Gly Ala Thr Pro Glu Arg Pro Arg Leu

1 5 10 15

<210> 991

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 991

Asp Asp Lys Leu Arg Met Glu Thr Cys Phe Gln Gln Ala Cys Lys

1 5 10 15

<210> 992

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 992

Asp Asp Pro Leu Leu Arg Thr Ala

1 5

<210> 993

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 993

Asp Asp Pro Leu Leu Arg Thr Ala Pro Gly Pro Gly Glu Val Trp

1 5 10 15

<210> 994

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 994

Asp Asp Pro Pro Ala Thr Val Tyr Arg Tyr Asp

1 5 10

<210> 995

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 995

Asp Asp Pro Pro Ala Thr Val Tyr Arg Tyr Asp Ser Arg Pro Pro Glu

1 5 10 15

Asp

<210

> 996

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 996

Asp Asp Pro Pro Ala Thr Val Tyr Arg Tyr Asp Ser Arg Pro Pro Glu

1 5 10 15

Asp Val

<210> 997

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 997

Asp Glu Val Gly Leu Arg Thr Pro Gln Arg Phe Thr Asp Leu Val

1 5 10 15

<210> 998

<211> 15

<212>

PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 998

Asp His Cys Pro Val Val Glu Val Asn Gly Val Thr Ile Gln Val

1 5 10 15

<210> 999

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 999

Asp Ile Asp Thr Ala Ser Glu Ser Ser Gln Asp Pro Gln

1 5 10

<210> 1000

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1000

Asp Ile Asn Lys Val Leu Glu Lys Leu Gly Tyr Ser Gly Gly Asp

1 5 10 15

<210> 1001

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1001

Asp Leu Ile Ala Tyr Lys Gln

1 5

<210> 1002

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1002

Asp Leu Ile Gly Gln Lys Leu Gly Leu Lys Leu Leu Arg Tyr Tyr

1 5 10 15

<210> 1003

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1003

Asp Leu Gln Lys Ala Leu Glu Ala Gln Ser Arg Ala Leu Arg Ala Glu

1 5 10 15

Leu Ala Ala

<210> 1004

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1004

Asp Leu Gln Tyr Val Leu Ala Thr Tyr Asp Thr Ser Arg Val Glu

1 5 10 15

<210> 1005

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1005

Asp Leu Ser Leu Arg Arg Phe Met Val

1 5

<210> 1006

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1006

Asp Leu Ser Asn Cys Met Val Ala Leu Gly Glu Leu Lys Leu Ala

1 5 10 15

<210> 1007

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1007

Asp Leu Val Glu Tyr Ile Met Asn Tyr Thr Gly Asn Gln Gln Ser Arg

1 5 10 15

Trp Gly Leu Gly Ser Pro Asn Cys

20

<210> 1008

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1008

Asp Leu Val Lys Phe Ile Ser Asp Lys Ile Lys Phe Leu Asn Pro

1 5 10 15

<210> 1009

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1009

Asp Leu Val Lys Phe Ile Ser Thr Lys Ile Lys Phe Leu Asn Pro

1 5 10 15

<210> 1010

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1010

Asp Leu Tyr Lys Ser Asn His Asn Asn Val

1 5 10

<210> 1011

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1011

Asp Asn Val Leu Asp His Leu Thr Gly Arg

1 5 10

<210> 1012

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1012

Asp Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys

1 5 10

<210> 1013

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1013

Asp Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys Gln

1 5 10

<210> 1014

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1014

Asp Pro Asp Lys Ile Leu Thr Tyr Ile Ala Ala

1 5 10

<210> 1015

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1015

Asp Pro Gly Asp Leu Val Glu Tyr Ile Met Asn Tyr Thr Gly Asn Gln

1 5 10 15

Gln Ser Arg

<210> 1016

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1016

Asp Pro Leu Leu Arg Thr Ala Pro

1 5

<210> 1017

<211> 22

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 1017

Asp Pro Leu Leu Arg Thr Ala Pro Gly Pro Gly Glu Val Trp Val Thr

1 5 10 15

Pro Val Ile Gly Ser Gln

20

<210> 1018

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1018

Asp Pro Gln Asp Ser Arg Arg Ser Ala Glu Pro Leu Leu

1 5 10

<210> 1019

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223>

Synthetic

<400> 1019

Asp Pro Val Ile Asp Arg Leu Tyr Leu Ser Ser His Arg Gly Val

1 5 10 15

<210> 1020

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1020

Asp Gln Ile Leu Arg Ser Met Lys Gly Leu Ser Ser Thr Ser Ile

1 5 10 15

<210> 1021

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1021

Asp Gln Tyr Cys Ala Asp Val Ala Ala Glu Glu Leu Met Asn Ala

1 5 10 15

<210> 1022

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1022

Asp Ser Glu Ser Gly Gly His Ile Thr His

1 5 10

<210> 1023

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1023

Asp Thr Ala Ser Glu Ser Ser Gln Asp Pro Gln Asp Ser

1 5 10

<210> 1024

<211> 15

<212> PRT

<213>

Artificial Sequence

<220><223> Synthetic

<400> 1024

Asp Thr Val Met Ser Val Phe Ala Leu Ala Ser Tyr Val Gln His

1 5 10 15

<210> 1025

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1025

Asp Val Phe Gln Asn Gly Phe Thr Ala Trp Gly Asn Asn Asp

1 5 10

<210> 1026

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1026

Asp Val Gly Ala Val Pro Pro Gly Lys Phe Val Thr Ala Ala Leu

1 5 10 15

<210> 1027

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1027

Asp Val Asn Lys Ser Lys Thr His Ile Ser Val Asn Gly Arg Lys Ile

1 5 10 15

<210> 1028

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1028

Asp Val Ser Cys Glu Gly Leu Gly Ala Trp Val Pro Ala Ala Pro

1 5 10 15

<210> 1029

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1029

Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser Pro Asp Cys Ser Arg

1 5 10 15

Leu Val Gly Ala Thr Pro Glu Arg Pro Arg Leu

20 25

<210> 1030

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1030

Glu Ala Arg Glu Ser Tyr Arg Glu Thr Gly Pro Ser Arg

1 5 10

<210> 1031

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1031

Glu Ala Val Glu Ala Glu Arg Ala Gly Arg Gly Thr Gly

1 5 10

<210> 1032

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1032

Glu Asp Ala Lys Glu Leu Leu Glu Ser Ser Asp Gln Ile Leu Arg

1 5 10 15

<210> 1033

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1033

Glu Asp Arg Arg Val Lys Gln Ser Arg Gly Glu Ala Arg

1 5 10

<210> 1034

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1034

Glu Asp Ser Ile Thr Ile Pro Tyr Gln Gly Ser Gly Lys Gly Val

1 5 10 15

<210> 1035

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1035

Glu Glu Ala Phe Thr Tyr Leu Cys Thr Ala Pro Gly Cys Ala Thr

1 5 10 15

<210> 1036

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1036

Glu Gly Phe Asn Met Ile Leu Gly Thr Ile Leu Ala Gln Ile Trp

1 5 10 15

<210> 1037

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1037

Glu Gly Phe Asn Met Ile Leu Gly Thr Ile Leu Ala Gln Ile Trp Val

1 5 10 15

Leu Leu Ala Lys

20

<210> 1038

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1038

Glu His Pro Phe Cys Asn Thr Pro His Gly Gln Leu Glu Val Gln

1 5 10 15

<210> 1039

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1039

Glu Ile Ser Asp Ile Glu Val Gln Asp Pro Glu Gly Phe Asn Met

1 5 10 15

<210> 1040

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1040

Glu Ile Ser Asp Ile Glu Val Gln Asp Pro Glu Gly Phe Asn Met Ile

1 5 10 15

Leu Gly Thr Ile

20

<210> 1041

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1041

Glu Lys Pro Asn Leu Ser Ser Lys Arg Ser Glu

1 5 10

<210> 1042

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1042

Glu Leu Lys Leu Ala Ala Leu Cys His Gly Glu Asp Ser Ile Thr

1 5 10 15

<210> 1043

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1043

Glu Leu Met Asn Ala Leu Val Asn Ser Thr Leu Leu Glu Thr Arg

1 5 10 15

<210> 1044

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1044

Glu Leu Pro Arg Leu

1 5

<210> 1045

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1045

Glu Asn Pro Glu Trp Ala Pro Leu Lys Asp Asn Arg Ile Pro Ser Tyr

1 5 10 15

Gly Val Leu Ser Val Asp Leu

20

<210> 1046

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1046

Glu Pro Ile Arg Asp Ala Leu Asn Ala Met Thr Gln Asn Ile Arg

1 5 10 15

<210> 1047

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1047

Glu Gln Val Gly Met Ser Arg Pro Gly Leu Lys Pro Asp Leu Thr

1 5 10 15

<210> 1048

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1048

Glu Arg Pro Arg Leu Arg Leu Val

1 5

<210> 1049

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1049

Glu Arg Pro Arg Leu Arg Leu Val Asp Ala Asp Asp Pro Leu Leu

1 5 10 15

<210> 1050

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1050

Glu Ser Pro Gly Gln Leu Ile Gln Arg Ile Thr Asp Asp Pro Asp Val

1 5 10 15

Ser

<210> 1051

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1051

Glu Ser Arg Gly Ile Lys Ala Arg Ile Thr His Val Asp Thr Glu

1 5 10 15

<210> 1052

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1052

Glu Ser Ser Cys Thr Phe Met Pro Glu Gly Thr Val Cys Ser Gln

1 5 10 15

<210> 1053

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1053

Glu Ser Ser Gln Asp Pro Gln Asp Ser Arg Arg Ser Ala

1 5 10

<210> 1054

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1054

Glu Thr Arg Thr Thr Asn Gln Phe Leu Ala Val Ser Lys Gly Asn

1 5 10 15

<210> 1055

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1055

Glu Val Asp Gly Asp Val Lys Leu Ser Ser Asn Leu Val Ile Leu

1 5 10 15

<210> 1056

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1056

Glu Val Asn Gly Val Thr Ile Gln Val

1 5

<210> 1057

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1057

Glu Val Trp Val Thr Pro Val Ile

1 5

<210> 1058

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1058

Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala

1 5 10

<210> 1059

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1059

Glu Trp Ala Ala Ala His Trp Trp Gln Leu Thr Leu Gly Ala Thr

1 5 10 15

<210> 1060

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1060

Glu Trp Ile Pro Arg Phe Lys Val Ser Pro Tyr Leu Phe Thr Val

1 5 10 15

<210> 1061

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1061

Phe Glu Tyr Val Asp Thr Tyr Gly Asp Asn Ala Gly

1 5 10

<210> 1062

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1062

Phe Gly Pro Leu Ile Thr His Gly Ser Gly Met Asp Leu Tyr Lys

1 5 10 15

<210> 1063

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1063

Phe Ile Phe Asp Ala Leu Ala Glu Val

1 5

<210> 1064

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1064

Phe Ile Ser Asp Lys Ile Lys Phe Leu

1 5

<210> 1065

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1065

Phe Lys Arg Asn Lys Asp Lys Pro Pro Ile Thr Ser Gly Ser Gly

1 5 10 15

<210> 1066

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1066

Phe Lys Arg Asn Lys Asp Lys Pro Pro Ile Thr Ser Gly Ser Gly Gly

1 5 10 15

Ala Ile Arg Gly

20

<210> 1067

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1067

Phe Lys Thr Val Arg Pro Val Ala Leu Pro Arg Thr Leu Ala Pro

1 5 10 15

<210> 1068

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 1068

Phe Leu Met Asp Arg His Ile Ile Val

1 5

<210> 1069

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1069

Phe Met Ala Val Leu Leu Thr Leu Gln Thr Pro Thr Gly Gln Ile

1 5 10 15

<210> 1070

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1070

Phe Met Pro Glu Gly Thr Val Cys Ser Gln Asn Ala Leu Tyr Pro

1 5 10 15

<210> 1071

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1071

Phe Met Tyr Met Ser Leu Leu Gly Val

1 5

<210> 1072

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1072

Phe Asn Val Pro Ile Lys Glu Ala Gly Glu Asp Cys His Ala Pro

1 5 10 15

<210> 1073

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1073

Phe Arg Asp Leu Thr Trp Cys Ile Asn Pro Pro Glu Arg Ile Lys

1 5 10 15

<210> 1074

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1074

Phe Ser His Asp Asp Pro Ile Ser Ser Asp Gln Ser Arg Phe Gly

1 5 10 15

<210> 1075

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1075

Phe Ser His Asp Asp Pro Ile Ser Ser Asp Gln Ser Arg Phe Gly Trp

1 5 10 15

Phe Glu Asn Lys

20

<210> 1076

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1076

Phe Thr Asp Leu Val Lys Phe Ile Ser Asp Lys Ile Lys Phe Leu

1 5 10 15

<210> 1077

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1077

Phe Thr Trp Asp Gln Lys Leu Trp Cys Arg His Phe Cys Val Leu

1 5 10 15

<210> 1078

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1078

Phe Val Arg Asp Gly Gln Ser Val Ile Gly Ala Cys Ala Ser Pro Tyr

1 5 10 15

Glu Gly Arg Tyr Arg Asp Leu Tyr Asp Ala Leu Arg Arg Leu Leu Tyr

20 25 30

<210> 1079

<211> 33

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1079

Phe Val Arg Ser Gly Gln Pro Val Ile Gly Ala Cys Thr Ser Pro Tyr

1 5 10 15

Asp Gly Lys Tyr Trp Ser Ile Leu Tyr Ser Arg Leu Arg Lys Met Leu

20 25 30

Tyr

<210> 1080

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1080

Phe Tyr Lys Asp Asn Pro His Pro Lys Gly Ser Arg Ile Val Ile

1 5 10 15

<210> 1081

<211> 14

<212> PRT

<213> Artificial Sequence

<

220><223> Synthetic

<400> 1081

Gly Ala Ala Ser Ser Tyr Phe Glu Tyr Val Asp Thr Tyr Gly

1 5 10

<210> 1082

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1082

Gly Ala Phe Asp Leu Lys Thr Thr Phe Cys Ile Met Thr Thr Arg Asn

1 5 10 15

Thr Gly Gln Pro Ala

20

<210> 1083

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 1083

Gly Ala Leu Ile Gly Ile Leu Ser Leu Phe Val Glu Ser Pro Gly

1 5 10 15

<210> 1084

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1084

Gly Ala Leu Ile Gly Ile Leu Ser Leu Phe Val Glu Ser Pro Gly Gln

1 5 10 15

Leu Ile Gln Arg

20

<210> 1085

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1085

Gly Ala Thr Pro Glu Arg Pro Arg

1 5

<210> 1086

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1086

Gly Ala Tyr Gly Arg Cys Pro Asn Gly Thr Arg Ala Leu Thr Val Ala

1 5 10 15

Glu Leu Arg Gly Asn Ala Glu Leu

20

<210> 1087

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1087

Gly Cys Phe Ala Pro Trp Asp Leu Glu Ala Thr Gly Ala Cys Ile

1 5 10 15

<210> 1088

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1088

Gly Asp Ile Asn Lys Val Leu Glu Lys Leu Gly Tyr Ser

1 5 10

<210> 1089

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1089

Gly Asp Ile Asn Lys Val Leu Glu Lys Leu Gly Tyr Ser Gly Gly Asp

1 5 10 15

Leu Leu Gly

<210> 1090

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1090

Gly Asp Leu Arg Ala Val His His Arg Pro Val Pro Ala

1 5 10

<210> 1091

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1091

Gly Asp Leu Val Glu Tyr Ile Met Asn Tyr Thr Gly Asn Gln Gln

1 5 10 15

<210> 1092

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1092

Gly Asp Ser Ser Ile Thr Thr Arg Ser Arg Leu Leu Asp Arg Leu

1 5 10 15

<210> 1093

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1093

Gly Asp Ser Ser Ile Thr Thr Arg Ser Arg Leu Leu Asp Arg Leu Val

1 5 10 15

Arg Leu Ile Gly

20

<210> 1094

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1094

Gly Glu Asp Cys His Ala Pro Thr Tyr Leu Pro Ala Glu Val Asp

1 5 10 15

<210> 1095

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1095

Gly Glu Leu Ser Thr Leu Glu Ser Leu Met Asn Leu Tyr Gln Gln

1 5 10 15

<210> 1096

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1096

Gly Glu Leu Ser Thr Leu Glu Ser Leu Met Asn Leu Tyr Gln Gln Met

1 5 10 15

Gly Lys Pro Ala

20

<210> 1097

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1097

Gly Glu Gln Ala Arg Tyr Leu Ala Leu Leu Glu Ala

1 5 10

<210> 1098

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1098

Gly Glu Val Trp Val Thr

1 5

<210> 1099

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1099

Gly Glu Val Trp Val Thr Pro Val

1 5

<210> 1100

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1100

Gly Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala Arg

1 5 10

<210> 1101

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1101

Gly Glu Tyr Gly Gly Val Ile Lys Asp Gly Thr Pro Gly Gly Ala

1 5 10 15

<210> 1102

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1102

Gly Phe Leu Ser Gly Val Gly Pro Met Arg Leu Arg His Gly Ala Asp

1 5 10 15

Thr

<210> 1103

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1103

Gly Phe Arg Ala Ser Asp Val Glu Thr Ala Glu Gly Gly Glu Ile His

1 5 10 15

Glu Leu Leu Arg Leu Gln

20

<210> 1104

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1104

Gly Gly Ala Val Pro Gly Gly Ala Val Pro Gly Gly Ala Val Pro Gly

1 5 10 15

Gly Phe Gly Pro Gly Gly Phe Gly Pro

20 25

<210> 1105

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1105

Gly Gly Ala Val Pro Gly Gly Ala Val Pro Gly Gly Phe Gly Pro Gly

1 5 10 15

Gly Phe Gly Pro Gly Gly Phe Gly Pro

20 25

<210> 1106

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1106

Gly Gly Ala Val Pro Gly Gly Ala Val Pro Gly Gly Phe Gly Pro Gly

1 5 10 15

Gly Phe Gly Pro Gly Gly Phe Gly Pro Gly Gly Phe Gly Pro

20 25 30

<210> 1107

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1107

Gly Gly His Ile Thr His Ser Gly Met Val Gly Met Gly Val Ser

1 5 10 15

<210> 1108

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1108

Gly Ile Leu Glu Ser Arg Gly Ile Lys Ala Arg Ile Thr His Val Asp

1 5 10 15

Thr Glu Ser Tyr

20

<210> 1109

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1109

Gly Ile Thr Gly Glu Thr Thr Thr Thr Glu Tyr Ser Asn Ala Arg Tyr

1 5 10 15

Val

<210> 1110

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1110

Gly Lys Glu Asp Arg Arg Val Lys Gln Ser Arg Gly Glu

1 5 10

<210> 1111

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 1111

Gly Lys Val Thr Asn Gly Ser

1 5

<210> 1112

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1112

Gly Leu Gly Ala Trp Val Pro Ala Ala Pro Cys Ala Arg Ile Trp

1 5 10 15

<210> 1113

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1113

Gly Leu Ile Gly Ile Pro Ala Leu Ile Cys Cys Cys Arg Gly Arg

1 5 10 15

<210> 1114

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1114

Gly Leu Leu Ala Cys Cys Ala Lys Cys Leu Tyr Tyr Leu Arg Gly Ala

1 5 10 15

Ile Ala Pro Arg

20

<210> 1115

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1115

Gly Leu Leu Ala Ile Ala Gly Ile Arg Leu His Arg Ala Ala Ile

1 5 10 15

<210> 1116

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1116

Gly Leu Gln Pro Arg Ala Asp Met Ala Ala Pro Pro Thr Leu Pro Gln

1 5 10 15

<210> 1117

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1117

Gly Met Gly Val Ser Cys Thr Val Thr Arg Glu Asp Gly Thr Asn Arg

1 5 10 15

Arg

<210> 1118

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1118

Gly Met Tyr Gly Gly Thr Tyr Leu Val Glu Lys Pro

1 5 10

<210> 1119

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1119

Gly Asn Ala Glu Leu Gln Thr Tyr Leu Arg Gln Ile Thr Pro Gly Trp
1 5 10 15

Ser Ile Tyr Gly Leu Tyr Asp Gly Thr Tyr Leu Gly
20 25

<210>

1120

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1120

Gly Asn Cys His Leu Thr Val Asn Gly Glu Asp Val Gly Ala Val
1 5 10 15

<210> 1121

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1121

Gly Asn Asn Asp Asn Val Leu Asp His Leu Thr Gly Arg
1 5 10

<210> 1122

<211> 15

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 1122

Gly Asn Asn Asp Asn Val Leu Asp His Leu Thr Gly Arg Ser Cys
1 5 10 15

<210> 1123

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1123

Gly Asn Arg Phe Ile Leu Ser Gln Gly Asn Leu Ile Ala Asn Cys

1 5 10 15

<210> 1124

<211> 34

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1124

Gly Asn Arg Gly Arg Gly Gln Arg Arg Asp Trp Ser Arg Ala Pro Pro

1 5 10 15

Pro Pro Glu Glu Arg Gln Glu Thr Arg Ser Gln Thr Pro Ala Pro Lys

20 25 30

Pro Ser

<210> 1125

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1125

Gly Pro Gly Glu Val Trp Val Thr

1 5

<210> 1126

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1126

Gly Pro Met Arg Leu Arg His Gly Ala Asp Thr Arg Cys Gly Arg Leu

1 5 10 15

Ile

<210> 1127

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1127

Gly Pro Asn His Thr Lys Val

1 5

<210> 1128

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1128

Gly Pro Arg Gln Ala Gln Val Ser Phe

1 5

<210> 1129

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1129

Gly Pro Arg Gln Ala Gln Val Ser Phe Leu Gln Gly Asp Gln Ser

1 5 10 15

<210> 1130

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1130

Gly Pro Arg Gln Ala Gln Val Ser Phe Leu Gln Gly Asp Gln Ser Glu

1 5 10 15

Asn Glu Leu Pro

20

<210> 1131

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1131

Gly Arg Gly Tyr Asn Val Ser Ser Ile Val Thr Met Thr Ser Gln

1 5 10 15

<210> 1132

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1132

Gly Arg Thr Pro Phe Ile Ile

1 5

<210> 1133

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1133

Gly Ser Pro Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys

1 5 10 15

<210> 1134

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1134

Gly Ser Gln Ala Arg Lys Cys Gly Leu His Ile Arg Ala Gly Pro

1 5 10 15

<210> 1135

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1135

Gly Ser Ser Asn Ser Ala Phe Val Ser Thr Ser Ser Ser Arg Arg

1 5 10 15

<210> 1136

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1136

Gly Ser Thr Lys Ser Cys Ala Arg Thr Leu Val Ser Gly Ser Phe

1 5 10 15

<210> 1137

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1137

Gly Ser Tyr Tyr Lys Gln Tyr His Pro Thr Ala Cys Glu Val Glu

1 5 10 15

<210> 1138

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1138

Gly Thr His Thr Thr Ala Val Ser Glu Thr Arg Gln Thr Trp Ala

1 5 10 15

<210> 1139

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1139

Gly Thr Ile Ile Asn Gln Asp Pro Asp Lys Ile Leu Thr Tyr Ile

1 5 10 15

<210> 1140

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1140

Gly Thr Leu Val Arg Met Ala Pro Val Ile Gly

1 5 10

<210> 1141

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1141

Gly Thr Pro Leu Asp Ile Asp Thr Ala Ser Glu Ser Ser

1 5 10

<210> 1142

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1142

Gly Thr Tyr Leu Gly Gln Ala Tyr Gly Gly Ile Ile Lys Asp Ala Pro

1 5 10 15

Pro Gly Ala Gly Phe Ile Tyr Arg Glu Thr Phe Cys

20 25

<210> 1143

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1143

Gly Val Ala Thr Lys Gly Leu Gly Val His Ala Lys Ser Ser Asp Trp

1 5 10 15

Gly

<210> 1144

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1144

Gly Val Leu Leu Pro Thr Ile Pro Gly Lys Leu Asp Val Asn Lys Ser

1 5 10 15

Lys Thr His Ile

20

<210> 1145

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1145

Gly Val Leu Ser Val Asp Leu Ser Leu Thr Val Glu Leu Lys Ile

1 5 10 15

<210> 1146

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1146

Gly Val Pro Ile Glu Leu Gln Val Glu Cys Phe Thr Trp Asp Gln

1 5 10 15

<210> 1147

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1147

Gly Val Ser Cys Thr Val Thr Arg Glu Asp Gly Thr Asn Arg Arg

1 5 10 15

<210> 1148

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1148

Gly Val Ser Tyr Asn Ile Gly Ser Gln Glu Trp Tyr Thr Thr Val

1 5 10 15

<210> 1149

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1149

Gly Tyr Asn Val Ser Ser Ile Val Thr Met Thr Ser Gln Gly Met

1 5 10 15

<210> 1150

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1150

His Phe Cys Val Leu Ala Asp Ser Glu Ser Gly Gly His Ile Thr

1 5 10 15

<210> 1151

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1151

His Gly Glu Asp Ser Ile Thr Ile Pro Tyr Gln Gly Ser Gly Lys

1 5 10 15

<210> 1152

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1152

His Gly Pro Asp Trp Ala Ser Pro

1 5

<210> 1153

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1153

His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser Pro

1 5 10 15

<210> 1154

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1154

His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser Pro Asp

1 5 10 15

Cys Ser Arg Leu Val Gly

20

<210> 1155

<211> 33

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1155

His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser Pro Asp

1 5 10 15

Cys Ser Arg Leu Val Gly Ala Thr Pro Glu Arg Pro Arg Leu Arg Leu

20 25 30

Val

<210> 1156

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1156

His Ile Thr His Ser Gly Met Glu Gly Met Gly Val Ser Cys Thr

1 5 10 15

<210> 1157

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1157

His Lys Ser Leu Ser Thr Asn Leu Asp Val Thr Asn Ser Ile Glu

1 5 10 15

<210> 1158

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1158

His Leu Met Ile Asp Arg Pro Tyr Val

1 5

<210> 1159

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1159

His Leu Pro Thr Gly Thr Pro Leu Asp Ile Asp Thr Ala

1 5 10

<210> 1160

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1160

His Leu Pro Thr Gly Thr Pro Leu Asp Ile Asp Thr Ala Thr Glu Ser

1 5 10 15

Ser Gln Asp Pro Gln Asp Ser Arg

20

<210> 1161

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1161

His Met Thr Asn Tyr Leu Glu Gln Pro Val Ser Asn Asp Leu Ser

1 5 10 15

<210> 1162

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1162

His Gln Ser Leu Val Ile Lys Leu Met Pro Asn Ile Thr Leu Leu

1 5 10 15

<210> 1163

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1163

His Arg Ala Ala Ile Tyr Thr Ala Glu Ile His Lys Ser Leu Ser

1 5 10 15

<210> 1164

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1164

His Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly Arg Gly Thr

1 5 10 15

Gly His

<210> 1165

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1165

His Val Asp Thr Glu Ser Tyr Phe Ile Val Leu Ser Ile Ala Tyr

1 5 10 15

<210> 1166

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1166

His Trp Gly Asn Leu Ser Lys Ile Gly Val Val Gly Ile Gly Ser

1 5 10 15

<210> 1167

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 1167

His Trp Trp Gln Leu Thr Leu Gly Ala Thr Cys Ala Leu Pro Leu

1 5 10 15

<210> 1168

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1168

His Tyr Arg Asn Ala Ser Asp Val Leu Pro Gly His Trp Leu Gln Gly

1 5 10 15

Gly Trp Gly Cys Tyr Asn Leu

20

<210> 1169

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1169

Ile Asp Leu Gly Pro Pro Ile Ser Leu Glu Arg Leu Asp Val Gly

1 5 10 15

<210> 1170

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1170

Ile Glu Ala Ile Arg Gln Ala Gly Gln Glu Met Ile Leu Ala Val

1 5 10 15

<210> 1171

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1171

Ile Glu Thr Arg Ser Ala Arg His Pro

1 5

<210> 1172

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1172

Ile Gly Ser Gln Glu Trp Tyr Thr Thr Val Pro Lys Tyr Val Ala

1 5 10 15

<210> 1173

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1173

Ile Lys Gly Val Ile Val His Arg Leu Glu Gly Val Ser Tyr Asn

1 5 10 15

<210> 1174

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1174

Ile Lys His Ile Ile Ile Val Pro Ile Pro Gly Asp Ser Ser Ile

1 5 10 15

<210> 1175

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1175

Ile Lys His Ile Ile Ile Val Pro Ile Pro Gly Asp Ser Ser Ile Thr

1 5 10 15

Thr Arg Ser Arg

20

<210> 1176

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1176

Ile Lys Leu Lys Asp Cys Pro

1 5

<210> 1177

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1177

Ile Lys Leu Met Pro Asn Ile Thr Leu Leu Asn Asn Cys Thr Arg

1 5 10 15

<210> 1178

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1178

Ile Leu Leu Glu Arg Leu Asp Val Gly Thr

1 5 10

<210> 1179

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1179

Ile Leu Pro Gly Gln Asp Leu Gln Tyr Val

1 5 10

<210> 1180

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1180

Ile Leu Thr Tyr Ile Ala Ala Asp His Cys Pro Val Val Glu Val

1 5 10 15

<210> 1181

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1181

Ile Asn Gln Asp Pro Asp Lys Ile Leu Thr Tyr

1 5 10

<210> 1182

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1182

Ile Pro Arg Phe Lys Val Ser Pro Tyr Leu Phe Asn Val Pro Ile

1 5 10 15

<210> 1183

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1183

Ile Gln Ala Leu Ser Tyr Ala Leu Gly Gly Asp Ile Asn Lys Val

1 5 10 15

<210> 1184

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1184

Ile Arg Ala Gly Pro Tyr Gly His Ala Thr Val Glu Met Pro Glu

1 5 10 15

<210> 1185

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1185

Ile Arg Met Gly Thr Asp Lys

1 5

<210> 1186

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1186

Ile Ser Asn Phe Asp Glu Ser Ser Cys Thr Phe Met Pro Glu Gly

1 5 10 15

<210> 1187

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1187

Ile Thr Ala Gly Ile Ala Leu His Gln Ser Met Leu Asn Ser Gln

1 5 10 15

<210> 1188

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1188

Ile Thr Asp Asp Pro Asp Val Ser Ile Arg Leu Leu Glu Val Val

1 5 10 15

<210> 1189

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1189

Ile Thr Asp Asp Pro Asp Val Ser Ile Arg Leu Leu Glu Val Val Gln

1 5 10 15

Ser Asp Gln Ser

20

<210> 1190

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1190

Ile Thr Thr Tyr Val

1 5

<210> 1191

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1191

Ile Val Glu Ala Gly Leu Ala Ser Phe Ile Leu Thr Ile Lys Phe

1 5 10 15

<210> 1192

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1192

Ile Val Glu Ala Gly Leu Ala Ser Phe Ile Leu Thr Ile Lys Phe Gly

1 5 10 15

Ile Glu Thr Met

20

<210> 1193

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1193

Ile Val Glu Ala Gly Leu Ala Ser Phe Ile Leu Thr Ile Lys Phe Gly

1 5 10 15

Ile Glu Thr Met Tyr Pro Ala Leu Gly

20 25

<210> 1194

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1194

Lys Ala Leu Glu Ala Gln Ser Arg Ala Leu Arg Ala Glu Leu Ala Ala

1 5 10 15

<210> 1195

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1195

Lys Ala Arg Ile Thr His Val Asp Thr Glu Ser Tyr Phe Ile Val

1 5 10 15

<210> 1196

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1196

Lys Cys Gly Leu His Ile Arg Ala Gly Pro Tyr Gly His Ala Thr

1 5 10 15

<210> 1197

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1197

Lys Cys Tyr Thr Thr Gly Thr Ile Ile Asn Gln Asp Pro Asp Lys

1 5 10 15

<210> 1198

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1198

Lys Asp Asn Pro His Pro Lys Gly Ser Arg

1 5 10

<210> 1199

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1199

Lys Asp Asn Arg Ile Pro Ser Tyr Gly Val Leu Ser Val Asp Leu

1 5 10 15

<210> 1200

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1200

Lys Phe Leu Asn Pro Asp Arg Glu Tyr Asp Phe Arg Asp Leu Thr

1 5 10 15

<210> 1201

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1201

Lys Phe Val Thr Ala Ala Leu Leu Asn

1 5

<210> 1202

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1202

Lys Gly Asn Cys Ser Gly Pro Thr Thr Ile Arg

1 5 10

<210> 1203

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1203

Lys Ile Lys Phe Leu Asn Pro Asp Arg Glu Tyr Asp Phe Arg Asp

1 5 10 15

<210> 1204

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1204

Lys Ile Val Asp Gly Gly Cys Phe Ala Pro Trp Asp Leu Glu Ala

1 5 10 15

<210> 1205

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1205

Lys Leu Gly Leu Lys Leu Leu Arg Tyr Tyr Thr Glu Ile Leu Ser

1 5 10 15

<210> 1206

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1206

Lys Leu Gly Val Trp Lys Ser Pro Thr Asp Met Gln Ser Trp Val

1 5 10 15

<210> 1207

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1207

Lys Leu Lys Glu Cys Pro Gln

1 5

<210> 1208

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1208

Lys Leu Leu Arg Tyr Tyr Thr Glu Ile

1 5

<210> 1209

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1209

Lys Leu Met Pro Phe Ser Gly Asp Phe Val

1 5 10

<210> 1210

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1210

Lys Leu Met Pro Asn Ile Thr Leu Leu

1 5

<210> 1211

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1211

Lys Leu Arg Met Glu Thr Cys Phe Gln Gln Ala Cys Lys Gly Lys Ile

1 5 10 15

Gln Ala Leu Cys Glu Asn Pro Glu Trp Ala

20 25

<210> 1212

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1212

Lys Leu Trp Cys Arg His Phe Cys Val

1 5

<210> 1213

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1213

Lys Leu Trp Cys Arg His Phe Cys Val Leu

1 5 10

<210> 1214

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1214

Lys Leu Trp Glu Ser Pro Gln Glu Ile

1 5

<210> 1215

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1215

Lys Met Ser Ser Ala Val Gly Phe Val

1 5

<210> 1216

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1216

Lys Met Ser Ser Ala Val Gly Phe Val Pro Asp Thr Gly Pro Ala Ser

1 5 10 15

Arg

<210> 1217

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1217

Lys Met Val Tyr Ala Thr Asn

1 5

<210> 1218

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1218

Lys Pro Asp Leu Thr Gly Thr Ser Lys Ser Tyr Val Arg Ser Leu

1 5 10 15

<210> 1219

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1219

Lys Pro Asn Leu Ser Ser Lys Arg Ser Glu Leu Ser Gln Leu Ser

1 5 10 15

<210> 1220

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1220

Lys Pro Asn Leu Ser Ser Lys Arg Ser Glu Leu Ser Gln Leu Ser Met

1 5 10 15

Tyr Arg Val Phe

20

<210> 1221

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1221

Lys Gln Ser Arg Gly Glu Ala Arg Glu Ser Tyr Arg Glu Thr Gly

1 5 10 15

<210> 1222

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1222

Lys Gln Ser Arg Gly Glu Ala Arg Glu Ser Tyr Arg Glu Thr Gly Pro

1 5 10 15

Ser Arg Ala Ser

20

<210> 1223

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1223

Lys Arg Phe Ala Gly Val Val Leu Ala Gly Ala Ala Leu Gly Val

1 5 10 15

<210> 1224

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1224

Lys Arg Thr Pro Gly Asn Lys Pro Arg Ile Ala Glu Met Ile Cys

1 5 10 15

<210> 1225

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1225

Lys Arg Thr Pro Gly Asn Lys Pro Arg Ile Ala Glu Met Ile Cys Asp

1 5 10 15

Ile Asp Thr Tyr

20

<210> 1226

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1226

Lys Ser Asn His Asn Asn Val Tyr Trp Leu Thr Ile Pro Pro Met Lys

1 5 10 15

Asn Leu Ala Leu Gly Val Ile Asn Thr Leu

20 25

<210> 1227

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1227

Lys Val Ser Pro Tyr Leu Phe Asn Val

1 5

<210> 1228

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1228

Lys Val Val Gln Leu Pro Lys Ile Ser Lys Asn Ala Leu Lys Ala Asn

1 5 10 15

Gly

<210> 1229

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1229

Lys Val Val Gln Leu Pro Lys Ile Ser Lys Asn Ala Leu Arg Asn Asp

1 5 10 15

Gly

<210> 1230

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 1230

Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile Arg

1 5 10

<210> 1231

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1231

Leu Ala Leu Ala Leu Trp Ala Gly Phe Ala Leu Ser

1 5 10

<210> 1232

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1232

Leu Ala Pro Gly Gly Gly Asn Cys His Leu Thr Val Asn Gly Glu

1 5 10 15

<210> 1233

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1233

Leu Ala Gln Ile Trp Val Leu Leu Ala Lys Ala Val Thr Ala Pro

1 5 10 15

<210> 1234

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1234

Leu Ala Gln Ile Trp Val Leu Leu Ala Lys Ala Val Thr Ala Pro Asp

1 5 10 15

Thr Ala Ala Asp

20

<210> 1235

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1235

Leu Ala Ser Tyr Phe Asn Pro Gly Gly Ser Tyr Tyr Lys Gln Tyr His

1 5 10 15

Pro Thr Ala Cys Glu Val Glu Pro Ala Phe Gly His Ser

20 25

<210> 1236

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1236

Leu Ala Val Ser Lys Gly Asn Cys Ser Gly Pro Thr Thr Ile Arg

1 5 10 15

<210> 1237

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1237

Leu Cys Glu Asn Pro Glu Trp Ala Pro Leu Lys Asp Asn Arg Ile

1 5 10 15

<210> 1238

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1238

Leu Asp Arg Leu Val Arg Leu Ile Gly

1 5

<210> 1239

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1239

Leu Glu Glu Glu Gly Val Thr Pro Leu

1 5

<210> 1240

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1240

Leu Glu His Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly Arg

1 5 10 15

Gly Thr Gly His Phe Ile

20

<210> 1241

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1241

Leu Glu Lys Leu Gly Tyr Ser Gly Gly Asp Leu Leu Gly Ile Leu

1 5 10 15

<210> 1242

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1242

Leu Glu Gln Pro Val Ser Asn Asp Leu Ser

1 5 10

<210> 1243

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1243

Leu Glu Arg Lys Trp Leu Asp Val Val Arg Asn Ile Ile Ala Glu

1 5 10 15

<210> 1244

<211> 20

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1244

Leu Glu Arg Lys Trp Leu Asp Val Val Arg Asn Ile Ile Ala Glu Asp

1 5 10 15

Leu Ser Leu Arg

20

<210> 1245

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1245

Leu Phe Gly Pro Ser Leu Arg Asp Pro Ile Ser Ala Glu Ile Ser

1 5 10 15

<210> 1246

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1246

Leu Gly Glu Leu Lys Leu Ala Ala Leu Cys His Gly Glu Asp Ser

1 5 10 15

<210> 1247

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1247

Leu Gly Gly Lys Glu Asp Arg Arg Val Lys Gln Ser Arg

1 5 10

<210> 1248

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1248

Leu Gly His Asp Gly His His Gly Gly Thr Leu Arg Val Gly Gln His

1 5 10 15

His Arg Asn Ala Ser Asp Val Leu

20

<210> 1249

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1249

Leu Gly Ser Pro Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys

1 5 10 15

Gln Arg His Ser

20

<210> 1250

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1250

Leu Gly Ser Pro Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys

1 5 10 15

Gln Arg His Ser Pro Asp Cys Ser Arg Leu Val

20 25

<210> 1251

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1251

Leu His Asp Pro Asp Thr Glu Ala Pro Thr Glu Ala Cys Val Thr Ser

1 5 10 15

Trp Leu

<210> 1252

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1252

Leu Ile Ala Asn Cys Ala Ser Ile Leu Cys Lys Cys Tyr Thr Thr

1 5 10 15

<210> 1253

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1253

Leu Ile Gly Leu Leu Ala Ile Ala Gly Ile Arg Leu His Arg Ala Ala

1 5 10 15

Ile Tyr Thr Ala Glu Ile His Lys

20

<210> 1254

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1254

Leu Ile Pro Ser Met Asn Gln Leu Ser Cys Asp Leu Ile Gly Gln

1 5 10 15

<210> 1255

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1255

Leu Lys Ile Lys Ile Ala Ser Gly Phe Gly Pro Leu Ile Thr His

1 5 10 15

<210> 1256

<211

> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1256

Leu Lys Ile Lys Ile Ala Ser Gly Phe Gly Pro Leu Ile Thr His Gly

1 5 10 15

Ser Gly Met Asp Leu Tyr Lys

20

<210> 1257

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1257

Leu Lys Leu Tyr Phe Glu Pro

1 5

<210> 1258

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1258

Leu Leu Ala Val Leu Phe Val Met Phe Leu

1 5 10

<210> 1259

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1259

Leu Leu Asp Arg Leu Val Arg Leu Ile Gly Asn Pro Asp Val Ser

1 5 10 15

<210> 1260

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1260

Leu Leu Asp Arg Leu Val Arg Leu Ile Gly Asn Pro Asp Val Ser Gly

1 5 10 15

Pro Lys Leu Thr

20

<210> 1261

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1261

Leu Leu Glu Ser Ser Asp Gln Ile Leu Arg Ser Met Lys Gly Leu

1 5 10 15

<210> 1262

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1262

Leu Leu Glu Thr Arg Thr Thr Asn Gln Phe Leu Ala Val Ser Lys

1 5 10 15

<210> 1263

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1263

Leu Leu Glu Val Val Gln Ser Asp Gln Ser Gln Ser Gly Leu Thr

1 5 10 15

<210> 1264

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1264

Leu Leu Glu Val Val Gln Ser Asp Gln Ser Gln Ser Gly Leu Thr Phe

1 5 10 15

Ala Ser Arg

<210> 1265

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1265

Leu Leu Glu Val Val Gln Ser Asp Gln Ser Gln Ser Gly Leu Thr Phe

1 5 10 15

Ala Ser Arg Gly

20

<210> 1266

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1266

Leu Leu Gly Ile Leu Glu Ser Arg Gly Ile Lys Ala Arg Ile Thr

1 5 10 15

<210> 1267

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1267

Leu Leu Arg Thr Ala Pro Gly Pro

1 5

<210> 1268

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1268

Leu Leu Arg Tyr Tyr Thr Glu Ile Leu Ser Leu Phe Gly Pro Ser

1 5 10 15

<210> 1269

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1269

Leu Leu Val Pro Trp Val Leu Ile Phe Met Val Cys Arg Arg Ala Cys

1 5 10 15

Arg Arg Arg Gly

20

<210> 1270

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1270

Leu Leu Trp Arg Ser Arg Cys Lys Ile Val

1 5 10

<210> 1271

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1271

Leu Leu Trp Ser Tyr Ala Met Gly Val

1 5

<210> 1272

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1272

Leu Leu Trp Ser Tyr Ala Met Gly Val Gly Val Glu Leu Glu Asn

1 5 10 15

<210> 1273

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1273

Leu Leu Trp Ser Tyr Ala Met Gly Val Gly Val Glu Leu Glu Asn Ser

1 5 10 15

Met Gly Gly Leu

20

<210> 1274

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1274

Leu Met Ile Asp Arg Pro Tyr Val Leu

1 5

<210> 1275

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1275

Leu Asn Ala Met Thr Gln Asn Ile Arg Pro Val Gln Ser Val Ala

1 5 10 15

<210> 1276

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1276

Leu Asn Thr Pro Pro Pro Tyr Gln Val Ser Cys Gly Gly Glu Ser

1 5 10 15

<210> 1277

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1277

Leu Arg Asp Pro Ile Ser Ala Glu Ile Ser Ile Gln Ala Leu Ser

1 5 10 15

<210> 1278

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1278

Leu Arg Gly Ser Gly Asp Leu Gln Glu Tyr Leu Arg His Val Thr Arg

1 5 10 15

<210> 1279

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1279

Leu Arg Leu Val Asp Ala Asp Asp

1 5

<210> 1280

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1280

Leu Arg Leu Val Asp Ala Asp Asp Pro Leu Leu Arg

1 5 10

<210> 1281

<211> 31

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1281

Leu Arg Leu Val Asp Ala Asp Asp Pro Leu Leu Arg Thr Ala Pro Gly

1 5 10 15

Pro Gly Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala Arg

20 25 30

<210> 1282

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1282

Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly Leu Ala Val Arg

1 5 10 15

Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr

20 25

<210> 1283

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1283

Leu Arg Thr Ala Pro Gly Pro Gly

1 5

<210> 1284

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1284

Leu Arg Val Gly Gln His Tyr Arg Asn Ala Ser Asp Val Leu Pro Gly

1 5 10 15

His Trp Leu Gln

20

<210> 1285

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1285

Leu Arg Tyr Leu Ala

1 5

<210> 1286

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1286

Leu Ser Cys Lys Pro Trp Gln Glu Ser Arg Lys Asn Lys

1 5 10

<210> 1287

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1287

Leu Ser Glu Ile Lys Gly Val Ile Val His Arg Leu Glu Gly Val

1 5 10 15

<210> 1288

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1288

Leu Ser Ile Ala Tyr Pro Thr Leu Ser Glu Ile Lys Gly Val Ile

1 5 10 15

<210> 1289

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1289

Leu Ser Leu Leu Asp Leu Tyr Leu Gly Arg Gly Tyr Asn Val Ser

1 5 10 15

<210> 1290

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1290

Leu Ser Gln Gly Asn Leu Ile Ala Asn Cys Ala Ser Ile Leu Cys

1 5 10 15

<210> 1291

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1291

Leu Ser Ser His Arg Gly Val Ile Ala Asp Asn Gln Ala Lys Trp

1 5 10 15

<210> 1292

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1292

Leu Ser Val Asp Leu Ser Leu Thr Val Glu Leu Lys Ile Lys Ile

1 5 10 15

<210> 1293

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1293

Leu Thr Gly Ile Ser Ile Cys Asn Pro Gly Ser Ser Leu Cys

1 5 10

<210> 1294

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1294

Leu Thr Ile Lys Phe Gly Ile Glu Thr Met Tyr Pro Ala Leu Gly

1 5 10 15

<210> 1295

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1295

Leu Thr Ile Lys Phe Gly Ile Glu Thr Met Tyr Pro Ala Leu Gly Leu

1 5 10 15

His Glu Phe Ala

20

<210> 1296

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1296

Leu Thr Leu Gln Thr Pro Thr Gly Gln Ile His Trp Gly Asn Leu

1 5 10 15

<210> 1297

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1297

Leu Val Asp Ala Asp Asp Pro Leu

1 5

<210> 1298

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1298

Leu Val Asp Ala Asp Asp Pro Leu Leu Arg

1 5 10

<210> 1299

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1299

Leu Val Glu Lys Pro Asn Leu Ser Ser Lys Arg Ser Glu Leu Ser

1 5 10 15

<210> 1300

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1300

Leu Val Gly Ala Thr Pro Glu

1 5

<210> 1301

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1301

Leu Val Gly Ala Thr Pro Glu Arg

1 5

<210> 1302

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1302

Leu Val Lys Leu Gly Val Trp Lys Ser Pro Thr Gly Met Gln Ser

1 5 10 15

<210> 1303

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1303

Leu Val Ser Gly Ser Phe Gly Asn Arg Phe Ile Leu Ser Gln Gly Asn

1 5 10 15

Leu Ile

<210> 1304

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1304

Leu Tyr Lys Ser Asn His Asn Asn Val Tyr Trp Leu Thr Ile Pro

1 5 10 15

<210> 1305

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1305

Leu Tyr Pro Met Ser Pro Leu Leu Gln Glu Cys Leu Arg Gly Ser Thr

1 5 10 15

Lys Ser Cys Ala Arg Thr Leu Val Ser

20 25

<210> 1306

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1306

Met Ala Ser Thr Thr Pro Ile Thr Met Glu Asp Leu Gln Lys Ala Leu

1 5 10 15

Glu Ala

<210> 1307

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1307

Met Ala Ser Thr Thr Pro Ile Thr Met Glu Asp Leu Gln Lys Ala Leu

1 5 10 15

Glu Ala Gln Ser Arg

20

<210> 1308

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1308

Met Ala Ser Thr Thr Pro Ile Thr Met Glu Asp Leu Gln Lys Ala Leu

1 5 10 15

Glu Ala Gln Ser Arg Ala Leu Arg Ala Glu Leu Ala Ala

20 25

<210> 1309

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1309

Met Ala Ser Thr Thr Pro Ile Thr Met Glu Asp Leu Gln Lys Ala Leu

1 5 10 15

Glu Ala Gln Ser Arg Ala Leu Arg Ala Gly Leu Ala Ala

20 25

<210> 1310

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1310

Met Ala Ser Thr Thr Pro Ile Thr Met Glu Asp Leu Gln Lys Ala Leu
1 5 10 15

Glu Thr Gln Ser Arg Val Leu Arg Ala Gly Leu Thr Ala
20 25

<210> 1311

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1311

Met Ala Thr Leu Leu Arg Ser Leu Ala Leu Phe Lys Arg Asn Lys
1 5 10 15

<210

> 1312

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1312

Met Ala Thr Leu Leu Arg Ser Leu Ala Leu Phe Lys Arg Asn Lys Asp
1 5 10 15

Lys Pro Pro Ile
20

<210> 1313

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1313

Met Asp Leu Tyr Lys Ser Asn His Asn Asn Val Tyr Trp Leu Thr

1 5 10 15

<210> 1314

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1314

Met Glu Asp Leu Gln Lys Ala Leu Glu Ala Gln Ser Arg Ala

1 5 10

<210> 1315

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1315

Met Glu Asp Leu Gln Lys Ala Leu Glu Ala Gln Ser Arg Ala Leu Arg

1 5 10 15

Ala Glu Leu Ala Ala

20

<210> 1316

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1316

Met Gly Leu Lys Val Asn Val Ser Ala Ile Phe Met Ala Val Leu

1 5 10 15

<210> 1317

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1317

Met Ile Asp Arg Pro Tyr Val Leu Leu Ala Val Leu Phe Val Met

1 5 10 15

<210> 1318

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 1318

Met Ile Leu Ala Val Gln Gly Val Gln Asp Tyr Ile Asn Asn Glu

1 5 10 15

<210> 1319

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1319

Met Leu Asn Ser Gln Ala Ile Asp Asn Leu Arg Ala Ser Leu Glu

1 5 10 15

<210> 1320

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1320

Met Asn Ala Leu Val Asn Ser Thr Leu Leu Glu Thr Arg Thr Thr

1 5 10 15

<210> 1321

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1321

Met Asn Tyr Thr Gly Asn Gln Gln Ser Arg Trp Gly Leu Gly Ser Pro

1 5 10 15

Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser

20 25 30

<210> 1322

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223>

Synthetic

<400> 1322

Met Gln Ser Trp Val Pro Leu Ser Thr Asp Asp Pro Val Ile Asp

1 5 10 15

<210> 1323

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1323

Met Ser Leu Ser Leu Leu Asp Leu Tyr Leu Gly Arg Gly Tyr Asn

1 5 10 15

<210> 1324

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1324

Met Ser Pro Leu Leu Gln Glu Cys Leu Arg Gly Ser Thr Lys Ser

1 5 10 15

<210> 1325

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1325

Met Ser Pro Gln Arg Asp Arg Ile Asn Ala Phe Tyr Lys Asp Asn
1 5 10 15

<210> 1326

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1326

Met Tyr Arg Val Phe Glu Val Ser Val Ile Arg Asn Pro Gly Leu
1 5 10 15

<210> 1327

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1327

Asn Ala Leu Tyr Pro Met Ser Pro Leu Leu Gln Glu Cys Leu Arg
1 5 10 15

<210> 1328

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1328

Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser
1 5 10 15

Pro Asp Cys Ser Arg Leu Val Gly Ala Thr
20 25

<210> 1329

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1329

Asn Phe Gly Arg Ser Tyr Phe Asp Pro Ala Tyr Phe Arg Leu Gly

1 5 10 15

<210> 1330

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1330

Asn Phe Gly Arg Ser Tyr Phe Asp Pro Ala Tyr Phe Arg Leu Gly Gln

1 5 10 15

Glu Met Val Arg

20

<210> 1331

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1331

Asn Gly Thr Gln Arg Ala Cys Thr Phe Trp Ala Val Asn Ala Tyr

1 5 10 15

<210> 1332

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1332

Asn Gly Val Thr Ile Gln Val Gly Ser Arg Arg Tyr Pro Asp Ala

1 5 10 15

<210> 1333

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1333

Asn Ile Ile Ala Glu Asp Leu Ser Leu Arg Arg Phe Met Val Ala

1 5 10 15

<210> 1334

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1334

Asn Ile Ile Ala Glu Asp Leu Ser Leu Arg Arg Phe Met Val Ala Leu

1 5 10 15

Ile Leu Asp Ile

20

<210> 1335

<211> 15

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 1335

Asn Ile Thr Leu Leu Asn Asn Cys Thr Arg Val Glu Ile Ala Glu

1 5 10 15

<210> 1336

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1336

Asn Leu Ala Leu Gly Val Ile Asn Thr Leu Glu Trp Ile Pro Arg

1 5 10 15

<210> 1337

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1337

Asn Leu Phe Gln Val Val His Trp Ser Tyr Asn Arg Pro Ala Tyr Ser

1 5 10 15

Pro Gly

<210> 1338

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1338

Asn Leu Val Ile Leu Pro Gly Gln Asp Leu Gln Tyr Val Leu Ala

1 5 10 15

<210> 1339

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1339

Asn Leu Tyr Gln Gln Met Gly Lys Pro Ala Pro Tyr Met Val Asn

1 5 10 15

<210> 1340

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1340

Asn Leu Tyr Gln Gln Met Gly Lys Pro Ala Pro Tyr Met Val Asn Leu

1 5 10 15

Glu Asn Ser Ile

20

<210> 1341

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1341

Asn Asn Cys Thr Arg Val Glu Ile Ala Glu Tyr Arg Arg Leu Leu

1 5 10 15

<210> 1342

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1342

Asn Pro Asp Arg Glu Tyr Asp Phe Arg Asp

1 5 10

<210> 1343

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1343

Asn Pro Asp Val Ser Gly Pro Lys Leu

1 5

<210> 1344

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1344

Asn Pro Asp Val Ser Gly Pro Lys Leu Thr Gly Ala Leu Ile Gly

1 5 10 15

<210> 1345

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1345

Asn Pro Asp Val Ser Gly Pro Lys Leu Thr Gly Ala Leu Ile Gly Ile

1 5 10 15

Leu Ser Leu Phe

20

<210> 1346

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1346

Asn Pro Pro Glu Arg Ile Lys Leu Asp Tyr Asp Gln Tyr Cys Ala

1 5 10 15

<210> 1347

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1347

Asn Gln Ala Lys Trp Ala Val Pro Thr Thr Arg Thr Asp Asp Lys

1 5 10 15

<210> 1348

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1348

Asn Gln Asp Pro Asp Lys Ile Leu Thr Tyr Ile Ala Ala Asp His

1 5 10 15

<210> 1349

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1349

Asn Gln Leu Ser Cys Asp Leu Ile Gly Gln Lys Leu Gly Leu Lys

1 5 10 15

<210> 1350

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1350

Asn Gln Gln Ser Arg Trp Gly Leu Gly Ser Pro Asn Cys His Gly Pro

1 5 10 15

Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser

20 25

<210> 1351

<211

> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1351

Asn Ser Thr Leu Gly Val Lys Ser Ala Arg Glu Phe

1 5 10

<210> 1352

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1352

Asn Thr Pro His Gly Gln Leu Glu Val Gln Val Pro Pro Asp Pro

1 5 10 15

<210> 1353

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<

400> 1353

Asn Val Ser Ala Ile Phe Met Ala Val Leu Leu Thr Leu Gln Thr

1 5 10 15

<210> 1354

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1354

Pro Ala Glu Val Asp Gly Asp Val Lys Leu Ser Ser Asn Leu Val

1 5 10 15

<210> 1355

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1355

Pro Ala Phe Gly His Ser Asp Ala Ala Cys Trp Gly Phe Pro Thr

1 5 10 15

<210> 1356

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1356

Pro Ala Leu Ile Cys Cys Cys Arg Gly Arg Cys Asn Lys Lys Gly

1 5 10 15

<210> 1357

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1357

Pro Asp Lys Ile Leu Thr Tyr Ile Ala Ala Asp His Cys

1 5 10

<210

> 1358

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1358

Pro Glu Arg Ile Lys Leu Asp Tyr Asp Gln Tyr Cys Ala Asp Val

1 5 10 15

<210> 1359

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1359

Pro Glu Arg Pro Arg Leu Arg Leu

1 5

<210> 1360

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1360

Pro Gly Cys Ala Thr Gln Ala Pro Val Pro Val Arg Leu Ala Gly

1 5 10 15

<210> 1361

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1361

Pro Gly Cys Ala Thr Gln Ala Pro Val Pro Val Arg Leu Ala Gly Val

1 5 10 15

Arg Phe Glu Ser Lys Ile Val Asp Gly Gly Cys Phe Ala

20 25

<210> 1362

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1362

Pro Gly Glu Val Trp Val Thr Pro

1 5

<210> 1363

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1363

Pro Gly Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala Arg

1 5 10 15

<210> 1364

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1364

Pro Gly Lys Leu Asp Val Asn Lys Ser Lys Thr His Ile Ser Val Asn

1 5 10 15

<210> 1365

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1365

Pro Gly Leu Gly Ala Pro Val Phe His Met Thr Asn Tyr Leu Glu

1 5 10 15

<210> 1366

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1366

Pro Gly Pro Gly Glu Val Trp Val

1 5

<210> 1367

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1367

Pro His Lys Thr Val Arg Val Lys Phe His Thr Glu Thr Arg Thr

1 5 10 15

<210> 1368

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1368

Pro Ile Glu Leu Gln Val Glu Cys Phe Thr Trp Asp Gln Lys Leu

1 5 10 15

<210> 1369

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1369

Pro Ile Ser Leu Glu Arg Leu Asp Val Gly

1 5 10

<210> 1370

<211> 15

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1370

Pro Ile Ser Leu Glu Arg Leu Asp Val Gly Thr Asn Leu Gly Asn

1 5 10 15

<210> 1371

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1371

Pro Lys Ala Leu Phe Thr Gln Gln Gly Gly Ala Tyr Gly Arg Cys

1 5 10 15

<210> 1372

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1372

Pro Lys Tyr Val Ala Thr Gln Gly Tyr Leu Ile Ser Asn Phe Asp

1 5 10 15

<210> 1373

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1373

Pro Leu Asp Ile Asp Thr Ala Ser Glu Ser Ser Gln Asp

1 5 10

<210> 1374

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1374

Pro Leu Gly Leu Lys Phe Lys Thr Val Arg Pro Val Ala Leu Pro

1 5 10 15

<210> 1375

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1375

Pro Leu Ile Thr His Gly Ser Gly Met Asp Leu Tyr Lys Ser Asn

1 5 10 15

<210> 1376

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1376

Pro Leu Lys Asp Asn Arg Ile Pro Ser Tyr Gly Val Leu Ser Val

1 5 10 15

<210> 1377

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1377

Pro Leu Leu Arg Thr Ala Pro Gly

1 5

<210> 1378

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1378

Pro Leu Leu Ser Cys Lys Pro Trp Gln Glu Ser Arg Lys

1 5 10

<210> 1379

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1379

Pro Pro Ala Thr Val Tyr Arg Tyr Asp Ser Arg Pro Pro Glu

1 5 10

<210> 1380

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1380

Pro Pro Ile Ser Leu Glu Arg Leu Asp Val Gly Thr

1 5 10

<210> 1381

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1381

Pro Pro Pro Pro Glu Glu Arg Gln Glu Thr Arg Ser Gln Thr Pro Ala

1 5 10 15

Pro Lys Pro Ser

20

<210> 1382

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1382

Pro Gln Glu Gln Ile Thr Gln His Gly Ser Pro Tyr Gly Arg Cys

1 5 10 15

<210> 1383

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1383

Pro Gln Glu Gln Ile Thr Gln His Gly Ser Pro Tyr Gly Arg Cys Ala

1 5 10 15

Asn Lys

<210

> 1384

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1384

Pro Gln Pro Gly Pro Gln Pro Pro Gln Pro Pro Gln Pro Gln Pro Glu

1 5 10 15

Ala Pro Ala Pro Gln Pro Pro Ala Gly

20 25

<210> 1385

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1385

Pro Arg Leu Gly Gly Lys Glu Asp Arg Arg Val Lys Gln

1 5 10

<210> 1386

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1386

Pro Arg Leu Arg Leu Val Asp Ala

1 5

<210> 1387

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1387

Pro Arg Asn Val Arg Val Thr Gly Cys Tyr Gln Cys Gly Thr Pro

1 5 10 15

<210> 1388

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1388

Pro Ser Arg Ala Ser Asp Ala Arg Ala Ala His Leu Pro

1 5 10

<210> 1389

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1389

Pro Thr Gly Gln Ile His Trp Gly Asn Leu Ser Lys Ile Gly Val

1 5 10 15

<210> 1390

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1390

Pro Thr Gly Thr Pro Leu Asp Ile Asp Thr Ala Ser Glu

1 5 10

<210> 1391

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1391

Pro Thr Leu Ser Glu Ile Lys Gly Val Ile Val His Arg Leu Glu

1 5 10 15

<210> 1392

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1392

Pro Thr Thr Ile Arg Gly Gln Phe Ser Asn Met Ser Leu Ser Leu

1 5 10 15

<210> 1393

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1393

Pro Thr Thr Arg Thr Asp Asp Lys Leu Arg

1 5 10

<210> 1394

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1394

Pro Thr Thr Arg Thr Asp Asp Lys Leu Arg Met Glu Thr Cys Phe Gln

1 5 10 15

Gln Ala Cys Lys Gly

20

<210> 1395

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1395

Pro Val Ala Leu Pro Arg Thr Leu Ala Pro Pro Arg Asn Val Arg

1 5 10 15

<210> 1396

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1396

Pro Val Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala

1 5 10 15

Gln Val Ile

<210> 1397

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1397

Pro Val Ile Gly Ser Gln Ala Arg

1 5

<210> 1398

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1398

Pro Val Val Glu Val Asn Gly Val Thr Ile Gln Val Gly Ser Arg

1 5 10 15

<210> 1399

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1399

Pro Trp Glu Leu Val Val Leu Thr Ala Arg Pro Glu Asp Gly Trp Thr

1 5 10 15

Cys Arg Gly Val

20

<210> 1400

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1400

Pro Trp Gln Glu Ser Arg Lys Asn Lys Ala Gln Thr Arg

1 5 10

<210> 1401

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1401

Pro Tyr Met Val Asn Leu Glu Asn Ser Ile Gln Asn Lys Phe Ser

1 5 10 15

<210> 1402

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1402

Pro Tyr Met Val Asn Leu Glu Asn Ser Ile Gln Asn Lys Phe Ser Ala

1 5 10 15

Gly Ser Tyr Pro

20

<210> 1403

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1403

Pro Tyr Gln Gly Ser Gly Lys Gly Val Ser

1 5 10

<210> 1404

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1404

Pro Tyr Gln Val Ser Cys Gly Gly Glu Ser Asp Arg Ala Ser Ala

1 5 10 15

<210> 1405

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1405

Gln Ala Cys Lys Gly Lys Ile Gln Ala Leu Cys Glu Asn

1 5 10

<210> 1406

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1406

Gln Ala Gly Gln Glu Met Ile Leu Ala Val Gln Gly Val Gln Asp

1 5 10 15

<210

> 1407

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1407

Gln Ala Leu Cys Glu Asn Pro Glu Cys Val Pro Leu Lys Asp Asn

1 5 10 15

<210> 1408

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1408

Gln Ala Leu Gly Ala Leu Lys

1 5

<210> 1409

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1409

Gln Ala Pro Val Pro Val Arg Leu Ala Gly Val Arg Phe Glu Ser

1 5 10 15

<210> 1410

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1410

Gln Cys Gly Thr Pro Ala Leu Val Glu Gly Leu Ala Pro Gly Gly

1 5 10 15

<210> 1411

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1411

Gln Asp Pro Asp Lys Ile Leu Thr Tyr Ile Ala Ala Asp His Cys

1 5 10 15

<210> 1412

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1412

Gln Glu Cys Leu Arg Gly Ser Thr Lys Ser Cys Ala Arg Thr Leu

1 5 10 15

<210> 1413

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1413

Gln Glu Gln Ile Thr Gln His Gly Ser Pro Tyr Gly Arg Cys

1 5 10

<210> 1414

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1414

Gln Glu Ser Arg Lys Asn Lys Ala Gln Thr Arg Thr Pro

1 5 10

<210> 1415

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1415

Gln Gly Asp Gln Ser Glu Asn Glu Leu Pro Arg Leu Gly Gly Lys

1 5 10 15

<210> 1416

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1416

Gln Gly Asp Gln Ser Glu Asn Glu Leu Pro Arg Leu Gly Gly Lys Glu

1 5 10 15

Asp Arg Arg Val

20

<210> 1417

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1417

Gln Gly Glu Ser Gly His Asp Ile Lys Ile Thr Ala Glu Asn Thr Pro

1 5 10 15

Leu Pro Ile Ala

20

<210> 1418

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1418

Gln Gly Ser Gly Lys Gly Val Ser Phe Gln Leu Val Lys Leu Gly

1 5 10 15

<210> 1419

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1419

Gln Gly Val Gln Asp Tyr Ile Asn Asn Glu Leu Ile Pro Ser Met

1 5 10 15

<210> 1420

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1420

Gln Leu Glu Val Gln Val Pro Pro Asp Pro Gly Asp Leu Val Glu

1 5 10 15

<210> 1421

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1421

Gln Leu Pro Glu Ala Thr Phe Met Val

1 5

<210> 1422

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1422

Gln Leu Pro Phe Leu Gly His Asp Gly His His Gly Gly Thr Leu Arg

1 5 10 15

Val Gly Gln His Tyr Arg Asn Ala Ser

20 25

<210> 1423

<

211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1423

Gln Leu Ser Met Tyr Arg Val Phe Glu Val

1 5 10

<210> 1424

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1424

Gln Leu Ser Asn Ile Thr

1 5

<210> 1425

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1425

Gln Asn Ile Arg Pro Val Gln Ser Val Ala Ser Ser Arg Arg His

1 5 10 15

<210> 1426

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1426

Gln Asn Lys Phe Ser Ala Gly Ser Tyr Pro Leu Leu Trp Ser Tyr

1 5 10 15

<210> 1427

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1427

Gln Asn Lys Phe Ser Ala Gly Ser Tyr Pro Leu Leu Trp Ser Tyr Ala

1 5 10 15

Met Gly Val Gly

20

<210> 1428

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1428

Gln Asn Lys Phe Ser Ala Gly Ser Tyr Pro Leu Leu Trp Ser Tyr Ala

1 5 10 15

Met Gly Val Gly Val Glu Leu Glu Asn

20 25

<210> 1429

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1429

Gln Gln Ala Cys Lys Gly Lys Ile Gln Ala Leu Cys Glu Asn Pro

1 5 10 15

<210> 1430

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1430

Gln Gln Arg Arg Val Val Gly Glu Phe Arg Leu Glu Arg Lys Trp

1 5 10 15

<210> 1431

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1431

Gln Gln Arg Arg Val Val Gly Glu Phe Arg Leu Glu Arg Lys Trp Leu

1 5 10 15

Asp Val Val Arg

20

<210> 1432

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1432

Gln Gln Thr Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val

1 5 10 15

Ala Ser

<210> 1433

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1433

Gln Arg His Ser Pro Asp Cys Ser Arg Leu Val Gly Ala Thr Pro

1 5 10 15

<210> 1434

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1434

Gln Ser Gly Leu Thr Phe Ala Ser Arg Gly Thr Asn Met Glu Asp

1 5 10 15

<210> 1435

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1435

Gln Ser Gly Leu Thr Phe Ala Ser Arg Gly Thr Asn Met Glu Asp Glu

1 5 10 15

Ala Asp Gln Tyr

20

<210> 1436

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1436

Gln Ser Ile Ala Leu Ser Ser Leu Met Val Ala Gln Ala Ile Pro Leu

1 5 10 15

Val Gly Glu Leu

20

<210> 1437

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1437

Gln Ser Arg Phe Gly Trp Phe Glu Asn Lys Glu Ile Ser Asp Ile

1 5 10 15

<210> 1438

<211> 20

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 1438

Gln Ser Arg Phe Gly Trp Phe Glu Asn Lys Glu Ile Ser Asp Ile Glu

1 5 10 15

Val Gln Asp Pro

20

<210> 1439

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1439

Gln Ser Arg Gly Glu Ala Arg

1 5

<210> 1440

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1440

Gln Ser Arg Gly Glu Ala Arg Glu Ser Tyr Arg Glu Thr Gly Pro Ser

1 5 10 15

Arg Ala

<210> 1441

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1441

Gln Thr Gly Arg Gly Gly Ser Ala Pro Arg Pro Glu Leu Gly Pro Pro

1 5 10 15

Thr Asn

<210> 1442

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1442

Gln Thr Pro Ala Pro Lys Pro Ser Arg Ala Pro Pro Gln Gln Pro Gln

1 5 10 15

Pro Pro Arg Met Gln Thr Gly Arg Gly

20 25

<210> 1443

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1443

Gln Val Gly Ser Arg Arg Tyr Pro Asp Ala Val Tyr Leu His Arg

1 5 10 15

<210> 1444

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1444

Gln Val Ser Phe Leu Gln Gly Asp Gln Ser Glu Asn Glu

1 5 10

<210> 1445

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1445

Gln Tyr His Pro Thr Ala Cys Glu Val Glu Pro Ala Phe Gly His

1 5 10 15

<210> 1446

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1446

Gln Tyr Val Leu Ala Thr Tyr Asp Thr Ser Arg Val Glu His Ala

1 5 10 15

<210> 1447

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1447

Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val

1 5 10

<210> 1448

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1448

Arg Ala Ser Asp Ala Arg Ala Ala His Leu Pro Thr Gly

1 5 10

<210> 1449

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1449

Arg Ala Ser Leu Glu Thr Thr Asn Gln Ala Ile Glu Ala Ile Arg

1 5 10 15

<210> 1450

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1450

Arg Cys Gly Arg Leu Ile Cys Gly Leu Ser Thr Thr Ala Gln Tyr Pro

1 5 10 15

Pro Thr Arg Phe

20

<210> 1451

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1451

Arg Asp Gly Gln Ser Val Ile Gly Ala Cys Ala Ser Pro Tyr Glu Gly

1 5 10 15

Arg Tyr Arg

<210> 1452

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1452

Arg Glu Ser Tyr Arg Glu Thr Gly Pro Ser Arg Ala Ser

1 5 10

<210> 1453

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1453

Arg Glu Thr Gly Pro Ser Arg Ala Ser Asp Ala Arg Ala

1 5 10

<210> 1454

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1454

Arg Phe Ala Lys Tyr Gln Gln Gln Gly Arg Leu Glu Ala Arg

1 5 10

<210> 1455

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1455

Arg Phe Gly Ala Pro Gln Ala Phe Leu Ala Gly Leu Leu Leu Ala Thr
1 5 10 15

Val Ala Val Gly Thr Ala Arg Ala
20

<210> 1456

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1456

Arg Phe Met Val Ala Leu Ile Leu Asp Ile Lys Arg Thr Pro Gly
1 5 10 15

<210> 1457

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1457

Arg Phe Met Val Ala Leu Ile Leu Asp Ile Lys Arg Thr Pro Gly Asn
1 5 10 15

Lys Pro Arg Ile
20

<210> 1458

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1458

Arg Gly Glu Ala Arg Glu Ser Tyr Arg Glu Thr Gly Pro

1 5 10

<210> 1459

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1459

Arg Gly Thr Thr Pro Pro Ala Tyr Gly

1 5

<210> 1460

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1460

Arg Ile Glu Thr Arg Ser Ala Arg His

1 5

<210> 1461

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1461

Arg Ile Leu Ala Gly Ala Leu Ala Thr Tyr Gln

1 5 10

<210> 1462

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1462

Arg Ile Pro Pro Glu Asn Ile Arg Arg Val Thr

1 5 10

<210> 1463

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1463

Arg Ile Ser Asn Leu Asn Asp

1 5

<210> 1464

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1464

Arg Leu Ala Asn Leu Asn Gly

1 5

<210> 1465

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1465

Arg Leu Asp Val Gly Thr Asn Leu Gly Asn Ala Ile Ala Lys Leu

1 5 10 15

<210> 1466

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1466

Arg Leu Glu Arg Lys Trp Leu Asp Val

1 5

<210> 1467

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1467

Arg Leu Gly Gly Lys Glu Asp Arg Arg Val Lys Gln Ser Arg Gly

1 5 10 15

<210> 1468

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1468

Arg Leu Gly Gly Lys Glu Asp Arg Arg Val Lys Gln Ser Arg Gly Glu

1 5 10 15

Ala Arg Glu Ser

20

<210> 1469

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1469

Arg Leu Leu Asp Arg Leu Val Arg Leu

1 5

<210> 1470

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1470

Arg Leu Arg Leu Val Asp Ala Asp

1 5

<210> 1471

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1471

Arg Leu Arg Leu Val Asp Ala Asp Asp Pro Leu Leu Arg

1 5 10

<210> 1472

<211> 31

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1472

Arg Leu Arg Leu Val Asp Ala Asp Asp Pro Leu Leu Arg Thr Ala Pro

1 5 10 15

Gly Pro Gly Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala

20 25 30

<210> 1473

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1473

Arg Leu Arg Leu Val Gln Asp Ala Asp Asp Pro Leu Leu Arg Ile Ala

1 5 10 15

Pro Gly Pro Gly Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala

20 25 30

<210> 1474

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223>

Synthetic

<400> 1474

Arg Leu Ser Asp Asn Gly Tyr Tyr Thr Val

1 5 10

<210> 1475

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1475

Arg Leu Val Asp Ala Asp Asp Pro

1 5

<210> 1476

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1476

Arg Leu Val Asp Ala Asp Asp Pro Leu Leu Arg Thr Ala Pro Gly

1 5 10 15

<210> 1477

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1477

Arg Leu Val Gly Ala Thr Pro Glu

1 5

<210> 1478

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1478

Arg Met Gln Thr Gly Arg Gly Gly Ser Ala Pro Arg Pro Glu Leu Gly

1 5 10 15

Pro Pro Thr Asn Pro Phe Gln Ala Ala Val Ala

20 25

<210> 1479

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1479

Arg Met Ser Lys Gly Val Phe Lys Val

1 5

<210> 1480

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1480

Arg Asn Pro Gly Leu Gly Ala Pro Val Phe His Met Thr Asn Tyr

1 5 10 15

<210> 1481

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1481

Arg Pro Arg Leu Arg Leu Val Asp

1 5

<210> 1482

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1482

Arg Gln Ala Glu Ser Ser Glu Ala Met Ala Ala Trp Ser Glu Arg Ala
1 5 10 15

Gly Glu Ala

<210> 1483

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1483

Arg Gln Thr Trp Ala Glu Trp Ala Ala Ala His Trp Trp Gln Leu
1 5 10 15

<210> 1484

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1484

Arg Arg Val Lys Gln Ser Arg Gly Glu Ala Arg Glu Ser
1 5 10

<210> 1485

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1485

Arg Arg Tyr Pro Asp Ala Val Tyr Leu
1 5

<210> 1486

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1486

Arg Arg Tyr Pro Asp Ala Val Tyr Leu His Arg Ile Asp Leu Gly

1 5 10 15

<210> 1487

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1487

Arg Ser Ala Gly Lys Val Ser Ser Thr Leu Ala Ser Glu Leu Gly

1 5 10 15

<210> 1488

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1488

Arg Ser Ala Gly Lys Val Ser Ser Thr Leu Ala Ser Glu Leu Gly Ile

1 5 10 15

Thr Ala Glu Asp

20

<210> 1489

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1489

Arg Ser Glu Leu Ser Gln Leu Ser

1 5

<210> 1490

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1490

Arg Ser Glu Leu Ser Gln Leu Ser Met Tyr Arg Val Phe Glu Val

1 5 10 15

<210> 1491

<211> 24

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 1491

Arg Ser Gln Thr Pro Ala Pro Lys Pro Ser Arg Ala Pro Pro Gln Gln

1 5 10 15

Pro Gln Pro Pro Arg Met Gln Thr

20

<210> 1492

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1492

Arg Thr Ala Pro Gly Pro Gly Glu

1 5

<210> 1493

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1493

Arg Thr Ala Pro Gly Pro Gly Glu Val Trp Val Thr Pro Val Ile

1 5 10 15

<210> 1494

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1494

Arg Thr Asp Asp Lys Leu Arg Met Glu Thr Cys Phe Gln Gln Ala

1 5 10 15

<210> 1495

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1495

Arg Thr Leu Ala Pro Pro Arg Asn Val Arg Val Thr Gly Cys Tyr

1 5 10 15

<210> 1496

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1496

Arg Thr Val Leu Glu Pro Ile Arg Asp Ala Leu Asn Ala Met Thr

1 5 10 15

<210> 1497

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1497

Arg Val Glu His Ala Val Val Tyr Tyr Val Tyr Ser Pro Ser Arg

1 5 10 15

<210> 1498

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1498

Arg Val Phe Glu Val Gly Val Ile Arg Asn Pro Gly Leu Gly Ala

1 5 10 15

<210> 1499

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1499

Arg Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala

1 5 10 15

Thr Phe Glu

<210> 1500

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1500

Arg Val Ile Asp Pro Ala Ala Gln

1 5

<210> 1501

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1501

Arg Val Lys Phe His Thr Glu Thr Arg Thr Val Trp Gln Leu Ser

1 5 10 15

<210> 1502

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1502

Arg Val Tyr His Asn Gly Ile Thr Gly Glu Thr

1 5 10

<210> 1503

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1503

Arg Tyr Asp Ser Arg Pro Pro Glu Asp Val Phe

1 5 10

<210> 1504

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1504

Arg Tyr Pro Asp Ala Val Tyr Leu His Arg Ile Asp Leu Gly Pro

1 5 10 15

<210> 1505

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1505

Ser Ala Glu Ile Ser Ile Gln Ala Leu Ser Tyr Ala Leu Gly Gly

1 5 10 15

<210> 1506

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1506

Ser Ala Glu Pro Leu Leu Ser Cys Lys Pro Trp Gln Glu Ser Arg

1 5 10 15

<210> 1507

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1507

Ser Ala Glu Pro Leu Leu Ser Cys Lys Pro Trp Gln Glu Ser Arg Lys

1 5 10 15

Asn Lys Ala Gln

20

<210> 1508

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1508

Ser Ala Gly Lys Val Ser Ser Thr Leu Ala Ser Glu Leu Gly

1 5 10

<210> 1509

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1509

Ser Ala Gly Lys Val Ser Ser Thr Leu Ala Ser Glu Leu Gly Ile Thr

1 5 10 15

Ala Glu Asp Ala Arg Leu Val Ser

20

<210> 1510

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1510

Ser Cys Thr Val Thr Arg Glu Asp Gly Thr

1 5 10

<210> 1511

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1511

Ser Asp Ala Ala Cys Trp Gly Phe Pro Thr Asp Thr Val Met Ser

1 5 10 15

<210> 1512

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1512

Ser Asp Ala Arg Ala Ala His Leu Pro Thr Gly Thr Pro

1 5 10

<210> 1513

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1513

Ser Asp Trp His Gln Gly Thr His Val Cys His Thr Lys His Met Asp

1 5 10 15

Phe Trp Cys Val Glu His Asp

20

<210> 1514

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1514

Ser Glu Leu Arg Arg Trp Ile Lys Tyr Thr Gln Gln Arg Arg Val

1 5 10 15

<210> 1515

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1515

Ser Glu Leu Arg Arg Trp Ile Lys Tyr Thr Gln Gln Arg Arg Val Val

1 5 10 15

Gly Glu Phe Arg

20

<210> 1516

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1516

Ser Glu Leu Ser Gln Leu

1 5

<210> 1517

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1517

Ser Glu Leu Ser Gln Leu Ser

1 5

<210> 1518

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1518

Ser Glu Tyr Leu Ala His Arg Arg Ile Pro Pro Glu Asn Ile Arg Arg

1 5 10 15

Val Thr Arg Val

20

<210> 1519

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1519

Ser Phe Leu Gln Gly Asp Gln Ser Glu Asn Glu Leu Pro

1 5 10

<210> 1520

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1520

Ser Gly Lys Gly Val Ser Phe Gln Leu Val Lys Leu Gly Val Trp

1 5 10 15

<210> 1521

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1521

Ser His Arg Gly Val Ile Ala Asp Asn Gln Ala Lys Trp Ala Val

1 5 10 15

<210> 1522

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1522

Ser Ile Glu His Gln Val Lys Asp Val Leu Thr Pro Leu Phe Lys

1 5 10 15

<210> 1523

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1523

Ser Lys Ile Gly Val Val Gly Ile Gly Ser Ala Ser Tyr Lys Val

1 5 10 15

<210> 1524

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1524

Ser Lys Arg Ser Glu Leu Ser Gln Leu Ser Met Tyr Arg Val Phe

1 5 10 15

<210> 1525

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1525

Ser Leu Phe Val Glu Ser Pro Gly Gln Leu Ile Gln Arg Ile Thr Asp

1 5 10 15

Asp Pro Asp Val Ser

20

<210> 1526

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1526

Ser Leu Ser Thr Asn Leu Asp Val Thr Asn Ser Ile Glu His Gln

1 5 10 15

<210> 1527

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1527

Ser Leu Ser Thr Asn Leu Asp Val Thr Asn Ser Ile Glu His Gln Val

1 5 10 15

Lys Asp Val Leu Thr Pro Leu Phe Lys

20 25

<210> 1528

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1528

Ser Leu Trp Gly Ser Gly Leu Leu Met Leu

1 5 10

<210> 1529

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1529

Ser Met Lys Gly Leu Ser Ser Thr Ser Ile Val Tyr Ile Leu Ile

1 5 10 15

<210> 1530

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1530

Ser Met Tyr Arg Val Phe Glu Val Gly Val

1 5 10

<210> 1531

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1531

Ser Asn Asp Leu Ser Asn Cys Met Val Ala Leu Gly Glu Leu Lys

1 5 10 15

<210> 1532

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1532

Ser Pro Gly Gln Leu Ile Gln Arg

1 5

<210> 1533

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1533

Ser Gln Asp Pro Gln Asp Ser Arg Arg Ser Ala Glu Pro

1 5 10

<210> 1534

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1534

Ser Arg Ile Val Ile Asn Arg Glu His Leu Met Ile Asp Arg Pro

1 5 10 15

<210> 1535

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1535

Ser Arg Lys Asn Lys Ala Gln Thr Arg Thr Pro Leu Gln

1 5 10

<210> 1536

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1536

Ser Arg Leu Val Gly Ala Thr Pro

1 5

<210> 1537

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1537

Ser Arg Leu Val Gly Ala Thr Pro Glu Arg Pro Arg Leu Arg Leu Val

1 5 10 15

Asp Ala Asp Asp Pro Leu Leu Arg

20

<210> 1538

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1538

Ser Arg Pro Gly Leu Lys Pro Asp Leu Thr Gly Thr Ser Lys Ser

1 5 10 15

<210> 1539

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1539

Ser Arg Arg Ser Val Ala Ser Ile Val Gly Thr Leu Val Arg Met

1 5 10 15

<210> 1540

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1540

Ser Arg Trp Gly Leu Gly Ser Pro Asn Cys His Gly Pro Asp Trp

1 5 10 15

<210> 1541

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1541

Ser Ser Ala Thr Lys

1 5

<210> 1542

<211> 15

<212> PRT

<213> Artificial Sequence

<

220><223> Synthetic

<400> 1542

Ser Ser Gly Gly Tyr Ala Gln Leu Ala Ser Tyr Phe Asn Pro Gly

1 5 10 15

<210> 1543

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1543

Ser Ser Leu Gly Asn Gly Val

1 5

<210> 1544

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1544

Ser Ser Asn Leu Val Ile Leu Pro Gly Gln Asp Leu Gln Tyr Val

1 5 10 15

<210> 1545

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1545

Ser Ser Gln Asp Pro Gln Asp Ser Arg Arg Ser Ala Glu Pro Leu

1 5 10 15

<210> 1546

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1546

Ser Ser Gln Asp Pro Gln Asp Ser Arg Arg Ser Ala Glu Pro Leu Leu

1 5 10 15

Ser Cys Lys Pro

20

<210> 1547

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1547

Ser Ser Arg Arg His Lys Arg Phe Ala Gly Val Val Leu Ala Gly

1 5 10 15

<210> 1548

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1548

Ser Ser Thr Ser Ile Val Tyr Ile Leu Ile Ala Val Cys Leu Gly

1 5 10 15

<210> 1549

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1549

Ser Thr Pro Gly Ile Val Ile

1 5

<210> 1550

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1550

Ser Thr Pro Gly Ile Val Ile Pro Pro Gln Glu Gln Ile Thr Gln His

1 5 10 15

Gly Ser Pro Tyr Gly Arg Cys

20

<210> 1551

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1551

Ser Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val Tyr

1 5 10

<210> 1552

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1552

Ser Tyr Phe Ile Val Leu Ser Ile Ala Tyr Pro Thr Leu Ser Glu

1 5 10 15

<210> 1553

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1553

Ser Tyr Arg Glu Thr Gly Pro Ser Arg Ala Ser Asp Ala

1 5 10

<210> 1554

<211

> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1554

Ser Tyr Val Lys

1

<210> 1555

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1555

Ser Tyr Val Gln His Pro His Lys Thr Val Arg Val Lys Phe His

1 5 10 15

<210> 1556

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1556

Thr Ala Pro Gly Pro Gly Glu Val

1 5

<210> 1557

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1557

Thr Ala Thr Arg Leu Leu Ser Ser Thr Asn Ser Arg Leu Cys

1 5 10

<210> 1558

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1558

Thr Asp Asp Pro Val Ile Asp Arg Leu Tyr Leu Ser Ser His Arg

1 5 10 15

<210> 1559

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1559

Thr Glu Ile Leu Ser Leu Phe Gly Pro Ser Leu Arg Asp Pro Ile

1 5 10 15

<210> 1560

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1560

Thr Glu Thr Arg Thr Val Trp Gln Leu Ser Val Ala Gly Val Ser

1 5 10 15

<210> 1561

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1561

Thr Glu Val Tyr Leu Glu His Arg Met Gln Glu Ala Val Glu

1 5 10

<210> 1562

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1562

Thr Phe Met Pro Glu Gly Thr Val Cys Ser Gln Asn Ala Leu Tyr

1 5 10 15

<210> 1563

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1563

Thr Gly Ala Cys Ile Cys Glu Ile Pro Thr Asp Val Ser Cys Glu

1 5 10 15

<210> 1564

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1564

Thr Gly Asp Leu Arg Ala Tyr

1 5

<210> 1565

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1565

Thr Gly Met Gln Ser Trp Val Pro Leu Ser Thr Asp Asp Pro Val

1 5 10 15

<210> 1566

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1566

Thr Gly Asn Gln Gln Ser Arg Trp Gly Leu Gly Ser Pro Asn Cys

1 5 10 15

<210> 1567

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1567

Thr Gly Pro Ser Arg Ala Ser Asp Ala Arg Ala Ala His

1 5 10

<210> 1568

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1568

Thr Gly Thr Ile Ile Asn Gln Asp Pro Asp Lys Ile Leu Thr Tyr

1 5 10 15

<210

> 1569

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1569

Thr Gly Val Val Tyr Gly Thr His Thr Thr Ala Val Ser Glu Thr

1 5 10 15

<210> 1570

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1570

Thr Ile Arg Gly Gln Phe Ser Asn Met Ser Leu Ser Leu Leu Asp

1 5 10 15

<210> 1571

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1571

Thr Leu Gly Ala Thr Cys Ala Leu Pro Leu Ala Gly Leu Leu Ala

1 5 10 15

<210> 1572

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1572

Thr Leu Leu Asn Asn Cys Thr Arg Val

1 5

<210> 1573

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1573

Thr Leu Asn Val Pro Pro Pro Pro Asp Pro Gly Arg

1 5 10

<210> 1574

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1574

Thr Leu Asn Val Pro Pro Pro Pro Asp Pro Gly Arg Ala Ser Thr Ser

1 5 10 15

Gly Thr Pro Ile Lys Lys

20

<210> 1575

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1575

Thr Met Thr Ser Gln Gly Met Tyr Gly Gly Thr Tyr Pro Val Glu

1 5 10 15

<210> 1576

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1576

Thr Asn Leu Gly Asn Ala Ile Ala Lys Leu Glu Asp Ala Lys Glu

1 5 10 15

<210> 1577

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1577

Thr Asn Met Glu Asp Glu Ala Asp Gln Tyr Phe Ser His Asp Asp

1 5 10 15

<210> 1578

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1578

Thr Asn Met Glu Asp Glu Ala Asp Gln Tyr Phe Ser His Asp Asp Pro
1 5 10 15

Ile Ser Ser Asp
20

<210> 1579

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1579

Thr Asn Pro Phe Gln Ala Ala Val Ala Arg Gly Leu Arg Pro Pro
1 5 10 15

<210> 1580

<211> 15

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 1580

Thr Asn Ser Ile Glu His Gln Val Lys Asp Val Leu Thr Pro Leu
1 5 10 15

<210> 1581

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1581

Thr Asn Tyr Leu Glu Gln Pro Val Ser Asn Asp Leu Ser Asn Cys
1 5 10 15

<210> 1582

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1582

Thr Asn Tyr Leu Glu Gln Pro Val Ser Asn Asp Leu Ser Asn Cys Met

1 5 10 15

Val Ala Leu Gly Glu Leu Lys Leu Ala Ala Leu

20 25

<210> 1583

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1583

Thr Pro Glu Arg Pro Arg Leu Arg

1 5

<210> 1584

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1584

Thr Pro Glu Arg Pro Arg Leu Arg Leu Val Asp Ala Asp Asp Pro Leu

1 5 10 15

Leu Arg Thr Ala

20

<210> 1585

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1585

Thr Pro Gly Asn Lys Pro Arg Ile Ala

1 5

<210> 1586

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1586

Thr Pro Leu Asp Ile Asp Thr Ala Ser Glu Ser Ser Gln Asp Pro

1 5 10 15

<210> 1587

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1587

Thr Pro Leu Asp Ile Asp Thr Ala Ser Glu Ser Ser Gln Asp Pro Gln

1 5 10 15

Asp Ser Arg Arg

20

<210> 1588

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1588

Thr Pro Leu Phe Lys Ile Ile Gly Asp Glu Val Gly Leu Arg Thr

1 5 10 15

<210> 1589

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 1589

Thr Pro Leu Gln Cys Thr Met

1 5

<210> 1590

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1590

Thr Pro Val Ile Gly Ser Gln Ala

1 5

<210> 1591

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1591

Thr Pro Val Ile Gly Ser Gln Ala Arg Lys

1 5 10

<210> 1592

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1592

Thr Gln Gly Tyr Leu Ile Ser Asn Phe Asp Glu Ser Ser Cys Thr

1 5 10 15

<210> 1593

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1593

Thr Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val Ala Ser

1 5 10 15

Ile Val Gly Thr Leu Val Arg Met

20

<210> 1594

<211> 13

<212> PRT

<213

> Artificial Sequence

<220><223> Synthetic

<400> 1594

Thr Arg Phe Gly Cys Ala Met Arg Trp Gly Leu Pro Pro

1 5 10

<210> 1595

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1595

Thr Arg Asn Thr Gly Gln Pro Ala Thr Asp His Tyr Tyr Ser Asn Val

1 5 10 15

<210> 1596

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1596

Thr Arg Thr Pro Leu Gln Cys Thr Met Thr Glu Ile Phe

1 5 10

<210> 1597

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1597

Thr Arg Trp His Arg Leu Leu Arg Met Pro Val Arg

1 5 10

<210> 1598

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1598

Thr Ser Gly Ser Gly Gly Ala Ile Arg Gly Ile Lys His Ile Ile

1 5 10 15

<210> 1599

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1599

Thr Ser Gly Ser Gly Gly Ala Ile Arg Gly Ile Lys His Ile Ile Ile

1 5 10 15

Val Pro Ile Pro

20

<210> 1600

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1600

Thr Ser Gln Gly Met Tyr Gly Gly Thr Tyr Leu Val Glu Lys Pro

1 5 10 15

<210> 1601

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1601

Thr Ser Arg Val Glu His Ala Val Val Tyr Tyr Val Tyr Ser Pro

1 5 10 15

<210> 1602

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1602

Thr Ser Ser Ser Arg Arg Tyr Thr Glu Val Tyr Leu

1 5 10

<210> 1603

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1603

Thr Ser Tyr Val Gly

1 5

<210> 1604

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1604

Thr Thr Glu Asp Lys Ile Ser Arg Ala Val Gly Pro Arg Gln Ala

1 5 10 15

<210> 1605

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1605

Thr Thr Glu Asp Lys Ile Ser Arg Ala Val Gly Pro Arg Gln Ala Gln

1 5 10 15

Val Ser Phe Leu

20

<210> 1606

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1606

Thr Thr Glu Arg Ile Glu Thr Arg Ser Ala Arg His Pro

1 5 10

<210> 1607

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1607

Thr Thr Asn Gln Ala Ile Glu Ala Ile Arg Gln Ala Gly Gln Glu

1 5 10 15

<210> 1608

<211> 15

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1608

Thr Thr Ser Asp Pro Trp His Pro Pro Gly Pro Leu Gly Leu Lys

1 5 10 15

<210> 1609

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1609

Thr Val Cys Ser Gln Asn Ala Leu Tyr Pro Met Ser Pro Leu Leu

1 5 10 15

<210> 1610

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1610

Thr Val Asn Gly Glu Asp Val Gly Ala Val Pro Pro Gly Lys Phe

1 5 10 15

<210> 1611

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1611

Thr Tyr Pro Val Glu Lys Pro Asn Leu Ser Ser Lys Arg Ser Glu

1 5 10 15

<210> 1612

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1612

Val Ala Gly Val Ser Cys Asn Val Thr Thr Glu His Pro Phe Cys

1 5 10 15

<210> 1613

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1613

Val Ala Pro Gly Ile Val Ile Pro Pro Lys Ala Leu Phe Thr Gln Gln

1 5 10 15

Gly Gly Ala Tyr Gly Arg Cys

20

<210> 1614

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1614

Val Cys His Thr Lys His Met Asp Phe Trp Cys Val Glu His Asp Arg

1 5 10 15

Pro Pro Pro Ala Thr Pro Thr Pro Leu

20 25

<210> 1615

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1615

Val Cys Gln Arg His Ser Pro Asp Cys Ser Arg Leu Val Gly Ala Thr

1 5 10 15

Pro Glu Arg

<210> 1616

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1616

Val Asp Ala Asp Asp Pro Leu Leu

1 5

<210> 1617

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1617

Val Asp Ala Asp Asp Pro Leu Leu Arg Thr Ala Pro Gly Pro Gly Glu

1 5 10 15

Val Trp Val Thr

20

<210> 1618

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1618

Val Asp Ile Gly Phe Ala Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn

1 5 10 15

Leu Phe Gln Val

20

<210> 1619

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1619

Val Glu Ile Ala Glu Tyr Arg Arg Leu Leu Arg Thr Val Leu Glu

1 5 10 15

<210> 1620

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1620

Val Glu Leu Glu Asn Ser Met Gly Gly Leu Asn Phe Gly Arg Ser

1 5 10 15

<210> 1621

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1621

Val Glu Leu Glu Asn Ser Met Gly Gly Leu Asn Phe Gly Arg Ser Tyr

1 5 10 15

Phe Asp Pro Ala

20

<210> 1622

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1622

Val Glu Leu Lys Ile Lys Ile Ala Ser Gly Phe Gly Pro Leu Ile

1 5 10 15

<210> 1623

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1623

Val Glu Met Asp Glu Trp Ile His Ala His Thr Thr Ser Asp

1 5 10

<210> 1624

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1624

Val Glu Met Pro Glu Trp Ile His Ala His Thr Thr Ser Asp Pro

1 5 10 15

<210> 1625

<211> 9

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 1625

Val Glu Arg Arg Leu Val Lys Val Leu

1 5

<210> 1626

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1626

Val Glu Ser Pro Gly Gln Leu Ile

1 5

<210> 1627

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1627

Val Glu Ser Pro Gly Gln Leu Ile Gln Arg Ile Thr Asp Asp Pro

1 5 10 15

<210> 1628

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1628

Val Glu Ser Pro Gly Gln Leu Ile Gln Arg Ile Thr Asp Asp Pro Asp

1 5 10 15

Val Ser Ile Arg

20

<210> 1629

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1629

Val Phe Ala Leu Ala Ser Tyr Val Gln His Pro His Lys Thr Val

1 5 10 15

<210> 1630

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1630

Val Gly Ala Thr Pro Glu Arg Pro

1 5

<210> 1631

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1631

Val Gly Ala Thr Pro Glu Arg Pro Arg Leu

1 5 10

<210> 1632

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1632

Val Gly Ala Thr Pro Glu Arg Pro Arg Leu Arg Leu Val Asp Ala

1 5 10 15

<210> 1633

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1633

Val Gly Ile Gly Ser Ala Ser Tyr Lys Val Met Thr Arg Ser Ser

1 5 10 15

<210> 1634

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1634

Val Gly Leu Arg Thr Pro Gln Arg Phe Thr Asp Leu Val Lys Phe

1 5 10 15

<210> 1635

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1635

Val His His Asn Thr Glu Glu Ile Val Ala Gln Ser Ile Ala Leu Ser

1 5 10 15

Ser Leu Met Val

20

<210> 1636

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1636

Val His Arg Leu Glu Gly Val Ser Tyr Asn Ile Gly Ser Gln Glu

1 5 10 15

<210> 1637

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1637

Val Ile Gly Ser Gln Ala Arg Lys

1 5

<210> 1638

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1638

Val Ile Thr Gly Ser Ile

1 5

<210> 1639

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1639

Val Ile Thr Gly Thr Ile

1 5

<210> 1640

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1640

Val Lys Gln Ser Arg Gly Glu Ala

1 5

<210> 1641

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1641

Val Leu Phe Val Met Phe Leu Ser Leu Ile

1 5 10

<210> 1642

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1642

Val Leu Phe Val Met Phe Leu Ser Leu Ile Gly Leu Leu Ala Ile

1 5 10 15

<210> 1643

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1643

Val Leu Thr Pro Leu Phe Lys Ile Ile Gly Asp Glu Val Gly Leu

1 5 10 15

<210> 1644

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1644

Val Met Thr Arg Ser Ser His Gln Ser Leu Val Ile Lys Leu Met

1 5 10 15

<210> 1645

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1645

Val Pro Ala Ala Pro Cys Ala Arg Ile Trp Asn Gly Thr Gln Arg

1 5 10 15

<210> 1646

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1646

Val Pro Pro Asp Pro Gly Asp Leu Val Glu Tyr Ile Met Asn Tyr

1 5 10 15

<210> 1647

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1647

Val Gln Ser Val Ala Ser Ser Arg Arg His Lys Arg Phe Ala Gly

1 5 10 15

<210> 1648

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1648

Val Gln Thr Gly Gly Thr Ser Arg Thr Val Thr Met Arg Tyr Leu Ala

1 5 10 15

Ser

<210> 1649

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1649

Val Gln Val Arg Ile

1 5

<210> 1650

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1650

Val Arg Ala Tyr Asn Gln Pro Ala Gly Asp Val

1 5 10

<210> 1651

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1651

Val Arg Phe Glu Ser Lys Ile Val Asp Gly Gly Cys Phe Ala Pro

1 5 10 15

<210> 1652

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1652

Val Arg Leu Ala Gly Val Arg Phe Glu Ser Lys Ile Val Asp Gly

1 5 10 15

<210> 1653

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1653

Val Ser Gly Ser Phe Gly Asn Arg Phe Ile Leu Ser Gln Gly Asn

1 5 10 15

<210> 1654

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1654

Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp Ala Thr

1 5 10

<210> 1655

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1655

Val Ser Lys Gly Asn Cys Ser Gly Pro Thr Thr Ile Arg Gly Gln

1 5 10 15

<210> 1656

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1656

Val Thr Ala Ala Leu Leu Asn Thr Pro Pro Pro Tyr Gln Val Ser

1 5 10 15

<210> 1657

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1657

Val Thr Gly Cys Tyr Gln Cys Gly Thr Pro Ala Leu Val Glu Gly

1 5 10 15

<210> 1658

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1658

Val Thr Pro Val Ile Gly Ser Gln

1 5

<210> 1659

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1659

Val Thr Thr Glu His Pro Phe Cys Asn Thr Pro His Gly Gln Leu Glu

1 5 10 15

Val Gln Val Pro Pro Asp

20

<210> 1660

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1660

Val Val Leu Ala Gly Ala Ala Leu Gly Val Ala Thr Ala Ala Gln

1 5 10 15

<210> 1661

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1661

Val Trp Gln Leu Ser Val Ala Gly Val Ser Cys Asn Val Thr Thr

1 5 10 15

<210> 1662

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1662

Val Trp Val Thr Pro Val Ile Gly

1 5

<210> 1663

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1663

Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala Arg

1 5 10

<210> 1664

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1664

Val Tyr Ile Leu Ile Ala Val Cys Leu Gly Gly Leu Ile Gly Ile

1 5 10 15

<210> 1665

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1665

Val Tyr Leu His Arg Ile Asp Leu Gly Pro Pro Ile Ser Leu Glu

1 5 10 15

<210> 1666

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1666

Val Tyr Arg Tyr Asp Ser Arg Pro

1 5

<210> 1667

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1667

Val Tyr Arg Tyr Asp Ser Arg Pro Pro Glu Asp Val

1 5 10

<210> 1668

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1668

Val Tyr Trp Leu Thr Ile Pro Pro Met Lys Asn Leu Ala Leu Gly

1 5 10 15

<210> 1669

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1669

Trp Asp Leu Glu Ala Thr Gly Ala Cys Ile Cys Glu Ile Pro Thr

1 5 10 15

<210> 1670

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1670

Trp Asp Gln Lys Leu Trp Cys Arg His Phe Cys Val Leu Ala Asp

1 5 10 15

<210> 1671

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1671

Trp Gly Phe Pro Thr Asp Thr Val Met Ser Val Phe Ala Leu Ala

1 5 10 15

<210> 1672

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1672

Trp His Pro Pro Gly Pro Leu Gly Leu Lys Phe Lys Thr Val Arg

1 5 10 15

<210> 1673

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1673

Trp Ile His Ala His Thr Thr Ser Asp Pro Trp His Pro Pro Gly

1 5 10 15

<210> 1674

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1674

Trp Leu Thr Ile Pro Pro Met Lys Asn Leu Ala Leu Gly Val Ile

1 5 10 15

<210> 1675

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1675

Trp Gln Glu Ser Arg Lys Asn Lys Ala Gln Thr Arg Thr Pro Leu Gln

1 5 10 15

Cys Thr Met Thr

20

<210> 1676

<211> 13

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic

<400> 1676

Trp Val Cys Ile Phe Met Val Cys Arg Arg Ala Cys Arg

1 5 10

<210> 1677

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1677

Trp Val Thr Pro Val Ile Gly Ser

1 5

<210> 1678

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1678

Trp Tyr Thr Thr Val Pro Lys Tyr Val Ala Thr Gln Gly Tyr Leu

1 5 10 15

<210> 1679

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1679

Tyr Ala Leu Gly Gly Asp Ile Asn Lys Val Leu Glu Lys Leu Gly

1 5 10 15

<210> 1680

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1680

Tyr Ala Met Gly Val Gly Val Glu Leu Glu

1 5 10

<210> 1681

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1681

Tyr Ala Met Gly Val Gly Val Glu Leu Glu Asn

1 5 10

<210> 1682

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1682

Tyr Cys Ala Asp Val Ala Ala Glu Glu Leu Met Asn Ala Leu Val

1 5 10 15

<210> 1683

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1683

Tyr Asp Phe Arg Asp Leu Thr Trp Cys Ile Asn Pro Pro Glu Arg

1 5 10 15

<210> 1684

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1684

Tyr Phe Glu Pro Gly Pro Thr

1 5

<210> 1685

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1685

Tyr Phe Asn Pro Gly Gly Ser Tyr Tyr Lys Gln Tyr His Pro Thr

1 5 10 15

<210> 1686

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1686

Tyr Phe Arg Leu Gly Gln Glu Met Val Arg Arg Ser Ala Gly Lys

1 5 10 15

<210> 1687

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1687

Tyr Phe Arg Leu Gly Gln Glu Met Val Arg Arg Ser Ala Gly Lys Val

1 5 10 15

Ser Ser Thr Leu

20

<210> 1688

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1688

Tyr Phe Tyr Pro Phe Arg Leu Pro Ile Lys Gly Val Pro Ile Glu

1 5 10 15

<210> 1689

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1689

Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly Ala Leu Ala Thr

1 5 10 15

<210> 1690

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1690

Tyr Gly His Ala Thr Val Glu Met Pro Glu Trp Ile His Ala His

1 5 10 15

<210> 1691

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1691

Tyr Ile Met Asn Tyr Thr Gly Asn Gln Gln Ser Arg Trp Gly Leu

1 5 10 15

<210> 1692

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1692

Tyr Ile Asn Asn Glu Leu Ile Pro Ser Met Asn Gln Leu Ser Cys

1 5 10 15

<210> 1693

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1693

Tyr Leu Cys Thr Ala Pro Gly Cys Ala Thr Gln Ala Pro Val Pro

1 5 10 15

<210> 1694

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1694

Tyr Leu Phe Thr Val Pro Ile Lys Glu Ala Gly Glu Asp Cys His

1 5 10 15

<210> 1695

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1695

Tyr Leu His Asp Pro Glu Phe Asn Leu

1 5

<210> 1696

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1696

Tyr Leu Asn Met Ser Arg Leu Phe Val

1 5

<210> 1697

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1697

Tyr Pro Ala Leu Gly Leu His Glu Phe

1 5

<210> 1698

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1698

Tyr Pro Ala Leu Gly Leu His Glu Phe Ala

1 5 10

<210> 1699

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1699

Tyr Pro Ala Leu Gly Leu His Glu Phe Ala Gly Glu Leu Ser Thr

1 5 10 15

<210> 1700

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1700

Tyr Pro Ala Leu Gly Leu His Glu Phe Ala Gly Glu Leu Ser Thr Leu

1 5 10 15

Glu Ser Leu Met

20

<210> 1701

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1701

Tyr Pro Phe Arg Leu Pro Ile Lys Gly Val Pro Ile Glu Leu Gln

1 5 10 15

<210> 1702

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1702

Tyr Pro Leu Leu Trp Ser Tyr Ala Met

1 5

<210> 1703

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1703

Tyr Gln Gln Gln Gly Arg Leu

1 5

<210> 1704

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1704

Tyr Gln Ser Glu Tyr Leu Ala His Arg Arg

1 5 10

<210> 1705

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1705

Tyr Arg Glu Thr Gly Pro Ser Arg Ala Ser Asp Ala Arg Ala Ala

1 5 10 15

<210> 1706

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1706

Tyr Arg Glu Thr Gly Pro Ser Arg Ala Ser Asp Ala Arg Ala Ala His

1 5 10 15

Leu Pro Thr Gly

20

<210> 1707

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1707

Tyr Arg Asn Ala Ser Asp Val Leu Pro Gly His Trp Leu Gln Gly Gly

1 5 10 15

Trp Gly Cys Tyr Asn Leu Ser Asp Trp

20 25

<210> 1708

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1708

Tyr Arg Arg Leu Leu Arg Thr Val Leu Glu Pro Ile Arg Asp Ala

1 5 10 15

<210> 1709

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1709

Tyr Arg Tyr Asp Ser Arg Pro Pro

1 5

<210> 1710

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1710

Tyr Ser Gly Gly Asp Leu Leu Gly Ile Leu Glu Ser Arg Gly Ile

1 5 10 15

<210> 1711

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1711

Tyr Ser Lys Val Thr Ala Thr Asx Leu Leu Ala Ser Thr Asn Ser Arg

1 5 10 15

Leu Cys Ala Val Phe Val Arg Asp Gly

20 25

<210> 1712

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1712

Tyr Ser Pro Ser Arg Ser Phe Ser Tyr Phe Tyr Pro Phe Arg Leu

1 5 10 15

<210> 1713

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1713

Tyr Thr Gly Asn Gln Gln Ser Arg Trp Gly Leu Gly Ser Pro Asn Cys

1 5 10 15

His Gly Pro Asp Trp Ala Ser Pro Val

20 25

<210> 1714

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1714

Tyr Val Leu Leu Ala Val Leu Phe Val

1 5

<210> 1715

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1715

Tyr Val Tyr Ser Pro Gly Arg Ser Phe Ser Tyr Phe Tyr Pro Phe

1 5 10 15

<210> 1716

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1716

Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe Gln Thr Tyr Ala Leu Thr Gly

1 5 10 15

Ile Ser Leu Cys Asn Pro Ala Ala Ser Ile Cys

20 25

<210> 1717

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1717

Gly Asp Leu Leu Gly Ile Leu Glu Ser Arg Gly Ile Lys Ala Arg

1 5 10 15

<210> 1718

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1718

Thr Val Pro Lys Tyr Val Ala Thr Gln Gly Tyr Leu Ile Ser Asn

1 5 10 15

<210

> 1719

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1719

Lys Pro Trp Asp Ser Pro Gln Glu Ile

1 5

<210> 1720

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1720

Lys Pro Trp Glu Ser Pro Gln Glu Ile

1 5

<210> 1721

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1721

Ala Thr Tyr Gln Ser Glu Tyr Leu Ala His Arg Arg Ile Pro Pro

1 5 10 15

<210> 1722

<211> 35

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1722

Cys Met Ala Arg Gln Ala Glu Ser Ser Glu Ala Met Ala Ala Trp Ser
1 5 10 15

Glu Arg Ala Gly Glu Ala Met Val Leu Val Tyr Tyr Glu Ser Ile Ala
20 25 30

Tyr Ser Phe
35

<210> 1723

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1723

Cys Gln Val Gly Ser Ser Asn Ser Ala Phe Val Ser Thr Ser Ser Ser
1 5 10 15

Arg Arg Tyr Thr Glu Val Tyr Leu
20

<210> 1724

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1724

Asp Asp Pro Pro Ala Thr Val Tyr Arg Tyr Asp Ser Arg Pro Pro
1 5 10 15

<210> 1725

<211> 18

<212> PRT

<213
> Artificial Sequence

<220><223> Synthetic

<400> 1725

Gly Ala Leu Ala Thr Tyr Gln Ser Glu Tyr Leu Ala His Arg Arg Ile

1 5 10 15

Pro Pro

<210> 1726

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1726

Met Ala Ala Trp Ser Glu Arg Ala Gly Glu Ala Met Val Leu Val Tyr

1 5 10 15

Tyr Glu Ser Ile Ala Tyr Ser Phe

20

<210> 1727

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1727

Met Val Leu Val Tyr Tyr Glu Ser Ile Ala Tyr Ser Phe

1 5 10

<210> 1728

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1728

Pro Ala Thr Val Tyr Arg Tyr Asp Ser Arg Pro Pro Glu Asp Val

1 5 10 15

<210> 1729

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1729

Tyr Asp Ser Arg Pro Pro Glu Asp Val

1 5

<210> 1730

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1730

Glu Pro Gly Ile Thr Thr Asn Tyr Asp Thr

1 5 10

<210> 1731

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1731

Gly Asp Leu Arg Ala Tyr Lys Met Val Tyr Ala Thr Asn Pro Gln Thr

1 5 10 15

Gln Leu Ser Asn

20

<210> 1732

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1732

Lys Asn Gly Asp Val Glu Ala Ser Ala Ile Thr Thr Tyr Val Gly Phe

1 5 10 15

Ser Val Val Tyr Pro

20

<210> 1733

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1733

Lys Val Thr Asn Gly Ser Lys Ser Tyr Thr Leu Arg Tyr Leu Ala Ser
1 5 10 15

Tyr Val Lys

<210> 1734

<211> 35

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1734

Gln Ala Leu Gly Ala Leu Lys Leu Tyr Phe Glu Pro Gly Ile Thr Thr
1 5 10 15

Asn Tyr Asp Thr Gly Asp Leu Ile Ala Tyr Lys Gln Thr Tyr Asn Ala
20 25 30

Ser Gly Asn
35

<210> 1735

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1735

Tyr Ala Thr Asn Pro Gln Thr Gln Leu Ser
1 5 10

<210> 1736

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1736

Asp Asn Glu Asn Pro Leu Ser Gly Lys Ala Gly Gly Val Val Lys Val
1 5 10 15

Thr Tyr Pro Gly Leu Thr Lys Val
20

<210> 1737

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1737

Glu Asn Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser
1 5 10 15

Ile

<210> 1738

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1738

Lys Val Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu
1 5 10 15

Thr Glu Pro

<210> 1739

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1739

Met Glu Asp Leu Gln Lys Ala Leu Glu Ala Gln Ser Arg Ala Leu Arg
1 5 10 15
Ala Gly Leu Ala Ala
20

<210> 1740
<211> 23
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 1740
Gln Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr Asp
1 5 10 15
Asp Asp Trp Lys Gly Phe Tyr
20

<210> 1741
<211> 34
<212> PRT

<213> Artificial Sequence
<220><223> Synthetic
<400> 1741
Arg Thr Gly Ala Trp Gln Arg Lys Asp Trp Ser Arg Ala Pro Pro Pro
1 5 10 15
Pro Glu Glu Arg Gln Glu Ser Arg Ser Gln Thr Pro Ala Pro Lys Pro
20 25 30
Ser Arg

<210> 1742
<211> 29
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 1742
Ala Ala Gly Ala Ser Gln Ser Arg Arg Pro Arg Pro Pro Arg His Ala

1 5 10 15

Arg Ala Gln His Leu Pro Glu Met Thr Pro Ala Val Thr

20 25

<210> 1743

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1743

Cys Val Thr Ser Trp Leu Trp Ser Glu Gly Glu Gly Ala Val Phe Tyr

1 5 10 15

Arg Val Asp Leu His Phe Ile Asn Leu Gly Thr Pro

20 25

<210> 1744

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1744

Phe Arg Val Gly Gly Thr Arg Trp His Arg Leu Leu Arg Met Pro Val

1 5 10 15

Arg Gly Leu Asp Gly Asp Thr Ala Pro Leu Pro

20 25

<210> 1745

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1745

Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile Asp Gly Asp Val Thr

1 5 10 15

Phe Cys Arg Pro Lys Ser Pro Val Tyr Val Gly Asn

20 25

<210> 1746

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1746

Gly Thr Pro Pro Leu Asp Glu Asp Gly Arg Trp Asp Pro Ala Leu Met
1 5 10 15

Tyr Asn Pro Cys Gly Pro Glu Pro Pro Ala His Val
20 25

<210> 1747

<211> 54

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 1747

Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser Glu
1 5 10 15

Lys Ile His Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu Gly
20 25 30

Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys Leu Ser Leu
35 40 45

Phe Phe Glu Ile Lys Ser
50

<210> 1748

<211> 30

<

212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1748

Met Ala Ser Thr Thr Pro Ile Thr Met Glu Asp Leu Gln Lys Ala Leu
1 5 10 15

Glu Ala Gln Ser Arg Ala Leu Arg Ala Gly Leu Ala Ala Gly

20 25 30

<210> 1749

<211> 34

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1749

Pro Glu Leu Gly Pro Pro Thr Asn Pro Phe Gln Ala Ala Val Ala Arg
1 5 10 15

Gly Leu Arg Pro Pro Leu His Asp Pro Asp Thr Glu Ala Pro Thr Glu
20 25 30

Ala Cys

<210> 1750

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1750

Pro Leu Pro Pro His Thr Thr Glu Arg Ile Glu Thr Arg Ser Ala Arg
1 5 10 15

His Pro Trp Arg Ile Arg Phe Gly Ala Pro
20 25

<210> 1751

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1751

Ser Arg Ala Pro Pro Pro Pro Glu Glu Arg Gln Glu Ser Arg Ser Gln
1 5 10 15

Thr Pro Ala Pro Lys Pro Ser Arg Ala Pro Pro
20 25

<210> 1752

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1752

Ser Arg Ala Pro Pro Gln Gln Pro Gln Pro Pro Arg Met Gln Thr Gly
1 5 10 15

Arg Gly Gly Ser Ala Pro Arg Pro Glu Leu Gly Pro
20 25

<210> 1753

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1753

Thr Pro Ala Val Thr Pro Glu Gly Pro Ala Pro Pro Arg Thr Gly Ala
1 5 10 15

Trp Gln Arg Lys Asp Trp Ser Arg Ala Pro Pro
20 25

<210> 1754

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1754

Val Arg Ala Tyr Asn Gln Pro Ala Gly Asp Val Arg Gly Val Trp Gly
1 5 10 15

Lys Gly Glu Arg Thr Tyr Ala Glu Gln Asp Phe Arg Val
20 25

<210> 1755

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1755

Ala Phe Gly His Ser Asp Ala Ala Cys Trp Gly Phe Pro Thr Asp Thr

1 5 10 15

Val Met Ser Val

20

<210> 1756

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1756

Cys Ala Arg Ile Trp Asn Gly Thr Gln Arg Ala Cys Thr Phe Trp Ala

1 5 10 15

Val Asn Ala Tyr Ser

20

<210> 1757

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1757

Glu Glu Ala Phe Thr Tyr Leu Cys Thr Ala Pro Gly Cys Ala Thr Gln

1 5 10 15

Thr Pro Val Pro Val Arg

20

<210> 1758

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1758

Phe Ala Pro Trp Asp Leu Glu Ala Thr Gly Ala Cys Ile Cys Glu Ile

1 5 10 15

Pro Thr Asp Val

20

<210> 1759

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1759

Gly Glu Asp Val Gly Ala Phe Pro Pro Gly Lys Phe Val Thr Ala Ala

1 5 10 15

Leu

<210> 1760

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1760

Gly Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala Arg Lys Cys

1 5 10 15

Gly Leu His Ile

20

<210> 1761

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1761

Gly Gln Leu Glu Val Gln Val Pro Pro Asp Pro Gly Asp Leu Val Glu

1 5 10 15

Tyr Ile Met Asn

20

<210> 1762

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1762

Gly Ser Tyr Tyr Lys Gln Tyr His Pro Thr Ala Cys Glu Val Glu Pro

1 5 10 15

Ala Phe Gly His

20

<210> 1763

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1763

Ile His Ala His Thr Thr Ser Asp Pro Trp His Pro Pro Gly Pro Leu

1 5 10 15

Gly Leu Lys Phe

20

<210> 1764

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1764

Ile Met Asn Tyr Thr Gly Asn Gln Gln Ser Arg Trp Gly Leu Gly Ser

1 5 10 15

Pro Asn Cys His

20

<210> 1765

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1765

Leu His Ile Arg Ala Gly Pro Tyr Gly His Ala Thr Val Glu Met Pro
1 5 10 15

Glu Trp Ile His
20

<210> 1766

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1766

Leu Lys Phe Lys Thr Val Arg Pro Val Ala Leu Pro Arg Ala Leu Ala
1 5 10 15

Pro Pro Arg Asn
20

<210> 1767

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1767

Leu Asn Thr Pro Pro Pro Tyr Gln Val Ser Cys Gly Gly Glu Ser Asp
1 5 10 15

Arg Ala Ser Ala Gly His
20

<210> 1768

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1768

Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln Arg His Ser

1 5 10 15

Pro Asp Cys Ser

20

<210> 1769

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1769

Pro Asp Cys Ser Arg Leu Val Gly Ala Thr Pro Glu Arg Pro Arg Leu

1 5 10 15

Arg Leu Val Asp

20

<210> 1770

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1770

Pro Arg Asn Val Arg Val Thr Gly Cys Tyr Gln Cys Gly Thr Pro Ala

1 5 10 15

Leu Val Glu Gly

20

<210> 1771

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1771

Pro Thr Asp Val Ser Cys Glu Gly Leu Gly Ala Trp Val Pro Thr Ala

1 5 10 15

Pro Cys Ala Arg Ile

20

<210> 1772

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1772

Arg Leu Val Asp Ala Asp Asp Pro Leu Leu Arg Thr Ala Pro Gly Pro

1 5 10 15

Gly Glu Val Trp

20

<210> 1773

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1773

Ser Val Phe Ala Leu Ala Ser Tyr Val Gln His Pro His Lys Thr Val

1 5 10 15

Arg Val Lys Phe

20

<210> 1774

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1774

Val Glu Gly Leu Ala Pro Gly Gly Gly Asn Cys His Leu Thr Val Asn

1 5 10 15

Gly Glu Asp Val

20

<210> 1775

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1775

Val Lys Phe His Thr Glu Thr Arg Thr Val Trp Gln Leu Ser Val Ala
1 5 10 15

Gly Val Ser Cys
20

<210> 1776

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1776

Val Pro Val Arg Leu Ala Gly Val Gly Phe Glu Ser Lys Ile Val Asp
1 5 10 15

Gly Gly Cys Phe
20

<210> 1777

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1777

Val Ser Cys Asn Val Thr Thr Glu His Pro Phe Cys Asn Thr Pro His
1 5 10 15

Gly Gln Leu Glu
20

<210> 1778

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1778

Ala Ala Ala Ser Ser Pro Asp Ala His Val Pro Phe

1 5 10

<210> 1779

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1779

Ala Ala Ser Ser Pro Asp Ala

1 5

<210> 1780

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1780

Ala Lys Leu Gly Ala Ala Ala Ser Ser Pro Asp Ala

1 5 10

<210> 1781

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1781

Ala Met Lys Pro Tyr Glu Val Thr Pro Thr Arg Met

1 5 10

<210> 1782

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1782

Ala Met Thr His Leu Ser Pro Ala Leu Ala Asp Val Pro Tyr Val Leu

1 5 10 15

Val Lys Thr Asn Met Val Val Thr Ser
20 25

<210> 1783

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1783

Ala Ser Ser Pro Asp Ala His Val Pro Phe Cys Phe
1 5 10

<210> 1784

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1784

Ala Ser Ser Pro Asp Ala His Val Pro Phe Cys Phe Gly Lys Asp Leu
1 5 10 15

Lys Arg Pro Gly Ser Ser Pro Met Glu
20 25

<210> 1785

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1785

Cys Phe Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser
1 5 10

<210> 1786

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1786

Cys Phe Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser Pro Met Glu Val
1 5 10 15

Met Leu Arg Ala Val Phe Met Gln Gln
20 25

<210> 1787

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1787

Cys Gly Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser
1 5 10

<210> 1788

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1788

Cys Gly Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser Ser Pro Asp Ala
1 5 10 15

His Val Pro Phe Cys Phe Gly Lys Asp
20 25

<210> 1789

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1789

Asp Ala His Val Pro Phe Cys Phe Gly Lys Asp Leu
1 5 10

<210> 1790

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1790

Asp Leu Lys Arg Pro Gly Ser Ser Pro Met Glu Val

1 5 10

<210> 1791

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1791

Asp Val Pro Tyr Val Leu Val Lys Thr Asn Met Val

1 5 10

<210> 1792

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1792

Asp Val Pro Tyr Val Leu Val Lys Thr Asn Met Val Val Thr Ser Val

1 5 10 15

Ala Met Lys Pro Tyr Glu Val Thr Pro Thr

20 25

<210> 1793

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1793

Glu Val Met Leu Arg Ala Val Phe Met Gln Gln Arg

1 5 10

<210> 1794

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1794

Phe Glu Gly Lys Pro Ala Leu Glu Leu Ile Arg Met

1 5 10

<210> 1795

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1795

Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu Gly Lys

1 5 10

<210> 1796

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1796

Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu Gly Lys Pro Ala Leu Glu

1 5 10 15

Leu Ile Arg Met Val Glu Cys Ser Gly

20 25

<210> 1797

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1797

Phe Met Gln Gln Arg Pro Leu Arg Met Phe Leu Gly Pro Lys Gln Leu

1 5 10 15

Thr

<210> 1798

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1798

Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser Pro Met

1 5 10

<210> 1799

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1799

Gly Lys Asp Leu Lys Arg Pro Gly Ser Ser Pro Met Glu

1 5 10

<210> 1800

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1800

Gly Lys Pro Ala Leu Glu Leu Ile Arg Met Val Glu

1 5 10

<210> 1801

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1801

Gly Pro Lys Gln Leu Thr Phe Glu Gly Lys Pro Ala

1 5 10

<210> 1802

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1802

His Val Pro Phe Cys Phe Gly Lys Asp Leu Lys Arg

1 5 10

<210> 1803

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1803

Ile Ala Ala Lys Leu Gly Ala Ala Ala Ser Ser Pro

1 5 10

<210> 1804

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1804

Lys Pro Tyr Glu Val Thr Pro Thr Arg Met Leu Val

1 5 10

<210> 1805

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1805

Lys Gln Leu Thr Phe Glu Gly Lys Pro Ala Leu Glu

1 5 10

<210> 1806

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1806

Lys Arg Pro Gly Ser Ser Pro Met Glu Val Met Leu

1 5 10

<210> 1807

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1807

Leu Glu Leu Ile Arg Met Val Glu Cys Ser Gly Lys

1 5 10

<210> 1808

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1808

Leu Gly Ala Ala Ala Ser Ser Pro Asp Ala His Val

1 5 10

<210> 1809

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1809

Leu Ile Arg Met Val Glu Cys Ser Gly Lys Gln Asp

1 5 10

<210> 1810

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1810

Leu Val Cys Gly Ile Ala Ala Lys Leu Gly Ala Ala

1 5 10

<210> 1811

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1811

Met Lys Pro Tyr Glu Val Thr Pro Thr Arg Met

1 5 10

<210> 1812

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1812

Met Leu Arg Ala Val Phe Met Gln Gln Arg Pro Leu

1 5 10

<210> 1813

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1813

Met Gln Gln Arg Pro Leu Arg Met

1 5

<210> 1814

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1814

Met Gln Gln Arg Pro Leu Arg Met Phe Leu Gly Pro

1 5 10

<210> 1815

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1815

Met Val Val Thr Ser Val Ala Met Lys Pro Tyr Glu

1 5 10

<210> 1816

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1816

Met Val Val Thr Ser Val Ala Met Lys Pro Tyr Glu Val Thr Pro Thr

1 5 10 15

Arg Met Leu Val Cys Gly Ile Ala Ala

20 25

<210> 1817

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1817

Pro Ala Leu Glu Leu Ile Arg Met Val Glu Cys Ser

1 5 10

<210> 1818

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1818

Pro Ala Leu Glu Leu Ile Arg Met Val Glu Cys Ser Gly Lys

1 5 10

<210> 1819

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1819

Pro Phe Cys Phe Gly Lys Asp Leu Lys Arg Pro Gly

1 5 10

<210> 1820

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1820

Pro Gly Ser Ser Pro Met Glu Val Met Leu Arg Ala

1 5 10

<210> 1821

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1821

Pro Gly Ser Ser Pro Met Glu Val Met Leu Arg Ala Val Phe

1 5 10

<210> 1822

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1822

Pro Lys Gln Leu Thr Phe Glu Gly Lys

1 5

<210> 1823

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1823

Pro Leu Arg Met Phe Leu Gly Pro Lys Gln Leu Thr

1 5 10

<210> 1824

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1824

Pro Thr Arg Met Leu Val Cys Gly Ile Ala Ala Lys

1 5 10

<210> 1825

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1825

Pro Tyr Val Leu Val Lys Thr Asn Met Val Val Thr

1 5 10

<210> 1826

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1826

Gln Leu Thr Phe Glu Gly Lys Pro Ala Leu Glu Leu Ile Arg Met Val

1 5 10 15

Glu Cys Ser Gly Lys Gln Asp Cys Pro

20 25

<210> 1827

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1827

Gln Arg Pro Leu Arg Met Phe Leu Gly Pro Lys Gln

1 5 10

<210> 1828

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1828

Arg Ala Val Phe Met Gln Gln Arg Pro Leu Arg Met

1 5 10

<210> 1829

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1829

Arg Met Phe Leu Gly Pro Lys Gln Leu Thr Phe Glu

1 5 10

<210> 1830

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1830

Arg Met Leu Val Cys Gly Ile Ala Ala Lys Leu Gly

1 5 10

<210> 1831

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1831

Arg Met Val Glu Cys Ser Gly Lys Gln Asp Cys Pro

1 5 10

<210> 1832

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1832

Ser Pro Asp Ala His Val Pro Phe Cys Phe Gly Lys

1 5 10

<210> 1833

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1833

Ser Ser Pro Met Glu Val Met Leu Arg Ala Val Phe

1 5 10

<210> 1834

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1834

Ser Ser Pro Met Glu Val Met Leu Arg Ala Val Phe Met Gln Gln Arg

1 5 10 15

Pro Leu Arg Met Phe Leu Gly Pro Lys

20 25

<210> 1835

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1835

Ser Val Ala Met Lys Pro Tyr Glu Val Thr Pro Thr

1 5 10

<210> 1836

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1836

Val Glu Cys Ser Gly Lys Gln Asp Cys Pro

1 5 10

<210> 1837

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1837

Val Phe Met Gln Gln Arg Pro Leu Arg Met Phe Leu

1 5 10

<210> 1838

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1838

Val Phe Met Gln Gln Arg Pro Leu Arg Met Phe Leu Gly Pro Lys Gln

1 5 10 15

Leu Thr Phe Glu Gly Lys Pro Ala Leu

20 25

<210> 1839

<211> 12

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1839

Val Lys Thr Asn Met Val Val Thr Ser Val Ala Met

1 5 10

<210> 1840

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1840

Val Leu Val Lys Thr Asn Met Val Val Thr Ser Val

1 5 10

<210> 1841

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1841

Val Thr Pro Thr Arg Met Leu Val Cys Gly Ile Ala

1 5 10

<210> 1842

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1842

Val Thr Ser Val Ala Met Lys Pro Tyr Glu Val Thr

1 5 10

<210> 1843

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1843

Tyr Glu Val Thr Pro Thr Arg Met Leu Val Cys Gly

1 5 10

<210> 1844

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1844

Tyr Glu Val Thr Pro Thr Arg Met Leu Val Cys Gly Ile Ala Ala Lys

1 5 10 15

Leu Gly Ala Ala Ala Ser Ser Pro Asp

20 25

<210> 1845

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1845

Cys Ala Ser Pro Tyr Glu Gly Arg Tyr Arg Asp Met Tyr Asp Ala Leu

1 5 10 15

Arg

<210> 1846

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1846

Cys Ala Val Phe Val Arg Asp Gly Gln Ser Val

1 5 10

<210> 1847

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1847

Cys Ile Thr Thr Ile Tyr Lys Thr Gly

1 5

<210> 1848

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1848

Cys Pro Asn Gly Thr Arg Ala Leu Thr Val

1 5 10

<210> 1849

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1849

Asp Ala Leu Arg Arg Leu Leu Tyr Met Ile Tyr Met Ser Gly

1 5 10

<210> 1850

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1850

Gly Asn Arg Gly Arg Gly Gln Arg Arg Asp Trp Ser Arg Ala Pro Pro

1 5 10 15

Pro Pro Glu Glu Arg Gln Glu Thr Arg Ser

20 25

<210> 1851

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1851

Gly Gln Pro Ala Ala Asp His Tyr Tyr Ser Lys Val Thr

1 5 10

<210> 1852

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1852

Gly Ser Pro Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln

1 5 10 15

Arg His Ser

<210> 1853

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1853

His Lys Ser Leu Ser Thr Asn Leu Asp Val Thr Asn Ser Ile Glu His

1 5 10 15

Gln

<210> 1854

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1854

Leu Phe Thr Gln Gln Gly Gly Ala Tyr Gly Arg Cys

1 5 10

<210> 1855

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1855

Leu Ile Gly Leu Leu Ala Ile Ala Gly Ile Arg Leu His Arg Ala Ala

1 5 10 15

Ile Tyr Thr Ala Glu Ile

20

<210> 1856

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1856

Pro Asp Thr Ala Ala Asp Ser Glu Leu Arg Arg Trp Ile Lys Tyr

1 5 10 15

<210> 1857

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1857

Pro Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val Cys Gln Arg His

1 5 10 15

Ser

<210> 1858

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1858

Gln Thr Pro Ala Pro Lys Pro Ser Arg Ala Pro Pro Gln Gln Pro Gln

1 5 10 15

Pro Pro Arg Met Gln Thr Gly Arg

20

<210> 1859

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1859

Arg Ala Gly Leu Thr Ala Gly Ala Ser Gln Ser Arg Arg Pro Arg Pro

1 5 10 15

Pro Arg

<210> 1860

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1860

Arg Phe Gly Ala Pro Gln Ala Phe Leu Ala Gly Leu Leu Leu Ala Ala

1 5 10 15

Val Ala Val Gly Thr Ala Arg Ala

20

<210> 1861

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1861

Arg Gly Asn Ala Glu Leu Gln Thr Tyr Leu Arg Gln Ile Thr Pro Gly

1 5 10 15

<210> 1862

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1862

Arg Val His Val Ser Lys Glu Glu Gln Tyr Tyr Asp Tyr Glu Asp

1 5 10 15

<210> 1863

<211> 11

<212> PRT

<213

> Artificial Sequence

<220><223> Synthetic

<400> 1863

Ser Ile Tyr Gly Leu Tyr Asp Gly Thr Tyr Leu

1 5 10

<210> 1864

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1864

Ser Lys Val Thr Ala Thr Arg Leu Leu Ala Ser Thr Asn Ser

1 5 10

<210> 1865

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1865

Thr Gln Gln Gly Gly Ala Tyr Gly Arg Cys Pro Asn Gly Thr Arg Ala

1 5 10 15

<210> 1866

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1866

Val Ala Pro Gly Ile Val Ile Pro Pro Lys Ala Leu

1 5 10

<210> 1867

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1867

Asp Ser Arg Pro Pro Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp

1 5 10 15

Gly

<210> 1868

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1868

Glu His Arg Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly Arg

1 5 10 15

<210> 1869

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1869

Glu Thr Cys Phe Gln Gln Ala Cys Lys Gly Lys Ile Gln Ala Leu Cys

1 5 10 15

Glu Asn Pro Glu Trp Ala

20

<210> 1870

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1870

Glu Tyr Val Asp Thr Tyr Gly Asp Asn Ala Gly Arg Ile Leu Ala Gly

1 5 10 15

Ala Leu Ala Thr Tyr Gln

20

<210> 1871

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1871

His Arg Arg Ile Pro Pro Glu Asn Ile Arg Arg Val Thr Arg

1 5 10

<210> 1872

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1872

Met Ala Arg Gln Ala Glu Ser Ser Glu

1 5

<210> 1873

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1873

Met Gln Glu Ala Val Glu Ala Glu Arg Ala Gly Arg

1 5 10

<210> 1874

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1874

Ser Gln Gln Thr Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg

1 5 10 15

<210> 1875

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1875

Thr Arg Ala Asn Pro Asn Pro Tyr Thr Ser Arg Arg Ser Val Ala Ser

1 5 10 15

Ile Val Gly Thr Leu Val His Gly

20

<210> 1876

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1876

Thr Val Tyr Arg Tyr Asp Ser Arg Pro Pro Glu Asp

1 5 10

<210> 1877

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1877

Asn Asp Arg Asn Leu Leu Asp

1 5

<210> 1878

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1878

Asn Met Glu Asp Glu Ala Asp Gln Tyr Phe Ser His Asp Asp Pro Ile

1 5 10 15

Ser Ser Asp Gln Ser Arg Phe Gly Trp Phe Glu Asn Lys

20 25

<210> 1879

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1879

Ser Arg Ala Ser Asp Ala Arg Ala Ala His Leu Pro Thr Gly Thr Pro

1 5 10 15

Leu Asp Ile Asp

20

<210> 1880

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1880

Glu Asp Val Phe Gln Asn Gly Phe Thr Ala Trp

1 5 10

<210> 1881

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1881

Ala Glu Gly Ser Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala

1 5 10 15

Lys

<210> 1882

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1882

Gly Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys Thr

1 5 10

<210> 1883

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic

<400> 1883

Gly Pro Lys Leu Thr Gly Ala Leu Ile Gly Ile Leu Ser Leu Phe Val

1 5 10 15

Glu Ser Pro Gly Gln Leu Ile Gln Arg Ile Thr Asp Asp Pro Asp Val

20 25 30

<210> 1884

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1884

Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys

1 5 10

<210> 1885

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1885

Lys Thr Val Asp His

1 5

<210> 1886

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1886

Ser Glu Ser Pro Asn Lys

1 5

<210> 1887

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1887

Ala Glu Gly Ser Ser Ser Val Glu Tyr Ile Asn Asn Trp Glu Gln Ala

1 5 10 15

Lys Ala Leu Ser
20

<210> 1888

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1888

Ala Gln Ala Ile Pro Leu Val Gly Glu Leu Val Asp Ile Gly Phe Ala
1 5 10 15

Ala Tyr Asn Phe
20

<210> 1889

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1889

Ala Ser Arg Val Val Leu Ser Leu Pro Phe Ala Glu Gly Ser Ser Ser
1 5 10 15

Val Glu Tyr Ile
20

<210> 1890

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1890

Cys Ile Asn Leu Asp Trp Asp Val Ile Arg Asp Lys Thr Lys Thr Lys
1 5 10 15

Ile Glu Ser Leu
20

<210> 1891

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1891

Cys Arg Ala Ile Asp Gly Asp Val Thr Phe Cys Arg Pro Lys Ser Pro

1 5 10 15

Val Tyr Val Gly

20

<210> 1892

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1892

Cys Arg Pro Lys Ser Pro Val Tyr Val Gly Asn Gly Val His Ala Asn

1 5 10 15

Leu His Val Ala

20

<210> 1893

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1893

Asp Ala Ala Gly Tyr Ser Val Asp Asn Glu Asn Pro Leu Ser Gly Lys

1 5 10 15

Ala Gly Gly Val

20

<210> 1894

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1894

Asp Lys Thr Lys Thr Lys Ile Glu Ser Leu Lys Glu His Gly Pro Ile
1 5 10 15

Lys Asn Lys Met
20

<210> 1895

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1895

Glu Glu Phe Ile Lys Arg Phe Gly Asp Gly Ala Ser Arg Val Val Leu
1 5 10 15

Ser Leu Pro Phe
20

<210> 1896

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1896

Glu Lys Ala Lys Gln Tyr Leu Glu Glu Phe His Gln Thr Ala Leu Glu
1 5 10 15

His Pro Glu Leu
20

<210> 1897

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1897

Phe His Arg Ser Ser Ser Glu Lys Ile His Ser Asn Glu Ile Ser Ser

1 5 10 15

Asp Ser Ile Gly
20

<210> 1898

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1898

Gly Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15

Phe Ser Ser Tyr
20

<210> 1899

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1899

Gly Lys Arg Gly Gln Asp Ala Met Tyr Glu Tyr Met Ala Gln Ala Cys
1 5 10 15

Ala Gly Asn Arg
20

<210> 1900

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1900

Gly Ser Val Met Gly Ile Ala Asp Gly Ala Val His His Asn Thr Glu
1 5 10 15

Glu Ile Val Ala
20

<210> 1901

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1901

Gly Thr Gln Gly Asn Tyr Asp Asp Asp Trp Lys Gly Phe Tyr Ser Thr
1 5 10 15

Asp Asn Lys Tyr
20

<210> 1902

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1902

His Asp Gly Tyr Ala Val Ser Trp Asn Thr Val Glu Asp Ser Ile Ile
1 5 10 15

Arg Thr Gly Phe
20

<210> 1903

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1903

His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln Lys Gly Ile Gln
1 5 10 15

Lys Pro Lys Ser
20

<210> 1904

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1904

His Gln Thr Ala Leu Glu His Pro Glu Leu Ser Glu Leu Lys Thr Val
1 5 10 15

Thr Gly Thr Asn
20

<210> 1905

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1905

Ile Gln Lys Gly Ile Gln Lys Pro Lys Ser Gly Thr Gln Gly Asn Tyr
1 5 10 15

Asp Asp Asp Trp
20

<210> 1906

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1906

Lys Glu His Gly Pro Ile Lys Asn Lys Met Ser Glu Ser Pro Asn Lys
1 5 10 15

Thr Val Ser Glu
20

<210> 1907

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1907

Lys Gly Phe Tyr Ser Thr Asp Asn Lys Tyr Asp Ala Ala Gly Tyr Ser

1 5 10 15

Val Asp Asn Glu
20

<210> 1908

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1908

Leu Asp Val Asn Lys Ser Lys Thr His Ile Ser Val Asn Gly Arg Lys
1 5 10 15

Ile Arg Met Arg
20

<210> 1909

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1909

Met Ala Ser Thr Ile Pro Ile Thr Met Glu Asp Leu Gln Lys Ala Leu
1 5 10 15

Glu Ala

<210> 1910

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1910

Asn Gly Val His Ala Asn Leu His Val Ala Phe His Arg Ser Ser Ser
1 5 10 15

Glu Lys Ile His
20

<210> 1911

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1911

Asn Asn Trp Glu Gln Ala Lys Ala Leu Ser Val Glu Leu Glu Ile Asn

1 5 10 15

Phe Glu Thr Arg

20

<210> 1912

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1912

Asn Pro Leu Ser Gly Lys Ala Gly Gly Val Val Lys Val Thr Tyr Pro

1 5 10 15

Gly Leu Thr Lys

20

<210> 1913

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1913

Pro Val Phe Ala Gly Ala Asn Tyr Ala Ala Trp Ala Val Asn Val Ala

1 5 10 15

Gln Val Ile Asp

20

<210> 1914

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1914

Ser Glu Leu Lys Thr Val Thr Gly Thr Asn Pro Val Phe Ala Gly Ala
1 5 10 15

Asn Tyr Ala Ala
20

<210> 1915

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1915

Ser Glu Ser Pro Asn Lys Thr Val Ser Glu Glu Lys Ala Lys Gln Tyr
1 5 10 15

Leu Glu Glu Phe
20

<210> 1916

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1916

Ser Glu Thr Ala Asp Asn Leu Glu Lys Thr Thr Ala Ala Leu Ser Ile
1 5 10 15

Leu Pro Gly Ile
20

<210> 1917

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1917

Ser Phe Val Met Glu Asn Phe Ser Ser Tyr His Gly Thr Lys Pro Gly

1 5 10 15

Tyr Val Asp Ser
20

<210> 1918

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1918

Ser Asn Glu Ile Ser Ser Asp Ser Ile Gly Val Leu Gly Tyr Gln Lys
1 5 10 15

Thr Val Asp His
20

<210> 1919

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1919

Ser Pro Gly His Lys Thr Gln Pro Phe Leu His Asp Gly Tyr Ala Val
1 5 10 15

Ser Trp Asn Thr
20

<210> 1920

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1920

Ser Val Asn Gly Arg Lys Ile Arg Met Arg Cys Arg Ala Ile Asp Gly
1 5 10 15

Asp Val Thr Phe
20

<210> 1921

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1921

Thr Ala Ala Leu Ser Ile Leu Pro Gly Ile Gly Ser Val Met Gly Ile

1 5 10 15

Ala Asp Gly Ala

20

<210> 1922

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1922

Thr Ala Glu Asn Thr Pro Leu Pro Ile Ala Gly Val Leu Leu Pro Thr

1 5 10 15

Ile Pro Gly Lys

20

<210> 1923

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1923

Thr Glu Pro Leu Met Glu Gln Val Gly Thr Glu Glu Phe Ile Lys Arg

1 5 10 15

Phe Gly Asp Gly

20

<210> 1924

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1924

Thr Ile Lys Lys Glu Leu Gly Leu Ser Leu Thr Glu Pro Leu Met Glu
1 5 10 15

Gln Val Gly Thr
20

<210> 1925

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1925

Thr Lys Val Asn Ser Lys Leu Ser Leu Phe Phe Glu Ile Lys Ser
1 5 10 15

<210> 1926

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1926

Val Glu Asp Ser Ile Ile Arg Thr Gly Phe Gln Gly Glu Ser Gly His
1 5 10 15

Asp Ile Lys Ile
20

<210> 1927

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1927

Val Glu Leu Glu Ile Asn Phe Glu Thr Arg Gly Lys Arg Gly Gln Asp
1 5 10 15

Ala Met Tyr Glu

20

<210> 1928

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1928

Val Glu Ser Ile Ile Asn Leu Phe Gln Val Val His Asn Ser Tyr Asn

1 5 10 15

Arg Pro Ala Tyr

20

<210> 1929

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1929

Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly His Lys Thr

1 5 10 15

Gln Pro Phe Leu

20

<210> 1930

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1930

Val Lys Val Thr Tyr Pro Gly Leu Thr Lys Val Leu Ala Leu Lys Val

1 5 10 15

Asp Asn Ala Glu

20

<210> 1931

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1931

Val Leu Ala Leu Lys Val Asp Asn Ala Glu Thr Ile Lys Lys Glu Leu

1 5 10 15

Gly Leu Ser Leu

20

<210> 1932

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1932

Val Leu Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val Asn Ser Lys

1 5 10 15

Leu Ser Leu Phe

20

<210> 1933

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1933

Val Arg Arg Ser Val Gly Ser Ser Leu Ser Cys Ile Asn Leu Asp Trp

1 5 10 15

Asp Val Ile Arg

20

<210> 1934

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1934

Trp Ala Val Asn Val Ala Gln Val Ile Asp Ser Glu Thr Ala Asp Asn

1 5 10 15

Leu Glu Lys Thr

20

<210> 1935

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1935

Tyr Met Ala Gln Ala Cys Ala Gly Asn Arg Val Arg Arg Ser Val Gly

1 5 10 15

Ser Ser Leu Ser

20

<210> 1936

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1936

Ala Lys Ala Pro Pro Ala Pro Lys Pro Ala Pro Gln Pro Gly Pro

1 5 10 15

<210> 1937

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1937

Ala Pro Lys Pro Ala Pro Gln Pro Gly Pro

1 5 10

<210> 1938

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1938

Ala Pro Lys Pro Ala Pro Gln Pro Gly Pro Gln Pro Pro Gln Pro

1 5 10 15

<210> 1939

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1939

Ala Gln Thr Arg Thr Pro Leu Gln Cys Thr Met Thr Glu Ile Phe

1 5 10 15

<210> 1940

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1940

Ala Ser Arg Gly Thr Asn Met Glu Asp Glu Ala Asp Gln Tyr Phe Ser

1 5 10 15

His Asp Asp

<210> 1941

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1941

Ala Thr Ile Arg Arg

1 5

<210> 1942

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1942

Asp Asn Arg Ala Gly

1 5

<210> 1943

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1943

Glu Ala Pro Ala Pro Gln Pro Pro Ala Gly Arg Glu Leu Ser Ala

1 5 10 15

<210> 1944

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1944

Glu Met Val Arg Arg Ser Ala Gly Lys Val Ser Ser Thr Leu Ala Ser

1 5 10 15

Glu Leu Gly Ile

20

<210> 1945

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1945

Gly Ala Ser Glu Leu

1 5

<210> 1946

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1946

Gly Asp Ala Leu Ala Gly Gly Ala Val Pro

1 5 10

<210> 1947

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1947

Gly Asp Ala Pro Ala Gly Gly Ala Val Pro

1 5 10

<210> 1948

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1948

Gly Asp Thr Trp Asp Asp Asp

1 5

<210> 1949

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1949

Gly Glu Arg Gln His

1 5

<210> 1950

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1950

Gly Gly Ala Val Pro

1 5

<210> 1951

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1951

Gly Gly Phe Gly Pro

1 5

<210> 1952

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1952

Gly Gly Phe Gly Pro Gly Gly Phe Gly Pro

1 5 10

<210> 1953

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1953

Gly Gly Phe Gly Pro Val Leu Asp Gly Trp

1 5 10

<210> 1954

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1954

Gly Gly Lys Glu Asp Arg Arg Val Lys Gln Ser Arg Gly Glu Ala Arg

1 5 10 15

Glu Ser Tyr Arg

20

<210> 1955

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1955

Gly Ile Leu Leu Glu Asn

1 5

<210> 1956

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1956

Gly Ile Arg Arg Phe Leu

1 5

<210> 1957

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1957

His Asp Asp Pro Ile Ser Ser Asp Gln Ser Arg Phe Gly Trp Phe Glu

1 5 10 15

Asn Lys Glu Ile

20

<210> 1958

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1958

His Glu Phe Ala Gly Glu Leu Ser Thr Leu Glu Ser Leu Met Asn Leu

1 5 10 15

Tyr

<210> 1959

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1959

His Leu Gly Gly Leu Ala Gly Tyr

1 5

<210> 1960

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1960

His Thr Thr Glu Asp Lys Ile Ser Arg Ala Val Gly Pro Arg Gln Ala

1 5 10 15

Gln Val Ser Phe Leu

20

<210> 1961

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1961

Ile Cys Asp Ile Asp Thr Tyr Ile Val Glu Ala Gly Leu Ala Ser Phe

1 5 10 15

Ile Leu Thr Ile
20

<210> 1962

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1962

Ile Ile Ile Val Pro Ile Pro Gly Asp Ser Ser Ile Thr Thr Arg Ser
1 5 10 15

Arg Leu Leu Asp
20

<210> 1963

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1963

Ile Leu Asp Ile Lys Arg Thr Pro Gly Asn Lys Pro Arg Ile Ala Glu
1 5 10 15

Met Ile Cys Asp
20

<210> 1964

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1964

Lys Ala Leu Leu Tyr Arg
1 5

<210> 1965

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1965

Lys Pro Pro Ile Thr Ser Gly Ser Gly Gly Ala Ile Arg Gly Ile Lys

1 5 10 15

His Ile Ile Ile

20

<210> 1966

<211> 6

<212> PRT

<213

> Artificial Sequence

<220><223> Synthetic

<400> 1966

Leu Ala Gly Ser Gly Leu

1 5

<210> 1967

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1967

Leu Gly Ile Thr Ala Glu Asp Ala Arg Leu Val Ser Glu Ile Ala Met

1 5 10 15

His Thr Thr Glu

20

<210> 1968

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1968

Leu Gly Thr Ile Leu Ala Gln Ile Trp Val Leu Leu Ala Lys Ala Val

1 5 10 15

Thr Ala

<210> 1969

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1969

Leu Pro Thr Gly Thr Pro Leu Asp Ile Asp Thr Ala Ser Glu Ser Ser

1 5 10 15

Gln Asp

<210> 1970

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1970

Met Ala Thr Leu Leu Arg Ser Leu Ala Leu Phe Lys Arg Asn Lys Asp

1 5 10 15

Lys

<210> 1971

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1971

Pro Ala Pro Gln Pro Pro

1 5

<210> 1972

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1972

Pro Asp Thr Ala Ala Asp Ser Glu Leu Arg Arg Trp Ile Lys Tyr Thr

1 5 10 15

Gln Gln Arg Arg

20

<210> 1973

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1973

Pro Lys Leu Thr Gly Ala Leu Ile Gly Ile Leu Ser Leu Phe Val Glu

1 5 10 15

Ser Pro Gly Gln

20

<210> 1974

<211> 3

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1974

Pro Gln Pro

1

<210> 1975

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1975

Pro Gln Pro Gly Pro

1 5

<210> 1976

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1976

Pro Gln Pro Gly Pro Gln Pro Pro Gln Pro Pro Gln Pro Gln Pro

1 5 10 15

<210> 1977

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1977

Gln Asp Pro Gln Asp Ser Arg Arg Ser Ala Glu Pro Leu Leu Ser Cys

1 5 10 15

Lys Pro Trp Gln

20

<210> 1978

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1978

Gln Arg Arg Val Val Gly Glu Phe Arg Leu Glu Arg Lys Trp Leu Asp

1 5 10 15

Val Val Arg

<210> 1979

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1979

Arg Glu Leu Ser Ala

1 5

<210> 1980

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1980

Arg Phe Ala Pro Gln

1 5

<210> 1981

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1981

Ser Ile Gln Asn Lys Phe Ser Ala Gly Ser Tyr Pro Leu Leu Trp Ser

1 5 10 15

Tyr Ala Met Gly

20

<210> 1982

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1982

Ser Ile Thr Leu Gln Ala Gly Ala His

1 5

<210> 1983

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1983

Ser Leu Gln Pro Glu Asp

1 5

<210> 1984

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1984

Ser Asn Ala Leu Ser Lys Arg Leu

1 5

<210> 1985

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1985

Ser Pro Gly Gln Leu Ile Gln Arg Ile Thr Asp Asp Pro Asp Val Ser

1 5 10 15

Ile Arg Leu Leu

20

<210> 1986

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1986

Thr Glu Leu Pro Ser Ile Pro Gly

1 5

<210> 1987

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1987

Thr Phe Thr Leu Ala Asn Lys

1 5

<210> 1988

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1988

Thr Trp Asp Asp Asp

1 5

<210> 1989

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1989

Val Ser Phe Leu Gln Gly Asp Gln Ser Glu Asn Glu Leu Pro Arg Leu

1 5 10 15

Gly Gly Lys Glu

20

<210> 1990

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1990

Trp Gln Glu Ser Arg Lys Asn Lys Ala Gln Thr Arg Thr Pro Leu Gln

1 5 10 15

Cys

<210> 1991

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1991

Tyr Ala Met Gly Val Gly Val Glu Leu Glu Asn Ser Met Gly Gly Leu

1 5 10 15

Asn Phe Gly Arg

20

<210> 1992

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1992

Tyr Gln Gln Met Gly Lys Pro Ala Pro Tyr Met Val Asn Leu Glu Asn

1 5 10 15

Ser Ile

<210> 1993

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1993

Met Thr Arg Ser Ser His Gln Ser Leu Val Ile Lys Leu Met Pro

1 5 10 15

<210> 1994

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1994

Pro Ile Arg Asp Ala Leu Asn Ala Met Thr Gln Asn Ile Arg Pro

1 5 10 15

<210> 1995

<211> 24

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1995

Ala Leu Leu Asn Thr Pro Pro Pro Tyr Gln Val Ser Cys Gly Gly Glu

1 5 10 15

Ser Asp Arg Ala Ser Ala Gly His

20

<210> 1996

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1996

Gly Leu Gly Ser Pro Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val

1 5 10 15

Cys Gln Arg His Ser

20

<210> 1997

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1997

Gly Leu Gly Ser Pro Asn Cys His Gly Pro Asp Trp Ala Ser Pro Val

1 5 10 15

Cys Gln Arg His Ser Pro Asp Cys Ser Arg Leu Val

20 25

<210> 1998

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1998

Asn Tyr Thr Gly Asn Gln Gln Ser Arg Trp Gly Leu Gly Ser Pro Asn

1 5 10 15

Cys His Gly Pro Asp Trp Ala Ser Pro Val

20 25

<210> 1999

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 1999

Thr Leu Pro Gln Pro Pro Cys Ala His Gly Gln His Tyr Gly His His

1 5 10 15

His His Gln Leu

20

<210> 2000

<211> 26

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2000

Thr Val Arg Val Lys Phe His Thr Glu Thr Arg Thr Val Trp Gln Leu

1 5 10 15

Ser Val Ala Gly Val Ser Cys Asn Val Thr

20 25

<210> 2001

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2001

Ala Ser Tyr Phe Asn Pro Gly Gly Ser Tyr Tyr Lys Gln Tyr His

1 5 10 15

<210> 2002

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2002

Phe Ala Leu Ala Ser Tyr Val Gln His Pro His Lys Thr Val Arg

1 5 10 15

<210> 2003

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2003

Gly Gly Glu Ser Asp Arg Ala Ser Ala Arg Val Ile Asp Pro Ala Ala

1 5 10 15

Gln Ser Phe Thr Gly

20

<210> 2004

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2004

Gly Pro Gly Glu Val Trp Val Thr Pro Val Ile Gly Ser Gln Ala Arg

1 5 10 15

Lys Cys

<210> 2005

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2005

Gly Ser Gln Ala Arg Lys Cys Gly Leu His Ile Arg Ala Gly

1 5 10

<210> 2006

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2006

Leu Val Gly Ala Thr Pro Glu Arg Pro Arg Leu Arg Leu Val Asp Ala

1 5 10 15

Asp Asp Pro Leu Leu Arg Thr Ala Pro

20 25

<210> 2007

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2007

Thr Ala Pro Gly Pro Gly Glu Val Trp Val Thr Pro Val Ile

1 5 10

<210> 2008

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2008

Phe Ile Val Leu Ser Ile Ala Tyr Pro Thr Leu Ser Glu Ile Lys

1 5 10 15

<210> 2009

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2009

Glu Thr Cys Phe Gln Gln Ala Cys Lys Gly Lys Ile Gln Ala Leu Cys

1 5 10 15

Glu Asn Pro Glu Trp Ala Pro Leu Lys Asp Asn Arg Ile Pro Ser Tyr

20 25 30

<210

> 2010

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2010

Leu Pro Arg Leu Gly Gly Lys Glu Asp Arg

1 5 10

<210> 2011

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2011

Met Ser Lys Thr Glu Trp Asn Ala Ser Gln

1 5 10

<210> 2012

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2012

Ser Arg Phe Gly Trp Phe Glu Asn Lys Glu

1 5 10

<210> 2013

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2013

Glu Ala Cys Val Thr Ser Trp Leu Trp Ser Glu Gly Glu Gly Ala Val

1 5 10 15

Phe Tyr Arg Val Asp Leu His Phe Ile Asn Leu Gly Thr

20 25

<210> 2014

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2014

Met Asp Phe Trp Cys Val Glu His Asp Arg Pro Pro Pro Ala Thr Pro

1 5 10 15

Thr Ser Leu Thr Thr

20

<210> 2015

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2015

Pro Phe Leu Gly His Asp Gly His His Gly Gly Thr Leu Arg Val Gly

1 5 10 15

Gln His His Arg Asn Ala Ser Asp Val

20 25

<210> 2016

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2016

Arg Val Lys Phe His Thr Glu Thr Arg Thr Val Trp Gln Leu Ser Val
1 5 10 15

Ala Gly Val Ser Cys
20

<210> 2017

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2017

Ala Glu Glu Gln Ala Arg His Val Lys Asn Gly Leu Glu
1 5 10

<210> 2018

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2018

Glu Ser Pro Gln Glu Ile Ser Lys His Gln Ala Leu Gly
1 5 10

<210> 2019

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2019

Gly Val Gly Val Glu Leu Glu Asn Ser Met Gly Gly Leu Asn Phe
1 5 10 15

<210> 2020

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2020

Ile Lys Gly Ala Asn Asp Leu Ala Lys Phe His Gln Met Leu Met Lys

1 5 10 15

Ile Ile Met Lys

20

<210> 2021

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2021

Met Ser Lys Thr Leu His Ala Gln Leu Gly Phe Lys Lys Thr

1 5 10

<210> 2022

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2022

Asn Ala Ser Gly Leu Ser Arg Pro Ser Pro Ser Ala His

1 5 10

<210> 2023

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2023

Val Arg Val Ile Asp Pro Ser Leu Gly Asp Arg Lys Asp Glu

1 5 10

<210> 2024

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2024

Asp Leu Ser Asp Gly Asp Leu Leu Val

1 5

<210> 2025

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2025

Glu Ala Glu Arg Ala Gly Arg Gly Thr Gly

1 5 10

<210> 2026

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2026

Tyr Arg Tyr Asp Ser Arg Pro Pro Glu Asp Val

1 5 10

<210> 2027

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2027

Tyr Val Asp Thr Tyr Gly Asp Asn Ala Gly

1 5 10

<210> 2028

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2028

Ser Phe Ser Tyr Phe Tyr Pro Phe Arg

1 5

<210> 2029

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2029

Asp Ile Phe Ile Val Ser Pro Arg

1 5

<210> 2030

<211> 37

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2030

Gln Asp Ser Arg Arg Ser Ala Asp Ala Leu Leu Arg Leu Gln Ala Met

1 5 10 15

Ala Gly Ile Ser Glu Glu Gln Gly Ser Asp Thr Asp Thr Pro Ile Val

20 25 30

Tyr Asn Asp Arg Asn

35

<210> 2031

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2031

Ser Ala Glu Ala Leu Leu Arg Leu Gln Ala

1 5 10

<210> 2032

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2032

Arg Ile Val Ile Asn Arg Glu His Leu

1 5

<210> 2033

<211> 5

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2033

Ile Pro Arg Phe Lys

1 5

<210> 2034

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2034

Ala Ala Ala Leu Ser Pro Met Glu Ile

1 5

<210> 2035

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2035

Ala Ala Ala Ser Val Val Gly Ala Pro Val

1 5 10

<210> 2036

<211>

9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2036

Ala Ala Leu Gly Arg Gln Asp Ser Ile

1 5

<210> 2037

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2037

Ala Ala Gln Arg Leu Val His Ala Ile Ala

1 5 10

<210> 2038

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2038

Ala Ala Val Glu Ala Ala Glu Leu

1 5

<210> 2039

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2039

Ala Gly Ala Asn Val Leu Asn Gly Leu

1 5

<210> 2040

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2040

Ala Gly Tyr Ala Asn Ala Ala Asp

1 5

<210> 2041

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2041

Ala Gly Tyr Glu Gln Phe Glu Phe Arg Val

1 5 10

<210> 2042

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2042

Ala Ile Thr Gly Asn Ala Asp Asn Leu

1 5

<210> 2043

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2043

Ala Lys Glu Lys Asn Ala Thr Leu Met

1 5

<210> 2044

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2044

Ala Lys Gly Val Phe Leu Ser Leu

1 5

<210> 2045

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2045

Ala Pro His Glu Tyr Gly Phe Gly Ile

1 5

<210> 2046

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2046

Ala Arg Gln Gly Asn Asp Leu Glu Ile

1 5

<210> 2047

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2047

Ala Ser Val Val Gly Ala Pro Val

1 5

<210> 2048

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2048

Ala Thr Leu Met Phe Arg Leu Val

1 5

<210> 2049

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2049

Ala Val Ala Ala Ala Gln Arg Leu

1 5

<210> 2050

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2050

Asp Ala Gly Ala Asn Val Leu Asn Gly Leu

1 5 10

<210> 2051

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2051

Asp Ala Leu Leu Ala Gln Leu Tyr Arg

1 5

<210> 2052

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2052

Asp Ala Asn Gly Val Leu Lys His Ser Ile

1 5 10

<210> 2053

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2053

Asp Gly Asp Met Asn Ile Gly Val Ile

1 5

<210> 2054

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2054

Asp His Val Lys Asn Ile Glu Asn Leu

1 5

<210> 2055

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2055

Asp Ile Asp Met Phe Ala Ile Met

1 5

<210> 2056

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2056

Asp Met Phe Ala Ile Met Pro His Leu

1 5

<210> 2057

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2057

Asp Asn Val Arg Asn Val Glu Asn Val

1 5

<210> 2058

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2058

Asp Asn Val Arg Asn Val Glu Asn Val Ile

1 5 10

<210> 2059

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2059

Asp Thr Val Asp Tyr Ser Ala Met

1 5

<210> 2060

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2060

Asp Thr Val Asp Tyr Ser Ala Met Ile

1 5

<210> 2061

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2061

Asp Tyr Leu Arg Gln Ala Gly Leu

1 5

<210> 2062

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2062

Asp Tyr Tyr Asp Asn Val Arg Asn Val

1 5

<210> 2063

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2063

Glu Phe Thr Thr Phe Val Glu Ile

1 5

<210> 2064

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2064

Glu Phe Thr Thr Phe Val Glu Ile Val

1 5

<210> 2065

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2065

Glu Gly Tyr Val Phe Tyr Glu Asn

1 5

<210> 2066

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2066

Glu Asn Val Gln Tyr Arg His Val

1 5

<210> 2067

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2067

Glu Gln Leu Ala Asn Ser Asp Gly Leu

1 5

<210> 2068

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2068

Phe Gly Val Gly Tyr Gly His Asp Thr Ile

1 5 10

<210> 2069

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2069

Phe Ser Pro Asp Val Leu Glu Thr Val Pro

1 5 10

<210> 2070

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2070

Phe Ser Val Asp His Val Lys Asn Ile

1 5

<210> 2071

<211

> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2071

Gly Asp Asp Thr Tyr Leu Phe Gly Val

1 5

<210> 2072

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2072

Gly Asp Asp Val Phe Leu Gln Asp Leu

1 5

<210> 2073

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2073

Gly Glu Asp Gly Asn Asp Ile Phe Leu

1 5

<210> 2074

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2074

Gly Glu Arg Phe Asn Val Arg Lys Gln Leu

1 5 10

<210> 2075

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2075

Gly Gly Ala Gly Asn Asp Thr Leu Val

1 5

<210> 2076

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2076

Gly Gly Asp Asp Phe Glu Ala Val

1 5

<210> 2077

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2077

Gly Lys Ser Glu Phe Thr Thr Phe Val

1 5

<210> 2078

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2078

Gly Lys Ser Leu Phe Asp Asp Gly Leu

1 5

<210> 2079

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2079

Gly Asn Ala Asp Asn Leu Lys Ser Val

1 5

<210> 2080

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2080

Gly Gln Leu Val Glu Val Asp Thr Leu

1 5

<210> 2081

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2081

Gly Arg Ser Lys Phe Ser Pro Asp Val

1 5

<210> 2082

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2082

Gly Ser Ser Ala Tyr Asp Thr Val

1 5

<210> 2083

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2083

Gly Thr Val Glu Lys Trp Pro Ala Leu

1 5

<210> 2084

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2084

Gly Val Asp Tyr Tyr Asp Asn Val

1 5

<210> 2085

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2085

Gly Tyr Glu Gln Phe Glu Phe Arg Val

1 5

<210> 2086

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2086

His Ala Val Gly Ala Gln Asp Val Val

1 5

<210> 2087

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2087

Ile Ala Ala Gly Arg Ile Gly Leu Gly Ile

1 5 10

<210> 2088

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2088

Ile Gly Asp Ala Gln Ala Asn Thr Leu

1 5

<210> 2089

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2089

Ile Gly Leu Gly Ile Leu Ala Asp Leu

1 5

<210> 2090

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2090

Ile Gly Asn Ala Ala Gly Ile Pro Leu

1 5

<210> 2091

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2091

Ile Gly Thr Ser Met Lys Asp Val Leu

1 5

<210> 2092

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2092

Ile Gly Val Ile Thr Asp Phe Glu Leu

1 5

<210> 2093

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2093

Ile Pro Leu Thr Ala Asp Ile Asp Met

1 5

<210> 2094

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2094

Ile Ser Lys Ser Ala Leu Glu Leu

1 5

<210> 2095

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2095

Ile Thr Gly Asn Ala Asp Asn Leu

1 5

<210> 2096

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2096

Lys Ile Phe Val Val Ser Ala Thr

1 5

<210> 2097

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2097

Lys Gln Leu Asn Asn Ala Asn Val Tyr Arg

1 5 10

<210> 2098

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2098

Lys Val Ile Gly Asn Ala Ala Gly Ile

1 5

<210> 2099

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2099

Leu Ala Lys Val Val Ser Gln Leu

1 5

<210> 2100

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2100

Leu Ala Asn Asp Tyr Ala Arg Lys Ile

1 5

<210> 2101

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2101

Leu Asp Tyr Leu Arg Gln Ala Gly Leu

1 5

<210> 2102

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2102

Leu Gly Lys Gly Phe Ala Ser Leu

1 5

<210> 2103

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2103

Leu Gly Lys Gly Phe Ala Ser Leu Met

1 5

<210> 2104

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2104

Leu Gly Val Asp Tyr Tyr Asp Asn

1 5

<210> 2105

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2105

Leu Gly Val Asp Tyr Tyr Asp Asn Val

1 5

<210> 2106

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2106

Leu Lys His Ser Ile Lys Leu Asp Val Ile

1 5 10

<210> 2107

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2107

Leu Gln Ala Gly Tyr Ile Pro Val

1 5

<210> 2108

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2108

Leu Gln Leu Thr Gly Gly Thr Val Glu

1 5

<210> 2109

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2109

Leu Ser Ala Ala Val Phe Gly Leu

1 5

<210> 2110

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2110

Leu Ser Leu Gly Lys Gly Phe Ala Ser Leu

1 5 10

<210> 2111

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2111

Leu Ser Pro Met Glu Ile Tyr Gly Leu

1 5

<210> 2112

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2112

Asn Ala His Asp Asn Phe Leu Ala Gly Gly

1 5 10

<210> 2113

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2113

Asn Ala Asn Val Tyr Arg Glu Gly Val

1 5

<210> 2114

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2114

Asn Asp Thr Leu Tyr Gly Gly Leu

1 5

<210> 2115

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 2115

Asn Gly Leu Ala Gly Asn Asp Val Leu

1 5

<210> 2116

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2116

Asn Asn Ala Asn Val Tyr Arg Glu Gly Val

1 5 10

<210> 2117

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2117

Asn Thr Val Ser Tyr Ala Ala Leu

1 5

<210> 2118

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2118

Asn Val Leu Arg Asn Ile Glu Asn Ala Val

1 5 10

<210> 2119

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2119

Pro Ala Leu Thr Phe Ile Thr Pro Leu

1 5

<210> 2120

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2120

Pro Glu Thr Ser Asn Val Leu Arg Asn Ile

1 5 10

<210> 2121

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2121

Pro Met Glu Ile Tyr Gly Leu Val

1 5

<210> 2122

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2122

Pro Gln Ala Tyr Phe Glu Lys Asn Leu

1 5

<210> 2123

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2123

Pro Val Asn Pro Asn Leu Ser Lys Leu

1 5

<210> 2124

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2124

Gln Ala Gly Trp Asn Ala Ser Ser Val

1 5

<210> 2125

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2125

Gln Ala Gly Trp Asn Ala Ser Ser Val Ile

1 5 10

<210> 2126

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2126

Gln Asp Ala Ala Asn Ala Gly Asn Leu

1 5

<210> 2127

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2127

Gln Asp Ala Ala Asn Ala Gly Asn Leu Leu

1 5 10

<210> 2128

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2128

Gln Asp Ser Gly Tyr Asp Ser Leu

1 5

<210> 2129

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2129

Gln Gln Ser His Tyr Ala Asp Gln Leu

1 5

<210> 2130

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2130

Arg Ala Leu Gln Gly Ala Gln Ala Val

1 5

<210> 2131

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2131

Arg Gly Gly Leu Gly Leu Asp Thr Leu

1 5

<210> 2132

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2132

Arg Lys Gln Leu Asn Asn Ala Asn Val

1 5

<210> 2133

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2133

Arg Gln Asp Ser Gly Tyr Asp Ser Leu

1 5

<210> 2134

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2134

Arg Gln Phe Arg Tyr Asp Gly Asp Met

1 5

<210> 2135

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2135

Arg Ser Lys Phe Ser Pro Asp Val Leu

1 5

<210> 2136

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2136

Ser Ala Gly Ala Ala Ala Gly Ala Leu

1 5

<210> 2137

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2137

Ser Ala His Trp Gly Gln Arg Ala Leu

1 5

<210> 2138

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2138

Ser Ala Met Ile His Pro Gly Arg Ile

1 5

<210> 2139

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2139

Ser Ala Met Ile His Pro Gly Arg Ile Val

1 5 10

<210> 2140

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2140

Ser Ala Tyr Asp Thr Val Ser Gly Ile

1 5

<210> 2141

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2141

Ser Ala Tyr Gly Tyr Glu Gly Asp

1 5

<210> 2142

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2142

Ser Gly Gly Ala Gly Asp Asp Val Leu

1 5

<210> 2143

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2143

Ser Gly Leu Gln Val Ala Gly Ala

1 5

<210> 2144

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2144

Ser Gly Tyr Asp Ser Leu Asp Gly Val

1 5

<210> 2145

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2145

Ser Leu Leu Thr Gly Ala Leu Asn Gly Ile

1 5 10

<210> 2146

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2146

Ser Pro Met Glu Ile Tyr Gly Leu

1 5

<210> 2147

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2147

Ser Gln Met Leu Thr Arg Gly Gln Leu

1 5

<210> 2148

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2148

Ser Ser Ala Tyr Asp Thr Val Ser Gly Ile

1 5 10

<210> 2149

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2149

Ser Ser Leu Ala His Gly His Thr Ala

1 5

<210> 2150

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2150

Ser Ser Val Thr Ser Gly Asp Ser Val

1 5

<210> 2151

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2151

Ser Val Ile Gly Val Gln Thr Thr Glu Ile

1 5 10

<210> 2152

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2152

Thr Asn Thr Val Ser Tyr Ala Ala Leu

1 5

<210> 2153

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2153

Thr Ser Leu Ile Ala Glu Gly Val

1 5

<210> 2154

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2154

Thr Ser Leu Leu Thr Gly Ala Leu

1 5

<210> 2155

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2155

Thr Val Pro Ala Ser Pro Gly Leu

1 5

<210> 2156

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2156

Val Ala Lys Glu Lys Asn Ala Thr Leu

1 5

<210> 2157

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2157

Val Ala Pro His Glu Tyr Gly Phe Gly Ile

1 5 10

<210> 2158

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2158

Val Ala Val Val Thr Ser Leu Leu

1 5

<210> 2159

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2159

Val Phe Tyr Glu Asn Arg Ala Tyr Gly

1 5

<210> 2160

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2160

Val Phe Tyr Glu Asn Arg Ala Tyr Gly Val

1 5 10

<210> 2161

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2161

Val Ile Thr Asp Phe Glu Leu Glu Val

1 5

<210> 2162

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2162

Val Asn Pro His Ser Thr Ser Leu

1 5

<210> 2163

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2163

Val Asn Pro His Ser Thr Ser Leu Ile

1 5

<210> 2164

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2164

Val Asn Pro Asn Leu Ser Lys Leu

1 5

<210> 2165

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2165

Val Asn Pro Asn Leu Ser Lys Leu Phe

1 5

<210> 2166

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2166

Val Gln Gln Pro Ile Ile Glu Lys Leu

1 5

<210> 2167

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2167

Val Gln Tyr Arg His Val Glu Leu

1 5

<210> 2168

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2168

Val Ser Ile Ala Ala Ala Ala Ser Val

1 5

<210> 2169

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2169

Val Ser Ile Ala Ala Ala Ala Ser Val Val

1 5 10

<210> 2170

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2170

Val Thr Ser Leu Leu Thr Gly Ala Leu
1 5

<210> 2171

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2171

Val Val Leu Ala Asn Ala Ser Arg Ile
1 5

<210> 2172

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2172

Trp Pro Ala Leu Asn Leu Phe Ser Val
1 5

<210> 2173

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 2173

Trp Val Arg Lys Ala Ser Ala Leu
1 5

<210> 2174

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2174

Tyr Ala Val Gln Tyr Arg Arg Lys Gly Gly

1 5 10

<210> 2175

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2175

Tyr Gly Gly Leu Gly Asp Asp Thr Leu

1 5

<210> 2176

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2176

Tyr Gly Leu Val Gln Gln Ser His Tyr Ala

1 5 10

<210> 2177

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2177

Tyr Gly Tyr Glu Gly Asp Ala Leu Leu

1 5

<210> 2178

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2178

Tyr Ile Pro Val Asn Pro Asn Leu

1 5

<210>

2179

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2179

Tyr Ser Ala Met Ile His Pro Gly Arg Ile

1 5 10

<210> 2180

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2180

Tyr Ser Gln Thr Gly Ala His Ala Gly Ile

1 5 10

<210> 2181

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2181

Ala Tyr Asn Phe Val Glu Ser Ile Ile Asn Leu Phe Gln Val Val His

1 5 10 15

Asn Ser Tyr Asn

20

<210> 2182

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2182

Ser Gly Thr Thr Ile Lys

1 5

<210> 2183

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2183

Arg Gly His Thr Leu Glu Ser Ala Glu Gly Arg Lys Ile Phe Gly

1 5 10 15

<210> 2184

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2184

Ala Gly Ala Met Thr Val Arg Asp Val Ala Ala Ala Ala Asp Leu Ala

1 5 10 15

Leu Gln Ala Gly Asp Ala

20

<210> 2185

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2185

Ala Gly Ala Met Thr Val Arg Asp Val Ala Ala Ala Ala Asp Leu Ala

1 5 10 15

Leu Gln Ala Gly Asp Ala Leu

20

<210> 2186

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2186

Ala Leu Ala Ala Val Leu Val Asn Pro His Ile Phe Thr Arg Ile Gly

1 5 10 15

Ala Ala Gln Thr Ser Leu Ala Asp Gly Ala Ala Gly Pro Ala

20 25 30

<210> 2187

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2187

Ala Leu Ser Ile Asp Ser Met Thr Ala Leu Gly Ala

1 5 10

<210> 2188

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2188

Asp Leu Ser Ala Ala Arg Gly Ala Asp Ile Ser Gly Glu Gly Arg

1 5 10 15

<210> 2189

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2189

Asp Gln Asn Arg Tyr Glu Tyr Ile Trp Gly Leu Tyr

1 5 10

<210> 2190

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2190

Asp Tyr Thr Val Ser Ala Asp Ala Ile Ala Leu Ala

1 5 10

<210> 2191

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2191

Gly Pro Ile Val Val Glu Ala Gly Glu Leu Val Ser His Ala Gly Gly

1 5 10 15

<210> 2192

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<

400> 2192

Gly Arg Pro Glu Gly Leu Lys Ile Gly Ala His Ser Ala Thr Ser Val

1 5 10 15

Ser Gly Ser Phe Asp Ala Leu

20

<210> 2193

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2193

Ile Thr Val Thr Ser Arg Gly Gly Phe Asp Asn Glu Gly Lys Met Glu

1 5 10 15

Ser Asn Lys

<210> 2194

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2194

Leu Asp Gln Asn Arg Tyr Glu Tyr Ile Trp Gly Leu Tyr Pro

1 5 10

<210> 2195

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2195

Leu Ser Ala Ala Arg Gly Ala Asp Ile Ser Gly

1 5 10

<210> 2196

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2196

Asn Lys Ile Arg Leu Met Gly Pro Leu Gln

1 5 10

<210> 2197

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2197

Asn Lys Leu Gly Arg Ile Arg Ala Gly Glu Asp Met

1 5 10

<210> 2198

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2198

Asn Lys Leu Gly Arg Ile Arg Ala Gly Glu Asp Met His Leu Asp Ala

1 5 10 15

Pro Arg Ile Glu

20

<210> 2199

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2199

Pro His Leu Arg Asn Thr Gly Gln Val Val Ala Gly

1 5 10

<210> 2200

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2200

Gln Val Asp Leu His Asp Leu Ser Ala Ala Arg Gly Ala Asp Ile Ser

1 5 10 15

Gly

<210> 2201

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2201

Arg Asp Val Ala Ala Ala Ala Asp Leu Ala Leu Gln

1 5 10

<210> 2202

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2202

Ser Ala Ala Arg Gly Ala Asp Ile Ser Gly Glu Gly

1 5 10

<210> 2203

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2203

Thr Lys Gly Glu Met Gln Ile Ala Gly Lys Gly Gly Gly Ser Pro

1 5 10 15

<210> 2204

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2204

Thr Val Ser Ala Asp Ala Ile Ala Leu Ala Ala Gln

1 5 10

<210> 2205

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2205

Val Val Ala Gly His Asp Ile His Ile

1 5

<210> 2206

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2206

Pro Ser Gly Pro Asn His Thr Lys Val Val Gln Leu Pro Lys Ile Ser

1 5 10 15

Lys Asn Ala Leu Lys Ala Asn Gly

20

<210> 2207

<211> 41

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2207

Leu Val Gly Ala Thr Pro Glu Arg Pro Arg Leu Arg Leu Val Asp Ala

1 5 10 15

Asp Asp Pro Leu Leu Arg Thr Ala Pro Gly Pro Gly Glu Val Trp Val

20 25 30

Thr Pro Val Ile Gly Ser Gln Ala Arg

35 40

<210> 2208

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2208

Gln Gln Ser Arg Trp Gly Leu Gly Ser Pro Asn Cys His Gly Pro Asp

1 5 10 15

Trp Ala Ser Pro Val Cys Gln Arg His Ser Pro

20 25

<210> 2209

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2209

Ala Gly Glu Ala Met Val Leu Val Tyr Tyr Glu Ser Ile Ala Tyr Ser
1 5 10 15

Phe

<210> 2210

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2210

Gly Gly Val Gly Leu Ala Ser Thr Leu Trp Tyr Ala Glu Ser Asn Ala
1 5 10 15

Leu Ser Lys Arg Leu Gly Glu Leu
20

<210> 2211

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2211

Gly Thr Leu Val Arg Ile Ala Pro Val Ile Gly Ala Cys Met Ala Arg
1 5 10 15

Gln Ala

<210> 2212

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2212

Ile Arg Arg Val Thr Arg Val Tyr His Asn Gly Ile Thr Gly Glu Thr

1 5 10 15

Thr Thr

<210> 2213

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2213

Ile Val Lys Thr Gly Glu Arg Gln His Gly Ile His Ile Gln Gly Ser

1 5 10 15

Asp Pro

<210> 2214

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2214

Ile Val Lys Thr Gly Glu Arg Gln His Gly Ile His Ile Gln Gly Ser

1 5 10 15

Asp Pro Gly Gly Val Arg Thr Ala

20

<210> 2215

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2215

Leu Arg Asp Thr Asn Val Thr Ala Val Pro Ala Ser Gly Ala Pro Ala

1 5 10 15

Ala Val Ser Val Leu Gly Ala Ser

20

<210> 2216

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2216

Pro Glu Ala Pro Ala Pro Gln Pro Pro Ala Gly Arg Glu Leu Ser Ala

1 5 10 15

Ala Ala Asn Ala Ala Val Asn Thr

20

<210> 2217

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2217

Ala Ala Ala Asp Phe Ala His Ala Glu

1 5

<210> 2218

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2218

Ala Ala Ala Glu Val Ala Gly Ala Leu

1 5

<210> 2219

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2219

Ala Ala Glu Ser Thr Phe Glu Ser Tyr

1 5

<210> 2220

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2220

Ala Ala Gly Phe Asp Pro Glu Val Gln

1 5

<210> 2221

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2221

Ala Ala Leu Gly Arg Gly His Ser Leu

1 5

<210> 2222

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2222

Ala Ala Met Gln Gly Ala Val Val His

1 5

<210> 2223

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2223

Ala Ala Pro Ala Ala His Ala Asp Trp

1 5

<210> 2224

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2224

Ala Ala Gln Ala Thr Val Val Gln Arg

1 5

<210> 2225

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2225

Ala Ala Arg Val Ala Gly Asp Asn Tyr

1 5

<210> 2226

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2226

Ala Ala Val Ala Leu Leu Asn Lys Leu

1 5

<210> 2227

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2227

Ala Asp Asp Pro Pro Ala Thr Val Tyr

1 5

<210> 2228

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2228

Ala Glu Ala Gly Arg Phe Lys Val Leu

1 5

<210> 2229

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2229

Ala Glu Ala Thr Gln Leu Val Thr Ala

1 5

<210> 2230

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2230

Ala Glu Gly Gly Ala Thr Leu Gly Ala

1 5

<210> 2231

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2231

Ala Glu His Gly Glu Val Ser Ile Gln

1 5

<210> 2232

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2232

Ala Glu Ile Ala Phe Tyr Pro Lys Glu

1 5

<210> 2233

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2233

Ala Glu Lys Val Thr Thr Pro Ala Val

1 5

<210> 2234

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2234

Ala Glu Leu Gln Thr Tyr Leu Arg Gln

1 5

<210> 2235

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2235

Ala Glu Gln Ser Leu Ile Glu Val Gly

1 5

<210> 2236

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2236

Ala Glu Ser Asn Ala Leu Ser Lys Arg

1 5

<210> 2237

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2237

Ala Glu Ser Ser Glu Ala Met Ala Ala

1 5

<210> 2238

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2238

Ala Glu Val Lys Val Gly Tyr Arg Ala

1 5

<210> 2239

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2239

Ala Glu Val Thr Asp Thr Ser Pro Ser

1 5

<210> 2240

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2240

Ala Gly Lys Ser Leu Lys Lys Lys Asn

1 5

<210> 2241

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2241

Ala Gly Leu Ala Gly Pro Ser Ala Val

1 5

<210> 2242

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2242

Ala Ile Arg Val Gly Arg Gly Ala Arg

1 5

<210> 2243

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2243

Ala Leu Ala Ala Ile Ala Ser Ala Ala

1 5

<210> 2244

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2244

Ala Leu Ala Asp Val Pro Tyr Val Leu

1 5

<210> 2245

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2245

Ala Leu Ala Asn Asp Gly Thr Ile Val

1 5

<210> 2246

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2246

Ala Leu Gly Arg Gly His Ser Leu Tyr

1 5

<210> 2247

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2247

Ala Leu Ile Leu Ala Ala Ser Pro Val

1 5

<210> 2248

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2248

Ala Leu Met Leu Ala Cys Thr Gly Leu

1 5

<210> 2249

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2249

Ala Met Gln Gly Ala Val Val His Leu

1 5

<210> 2250

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2250

Ala Met Thr His Leu Ser Pro Ala Leu

1 5

<210> 2251

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2251

Ala Met Tyr Gly Lys His Ile Thr Leu

1 5

<210> 2252

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2252

Ala Asn Glu Ala Asn Ala Leu Leu Trp

1 5

<210> 2253

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2253

Ala Pro Leu Ser Ile Thr Leu Gln Ala

1 5

<210> 2254

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2254

Ala Pro Asn Ala Leu Ala Trp Ala Leu

1 5

<210> 2255

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2255

Ala Pro Pro Ala Pro Lys Pro Ala Pro

1 5

<210> 2256

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2256

Ala Pro Pro Gly Ala Gly Phe Ile Tyr

1 5

<210> 2257

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2257

Ala Pro Gln Ala Ala Pro Leu Ser Ile

1 5

<210> 2258

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2258

Ala Pro Arg Ile Glu Asn Thr Ala Lys

1 5

<210> 2259

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2259

Ala Gln Gly Lys Ala Leu Leu Tyr Arg

1 5

<210> 2260

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2260

Ala Gln Ile Thr Ser Tyr Val Gly Phe

1 5

<210> 2261

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2261

Ala Gln Leu Glu Val Arg Gly Gln Arg

1 5

<210> 2262

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2262

Ala Gln Gln Leu Lys Gln Ala Asp Arg

1 5

<210> 2263

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2263

Ala Gln Val Thr Val Ala Gly Arg Tyr

1 5

<210> 2264

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2264

Ala Arg Arg Ser Arg Val Arg Ala Leu

1 5

<210> 2265

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2265

Ala Ser Pro Arg Arg Ala Arg Arg Ala

1 5

<210> 2266

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2266

Ala Ser Ser Pro Asp Ala His Val Pro

1 5

<210> 2267

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2267

Ala Ser Val Ser Asn Pro Gly Thr Phe

1 5

<210> 2268

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2268

Ala Thr Trp Asn Phe Gln Ser Thr Tyr

1 5

<210> 2269

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2269

Ala Thr Tyr Ile Ala Asp Ser Gly Phe

1 5

<210> 2270

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2270

Ala Val Ala Ala Pro Ala Val Gly Ala

1 5

<210> 2271

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2271

Ala Val Phe Met Gln Gln Arg Pro Leu

1 5

<210> 2272

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2272

Ala Val Leu Val Asn Pro His Ile Phe

1 5

<210> 2273

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2273

Cys Phe Gly Lys Asp Leu Lys Arg Pro

1 5

<210> 2274

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2274

Cys Pro Ser Ser Leu Gly Asn Gly Val

1 5

<210> 2275

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2275

Asp Ala Gly His Glu His Asp Thr Trp

1 5

<210> 2276

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2276

Asp Ala Lys His Asp Leu Thr Val Thr

1 5

<210> 2277

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2277

Asp Ala Ser Gly Gln His Arg Leu Trp

1 5

<210> 2278

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2278

Asp Ala Thr Phe Glu Thr Tyr Ala Leu

1 5

<210> 2279

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2279

Asp Ala Thr Phe Gln Thr Tyr Ala Leu

1 5

<210> 2280

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2280

Asp Ala Thr Leu Val Gly Ala Lys Phe

1 5

<210> 2281

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2281

Asp Asp Glu Val Asp Val Ser Gly Arg

1 5

<210> 2282

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2282

Asp Glu Asn Gly Lys Pro Gln Thr Tyr

1 5

<210> 2283

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2283

Asp Gly Pro Pro Ser Arg Pro Thr Thr

1 5

<210> 2284

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2284

Asp His Leu Thr Gly Arg Ser Cys Gln

1 5

<210> 2285

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2285

Asp Asn Glu Gly Lys Met Glu Ser Asn

1 5

<210> 2286

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2286

Asp Pro Pro Ala Thr Val Tyr Arg Tyr

1 5

<210> 2287

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2287

Glu Ala Thr Glu Gly Asp Ala Thr Leu

1 5

<210> 2288

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2288

Glu Ala Thr Gln Gln Ala Ala Gly Phe

1 5

<210> 2289

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2289

Glu Cys Ser Gly Lys Gln Asp Cys Pro

1 5

<210> 2290

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2290

Glu Gly Gly Lys Leu Arg Gly Lys Asp

1 5

<210> 2291

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2291

Glu Gly Lys Met Glu Ser Asn Lys Asp

1 5

<210> 2292

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2292

Glu His Arg Met Gln Glu Ala Val Glu

1 5

<210> 2293

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2293

Glu Lys Arg Leu Asp Ile Asp Asp Ala

1 5

<210> 2294

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2294

Glu Pro Ala Ser Ala Asn Thr Leu Leu

1 5

<210> 2295

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2295

Glu Pro Gln Ala Glu Leu Ala Val Phe

1 5

<210> 2296

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2296

Glu Pro Val Lys Leu Thr Leu Thr Gly

1 5

<210> 2297

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2297

Glu Ser Ala Glu Gly Arg Lys Ile Phe

1 5

<210> 2298

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2298

Glu Ser Tyr Ser Glu Ser His Asn Phe

1 5

<210> 2299

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2299

Glu Thr Phe Cys Ile Thr Thr Ile Tyr

1 5

<210> 2300

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2300

Glu Val Ala Gly Ala Leu Glu Leu Ser

1 5

<210> 2301

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2301

Glu Val Ala Lys Val Glu Val Val Pro

1 5

<210> 2302

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2302

Glu Val Asp Gly Ile Ile Gln Glu Phe

1 5

<210> 2303

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2303

Glu Val Arg Ala Asp Asn Asn Phe Tyr

1 5

<210> 2304

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2304

Phe Ala Ile Leu Ser Ser Thr Thr Glu

1 5

<210> 2305

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2305

Phe Ala Ile Ser Ala Tyr Ala Leu Lys

1 5

<210> 2306

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2306

Phe Ala Leu Tyr Asp Gly Thr Tyr Leu

1 5

<210> 2307

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2307

Phe Asp Thr Met Leu Gly Phe Ala Ile

1 5

<210> 2308

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2308

Phe Glu Gly Lys Pro Ala Leu Glu Leu

1 5

<210> 2309

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2309

Phe Glu Leu Gly Ala Asp His Ala Val

1 5

<210> 2310

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2310

Phe Glu Pro Gly Ile Thr Thr Asn Tyr

1 5

<210> 2311

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2311

Phe Glu Thr Tyr Ala Leu Thr Gly Ile

1 5

<210> 2312

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2312

Phe Ile Tyr Arg Glu Thr Phe Cys Ile

1 5

<210> 2313

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2313

Phe Pro Thr Arg Thr Thr Ala Pro Gly

1 5

<210> 2314

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2314

Phe Gln Thr Tyr Ala Leu Thr Gly Ile

1 5

<210> 2315

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2315

Phe Thr His Ala Asp Gly Trp Phe Leu

1 5

<210> 2316

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2316

Phe Val Arg Asp Gly Gln Ser Val Ile

1 5

<210> 2317

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2317

Phe Val Arg Ser Gly Gln Pro Val Ile

1 5

<210> 2318

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2318

Phe Val Trp Tyr Val Asp Thr Val Ile

1 5

<210> 2319

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2319

Gly Ala Ala Ser Ser Arg Gln Ala Leu

1 5

<210> 2320

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2320

Gly Ala Ala Ser Ser Tyr Phe Glu Tyr

1 5

<210> 2321

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2321

Gly Ala Phe Asp Leu Lys Thr Thr Phe

1 5

<210> 2322

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2322

Gly Ala Pro Ala Ala Val Ser Val Leu

1 5

<210> 2323

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2323

Gly Ala Thr Arg Ala Val Asp Ser Leu

1 5

<210> 2324

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2324

Gly Ala Val Pro Gly Gly Ala Val Pro

1 5

<210> 2325

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2325

Gly Glu Ala Met Val Leu Val Tyr Tyr

1 5

<210> 2326

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2326

Gly Glu Ile Ala Leu Gly Asp Ala Thr

1 5

<210> 2327

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2327

Gly Glu Leu Met Ala Ala Gln Val Ala

1 5

<210> 2328

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2328

Gly Gly Val Pro Gly Gly Ala Val Pro

1 5

<210> 2329

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2329

Gly His Glu His Asp Thr Trp Phe Asp

1 5

<210> 2330

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2330

Gly Ile Gly Ala Leu Lys Ala Gly Ala

1 5

<210> 2331

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2331

Gly Ile Val Ile Pro Pro Lys Ala Leu

1 5

<210> 2332

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2332

Gly Lys Asp Leu Lys Arg Pro Gly Ser

1 5

<210> 2333

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2333

Gly Lys Leu Pro Lys Pro Val Thr Val

1 5

<210> 2334

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2334

Gly Lys Ser Leu Lys Lys Lys Asn Gln

1 5

<210> 2335

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2335

Gly Leu Asp Val Gln Gln Gly Thr Val

1 5

<210> 2336

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2336

Gly Leu Thr Asp Gly Val Ser Arg Ile

1 5

<210> 2337

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2337

Gly Leu Tyr Pro Thr Tyr Thr Glu Trp

1 5

<210> 2338

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2338

Gly Leu Tyr Gln Thr Tyr Thr Glu Trp

1 5

<210> 2339

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2339

Gly Pro Pro Ser Arg Pro Thr Thr Pro

1 5

<210> 2340

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2340

Gly Pro Ser Ala Val Ala Ala Pro Ala

1 5

<210> 2341

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2341

Gly Val Ala Pro Thr Ala Gln Gln Leu

1 5

<210> 2342

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2342

Gly Val Gly Leu Ala Ser Thr Leu Trp

1 5

<210> 2343

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2343

Gly Tyr Glu Ala Gly Phe Ser Leu Gly

1 5

<210> 2344

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2344

His Ala Asp Asp Gly Thr Ile Val Ile

1 5

<210> 2345

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2345

His Ala Asp Trp Asn Asn Gln Ser Ile

1 5

<210> 2346

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2346

His Ala Glu His Glu Lys Asp Val Arg

1 5

<210> 2347

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2347

His Ala Asn His Tyr Gly Thr Arg Ile

1 5

<210> 2348

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2348

His Ala Gln Gly Lys Ala Leu Leu Tyr

1 5

<210> 2349

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2349

His Phe Ile Gly Tyr Ile Tyr Glu Val

1 5

<210> 2350

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2350

His Leu Ser Pro Ala Leu Ala Asp Val

1 5

<210> 2351

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2351

His Ser Leu Tyr Ala Ser Tyr Glu Tyr

1 5

<210> 2352

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2352

His Val Arg Gly Met Leu Val Pro Val

1 5

<210> 2353

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2353

His Val Ser Lys Glu Glu Gln Tyr Tyr

1 5

<210> 2354

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2354

His Val Thr Arg Gly Trp Ser Ile Phe

1 5

<210> 2355

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2355

Ile Ala Asp Ser Gly Phe Tyr Leu Asp

1 5

<210> 2356

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2356

Ile Ala His Arg Thr Glu Leu Arg Gly

1 5

<210> 2357

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2357

Ile Glu Asn Thr Ala Lys Leu Ser Gly

1 5

<210> 2358

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2358

Ile Glu Ser Lys Ile Ser Gln Ser Val

1 5

<210> 2359

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2359

Ile Glu Thr Gly Gly Ala Arg Arg Phe

1 5

<210> 2360

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2360

Ile Ile Lys Asp Ala Pro Pro Gly Ala

1 5

<210> 2361

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2361

Ile Ile Gln Glu Phe Ala Ala Asp Leu

1 5

<210> 2362

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2362

Ile Leu Ala Gly Ala Leu Ala Thr Tyr

1 5

<210> 2363

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2363

Ile Leu Leu Glu Asn Pro Ala Ala Glu

1 5

<210> 2364

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2364

Ile Leu Pro Ile Leu Val Leu Ala Leu

1 5

<210> 2365

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2365

Ile Pro Phe Gln Arg Ala Leu Arg Leu

1 5

<210> 2366

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2366

Ile Ser Val Arg Val His Val Ser Lys

1 5

<210> 2367

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2367

Ile Thr Asn Glu Thr Gly Lys Thr Tyr

1 5

<210> 2368

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2368

Ile Thr Asn Lys Arg Ala Ala Leu Ile

1 5

<210> 2369

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2369

Ile Thr Ser Tyr Val Gly Phe Ser Val

1 5

<210> 2370

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2370

Ile Thr Val Thr Ser Arg Gly Gly Phe

1 5

<210> 2371

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2371

Ile Val Ile Pro Pro Lys Ala Leu Phe

1 5

<210> 2372

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2372

Ile Val Val Glu Ala Gly Glu Leu Val

1 5

<210> 2373

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2373

Lys Ala Ala Lys Ser Val Asn Leu Met

1 5

<210> 2374

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2374

Lys Ala Ala Pro Leu Arg Arg Thr Thr

1 5

<210> 2375

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2375

Lys Ala Gly Lys Leu Ser Ala Thr Gly

1 5

<210> 2376

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2376

Lys Ala Gly Thr Ile Ala Ala Pro Trp

1 5

<210> 2377

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2377

Lys Ala Lys Ser Leu Thr Thr Glu Ile

1 5

<210> 2378

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2378

Lys Ala Thr Val Thr Thr Val Gln Val

1 5

<210> 2379

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2379

Lys Asp Tyr Arg Asp Lys Asp Gly Gly

1 5

<210> 2380

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2380

Lys Glu Ala Ala Thr Ile Val Ala Ala

1 5

<210> 2381

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2381

Lys Glu Asp Val Asp Ala Ala Gln Ile

1 5

<210> 2382

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2382

Lys Glu Val Asp Gly Ile Ile Gln Glu

1 5

<210> 2383

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2383

Lys Gly Pro Lys Leu Ala Met Pro Trp

1 5

<210> 2384

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2384

Lys Leu Ala Ser Gly Gly Gly Ala Val

1 5

<210> 2385

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2385

Lys Leu Lys Gly Lys Asn Gln Glu Phe

1 5

<210> 2386

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2386

Lys Leu Leu His His Ile Leu Pro Ile

1 5

<210> 2387

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2387

Lys Pro Ala Pro Gln Pro Gly Pro Gln

1 5

<210> 2388

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2388

Lys Pro Ala Pro Thr Ala Pro Pro Met

1 5

<210> 2389

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2389

Lys Pro Ala Val Ser Val Lys Val Ala

1 5

<210> 2390

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2390

Lys Pro Asp Arg Ala Ala Arg Val Ala

1 5

<210> 2391

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2391

Lys Pro Leu Ala Asp Ile Ala Val Ile

1 5

<210> 2392

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2392

Lys Pro Leu Ala Asp Ile Ala Val Val

1 5

<210> 2393

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2393

Lys Pro Leu Pro Lys Pro Leu Pro Val

1 5

<210> 2394

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2394

Lys Gln Ala Asp Arg Asp Phe Val Trp

1 5

<210> 2395

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2395

Lys Ser Leu Pro Gly Gly Lys Leu Pro

1 5

<210> 2396

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2396

Lys Ser Tyr Thr Leu Arg Tyr Leu Ala

1 5

<210> 2397

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2397

Lys Thr Asn Met Val Val Thr Ser Val

1 5

<210> 2398

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2398

Lys Val Leu Ala Pro Arg Leu Tyr Leu

1 5

<210> 2399

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2399

Lys Val Leu Ser Thr Lys Thr Thr Leu

1 5

<210> 2400

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2400

Leu Ala Ala Gly Ala Gly Leu Thr Leu

1 5

<210> 2401

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2401

Leu Ala Ala Ile Ala Ser Ala Ala His

1 5

<210> 2402

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2402

Leu Ala Ala Asn Gly Asn Gly Gln Trp

1 5

<210> 2403

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2403

Leu Ala Ala Gln Val Thr Gln Arg Gly

1 5

<210> 2404

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2404

Leu Ala Ala Arg Gly Asp Gly Ala Leu

1 5

<210> 2405

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2405

Leu Ala Ala Val Leu Val Asn Pro His

1 5

<210> 2406

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2406

Leu Ala Gly Ser Gly Leu Phe Arg Met

1 5

<210> 2407

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2407

Leu Ala Lys Ala Leu Ser Ala Ala Leu

1 5

<210> 2408

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2408

Leu Ala Leu Gln Ala Gly Asp Ala Leu

1 5

<210> 2409

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2409

Leu Ala Met Pro Trp Thr Phe His Ala

1 5

<210> 2410

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2410

Leu Ala Asn Asp Gly Thr Ile Val Ile

1 5

<210> 2411

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2411

Leu Ala Pro Thr Val Gly Val Ala Phe

1 5

<210> 2412

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2412

Leu Ala Ser Asp Gly Ser Val Asp Phe

1 5

<210> 2413

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2413

Leu Ala Thr Tyr Gln Ser Glu Tyr Leu

1 5

<210> 2414

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2414

Leu Ala Trp Ala Leu Met Leu Ala Cys

1 5

<210> 2415

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2415

Leu Glu Ala Gly Arg Arg Phe Thr His

1 5

<210> 2416

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2416

Leu Phe Thr Gln Gln Gly Gly Ala Tyr

1 5

<210> 2417

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2417

Leu Lys Arg Pro Gly Ser Ser Pro Met

1 5

<210> 2418

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2418

Leu Leu Gly Ser His Val Ala Arg Ala

1 5

<210> 2419

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2419

Leu Leu His His Ile Leu Pro Ile Leu

1 5

<210> 2420

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2420

Leu Leu Asn Ala Gly Gly Thr Leu Leu

1 5

<210> 2421

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2421

Leu Met Gly Pro Leu Gln Val Asn Ala

1 5

<210> 2422

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2422

Leu Asn Asp Ser Lys Ile Thr Met Gly

1 5

<210> 2423

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2423

Leu Pro Glu Pro Val Lys Leu Thr Leu

1 5

<210> 2424

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2424

Leu Pro Ile Leu Val Leu Ala Leu Leu

1 5

<210> 2425

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2425

Leu Pro Lys Ile Ser Lys Asn Ala Leu

1 5

<210> 2426

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2426

Leu Pro Lys Pro Val Thr Val Lys Leu

1 5

<210> 2427

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2427

Leu Pro Leu Lys Ala Asn Pro Met His

1 5

<210> 2428

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2428

Leu Pro Pro Arg Pro Val Val Ala Glu

1 5

<210> 2429

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2429

Leu Pro Ser Ile Pro Gly Thr Ser Ile

1 5

<210> 2430

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2430

Leu Pro Thr His Leu Tyr Lys Asn Phe

1 5

<210> 2431

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2431

Leu Pro Val Arg Gly Val Ala Leu Val

1 5

<210> 2432

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2432

Leu Pro Val Ser Leu Thr Ala Leu Asp

1 5

<210> 2433

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2433

Leu Gln Gly Gly Asn Lys Val Pro Val

1 5

<210> 2434

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2434

Leu Ser Ala Ala Leu Gly Ala Asp Trp

1 5

<210> 2435

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2435

Leu Ser Asp Ala Gly His Glu His Asp

1 5

<210> 2436

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2436

Leu Ser Gly Glu Val Gln Arg Lys Gly

1 5

<210> 2437

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2437

Leu Ser Ser Pro Ser Ala Ile Thr Val

1 5

<210> 2438

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2438

Leu Thr Trp Leu Ala Ile Leu Ala Val

1 5

<210> 2439

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2439

Leu Val Phe Ser His Val Arg Gly Met

1 5

<210> 2440

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2440

Leu Val Ser Asp Ala Gly Ala Asp Leu

1 5

<210> 2441

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2441

Leu Val Tyr Tyr Glu Ser Ile Ala Tyr

1 5

<210> 2442

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2442

Met Ala Ala Glu Ser Thr Phe Glu Ser

1 5

<210> 2443

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2443

Met Ala Ala Gly His Asp Ala Thr Leu

1 5

<210> 2444

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2444

Met Ala Ala Trp Ser Glu Arg Ala Gly

1 5

<210> 2445

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2445

Met Ala Leu Gly Ala Leu Gly Ala Ala

1 5

<210> 2446

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2446

Met Ala Pro Val Met Gly Ala Cys Met

1 5

<210> 2447

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2447

Met Ala Thr Lys Gly Glu Met Gln Ile

1 5

<210> 2448

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2448

Met Asp Ala Lys Gly Gly Thr Leu Leu

1 5

<210> 2449

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2449

Met Glu Ser Asn Lys Asp Ile Val Ile

1 5

<210> 2450

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2450

Met Glu Val Met Leu Arg Ala Val Phe

1 5

<210> 2451

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2451

Met Glu Tyr Phe Lys Thr Pro Leu Pro

1 5

<210> 2452

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2452

Met His Thr Ile Ala Ser Ile Leu Leu

1 5

<210> 2453

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2453

Met Ile Tyr Met Ser Gly Leu Ala Val

1 5

<210> 2454

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2454

Met Leu Ala Cys Thr Gly Leu Pro Leu

1 5

<210> 2455

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2455

Met Pro Ile Asp Arg Lys Thr Leu Cys

1 5

<210> 2456

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2456

Met Pro Lys Ala Pro Glu Leu Asp Leu

1 5

<210> 2457

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2457

Met Gln Arg Gln Ala Gly Leu Pro Leu

1 5

<210> 2458

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2458

Asn Ala Leu Ala Trp Ala Leu Met Leu

1 5

<210> 2459

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2459

Asn Ile Thr Asn Lys Arg Ala Ala Leu

1 5

<210> 2460

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2460

Asn Leu Met Ala Ala Glu Ser Thr Phe

1 5

<210> 2461

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2461

Asn Asn Glu Thr Met Ser Gly Arg Gln

1 5

<210> 2462

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2462

Asn Pro Gly Ser Leu Ile Ala Glu Val

1 5

<210> 2463

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2463

Asn Pro Met His Thr Ile Ala Ser Ile

1 5

<210> 2464

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2464

Asn Pro Gln Thr Gln Leu Ser Asn Ile

1 5

<210> 2465

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2465

Asn Pro Tyr Thr Ser Arg Arg Ser Val

1 5

<210> 2466

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2466

Pro Ala Pro Thr Ala Pro Pro Met Pro

1 5

<210> 2467

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2467

Pro Ala Ser Ala Asn Thr Leu Leu Leu

1 5

<210> 2468

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2468

Pro Ala Val Ala Leu Pro Arg Pro Leu

1 5

<210> 2469

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2469

Pro Asp Ala His Val Pro Phe Cys Phe

1 5

<210> 2470

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2470

Pro Glu Leu Gly Ala Ala Ile Arg Val

1 5

<210> 2471

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2471

Pro Phe Ile Ile Lys Leu Lys Asp Cys

1 5

<210> 2472

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2472

Pro Gly Pro Gln Pro Pro Gln Pro Pro

1 5

<210> 2473

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2473

Pro Gly Pro Gln Pro Pro Gln Pro Gln

1 5

<210> 2474

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2474

Pro Gly Pro Thr Thr Asp Tyr Ser Thr

1 5

<210> 2475

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2475

Pro Gly Thr Phe Thr Ala Gly Lys Asp

1 5

<210> 2476

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2476

Pro Gly Thr Pro Gly Asp Leu Leu Glu

1 5

<210> 2477

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2477

Pro Lys Pro Lys Pro Lys Ala Glu Arg

1 5

<210> 2478

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2478

Pro Lys Pro Lys Pro Lys Pro Lys Ala

1 5

<210> 2479

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2479

Pro Lys Pro Lys Pro Lys Pro Lys Pro

1 5

<210> 2480

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2480

Pro Leu Pro Pro Arg Pro Val Val Ala

1 5

<210> 2481

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2481

Pro Pro Ala Pro Lys Pro Ala Pro Gln

1 5

<210> 2482

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2482

Pro Pro Lys Pro Ala Pro Val Ala Lys

1 5

<210> 2483

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2483

Pro Pro Arg Pro Val Ala Ala Gln Val

1 5

<210> 2484

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2484

Pro Pro Arg Pro Val Val Ala Glu Lys

1 5

<210> 2485

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2485

Pro Pro Ser Arg Pro Thr Thr Pro Pro

1 5

<210> 2486

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2486

Pro Arg Arg Ala Arg Arg Ala Leu Arg

1 5

<210> 2487

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2487

Gln Ala Ala Pro Leu Ser Ile Thr Leu

1 5

<210> 2488

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2488

Gln Ala Asp Arg Asp Phe Val Trp Tyr

1 5

<210> 2489

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2489

Gln Ala Ile Val Val Gly Lys Asp Leu

1 5

<210> 2490

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2490

Gln Ala Leu Gly Ala Leu Lys Leu Tyr

1 5

<210> 2491

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2491

Gln Glu Leu Ala Leu Lys Leu Lys Gly

1 5

<210> 2492

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2492

Gln Ile Thr Gln His Gly Gly Pro Tyr

1 5

<210> 2493

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2493

Gln Ile Thr Gln His Gly Ser Pro Tyr

1 5

<210> 2494

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2494

Gln Pro Leu Pro Pro Arg Pro Val Ala

1 5

<210> 2495

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2495

Gln Pro Pro Ala Gly Arg Glu Leu Ser

1 5

<210> 2496

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2496

Gln Gln Leu Lys Gln Ala Asp Arg Asp

1 5

<210> 2497

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2497

Gln Gln Val Gln Val Leu Gln Arg Gln

1 5

<210> 2498

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2498

Gln Ser Ile Val Glu Ala Pro Glu Leu

1 5

<210> 2499

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2499

Gln Val Gly Ser Ser Asn Ser Ala Phe

1 5

<210> 2500

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2500

Arg Ala Gly Leu Ser Pro Ala Thr Trp

1 5

<210> 2501

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2501

Arg Ala Arg Arg Ala Leu Arg Gln Asp

1 5

<210> 2502

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2502

Arg Ala Ser Ala Ser Arg Ala Arg Ile

1 5

<210> 2503

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2503

Arg Glu Leu Ser Ala Ala Ala Asn Ala

1 5

<210> 2504

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2504

Arg Glu Thr Phe Cys Ile Thr Thr Ile

1 5

<210> 2505

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2505

Arg Gly Phe Ala Gln Arg Gln Gln Leu

1 5

<210> 2506

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2506

Arg Gly Ser Ala Ala Thr Phe Thr Leu

1 5

<210> 2507

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2507

Arg Gly Trp Ser Ile Phe Ala Leu Tyr

1 5

<210> 2508

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2508

Arg Lys Met Leu Tyr Leu Ile Tyr Val

1 5

<210> 2509

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2509

Arg Leu Arg Lys Met Leu Tyr Leu Ile

1 5

<210> 2510

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2510

Arg Pro Gln Ile Thr Asp Ala Val Thr

1 5

<210> 2511

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2511

Arg Pro Ser Val Asn Gly Gly Arg Ile

1 5

<210> 2512

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2512

Arg Arg Phe Thr His Ala Asp Gly Trp

1 5

<210> 2513

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2513

Arg Ser Gly Ala Arg Ala Thr Ser Leu

1 5

<210> 2514

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2514

Arg Ser Arg Val Arg Ala Leu Ala Trp

1 5

<210> 2515

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2515

Arg Ser Arg Val Arg Ala Leu Ser Trp

1 5

<210> 2516

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2516

Arg Thr His Gly Val Gly Ala Ser Leu

1 5

<210> 2517

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2517

Arg Thr Arg Gly Gln Ala Arg Ser Gly

1 5

<210> 2518

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2518

Arg Val Ala Pro Pro Ala Val Ala Leu

1 5

<210> 2519

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2519

Arg Val Leu Pro Glu Pro Val Lys Leu

1 5

<210> 2520

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2520

Arg Val Arg Ala Leu Ala Trp Leu Leu

1 5

<210> 2521

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2521

Arg Val Arg Ala Leu Ser Trp Leu Leu

1 5

<210> 2522

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2522

Arg Val Thr Val Ser Gly Gly Ser Leu

1 5

<210> 2523

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2523

Ser Glu Ala Met Ala Ala Trp Ser Glu

1 5

<210> 2524

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2524

Ser Glu Ser His Asn Phe His Ala Ser

1 5

<210> 2525

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2525

Ser Gly Glu Gly Arg Val Asn Ile Gly

1 5

<210> 2526

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2526

Ser Gly Leu Ala Val Arg Val His Val

1 5

<210> 2527

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2527

Ser Leu Ala Asp Ile Ser Leu Gly Ala

1 5

<210> 2528

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2528

Ser Leu Phe Ala Ile Leu Ser Ser Thr

1 5

<210> 2529

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2529

Ser Leu Phe Ala Pro His Gly Asn Val

1 5

<210> 2530

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2530

Ser Leu Ser Ile Asp Asn Ala Thr Trp

1 5

<210> 2531

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2531

Ser Pro Met Glu Val Met Leu Arg Ala

1 5

<210> 2532

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2532

Ser Pro Gln Pro Ile Arg Ala Thr Val

1 5

<210> 2533

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2533

Ser Pro Arg Arg Ala Arg Arg Ala Leu

1 5

<210> 2534

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2534

Ser Pro Ser Arg Leu Ala Gly Thr Leu

1 5

<210> 2535

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2535

Ser Ser Thr Pro Leu Gly Ser Leu Phe

1 5

<210> 2536

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2536

Ser Thr Tyr Glu Leu Leu Asp Tyr Leu

1 5

<210> 2537

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2537

Ser Val Ala Met Lys Pro Tyr Glu Val

1 5

<210> 2538

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2538

Ser Val Ala Pro Asn Ala Leu Ala Trp

1 5

<210> 2539

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2539

Ser Val Lys Val Ala Lys Lys Leu Phe

1 5

<210> 2540

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2540

Thr Ala Phe Met Ser Gly Arg Ser Leu

1 5

<210> 2541

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2541

Thr Ala Gly Ala Thr Pro Phe Asp Ile

1 5

<210> 2542

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2542

Thr Ala Pro Val Thr Ser Pro Ala Trp

1 5

<210> 2543

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2543

Thr Ala Arg Thr Gly Trp Leu Thr Trp

1 5

<210> 2544

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2544

Thr Glu Ala Gln Gly Val Gln Val Arg

1 5

<210> 2545

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2545

Thr Glu Val Tyr Leu Glu His Arg Met

1 5

<210> 2546

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2546

Thr Phe Glu Gly Lys Pro Ala Leu Glu

1 5

<210> 2547

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2547

Thr Phe Thr Gly Lys Val Thr Asn Gly

1 5

<210> 2548

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2548

Thr Gly Asp Gly Gly Gly His Thr Asp

1 5

<210> 2549

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2549

Thr Leu Ala Lys Ala Leu Ser Ala Ala

1 5

<210> 2550

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2550

Thr Leu Ala Asn Val Gly Asp Thr Trp

1 5

<210> 2551

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2551

Thr Leu Asn Ala Ser Asn Leu Thr Leu

1 5

<210> 2552

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2552

Thr Leu Ser Ser Ala His Gly Asn Val

1 5

<210> 2553

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2553

Thr Pro Phe Asp Ile Lys Leu Lys Glu

1 5

<210> 2554

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2554

Thr Pro Phe Ile Ile Lys Leu Lys Asp

1 5

<210> 2555

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2555

Thr Pro Gly Trp Ser Ile Tyr Gly Leu

1 5

<210> 2556

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2556

Thr Pro Leu Gly Ser Ala Ala Thr Phe

1 5

<210> 2557

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2557

Thr Pro Leu Gly Ser Leu Phe Ala Ile

1 5

<210> 2558

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2558

Thr Pro Leu Pro Val Ser Leu Thr Ala

1 5

<210> 2559

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2559

Thr Arg Gln Gly Ile Met Asp Gln Tyr

1 5

<210> 2560

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2560

Thr Ser Lys Gln Asp Glu Arg Asn Tyr

1 5

<210> 2561

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2561

Thr Ser Pro Tyr Asp Gly Lys Tyr Trp

1 5

<210> 2562

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2562

Thr Ser Arg Arg Ser Val Ala Ser Ile

1 5

<210> 2563

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2563

Thr Ser Arg Thr Val Thr Met Arg Tyr

1 5

<210> 2564

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2564

Thr Thr Glu Tyr Pro Asn Ala Arg Tyr

1 5

<210> 2565

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2565

Thr Thr Glu Tyr Ser Asn Ala Arg Tyr

1 5

<210> 2566

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2566

Thr Thr Leu Gly Leu Glu Gln Thr Phe

1 5

<210> 2567

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2567

Thr Val Leu Ala Ala Gly Ala Gly Leu

1 5

<210> 2568

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2568

Thr Val Gln Glu Leu Ala Leu Lys Leu

1 5

<210> 2569

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2569

Thr Val Val Gln Arg Asn Lys His Trp

1 5

<210> 2570

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2570

Val Ala Ala Ala Ala Asp Leu Ala Leu

1 5

<210> 2571

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2571

Val Ala Leu Ala Ser Gln Ala Arg Trp

1 5

<210> 2572

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2572

Val Ala Met Lys Pro Tyr Glu Val Thr

1 5

<210> 2573

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2573

Val Ala Arg Leu Val Lys Leu Gln Gly

1 5

<210> 2574

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2574

Val Ala Val Ala Gly Gly Arg Trp His

1 5

<210> 2575

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2575

Val Glu Ala Ser Ala Ile Thr Thr Tyr

1 5

<210> 2576

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2576

Val Glu Asp Ile Gly Gly Lys Asn Tyr

1 5

<210> 2577

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2577

Val Glu Val Ser Ser Pro Pro Pro Val

1 5

<210> 2578

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2578

Val Gly Ala Gly Gly Tyr Glu Ala Gly

1 5

<210> 2579

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2579

Val Gly Gly Gly Glu His Gly Arg Trp

1 5

<210> 2580

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2580

Val His Val Ser Lys Glu Glu Gln Tyr

1 5

<210> 2581

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2581

Val Ile Asp Gly Gln Lys Val Leu Ala

1 5

<210> 2582

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2582

Val Ile Gly Ala Cys Thr Ser Pro Tyr

1 5

<210> 2583

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2583

Val Lys Leu Gly Gly Val Tyr Glu Ala

1 5

<210> 2584

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2584

Val Leu Ala Pro Arg Leu Tyr Leu Thr

1 5

<210> 2585

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2585

Val Leu Val Lys Thr Asn Met Val Val

1 5

<210> 2586

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2586

Val Pro Ala Ser Gly Ala Pro Ala Ala

1 5

<210> 2587

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2587

Val Pro Phe Cys Phe Gly Lys Asp Leu

1 5

<210> 2588

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2588

Val Pro Val Ser Glu His Cys Thr Val

1 5

<210> 2589

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2589

Val Pro Val Thr Pro Pro Lys Val Glu

1 5

<210> 2590

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2590

Val Arg Thr Val Ser Ala Met Glu Tyr

1 5

<210> 2591

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2591

Val Ser Gly Gly Ser Leu Ser Ala Pro

1 5

<210> 2592

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2592

Val Ser Ser Ala Thr Lys Ala Lys Gly

1 5

<210> 2593

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2593

Val Ser Ser Pro Pro Pro Val Ser Val

1 5

<210> 2594

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2594

Val Ser Val Lys Val Ala Lys Lys Leu

1 5

<210> 2595

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2595

Val Thr Met Arg Tyr Leu Ala Ser Tyr

1 5

<210> 2596

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2596

Val Thr Ser Val Ala Met Lys Pro Tyr

1 5

<210> 2597

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2597

Val Val Ala Glu Lys Val Thr Thr Pro

1 5

<210> 2598

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2598

Val Val Asp Gly Pro Pro Ser Arg Pro

1 5

<210> 2599

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2599

Val Val Glu Thr Ala Gln Pro Leu Pro

1 5

<210> 2600

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2600

Trp Leu Thr Trp Leu Ala Ile Leu Ala

1 5

<210> 2601

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2601

Trp Thr Phe His Ala Gly Tyr Arg Tyr

1 5

<210> 2602

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2602

Trp Val Met Thr Asp Asn Ser Asn Val

1 5

<210> 2603

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2603

Tyr Ala Glu His Gly Glu Val Ser Ile

1 5

<210> 2604

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2604

Tyr Ala Ile Asp Gly Thr Ala Ala Gly

1 5

<210> 2605

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2605

Tyr Ala Leu Lys Ser Arg Ile Ala Leu

1 5

<210> 2606

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2606

Tyr Ala Thr Asn Pro Gln Thr Gln Leu

1 5

<210> 2607

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2607

Tyr Asp Thr Gly Asp Leu Ile Ala Tyr

1 5

<210> 2608

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2608

Tyr Glu Ala Gly Phe Ser Leu Gly Ser

1 5

<210> 2609

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2609

Tyr Glu Asp Ala Thr Phe Glu Thr Tyr

1 5

<210> 2610

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2610

Tyr Glu Asn Lys Ser Ser Thr Pro Leu

1 5

<210> 2611

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2611

Tyr Glu Val Thr Pro Thr Arg Met Leu

1 5

<210> 2612

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2612

Tyr Glu Tyr Ile Trp Gly Leu Tyr Pro

1 5

<210> 2613

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2613

Tyr Glu Tyr Ile Trp Gly Leu Tyr Gln

1 5

<210> 2614

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2614

Tyr Glu Tyr Ser Lys Gly Pro Lys Leu

1 5

<210> 2615

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2615

Tyr Phe Glu Pro Gly Pro Thr Thr Asp

1 5

<210> 2616

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2616

Tyr Ile Trp Gly Leu Tyr Pro Thr Tyr

1 5

<210> 2617

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2617

Tyr Ile Trp Gly Leu Tyr Gln Thr Tyr

1 5

<210> 2618

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2618

Tyr Leu Arg Gln Ile Thr Pro Gly Trp

1 5

<210> 2619

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2619

Tyr Met Ile Tyr Met Ser Gly Leu Ala

1 5

<210> 2620

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2620

Tyr Pro Ala Leu Arg Ala Ala Leu Ile

1 5

<210> 2621

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2621

Tyr Pro Gly Thr Pro Gly Asp Leu Leu

1 5

<210> 2622

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2622

Tyr Pro Thr Tyr Thr Glu Trp Ser Val

1 5

<210> 2623

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2623

Tyr Gln Thr Tyr Thr Glu Trp Ser Val

1 5

<210> 2624

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2624

Tyr Ser Thr Gly Asp Leu Arg Ala Tyr

1 5

<210> 2625

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2625

Tyr Thr Leu Arg Tyr Leu Ala Ser Tyr

1 5

<210> 2626

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2626

Tyr Val Leu Val Lys Thr Asn Met Val

1 5

<210> 2627

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2627

Tyr Tyr Asp Tyr Glu Asp Ala Thr Phe

1 5

<210> 2628

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2628

Ala Ala Phe Ile Ala Leu Tyr Pro Asn Ser Gln Leu Ala Pro Thr

1 5 10 15

<210> 2629

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2629

Gly Gly Ala Glu Tyr Asn Leu Ala Leu Gly Gln Arg Arg Ala

1 5 10

<210> 2630

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2630

Gly Gly Ala Glu Tyr Asn Leu Ala Leu Gly Gln Arg Arg Ala Asp Ala

1 5 10 15

<210> 2631

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2631

Ile Ala Leu Tyr Pro Asn Ser Gln Leu Ala Pro Thr

1 5 10

<210> 2632

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2632

Lys Pro Asp Gln Gly Glu Val Val Ala Val Gly Pro Gly Lys Lys Thr

1 5 10 15

Glu Asp

<210> 2633

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2633

Lys Pro Asp Gln Gly Glu Val Val Ala Val Gly Pro Gly Lys Lys Thr

1 5 10 15

Glu Asp Gly

<210> 2634

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2634

Leu Ala Glu Val Leu Asp Tyr His Asn Phe Val Leu Thr Gln

1 5 10

<210> 2635

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2635

Gln Ser Ile Ala Leu Ser Ser Leu Met Val Ala Gln Ala Ile Pro

1 5 10 15

<210> 2636

<211> 23

<212> PRT

<213>

Artificial Sequence

<220><223> Synthetic

<400> 2636

Ser Ile Gly Val Leu Gly Tyr Gln Lys Thr Val Asp His Thr Lys Val

1 5 10 15

Asn Ser Lys Leu Ser Leu Phe

20

<210> 2637

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2637

Ala Ala His Ala Asp Trp Asn Asn Gln Ser Ile Val Lys Thr

1 5 10

<210> 2638

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2638

Ala Ala Leu Gly Arg Gly

1 5

<210> 2639

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2639

Ala Ala Leu Gly Arg Gly His Ser Leu Tyr Ala Ser Tyr Glu

1 5 10

<210> 2640

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2640

Ala Ala Pro Leu Arg Arg Thr Thr Leu Ala Met Ala Leu Gly

1 5 10

<210> 2641

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2641

Ala Ala Pro Leu Ser Ile Thr Leu Gln Ala Gly Ala His Ala

1 5 10

<210> 2642

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2642

Ala Asp Ala Gln Gly Asp Ile Val Ala Thr Glu Leu Pro Ser

1 5 10

<210> 2643

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2643

Ala Asp Ser Gly Phe Tyr Leu Asp Ala Thr Leu Arg Ala Ser

1 5 10

<210> 2644

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2644

Ala Glu Leu Ala

1

<210> 2645

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2645

Ala Glu Leu Ala Val Phe Arg Ala Gly Gly Gly Ala Tyr Arg

1 5 10

<210> 2646

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2646

Ala Glu Leu Gln Phe Arg Asn Gly Ser Val Thr Ser Ser Gly

1 5 10

<210> 2647

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2647

Ala Gly Gly Arg Trp His Leu Gly Gly Leu Ala Gly Tyr Thr

1 5 10

<210> 2648

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2648

Ala Gly Val Ala Ala Met Gln Gly Ala Val Val His Leu Gln

1 5 10

<210> 2649

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2649

Ala Gly Tyr Thr Arg Gly Asp Arg Gly Phe Thr Gly Asp Gly

1 5 10

<210> 2650

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2650

Ala Leu Ala Ser Gln Ala Arg Trp Thr Gly Ala Thr Arg Ala

1 5 10

<210> 2651

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2651

Ala Met Pro Trp Thr Phe His Ala Gly Tyr Arg Tyr Ser Trp

1 5 10

<210> 2652

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2652

Ala Met Gln Gly Ala Val Val His Leu Gln Arg Ala Thr Ile Arg Arg

1 5 10 15

Gly Asp Ala Pro

20

<210> 2653

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2653

Ala Asn Gly Leu Arg Val Arg Asp Glu

1 5

<210> 2654

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2654

Ala Asn Gly Leu Arg Val Arg Asp Glu Gly Gly Ser Ser Val

1 5 10

<210> 2655

<211> 14

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 2655

Ala Asn Lys Asp Gly Lys Val Asp Ile Gly Thr Tyr Arg Tyr

1 5 10

<210> 2656

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2656

Ala Pro Ala Ala Val Ser Val Leu Gly Ala Ser Glu Leu Thr

1 5 10

<210> 2657

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2657

Ala Pro Pro Ala Pro Lys Pro Ala Pro Gln Pro Gly Pro Gln

1 5 10

<210> 2658

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2658

Ala Gln Gly Ile Leu Leu Glu Asn Pro Ala Ala Glu Leu Gln

1 5 10

<210> 2659

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2659

Ala Arg Trp Thr Gly Ala Thr Arg Ala Val Asp Ser Leu Ser

1 5 10

<210> 2660

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2660

Ala Ser Leu Glu Ala Gly Arg Arg Phe Thr His Ala Asp Gly

1 5 10

<210> 2661

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2661

Ala Ser Tyr Glu Tyr Ser Lys Gly Pro Lys Leu Ala Met Pro

1 5 10

<210> 2662

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2662

Ala Thr Phe Thr Leu Ala Asn Lys Asp

1 5

<210> 2663

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2663

Ala Thr Phe Thr Leu Ala Asn Lys Asp Gly Lys Val Asp Ile

1 5 10

<210> 2664

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2664

Ala Thr Arg Ala Val Asp Ser Leu Ser Ile Asp Asn Ala Thr

1 5 10

<210> 2665

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2665

Asp Asp Asp Gly Ile Ala Leu Tyr Val Ala Gly Glu Gln Ala Gln

1 5 10 15

<210> 2666

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2666

Asp Asp Gly Ile Ala Leu Tyr Val Ala Gly Glu Gln Ala Gln

1 5 10

<210> 2667

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2667

Asp Gly Gly His Ile Thr Gly Gly Arg Ala Ala Gly Val Ala

1 5 10

<210> 2668

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2668

Asp Gly Ile Arg Arg Phe Leu Gly Thr Val Thr Val Lys Ala Gly Lys

1 5 10 15

<210> 2669

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2669

Asp Gly Ser Val Asp Phe Gln Gln Pro Ala Glu Ala Gly Arg

1 5 10

<210> 2670

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2670

Asp Gly Tyr Ala Val Lys Gly Lys Tyr Arg Thr His Gly Val

1 5 10

<210> 2671

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2671

Asp Ile Val Ala Thr Glu Leu Pro Ser Ile Pro Gly Thr Ser

1 5 10

<210> 2672

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2672

Asp Lys Leu Val Val Met Gln Asp Ala Ser Gly Gln His Arg

1 5 10

<210> 2673

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2673

Asp Leu Gly Leu Ser Asp Lys Leu Val Val Met Gln Asp Ala

1 5 10

<210> 2674

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2674

Asp Asn Ala Thr Trp Val Met Thr Asp Asn Ser Asn Val Gly Ala

1 5 10 15

<210> 2675

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2675

Asp Asn Ala Thr Trp Val Met Thr Asp Asn Ser Asn Val Gly Ala Leu
1 5 10 15
Arg Leu Ala

<210> 2676

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2676

Asp Asn Arg Ala Gly Arg Arg Phe Asp Gln Lys Val Ala Gly
1 5 10

<210> 2677

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2677

Glu Ala Gly Arg Phe Lys Val Leu Thr Val Asn Thr Leu Ala
1 5 10

<210> 2678

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2678

Glu Leu Ala Gln Ser Ile Val Glu Ala Pro Glu Leu Gly Ala
1 5 10

<210> 2679

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2679

Glu Leu Gly Ala Ala Ile Arg Val Gly Arg Gly Ala Arg Val

1 5 10

<210> 2680

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2680

Glu Leu Gly Ala Asp His Ala Val Ala Val Ala Gly Gly Arg

1 5 10

<210> 2681

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2681

Glu Leu Pro Ser Ile Pro Gly Thr Ser Ile Gly Pro Leu Asp

1 5 10

<210> 2682

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2682

Glu Pro Val Lys Leu Thr Leu Thr Gly Gly Ala Asp Ala Gln

1 5 10

<210> 2683

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2683

Glu Gln Ala Gln Ala Ser Ile Ala Asp Ser Thr Leu Gln Gly

1 5 10

<210> 2684

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2684

Glu Arg Gly Ala Asn Val Thr Val Gln Arg Ser Ala Ile Val

1 5 10

<210> 2685

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2685

Glu Arg Gln His Gly Ile His Ile Gln Gly Ser Asp Pro Gly

1 5 10

<210> 2686

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2686

Glu Val Gly Lys Arg Ile Glu Leu Ala Gly Gly Arg Gln Val

1 5 10

<210> 2687

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2687

Phe Asp Gly Ala Gly Thr Val His Thr Asn Gly Ile Ala His

1 5 10

<210> 2688

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2688

Phe Gln Gln Pro Ala Glu Ala Gly Arg Phe Lys Val Leu Thr

1 5 10

<210> 2689

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2689

Phe Arg Ala Gly Gly Gly Ala Tyr Arg Ala Ala Asn Gly Leu

1 5 10

<210> 2690

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2690

Gly Ala His Ala Gln Gly Lys Ala Leu Leu Tyr Arg Val Leu

1 5 10

<210> 2691

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2691

Gly Ala Arg Val Thr Val Ser Gly Gly Ser Leu Ser Ala Pro

1 5 10

<210> 2692

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2692

Gly Ala Tyr Arg Ala Ala Asn Gly Leu Arg Val Arg Asp Glu

1 5 10

<210> 2693

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2693

Gly Asp Ala Pro Ala Gly Gly Ala Val Pro Gly Gly Ala Val

1 5 10

<210> 2694

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2694

Gly Gly Ala Val Pro Gly Gly Ala Val Pro Gly Gly Phe Gly

1 5 10

<210> 2695

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2695

Gly Gly Ala Val Pro Gly Gly Phe Gly Pro Val Leu Asp Gly

1 5 10

<210> 2696

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2696

Gly Gly Phe Gly Pro Val Leu Asp Gly Trp Tyr Gly Val Asp

1 5 10

<210> 2697

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2697

Gly Gly Leu His Ile Gly Ala Leu Gln Ser Leu Gln Pro Glu

1 5 10

<210> 2698

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2698

Gly Gly Val Gln Ile Glu Arg Gly Ala Asn Val Thr Val Gln

1 5 10

<210> 2699

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2699

Gly His Ser Leu Tyr Ala Ser Tyr Glu Tyr Ser Lys Gly Pro

1 5 10

<210> 2700

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2700

Gly His Thr Asp Ser Val His Val Gly Gly Tyr Ala Thr Tyr

1 5 10

<210> 2701

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2701

Gly Ile Ala His Arg Thr Glu Leu Arg Gly Thr Arg Ala Glu

1 5 10

<210> 2702

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2702

Gly Lys Ala Leu Leu Tyr Arg Val Leu Pro Glu Pro Val Lys

1 5 10

<210> 2703

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2703

Gly Leu Gly Met Ala Ala Ala Leu Gly Arg Gly His Ser Leu

1 5 10

<210> 2704

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2704

Gly Asn Val Ile Glu Thr Gly Gly Ala Arg Arg Phe Ala Pro

1 5 10

<210> 2705

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2705

Gly Pro Leu Asp Val Ala Leu Ala Ser Gln Ala Arg Trp Thr

1 5 10

<210> 2706

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2706

Gly Gln His Arg Leu Trp Val Arg Asn

1 5

<210> 2707

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2707

Gly Arg Gly Phe Ala Gln Arg Gln Gln Leu Asp Asn Arg Ala

1 5 10

<210> 2708

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2708

Gly Arg Leu Gly Leu Glu Val Gly Lys Arg Ile Glu Leu Ala

1 5 10

<210> 2709

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2709

Gly Arg Gln Val Gln Pro Tyr Ile Lys Ala Ser Val Leu Gln

1 5 10

<210> 2710

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2710

Gly Arg Arg Phe Thr His Ala Asp Gly Trp Phe Leu Glu Pro Gln Ala

1 5 10 15

Glu Leu Ala

<210> 2711

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2711

Gly Ser Glu Pro Ala Ser Ala Asn Thr Leu Leu Leu Val Gln

1 5 10

<210> 2712

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2712

Gly Ser Ser Val Leu Gly Arg Leu Gly Leu Glu Val Gly Lys

1 5 10

<210> 2713

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2713

Gly Thr Thr Ile Lys Val Ser Gly Arg Gln Ala Gln Gly Ile

1 5 10

<210> 2714

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2714

Gly Thr Val Thr Val Lys Ala Gly Lys Leu Val Ala Asp His

1 5 10

<210> 2715

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2715

His Ala Val Ala Val Ala Gly Gly Arg Trp His Leu Gly Gly

1 5 10

<210> 2716

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2716

Ile Glu Leu Ala Gly Gly Arg Gln Val Gln Pro Tyr Ile Lys

1 5 10

<210> 2717

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2717

Ile His Ile Gln Gly Ser Asp Pro Gly Gly Val Arg Thr Ala

1 5 10

<210> 2718

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2718

Ile Arg Arg Phe Leu Gly Thr Val Thr Val Lys Ala Gly Lys

1 5 10

<210> 2719

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2719

Ile Arg Val Gly Arg Gly Ala Arg Val Thr Val Ser Gly Gly

1 5 10

<210> 2720

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2720

Ile Thr Leu Gln Ala Gly Ala His Ala

1 5

<210> 2721

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2721

Ile Thr Leu Gln Ala Gly Ala His Ala Gln Gly Lys Ala Leu

1 5 10

<210> 2722

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2722

Ile Val Glu Ala Pro Glu Leu Gly Ala Ala Ile Arg Val Gly

1 5 10

<210> 2723

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2723

Ile Val Lys Thr Gly Glu Arg Gln His Gly Ile His Ile Gln

1 5 10

<210> 2724

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2724

Lys Ala Gly Lys Leu Val Ala Asp His Ala Thr Leu Ala Asn

1 5 10

<210> 2725

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2725

Lys Gly Lys Tyr Arg Thr His Gly Val Gly Ala Ser Leu Glu

1 5 10

<210> 2726

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2726

Lys Pro Ala Pro Gln Pro Gly Pro Gln Pro Pro Gln Pro Pro

1 5 10

<210> 2727

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2727

Lys Val Ala Gly Phe Glu Leu Gly Ala Asp His Ala Val Ala

1 5 10

<210> 2728

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2728

Lys Val Ala Gly Ser Asp Gly Tyr Ala Val Lys Gly Lys Tyr

1 5 10

<210> 2729

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2729

Lys Val Asp Ile Gly Thr Tyr Arg Tyr Arg Leu Ala Ala Asn

1 5 10

<210> 2730

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2730

Lys Val Leu Thr Val Asn Thr Leu Ala Gly Ser Gly Leu Phe

1 5 10

<210> 2731

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2731

Leu Ala Ala Asn Gly Asn Gly Gln Trp Ser Leu Val Gly Ala

1 5 10

<210> 2732

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2732

Leu Ala Met Pro Trp Thr Phe His Ala Gly Tyr Arg Tyr Ser

1 5 10

<210> 2733

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2733

Leu Ala Ser Thr Leu Trp Tyr Ala Glu Ser Asn Ala Leu Ser

1 5 10

<210> 2734

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2734

Leu Glu Asn Asp Phe Lys Val Ala Gly Ser Asp Gly Tyr Ala

1 5 10

<210> 2735

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2735

Leu Glu Asn Pro Ala Ala Glu Leu Gln Phe Arg Asn Gly Ser

1 5 10

<210> 2736

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2736

Leu Gly Ala Ala Pro Ala Ala His Ala Asp Trp Asn Asn Gln

1 5 10

<210> 2737

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2737

Leu Gly Gly Leu Ala Gly Tyr Thr Arg Gly Asp Arg Gly Phe Thr Gly

1 5 10 15

Asp Gly

<210> 2738

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2738

Leu Leu Glu Asn Pro

1 5

<210> 2739

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2739

Leu Leu Val Gln Thr Pro Leu Gly Ser Ala Ala Thr Phe Thr

1 5 10

<210> 2740

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2740

Leu Pro Pro Ser Arg Val Val Leu Arg Asp Thr Asn Val Thr

1 5 10

<210> 2741

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2741

Leu Gln Pro Glu Asp Leu Pro Pro Ser

1 5

<210> 2742

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2742

Leu Gln Pro Glu Asp Leu Pro Pro Ser Arg Val Val Leu Arg

1 5 10

<210> 2743

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2743

Leu Arg Ala Ser Arg Leu Glu Asn Asp Phe Lys Val Ala Gly

1 5 10

<210> 2744

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2744

Leu Arg Leu Ala Ser Asp Gly Ser Val Asp Phe Gln Gln Pro

1 5 10

<210> 2745

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2745

Leu Ser Ala Ala Ala Asn Ala Ala Val Asn Thr Gly Gly Val

1 5 10

<210> 2746

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2746

Leu Ser Ala Pro His Gly Asn Val Ile Glu Thr Gly Gly Ala

1 5 10

<210> 2747

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2747

Leu Ser Asp Asp Gly Ile Arg Arg Phe Leu Gly Thr Val Thr

1 5 10

<210> 2748

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2748

Leu Val Gly Ala Lys Ala Pro Pro Ala Pro Lys Pro Ala Pro

1 5 10

<210> 2749

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2749

Leu Tyr Val Ala Gly Glu Gln Ala Gln Ala Ser Ile Ala Asp

1 5 10

<210> 2750

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2750

Met Ala Leu Gly Ala Leu Gly Ala Ala Pro Ala Ala His Ala

1 5 10

<210> 2751

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2751

Met Asn Met Ser Leu Ser Arg Ile Val Lys Ala Ala Pro Leu

1 5 10

<210> 2752

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2752

Met Gln Asp Ala Ser Gly Gln His Arg

1 5

<210> 2753

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2753

Met Gln Gly Ala Val Val His Leu Gln Arg Ala Thr Ile Arg

1 5 10

<210> 2754

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2754

Asn Ala Ala Val Asn Thr Gly Gly Val Gly Leu Ala Ser Thr

1 5 10

<210> 2755

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2755

Asn Ala Leu Ser Lys Arg Leu Gly Glu Leu Arg Leu Asn Pro

1 5 10

<210> 2756

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2756

Asn Gly Gln Trp Ser Leu Val Gly Ala Lys Ala Pro Pro Ala

1 5 10

<210> 2757

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2757

Asn Thr Leu Ala Gly Ser Gly Leu Phe Arg Met Asn Val Phe

1 5 10

<210> 2758

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2758

Pro Ala Gly Arg Glu Leu Ser Ala Ala Ala Asn Ala Ala Val

1 5 10

<210> 2759

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2759

Pro Ala Pro Gln Pro Pro Ala Gly Arg Glu Leu Ser Ala Ala

1 5 10

<210> 2760

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2760

Pro Gly Pro Gln Pro Pro Gln Pro Pro Gln Pro Gln Pro Glu

1 5 10

<210> 2761

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2761

Pro Gly Thr Ser Ile Gly Pro Leu Asp Val Ala Leu Ala Ser

1 5 10

<210> 2762

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2762

Pro Leu Gly Ser Ala Ala Thr Phe Thr Leu Ala Asn Lys Asp

1 5 10

<210> 2763

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2763

Pro Gln Pro Glu Ala Pro Ala Pro Gln Pro Pro Ala Gly Arg

1 5 10

<210> 2764

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2764

Pro Gln Pro Pro Gln Pro Gln Pro Glu Ala Pro Ala Pro Gln

1 5 10

<210> 2765

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2765

Pro Tyr Ile Lys Ala Ser Val Leu Gln Glu Phe Asp Gly Ala

1 5 10

<210> 2766

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2766

Arg Phe Ala Pro Gln Ala Ala Pro Leu Ser Ile Thr Leu Gln

1 5 10

<210> 2767

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2767

Arg Leu Gly Glu Leu Arg Leu Asn Pro Asp Ala Gly Gly Ala

1 5 10

<210> 2768

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2768

Arg Leu Asn Pro Asp Ala Gly Gly Ala Trp Gly Arg Gly Phe

1 5 10

<210> 2769

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2769

Arg Asn Gly Ser Val Thr Ser Ser Gly Gln Leu Ser Asp Asp

1 5 10

<210> 2770

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2770

Arg Arg Phe Asp Gln Lys Val Ala Gly Phe Glu Leu Gly Ala

1 5 10

<210> 2771

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2771

Arg Thr Thr Leu Ala Met Ala Leu Gly Ala Leu Gly Ala Ala

1 5 10

<210> 2772

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2772

Ser Ala Ile Val Asp Gly Gly Leu His Ile Gly Ala Leu Gln

1 5 10

<210> 2773

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2773

Ser Ala Asn Thr Leu Leu Leu Val Gln Thr Pro Leu Gly Ser

1 5 10

<210> 2774

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2774

Ser Asp Pro Gly Gly Val Arg Thr Ala Ser Gly Thr Thr Ile

1 5 10

<210> 2775

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2775

Ser Glu Leu Thr Leu Asp Gly Gly His Ile Thr Gly Gly Arg

1 5 10

<210> 2776

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2776

Ser Gly Leu Phe Arg Met Asn Val Phe

1 5

<210> 2777

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2777

Ser Gly Leu Phe Arg Met Asn Val Phe Ala Asp Leu Gly Leu

1 5 10

<210> 2778

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2778

Ser Gly Ser Ser Val Glu Leu Ala Gln Ser Ile Val Glu Ala

1 5 10

<210> 2779

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2779

Ser Ile Ala Asp Ser Thr Leu Gln Gly Ala Gly Gly Val Gln

1 5 10

<210> 2780

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2780

Ser Lys Gly Pro Lys Leu Ala Met Pro Trp Thr Phe His Ala

1 5 10

<210> 2781

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2781

Ser Asn Val Gly Ala Leu Arg Leu Ala Ser Asp Gly Ser Val

1 5 10

<210> 2782

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2782

Ser Arg Ile Val Lys Ala Ala Pro Leu Arg Arg Thr Thr Leu

1 5 10

<210> 2783

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2783

Ser Val Leu Gly Ala Ser Glu Leu Thr Leu Asp Gly Gly His

1 5 10

<210> 2784

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2784

Ser Val Leu Gln Glu Phe Asp Gly Ala

1 5

<210> 2785

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2785

Ser Val Leu Gln Glu Phe Asp Gly Ala Gly Thr Val His Thr

1 5 10

<210> 2786

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2786

Thr Glu Leu Arg

1

<210> 2787

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2787

Thr Glu Leu Arg Gly Thr Arg Ala Glu Leu Gly Leu Gly Met

1 5 10

<210> 2788

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2788

Thr Gly Asp Gly Gly Gly His Thr Asp Ser Val His Val Gly

1 5 10

<210> 2789

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2789

Thr Gly Gly Ala Arg Arg Phe Ala Pro Gln Ala Ala Pro Leu

1 5 10

<210> 2790

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2790

Thr Gly Gly Arg Ala Ala Gly Val Ala Ala Met Gln Gly Ala

1 5 10

<210> 2791

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2791

Thr Gly Gly Val Gly Leu Ala Ser Thr Leu Trp Tyr Ala Glu

1 5 10

<210> 2792

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2792

Thr His Gly Val Gly Ala Ser Leu Glu Ala Gly Arg Arg Phe

1 5 10

<210> 2793

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2793

Thr Ile Arg Arg Gly Asp Ala Pro Ala

1 5

<210> 2794

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2794

Thr Leu Ala Asn Val Gly Asp Thr Trp Asp Asp Asp Gly Ile

1 5 10

<210> 2795

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2795

Thr Leu Gln Gly Ala Gly Gly Val Gln Ile Glu Arg Gly Ala

1 5 10

<210> 2796

<211> 14

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2796

Thr Leu Thr Gly Gly Ala Asp Ala Gln Gly Asp Ile Val Ala

1 5 10

<210> 2797

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2797

Thr Asn Val Thr Ala Val Pro Ala Ser Gly Ala Pro Ala Ala

1 5 10

<210> 2798

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400>

2798

Thr Arg Ala Glu Leu Gly Leu Gly Met Ala Ala Ala Leu Gly

1 5 10

<210> 2799

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2799

Thr Ser Ser Gly Gln Leu Ser Asp Asp Gly Ile Arg Arg Phe

1 5 10

<210> 2800

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2800

Thr Val His Thr Asn Gly Ile Ala His Arg Thr Glu Leu Arg

1 5 10

<210> 2801

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2801

Thr Tyr Arg Tyr Arg Leu Ala Ala Asn Gly Asn Gly Gln Trp

1 5 10

<210> 2802

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2802

Val Ala Asp His Ala Thr Leu Ala Asn Val Gly Asp Thr Trp

1 5 10

<210> 2803

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2803

Val His Val Gly Gly Tyr Ala Thr Tyr Ile Ala Asp Ser Gly

1 5 10

<210> 2804

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2804

Val Leu Asp Gly Trp Tyr Gly Val Asp

1 5

<210> 2805

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2805

Val Pro Ala Ser Gly Ala Pro Ala Ala Val Ser Val Leu Gly

1 5 10

<210> 2806

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2806

Val Arg Asp Glu Gly Gly Ser Ser Val Leu Gly Arg Leu Gly

1 5 10

<210> 2807

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2807

Val Arg Thr Ala Ser Gly Thr Thr Ile Lys Val Ser Gly Arg

1 5 10

<210> 2808

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2808

Val Ser Gly Gly Ser Leu Ser Ala Pro His Gly Asn Val Ile

1 5 10

<210> 2809

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2809

Val Ser Gly Arg Gln Ala Gln Gly Ile Leu Leu Glu Asn Pro

1 5 10

<210> 2810

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2810

Val Thr Val Gln Arg Ser Ala Ile Val Asp Gly Gly Leu His

1 5 10

<210> 2811

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2811

Val Val Leu Arg Asp Thr Asn Val Thr Ala Val Pro Ala Ser

1 5 10

<210> 2812

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2812

Trp Asn Asn Gln Ser Ile Val Lys Thr Gly Glu Arg Gln His

1 5 10

<210> 2813

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2813

Trp Val Arg Asn Ser Gly Ser Glu Pro Ala Ser Ala Asn Thr

1 5 10

<210> 2814

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2814

Trp Tyr Ala Glu Ser Asn Ala Leu Ser Lys Arg Leu Gly Glu

1 5 10

<210> 2815

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2815

Tyr Ala Thr Tyr Ile Ala Asp Ser Gly Phe Tyr Leu Asp Ala

1 5 10

<210> 2816

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2816

Tyr Gly Val Asp Val Ser Gly Ser Ser

1 5

<210> 2817

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2817

Tyr Gly Val Asp Val Ser Gly Ser Ser Val Glu Leu Ala Gln

1 5 10

<210> 2818

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2818

Tyr Leu Asp Ala Thr Leu Arg Ala Ser Arg Leu Glu Asn Asp

1 5 10

<210> 2819

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2819

Tyr Arg Val Leu Pro Glu Pro Val Lys Leu Thr Leu Thr Gly

1 5 10

<210> 2820

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2820

Val Lys Ala Gln Asn Ile Thr Asn Lys Arg Ala Ala Leu Ile Glu Ala

1 5 10 15

<210> 2821

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2821

Tyr Tyr Ser Asn Val Thr Ala Thr Arg Leu Leu Ser Ser Thr Asn Ser

1 5 10 15

<210> 2822

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2822

Ser Pro Asn Leu Thr Asp Glu Arg Ala Ala Gln Ala Gly Val Thr

1 5 10 15

<210> 2823

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2823

Ser Ser Arg Ala Ser Asp Glu Arg Ala Ala His Leu Pro Thr Ser

1 5 10 15

<210> 2824

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2824

Gln Val Val His Asn Ser Tyr Asn Arg Pro Ala Tyr Ser Pro Gly

1 5 10 15

<210> 2825

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2825

Ala Glu Gly Gly Glu Ile His Glu Leu

1 5

<210> 2826

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2826

Ala Glu Asn Leu Ile Ser Asn Gly Ile Gly Lys Tyr

1 5 10

<210> 2827

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2827

Ala Glu Val Asp Gly Asp Val Lys Leu

1 5

<210> 2828

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2828

Ala Ile Tyr Thr Ala Glu Ile His Lys

1 5

<210> 2829

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2829

Ala Pro Val Phe His Met Thr Asn Tyr

1 5

<210> 2830

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2830

Ala Pro Val Phe His Met Thr Asn Tyr Leu Glu Gln Pro Val Ser Asn

1 5 10 15

<210> 2831

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2831

Ala Gln Arg Leu Asn Glu Ile Tyr

1 5

<210> 2832

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2832

Ala Arg Val Pro His Ala Tyr Ser Leu

1 5

<210> 2833

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2833

Ala Val Arg Asp Leu Glu Arg Ala Met

1 5

<210> 2834

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2834

Ala Val Arg Asp Leu Glu Arg Ala Met Thr Thr Leu Lys

1 5 10

<210> 2835

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223>

Synthetic

<400> 2835

Asp Ala Leu Leu Arg Leu Gln Ala Met

1 5

<210> 2836

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2836

Asp Ile Lys Glu Lys Val Ile Asn Leu

1 5

<210> 2837

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2837

Asp Gln Gly Leu Phe Lys Val Leu

1 5

<210> 2838

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2838

Asp Thr Gly Val Asp Thr Arg Ile Trp

1 5

<210> 2839

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2839

Glu Pro Ile Gly Ser Leu Ala Ile Glu Glu Ala Met

1 5 10

<210> 2840

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2840

Glu Pro Ile Arg Asp Ala Leu Asn Ala Met

1 5 10

<210> 2841

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2841

Phe Pro Lys Leu Gly Lys Thr Leu

1 5

<210> 2842

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2842

Phe Arg Ser Val Asn Ala Val Ala Phe

1 5

<210> 2843

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2843

Gly Lys Ile Ile Asp Asn Thr Glu Gln Leu

1 5 10

<210> 2844

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2844

Gly Leu Asn Glu Lys Leu Val Phe Tyr

1 5

<210> 2845

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2845

Gly Met Tyr Gly Gly Thr Tyr Leu Val Glu Lys

1 5 10

<210> 2846

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2846

Gly Pro Pro Ile Ser Leu Glu Arg Leu Asp Val Gly Thr Asn

1 5 10

<210> 2847

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2847

Gly Pro Arg Gln Ala Gln Val Ser Phe Leu

1 5 10

<210> 2848

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2848

Gly Arg Leu Val Pro Gln Val Arg Val Ile Asp

1 5 10

<210> 2849

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2849

Gly Ser Ala Pro Ile Ser Met Gly Phe Arg

1 5 10

<210> 2850

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2850

His Ile Leu Ala Lys Ser Thr Ala Leu

1 5

<210> 2851

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2851

His Tyr Arg Glu Val Asn Leu Val Tyr

1 5

<210> 2852

<211> 9

<

212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2852

Ile Pro Pro Met Lys Asn Leu Ala Leu

1 5

<210> 2853

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2853

Ile Pro Tyr Gln Gly Ser Gly Lys Gly Val Ser Phe

1 5 10

<210> 2854

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2854

Ile Ser Lys Glu Ser Gln His Val Tyr

1 5

<210> 2855

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2855

Ile Val Ser Ser His Phe Phe Val Tyr

1 5

<210> 2856

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2856

Lys Glu Ile Lys Glu Thr Gly Arg Leu Phe

1 5 10

<210> 2857

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2857

Lys Glu Ser Gln His Val Tyr Tyr Leu

1 5

<210> 2858

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2858

Lys Ile Ile Asp Asn Thr Glu Gln Leu

1 5

<210> 2859

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2859

Lys Lys Gln Ile Asn Arg Gln Asn

1 5

<210> 2860

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2860

Lys Lys Val Asp Thr Asn Phe Ile Tyr Gln

1 5 10

<210> 2861

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2861

Lys Leu Ile Asp Gly Phe Phe Pro Ala

1 5

<210> 2862

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2862

Lys Pro Asn Leu Ser Ser Lys Arg Ser Glu Leu

1 5 10

<210> 2863

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2863

Lys Val Ser Pro Tyr Leu Phe Thr Val

1 5

<210> 2864

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2864

Leu Glu Thr Arg Thr Thr Asn Gln Phe Leu

1 5 10

<210> 2865

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2865

Leu Leu Lys Glu Ala Thr Glu Leu

1 5

<210> 2866

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2866

Leu Leu Lys Lys Gly Asn Ser Leu Tyr

1 5

<210> 2867

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2867

Leu Pro Ala Pro Ile Gly Gly Met Asn Tyr

1 5 10

<210> 2868

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2868

Met Pro Glu Glu Thr Leu His Gln Val Met

1 5 10

<210> 2869

<211

> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2869

Pro Thr Thr Ile Arg Gly Gln Phe Ser

1 5

<210> 2870

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2870

Gln Glu Ile Ser Arg His Gln Ala Leu Gly Tyr

1 5 10

<210> 2871

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2871

Arg Ile Thr His Val Asp Thr Glu Ser Tyr

1 5 10

<210> 2872

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2872

Arg Pro Gly Leu Lys Pro Asp Leu

1 5

<210> 2873

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2873

Arg Pro Ile Tyr Gly Leu Glu Val

1 5

<210> 2874

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2874

Arg Gln Ala Gly Gln Glu Met Ile Leu Ala Val

1 5 10

<210> 2875

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2875

Ser Ala Val Arg Ile Ala Thr Val Tyr

1 5

<210> 2876

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2876

Ser Leu Met Pro Glu Glu Thr Leu His Gln Val

1 5 10

<210> 2877

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2877

Ser Met Ile Asp Leu Val Thr Lys Phe

1 5

<210> 2878

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2878

Ser Met Leu Asn Ser Gln Ala Ile Asp Asn Leu Arg Ala

1 5 10

<210> 2879

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2879

Ser Met Tyr Arg Val Phe Glu Val

1 5

<210> 2880

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2880

Ser Gln Gln Gly Met Phe His Ala Tyr

1 5

<210> 2881

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2881

Thr Asp Thr Pro Ile Val Tyr Asn Asp Arg Asn Leu Leu Asp

1 5 10

<210> 2882

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2882

Val Ile Ile Asn Asp Asp Gln Gly Leu Phe Lys Val

1 5 10

<210> 2883

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2883

Tyr Glu Ser Gly Val Arg Ile Ala Ser Leu

1 5 10

<210> 2884

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2884

Tyr Leu Lys Asp Lys Ala Leu Ala

1 5

<210> 2885

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2885

Tyr Val Tyr Asp His Ser Gly Glu Ala Val Lys

1 5 10

<210> 2886

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2886

Ala Arg Val Ile Asp Pro Ala Ala Gln Ser Phe Thr Gly Val Val

1 5 10 15

<210> 2887

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2887

Ser Asp Arg Ala Ser Ala Arg Val Ile Asp Pro Ala Ala Gln Ser

1 5 10 15

<210

> 2888

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2888

Val Pro Pro Gly Lys Phe Val Thr Ala Ala Leu Leu Asn Thr Pro

1 5 10 15

<210> 2889

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2889

Trp Val Thr Pro Val Ile Gly Ser Gln Ala Arg Lys Cys Gly Leu

1 5 10 15

<210> 2890

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2890

ggcccagact gagcacgtga 20

<210> 2891

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2891

ggcccagact gagcacgtga 20

<210> 2892

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2892

ggcccagact gagcacgtga 20

<210> 2893

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2893

ggcccagact gagcacgtga 20

<210> 2894

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2894

ggcccagact gagcacgtga 20

<210> 2895

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2895

ggcccagact gagcacgtga 20

<210

> 2896

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2896

ggcccagact gagcacgtga 20

<210> 2897

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2897

ggcccagact gagcacgtga 20

<210> 2898

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2898

ggcccagact gagcacgtga 20

<210> 2899

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2899

ggcccagact gagcacgtga 20

<210> 2900

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2900

gagtccgagc agaagaagaa 20

<210> 2901

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2901

gagtccgagc agaagaagaa 20

<210> 2902

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2902

gagtccgagc agaagaagaa 20

<210> 2903

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2903

gagtccgagc agaagaagaa 20

<210> 2904

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2904

gagtccgagc agaagaagaa 20

<210> 2905

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2905

gagtccgagc agaagaagaa 20

<210> 2906

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2906

gagtccgagc agaagaagaa 20

<210> 2907

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2907

gagtccgagc agaagaagaa 20

<210> 2908

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2908

gagtccgagc agaagaagaa 20

<210> 2909

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2909

ggaatccctt ctgcagcacc 20

<210> 2910

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2910

ggaatccctt ctgcagcacc 20

<210> 2911

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2911

ggaatccctt ctgcagcacc 20

<210> 2912

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2912

ggaatccctt ctgcagcacc 20

<210> 2913

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2913

ggaatccctt ctgcagcacc 20

<210> 2914

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2914

ggaatccctt ctgcagcacc 20

<210> 2915

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2915

ggaatccctt ctgcagcacc 20

<210> 2916

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2916

ggaatccctt ctgcagcacc 20

<210> 2917

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2917

ggaatccctt ctgcagcacc 20

<210> 2918

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2918

ggaatccctt ctgcagcacc 20

<210> 2919

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2919

gtcatcttag tcattacctg 20

<210> 2920

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2920

gtcatcttag tcattacctg 20

<210> 2921

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2921

gtcatcttag tcattacctg 20

<210> 2922

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2922

gtcatcttag tcattacctg 20

<210> 2923

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2923

gtcatcttag tcattacctg 20

<210> 2924

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2924

gtcatcttag tcattacctg 20

<210> 2925

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2925

gtcatcttag tcattacctg 20

<210> 2926

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2926

gtcatcttag tcattacctg 20

<210> 2927

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2927

gtcatcttag tcattacctg 20

<210> 2928

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2928

ggcactgcgg ctggaggtgg 20

<210> 2929

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2929

ggcactgcgg ctggaggtgg 20

<210> 2930

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2930

ggcactgcgg ctggaggtgg 20

<210> 2931

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2931

ggcactgcgg ctggaggtgg 20

<210> 2932

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2932

ggcactgcgg ctggaggtgg 20

<210> 2933

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2933

ggcactgcgg ctggaggtgg 20

<210> 2934

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2934

ggcactgcgg ctggaggtgg 20

<210> 2935

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2935

ggcactgcgg ctggaggtgg 20

<210> 2936

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2936

ggcactgcgg ctggaggtgg 20

<210> 2937

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2937

ggcccagact gagcacgtga 20

<210> 2938

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2938

ggcccagact gagcacgtga 20

<210> 2939

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2939

ggcccagact gagcacgtga 20

<210> 2940

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2940

ggcccagact gagcacgtga 20

<210> 2941

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2941

ggcccagact gagcacgtga 20

<210> 2942

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2942

ggcccagact gagcacgtga 20

<210> 2943

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2943

ggcccagact gagcacgtga 20

<210> 2944

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2944

ggcccagact gagcacgtga 20

<210> 2945

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2945

ggcccagact gagcacgtga 20

<210> 2946

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2946

ggcccagact gagcacgtga 20

<210> 2947

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2947

ggcccagact gagcacgtga 20

<210> 2948

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2948

ggcccagact gagcacgtga 20

<210> 2949

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2949

ggcccagact gagcacgtga 20

<210> 2950

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2950

ggcccagact gagcacgtga 20

<210> 2951

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2951

gagtccgagc agaagaagaa 20

<210> 2952

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2952

gagtccgagc agaagaagaa 20

<210> 2953

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2953

gagtccgagc agaagaagaa 20

<210> 2954

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2954

gagtccgagc agaagaagaa 20

<210> 2955

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2955

gagtccgagc agaagaagaa 20

<210> 2956

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2956

gagtccgagc agaagaagaa 20

<210> 2957

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2957

gagtccgagc agaagaagaa 20

<210> 2958

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2958

gagtccgagc agaagaagaa 20

<210> 2959

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2959

gagtccgagc agaagaagaa 20

<210> 2960

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2960

gagtccgagc agaagaagaa 20

<210> 2961

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2961

gagtccgagc agaagaagaa 20

<210> 2962

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2962

ggaatccctt ctgcagcacc 20

<210> 2963

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2963

ggaatccctt ctgcagcacc 20

<210> 2964

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2964

ggaatccctt ctgcagcacc 20

<210> 2965

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2965

ggaatccctt ctgcagcacc 20

<210> 2966

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2966

ggaatccctt ctgcagcacc 20

<210> 2967

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2967

ggaatccctt ctgcagcacc 20

<210> 2968

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2968

ggaatccctt ctgcagcacc 20

<210> 2969

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2969

ggaatccctt ctgcagcacc 20

<210> 2970

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2970

ggaatccctt ctgcagcacc 20

<210> 2971

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2971

ggaatccctt ctgcagcacc 20

<210> 2972

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2972

ggaatccctt ctgcagcacc 20

<210> 2973

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2973

gtcatcttag tcattacctg 20

<210> 2974

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2974

gtcatcttag tcattacctg 20

<210> 2975

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2975

gtcatcttag tcattacctg 20

<210> 2976

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2976

gtcatcttag tcattacctg 20

<210> 2977

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2977

gtcatcttag tcattacctg 20

<210> 2978

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2978

gtcatcttag tcattacctg 20

<210> 2979

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2979

gtcatcttag tcattacctg 20

<210> 2980

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2980

gtcatcttag tcattacctg 20

<210> 2981

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2981

gtcatcttag tcattacctg 20

<210> 2982

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2982

gtcatcttag tcattacctg 20

<210> 2983

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2983

gtcatcttag tcattacctg 20

<210> 2984

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2984

ggcactgcgg ctggaggtgg 20

<210> 2985

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2985

ggcactgcgg ctggaggtgg 20

<210> 2986

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2986

ggcactgcgg ctggaggtgg 20

<210> 2987

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2987

ggcactgcgg ctggaggtgg 20

<210> 2988

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2988

ggcactgcgg ctggaggtgg 20

<210> 2989

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2989

ggcactgcgg ctggaggtgg 20

<210> 2990

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2990

ggcactgcgg ctggaggtgg 20

<210> 2991

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2991

ggcactgcgg ctggaggtgg 20

<210> 2992

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2992

ggcactgcgg ctggaggtgg 20

<210> 2993

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2993

ggcactgcgg ctggaggtgg 20

<210> 2994

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2994

ggcactgcgg ctggaggtgg 20

<210> 2995

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2995

ggcactgcgg ctggaggtgg 20

<210> 2996

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2996

ggcactgcgg ctggaggtgg 20

<210> 2997

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2997

tctgccatct cgtgctca 18

<210> 2998

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2998

tctgccatct cgtgctcag 19

<210> 2999

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 2999

tctgccatct cgtgctcagt 20

<210> 3000

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3000

tctgccatct cgtgctcagt c 21

<210> 3001

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3001

tctgccatct cgtgctcagt ct 22

<210> 3002

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3002

tctgccatct cgtgctcagt ctg 23

<210> 3003

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3003

tctgccatct cgtgctcagt ctgg 24

<210> 3004

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3004

tctgccatct cgtgctcagt ctggg 25

<210> 3005

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3005

tctgccatct cgtgctcagt ctgggc 26

<210> 3006

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3006

tctgccatct cgtgctcagt ctgggcc 27

<210> 3007

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3007

atgggagcac ttcttcttct gc 22

<210> 3008

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3008

atgggagcac ttcttcttct gct 23

<210> 3009

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3009

atgggagcac ttcttcttct gctc 24

<210> 3010

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3010

atgggagcac ttcttcttct gctcg 25

<210> 3011

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3011

atgggagcac ttcttcttct gctcgg 26

<210> 3012

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3012

atgggagcac ttcttcttct gctcgga 27

<210> 3013

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3013

atgggagcac ttcttcttct gctcggac 28

<210> 3014

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3014

atgggagcac ttcttcttct gctcggact 29

<210> 3015

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3015

atgggagcac ttcttcttct gctcggactc 30

<210> 3016

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3016

ggaaaagcga tcaaggtgct gcaga 25

<210> 3017

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3017

ggaaaagcga tcaaggtgct gcagaa 26

<210> 3018

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3018

ggaaaagcga tcaaggtgct gcagaag 27

<210> 3019

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3019

ggaaaagcga tcaaggtgct gcagaagg 28

<210> 3020

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3020

ggaaaagcga tcaaggtgct gcagaaggg 29

<210> 3021

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3021

ggaaaagcga tcaaggtgct gcagaaggga 30

<210> 3022

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3022

ggaaaagcga tcaaggtgct gcagaaggga t 31

<210> 3023

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3023

ggaaaagcga tcaaggtgct gcagaaggga tt 32

<210> 3024

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3024

ggaaaagcga tcaaggtgct gcagaaggga ttc 33

<210> 3025

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3025

ggaaaagcga tcaaggtgct gcagaaggga ttcc 34

<210> 3026

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3026

gaacacctca tgtaatgact 20

<210> 3027

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3027

gaacacctca tgtaatgact a 21

<210> 3028

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3028

gaacacctca tgtaatgact aa 22

<210> 3029

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3029

gaacacctca tgtaatgact aag 23

<210> 3030

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3030

gaacacctca tgtaatgact aaga 24

<210> 3031

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3031

gaacacctca tgtaatgact aagat 25

<210> 3032

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3032

gaacacctca tgtaatgact aagatg 26

<210> 3033

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3033

gaacacctca tgtaatgact aagatga 27

<210> 3034

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3034

gaacacctca tgtaatgact aagatgac 28

<210> 3035

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3035

gctttaaccc caacctccag 20

<210> 3036

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3036

gctttaaccc caacctccag c 21

<210> 3037

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3037

gctttaaccc caacctccag cc 22

<210> 3038

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3038

gctttaaccc caacctccag ccg 23

<210> 3039

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3039

gctttaaccc caacctccag ccgc 24

<210> 3040

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3040

gctttaaccc caacctccag ccgca 25

<210> 3041

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3041

gctttaaccc caacctccag ccgcag 26

<210> 3042

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3042

gctttaaccc caacctccag ccgcagt 27

<210> 3043

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3043

gctttaaccc caacctccag ccgcagtg 28

<210> 3044

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3044

tctgccatcc gtgctcagtc tg 22

<210> 3045

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3045

tctgccatca tcgtgctcag tctg 24

<210> 3046

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3046

tctgccatca aagcgtgctc agtctg 26

<210> 3047

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3047

tctgccatct cgtgctcagt ctg 23

<210> 3048

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3048

ctctgccatc tcgtgctcag tctg 24

<210> 3049

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3049

cctctgccat ctcgtgctca gtctg 25

<210> 3050

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3050

tcctctgcca tctcgtgctc agtctg 26

<210> 3051

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3051

ttcctctgcc atctcgtgct cagtctg 27

<210> 3052

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3052

tttcctctgc catctcgtgc tcagtctg 28

<210> 3053

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3053

ctttcctctg ccatctcgtg ctcagtctg 29

<210> 3054

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3054

cctttcctct gccatctcgt gctcagtctg 30

<210> 3055

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3055

tcctttcctc tgccatctcg tgctcagtct g 31

<210> 3056

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3056

ttcctttcct ctgccatctc gtgctcagtc tg 32

<210> 3057

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3057

cttcctttcc tctgccatct cgtgctcagt ctg 33

<210> 3058

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3058

ggagcccttg ttcttctgct cgg 23

<210> 3059

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3059

gggagccctt gttcttctgc tcgg 24

<210> 3060

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3060

tgggagccct tgttcttctg ctcgg 25

<210> 3061

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3061

atgggagccc ttgttcttct gctcgg 26

<210> 3062

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3062

gatgggagcc cttgttcttc tgctcgg 27

<210> 3063

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3063

tgatgggagc ccttgttctt ctgctcgg 28

<210> 3064

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3064

gtgatgggag cccttgttct tctgctcgg 29

<210> 3065

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3065

tgtgatggga gcccttgttc ttctgctcgg 30

<210> 3066

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3066

atgtgatggg agcccttgtt cttctgctcg g 31

<210> 3067

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3067

gatgtgatgg gagcccttgt tcttctgctc gg 32

<210> 3068

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3068

tgatgtgatg ggagcccttg ttcttctgct cgg 33

<210> 3069

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3069

cgatcaaggt gctgcagaag gga 23

<210> 3070

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3070

gcgatcaagg tgctgcagaa ggga 24

<210> 3071

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3071

agcgatcaag gtgctgcaga aggga 25

<210> 3072

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3072

aagcgatcaa ggtgctgcag aaggga 26

<210> 3073

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3073

aaagcgatca aggtgctgca gaaggga 27

<210> 3074

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3074

aaaagcgatc aaggtgctgc agaaggga 28

<210> 3075

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3075

gaaaagcgat caaggtgctg cagaaggga 29

<210> 3076

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3076

ggaaaagcga tcaaggtgct gcagaaggga 30

<210> 3077

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3077

cggaaaagcg atcaaggtgc tgcagaaggg a 31

<210> 3078

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3078

tcggaaaagc gatcaaggtg ctgcagaagg ga 32

<210> 3079

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3079

ctcggaaaag cgatcaaggt gctgcagaag gga 33

<210> 3080

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3080

aacacctcat gtaatgacta agatg 25

<210> 3081

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3081

gaacacctca tgtaatgact aagatg 26

<210> 3082

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3082

cgaacacctc atgtaatgac taagatg 27

<210> 3083

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3083

acgaacacct catgtaatga ctaagatg 28

<210> 3084

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3084

aacgaacacc tcatgtaatg actaagatg 29

<210> 3085

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3085

caacgaacac ctcatgtaat gactaagatg 30

<210> 3086

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3086

acaacgaaca cctcatgtaa tgactaagat g 31

<210> 3087

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3087

tacaacgaac acctcatgta atgactaaga tg 32

<210> 3088

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3088

ttacaacgaa cacctcatgt aatgactaag atg 33

<210> 3089

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3089

gttacaacga acacctcatg taatgactaa gatg 34

<210> 3090

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3090

agttacaacg aacacctcat gtaatgacta agatg 35

<210> 3091

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3091

accccaacct ccagccgc 18

<210> 3092

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3092

aaccccaacc tccagccgc 19

<210> 3093

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3093

taaccccaac ctccagccgc 20

<210> 3094

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3094

ttaaccccaa cctccagccg c 21

<210> 3095

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3095

tttaacccca acctccagcc gc 22

<210> 3096

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3096

ctttaacccc aacctccagc cgc 23

<210> 3097

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3097

gctttaaccc caacctccag ccgc 24

<210> 3098

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3098

cgctttaacc ccaacctcca gccgc 25

<210> 3099

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3099

ccgctttaac cccaacctcc agccgc 26

<210> 3100

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3100

tccgctttaa ccccaacctc cagccgc 27

<210> 3101

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3101

ctccgcttta accccaacct ccagccgc 28

<210> 3102

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3102

ctccgcttta accccaacct ccagccgc 28

<210> 3103

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3103

ctccgcttta accccaacct ccagccgc 28

<210> 3104

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3104

gtcatcttag tcattacctg 20

<210> 3105

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3105

gagtccgagc agaagaagaa 20

<210> 3106

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3106

ggaatccctt ctgcagcacc 20

<210> 3107

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3107

ggcccagact gagcacgtga 20

<210> 3108

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3108

ggcactgcgg ctggaggtgg 20

<210> 3109

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3109

gtcatcttag tcattacctg 20

<210> 3110

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3110

gtcatcttag tcattacctg 20

<210> 3111

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3111

ggaatccctt ctgcagcacc 20

<210> 3112

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3112

ggaatccctt ctgcagcacc 20

<210> 3113

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3113

aacgaacacc tcatgtaatg actaagatg 29

<210> 3114

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3114

atgggagcac ttcttcttct gctcggac 28

<210> 3115

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3115

ggaaaagcga tcaaggtgct gcagaaggga tt 32

<210> 3116

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3116

tctgccatga cgtgctcagt ctg 23

<210> 3117

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3117

ttaaccccaa cctccagcc 19

<210> 3118

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3118

aacgaacacc gcaggtaatg actaagatg 29

<210> 3119

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3119

aacgaacacc ccaggtaatg actaagatg 29

<210> 3120

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3120

ggaaaagcga tcaaggtgct gcagaaggga 30

<210> 3121

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3121

ggaaaagcga gccaggtgct gcagaaggga t 31

<210> 3122

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3122

gtcaaccatt aagcaaaaca t 21

<210> 3123

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3123

gtctcaggct gtgcagacaa a 21

<210> 3124

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3124

ggggcacaga tgagaaactc 20

<210> 3125

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3125

gccgtttgta ctttgtcctc 20

<210> 3126

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3126

gcgccaccgg ttgatgtgat 20

<210> 3127

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3127

gcttcgtggc aatgcgccac 20

<210> 3128

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3128

gacatcgatg tcctccccat 20

<210> 3129

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3129

gtggttgccc accctagtca t 21

<210> 3130

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3130

gcgactctct gcgtactgat 20

<210> 3131

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3131

gccctacttc cgctttcacc t 21

<210> 3132

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3132

ggattccatg aggtgcgcga 20

<210> 3133

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3133

gctgcagaag ggattccatg 20

<210> 3134

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3134

gcttgagacc gccagaagct 20

<210> 3135

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3135

ggggtcccag gtgctgacgt 20

<210> 3136

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3136

gcagaaatag actaattgca 20

<210> 3137

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3137

ggattgaccc aggccagggc 20

<210> 3138

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3138

gacgccctct ggaggaagca 20

<210> 3139

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3139

gctgtcctgc gacgccctc 19

<210> 3140

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3140

gcacatacta gcccctgtct 20

<210> 3141

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3141

gtcaaccagt atcccggtgc 20

<210> 3142

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3142

tcccttcctt ccacccagcc 20

<210> 3143

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3143

ccctgcctgt catcctgctt 20

<210> 3144

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3144

gcagtgccac cggggcgccg 20

<210> 3145

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3145

gcgggggctc agagagggca 20

<210> 3146

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3146

gagacacaca cacaggcctg g 21

<210> 3147

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3147

gtcaaccatt aagcaaaaca t 21

<210> 3148

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3148

gtgagttaca acgaacaccg c 21

<210> 3149

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3149

gtgagttaca acgaacaccc c 21

<210> 3150

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3150

ggggtcccag gtgctgacgt 20

<210> 3151

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3151

gaagctcgga aaagcgatca 20

<210> 3152

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3152

gaagctcgga aaagcgagcc 20

<210> 3153

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3153

gtcaaccagt atcccggtgc 20

<210> 3154

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3154

ggcccagact gagcacgtga 20

<210> 3155

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3155

ggcccagact gagcacgtga 20

<210> 3156

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3156

ggcccagact gagcacgtga 20

<210> 3157

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3157

ggcccagact gagcacgtga 20

<210> 3158

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3158

ggcccagact gagcacgtga 20

<210> 3159

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3159

ggcccagact gagcacgtga 20

<210> 3160

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3160

ggcccagact gagcacgtga 20

<210> 3161

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3161

ggcccagact gagcacgtga 20

<210> 3162

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3162

ggcccagact gagcacgtga 20

<210> 3163

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3163

ggcccagact gagcacgtga 20

<210> 3164

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3164

ggcccagact gagcacgtga 20

<210> 3165

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3165

ggcccagact gagcacgtga 20

<210> 3166

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3166

ggcccagact gagcacgtga 20

<210> 3167

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3167

ggcccagact gagcacgtga 20

<210> 3168

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3168

ggcccagact gagcacgtga 20

<210> 3169

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3169

ggcccagact gagcacgtga 20

<210> 3170

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3170

ggcccagact gagcacgtga 20

<210> 3171

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3171

ggcccagact gagcacgtga 20

<210> 3172

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3172

ggcccagact gagcacgtga 20

<210> 3173

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3173

ggcccagact gagcacgtga 20

<210> 3174

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3174

ggcccagact gagcacgtga 20

<210> 3175

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3175

ggcccagact gagcacgtga 20

<210> 3176

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3176

ggcccagact gagcacgtga 20

<210> 3177

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3177

ggcccagact gagcacgtga 20

<210> 3178

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3178

ggcccagact gagcacgtga 20

<210> 3179

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3179

ggcccagact gagcacgtga 20

<210> 3180

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3180

ggcccagact gagcacgtga 20

<210> 3181

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3181

ggcccagact gagcacgtga 20

<210> 3182

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3182

ggcccagact gagcacgtga 20

<210> 3183

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3183

ggcccagact gagcacgtga 20

<210> 3184

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3184

ggcccagact gagcacgtga 20

<210> 3185

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3185

ggcccagact gagcacgtga 20

<210> 3186

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3186

ggcccagact gagcacgtga 20

<210> 3187

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3187

ggcccagact gagcacgtga 20

<210> 3188

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3188

gtcatcttag tcattacctg 20

<210> 3189

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3189

gtcatcttag tcattacctg 20

<210> 3190

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3190

gtcatcttag tcattacctg 20

<210> 3191

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3191

gtcatcttag tcattacctg 20

<210> 3192

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3192

gtcatcttag tcattacctg 20

<210> 3193

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3193

gtcatcttag tcattacctg 20

<210> 3194

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3194

gtcatcttag tcattacctg 20

<210> 3195

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3195

gtcatcttag tcattacctg 20

<210> 3196

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3196

gtcatcttag tcattacctg 20

<210> 3197

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3197

gtcatcttag tcattacctg 20

<210> 3198

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3198

gtcatcttag tcattacctg 20

<210> 3199

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3199

gtcatcttag tcattacctg 20

<210> 3200

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3200

ggaatccctt ctgcagcacc 20

<210> 3201

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3201

ggaatccctt ctgcagcacc 20

<210> 3202

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3202

ggaatccctt ctgcagcacc 20

<210> 3203

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3203

ggaatccctt ctgcagcacc 20

<210> 3204

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3204

ggaatccctt ctgcagcacc 20

<210> 3205

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3205

ggaatccctt ctgcagcacc 20

<210> 3206

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3206

ggaatccctt ctgcagcacc 20

<210> 3207

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3207

ggaatccctt ctgcagcacc 20

<210> 3208

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3208

ggaatccctt ctgcagcacc 20

<210> 3209

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3209

ggaatccctt ctgcagcacc 20

<210> 3210

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3210

ggaatccctt ctgcagcacc 20

<210> 3211

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3211

ggaatccctt ctgcagcacc 20

<210> 3212

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3212

gagtccgagc agaagaagaa 20

<210> 3213

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3213

gagtccgagc agaagaagaa 20

<210> 3214

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3214

gagtccgagc agaagaagaa 20

<210> 3215

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3215

gagtccgagc agaagaagaa 20

<210> 3216

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3216

gagtccgagc agaagaagaa 20

<210> 3217

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3217

gagtccgagc agaagaagaa 20

<210> 3218

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3218

gagtccgagc agaagaagaa 20

<210> 3219

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3219

gagtccgagc agaagaagaa 20

<210> 3220

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3220

gagtccgagc agaagaagaa 20

<210> 3221

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3221

gagtccgagc agaagaagaa 20

<210> 3222

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3222

gagtccgagc agaagaagaa 20

<210> 3223

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3223

gagtccgagc agaagaagaa 20

<210> 3224

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3224

gcattttcag gaggaagcga 20

<210> 3225

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3225

gcattttcag gaggaagcga 20

<210> 3226

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3226

gcattttcag gaggaagcga 20

<210> 3227

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3227

gcattttcag gaggaagcga 20

<210> 3228

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3228

gcattttcag gaggaagcga 20

<210> 3229

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3229

gcattttcag gaggaagcga 20

<210> 3230

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3230

gcattttcag gaggaagcga 20

<210> 3231

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3231

gcattttcag gaggaagcga 20

<210> 3232

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3232

gcattttcag gaggaagcga 20

<210> 3233

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3233

gcattttcag gaggaagcga 20

<210> 3234

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3234

gcattttcag gaggaagcga 20

<210> 3235

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3235

gcattttcag gaggaagcga 20

<210> 3236

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3236

gatgtctgca ggccagatga 20

<210> 3237

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3237

gatgtctgca ggccagatga 20

<210> 3238

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3238

gatgtctgca ggccagatga 20

<210> 3239

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3239

gatgtctgca ggccagatga 20

<210> 3240

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3240

gatgtctgca ggccagatga 20

<210> 3241

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3241

gatgtctgca ggccagatga 20

<210> 3242

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3242

gatgtctgca ggccagatga 20

<210> 3243

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3243

gatgtctgca ggccagatga 20

<210> 3244

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3244

gatgtctgca ggccagatga 20

<210> 3245

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3245

gatgtctgca ggccagatga 20

<210> 3246

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3246

gatgtctgca ggccagatga 20

<210> 3247

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3247

gatgtctgca ggccagatga 20

<210> 3248

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3248

gattcctggt gccagaaaca 20

<210> 3249

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3249

gattcctggt gccagaaaca 20

<210> 3250

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3250

gattcctggt gccagaaaca 20

<210> 3251

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3251

gattcctggt gccagaaaca 20

<210> 3252

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3252

gattcctggt gccagaaaca 20

<210> 3253

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3253

gattcctggt gccagaaaca 20

<210> 3254

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3254

gattcctggt gccagaaaca 20

<210> 3255

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3255

gattcctggt gccagaaaca 20

<210> 3256

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3256

gattcctggt gccagaaaca 20

<210> 3257

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3257

gattcctggt gccagaaaca 20

<210> 3258

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3258

gattcctggt gccagaaaca 20

<210> 3259

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3259

gattcctggt gccagaaaca 20

<210> 3260

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3260

ggcccagact gagcacgtga 20

<210> 3261

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3261

ggcccagact gagcacgtga 20

<210> 3262

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3262

ggcccagact gagcacgtga 20

<210> 3263

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3263

ggcccagact gagcacgtga 20

<210> 3264

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3264

ggcccagact gagcacgtga 20

<210> 3265

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3265

ggcccagact gagcacgtga 20

<210> 3266

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3266

ggcccagact gagcacgtga 20

<210> 3267

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3267

ggcccagact gagcacgtga 20

<210> 3268

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3268

ggcccagact gagcacgtga 20

<210> 3269

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3269

ggcccagact gagcacgtga 20

<210> 3270

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3270

gtcatcttag tcattacctg 20

<210> 3271

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3271

gtcatcttag tcattacctg 20

<210> 3272

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3272

gtcatcttag tcattacctg 20

<210> 3273

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3273

gtcatcttag tcattacctg 20

<210> 3274

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3274

ggaatccctt ctgcagcacc 20

<210> 3275

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3275

ggaatccctt ctgcagcacc 20

<210> 3276

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3276

ggaatccctt ctgcagcacc 20

<210> 3277

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3277

ggaatccctt ctgcagcacc 20

<210> 3278

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3278

gagtccgagc agaagaagaa 20

<210> 3279

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3279

gagtccgagc agaagaagaa 20

<210> 3280

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3280

gagtccgagc agaagaagaa 20

<210> 3281

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3281

gagtccgagc agaagaagaa 20

<210> 3282

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3282

gcattttcag gaggaagcga 20

<210> 3283

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3283

gcattttcag gaggaagcga 20

<210> 3284

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3284

gcattttcag gaggaagcga 20

<210> 3285

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3285

gcattttcag gaggaagcga 20

<210> 3286

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3286

gatgtctgca ggccagatga 20

<210> 3287

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3287

gatgtctgca ggccagatga 20

<210> 3288

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3288

gatgtctgca ggccagatga 20

<210> 3289

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3289

gatgtctgca ggccagatga 20

<210> 3290

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3290

gattcctggt gccagaaaca 20

<210> 3291

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3291

gattcctggt gccagaaaca 20

<210> 3292

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3292

gattcctggt gccagaaaca 20

<210> 3293

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3293

gattcctggt gccagaaaca 20

<210> 3294

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3294

ggcccagact gagcacgtga 20

<210> 3295

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3295

ggcccagact gagcacgtga 20

<210> 3296

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3296

ggcccagact gagcacgtga 20

<210> 3297

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3297

ggcccagact gagcacgtga 20

<210> 3298

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3298

gtcatcttag tcattacctg 20

<210> 3299

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3299

gtcatcttag tcattacctg 20

<210> 3300

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3300

gtcatcttag tcattacctg 20

<210> 3301

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3301

gtcatcttag tcattacctg 20

<210> 3302

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3302

ggaatccctt ctgcagcacc 20

<210> 3303

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3303

ggaatccctt ctgcagcacc 20

<210> 3304

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3304

ggaatccctt ctgcagcacc 20

<210> 3305

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3305

tctgccatct cgtgctcagt ctg 23

<210> 3306

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3306

tctgccatcg cgtgctcagt ctg 23

<210> 3307

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3307

tctgccatcc cgtgctcagt ctg 23

<210> 3308

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3308

tctgccatta cgtgctcagt ctg 23

<210> 3309

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3309

tctgccatga cgtgctcagt ctg 23

<210> 3310

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3310

tctgccataa cgtgctcagt ctg 23

<210> 3311

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3311

tctgccagca cgtgctcagt ctg 23

<210> 3312

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3312

tctgccacca cgtgctcagt ctg 23

<210> 3313

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3313

tctgccaaca cgtgctcagt ctg 23

<210> 3314

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3314

tctgccttca cgtgctcagt ctg 23

<210> 3315

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3315

tctgccgtca cgtgctcagt ctg 23

<210> 3316

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3316

tctgccctca cgtgctcagt ctg 23

<210> 3317

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3317

tctgctatca cgtgctcagt ctg 23

<210> 3318

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3318

tctgcgatca cgtgctcagt ctg 23

<210> 3319

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3319

tctgcaatca cgtgctcagt ctg 23

<210> 3320

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3320

tctgtcatca cgtgctcagt ctg 23

<210> 3321

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3321

tctggcatca cgtgctcagt ctg 23

<210> 3322

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3322

tctgacatca cgtgctcagt ctg 23

<210> 3323

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3323

tcttccatca cgtgctcagt ctg 23

<210> 3324

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3324

tctcccatca cgtgctcagt ctg 23

<210> 3325

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3325

tctaccatca cgtgctcagt ctg 23

<210> 3326

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3326

tcggccatca cgtgctcagt ctg 23

<210> 3327

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3327

tccgccatca cgtgctcagt ctg 23

<210> 3328

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3328

tcagccatca cgtgctcagt ctg 23

<210> 3329

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3329

tggaggaagc agggcttcct ttcctctgcc atctcgtgct cagtctg 47

<210> 3330

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3330

tggaggaagc agggcttcct ttgctctgcc atcacgtgct cagtctg 47

<210> 3331

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3331

tggaggaagc agggcttcct atcctctgcc atcacgtgct cagtctg 47

<210> 3332

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3332

tggaggaagc agggcttgct ttcctctgcc atcacgtgct cagtctg 47

<210> 3333

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3333

tggaggaagc aggggttcct ttcctctgcc atcacgtgct cagtctg 47

<210> 3334

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3334

tggaggaagc acggcttcct ttcctctgcc atcacgtgct cagtctg 47

<210> 3335

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3335

tggaggaagc tgggcttcct ttcctctgcc atcacgtgct cagtctg 47

<210> 3336

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3336

tggaggaacc agggcttcct ttcctctgcc atcacgtgct cagtctg 47

<210> 3337

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3337

tggacgaagc agggcttcct ttcctctgcc atcacgtgct cagtctg 47

<210> 3338

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3338

tcgaggaagc agggcttcct ttcctctgcc atcacgtgct cagtctg 47

<210> 3339

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3339

aacgaacacc tcatgtaatg actaagatg 29

<210> 3340

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3340

aacgaacacc tcacgtaatg actaagatg 29

<210> 3341

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3341

aacgaacacc tcaagtaatg actaagatg 29

<210> 3342

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3342

aacgaacacc tctggtaatg actaagatg 29

<210> 3343

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3343

aacgaacacc tccggtaatg actaagatg 29

<210> 3344

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3344

aacgaacacc tgaggtaatg actaagatg 29

<210> 3345

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3345

aacgaacacc gcaggtaatg actaagatg 29

<210> 3346

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3346

aacgaacacc acaggtaatg actaagatg 29

<210> 3347

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3347

aacgaacacc ccaggtaatg actaagatg 29

<210> 3348

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3348

aacgaacaca tcaggtaatg actaagatg 29

<210> 3349

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3349

aacgaacatc tcaggtaatg actaagatg 29

<210> 3350

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3350

aacgaacgcc tcaggtaatg actaagatg 29

<210> 3351

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3351

ggaaaagcga tccaggcgct gcagaaggga t 31

<210> 3352

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3352

ggaaaagcga tccaggagct gcagaaggga t 31

<210> 3353

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3353

ggaaaagcga tccagttgct gcagaaggga t 31

<210> 3354

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3354

ggaaaagcga tccacgtgct gcagaaggga t 31

<210> 3355

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3355

ggaaaagcga tccaagtgct gcagaaggga t 31

<210> 3356

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3356

ggaaaagcga tcctggtgct gcagaaggga t 31

<210> 3357

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3357

ggaaaagcga tcccggtgct gcagaaggga t 31

<210> 3358

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3358

ggaaaagcga tctaggtgct gcagaaggga t 31

<210> 3359

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3359

ggaaaagcga tgcaggtgct gcagaaggga t 31

<210> 3360

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3360

ggaaaagcga gccaggtgct gcagaaggga t 31

<210> 3361

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3361

ggaaaagcgg tccaggtgct gcagaaggga t 31

<210> 3362

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3362

ggaaaagaga tccaggtgct gcagaaggga t 31

<210> 3363

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3363

gtgatgggag ccctgcttct tctgctcgga 30

<210> 3364

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3364

gtgatgggag ccctacttct tctgctcgga 30

<210> 3365

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3365

gtgatgggag cccctcttct tctgctcgga 30

<210> 3366

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3366

gtgatgggag ccgttcttct tctgctcgga 30

<210> 3367

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3367

gtgatgggag ctcttcttct tctgctcgga 30

<210> 3368

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3368

gtgatgggag cacttcttct tctgctcgga 30

<210> 3369

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3369

gtgatgggat cccttcttct tctgctcgga 30

<210> 3370

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3370

gtgatgggtg cccttcttct tctgctcgga 30

<210> 3371

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3371

gtgatggggg cccttcttct tctgctcgga 30

<210> 3372

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3372

gtgatgggcg cccttcttct tctgctcgga 30

<210> 3373

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3373

gtgatggcag cccttcttct tctgctcgga 30

<210> 3374

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3374

gtgatggaag cccttcttct tctgctcgga 30

<210> 3375

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3375

tgtctgaagc catctcttcc tcctgaaaat 30

<210> 3376

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3376

tgtctgaagc catcccttcc tcctgaaaat 30

<210> 3377

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3377

tgtctgaagc catcacttcc tcctgaaaat 30

<210> 3378

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3378

tgtctgaagc cattgcttcc tcctgaaaat 30

<210> 3379

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3379

tgtctgaagc cagcgcttcc tcctgaaaat 30

<210> 3380

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3380

tgtctgaagc caccgcttcc tcctgaaaat 30

<210> 3381

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3381

tgtctgaagc caacgcttcc tcctgaaaat 30

<210> 3382

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3382

tgtctgaagc cttcgcttcc tcctgaaaat 30

<210> 3383

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3383

tgtctgaagc cgtcgcttcc tcctgaaaat 30

<210> 3384

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3384

tgtctgaagc cctcgcttcc tcctgaaaat 30

<210> 3385

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3385

tgtctgaagc aatcgcttcc tcctgaaaat 30

<210> 3386

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3386

tgtctgaagg catcgcttcc tcctgaaaat 30

<210> 3387

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3387

aatgtgccat ctggagccct cttctggcct gcaga 35

<210> 3388

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3388

aatgtgccat ctggagccct cgtctggcct gcaga 35

<210> 3389

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3389

aatgtgccat ctggagccct cctctggcct gcaga 35

<210> 3390

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3390

aatgtgccat ctggagccct tatctggcct gcaga 35

<210> 3391

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3391

aatgtgccat ctggagcccg catctggcct gcaga 35

<210> 3392

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3392

aatgtgccat ctggagcccc catctggcct gcaga 35

<210> 3393

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3393

aatgtgccat ctggagccca catctggcct gcaga 35

<210> 3394

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3394

aatgtgccat ctggagcact catctggcct gcaga 35

<210> 3395

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3395

aatgtgccat ctggaggcct catctggcct gcaga 35

<210> 3396

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3396

aatgtgccat ctggatccct catctggcct gcaga 35

<210> 3397

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3397

aatgtgccat ctggaaccct catctggcct gcaga 35

<210> 3398

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3398

aatgtgccat ctgcagccct catctggcct gcaga 35

<210> 3399

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3399

gtcacccctg gttctggcac cagg 24

<210> 3400

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3400

gtcacccctg cttctggcac cagg 24

<210> 3401

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3401

gtcacccctt tttctggcac cagg 24

<210> 3402

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3402

gtcaccctc tttctggcac cagg 24

<210> 3403

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3403

gtcaccccta tttctggcac cagg 24

<210> 3404

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3404

gtcaccccag tttctggcac cagg 24

<210> 3405

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3405

gtcacccttg tttctggcac cagg 24

<210> 3406

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3406

gtcaccactg tttctggcac cagg 24

<210> 3407

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3407

gtcacgcctg tttctggcac cagg 24

<210> 3408

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3408

gccctcccgt ctcccctgtt tctggcacca gg 32

<210> 3409

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3409

gccctcccgt cgcccctgtt tctggcacca gg 32

<210> 3410

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3410

gccctcccgt ccccccctgtt tctggcacca gg 32

<210> 3411

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3411

tggaggaagc agggcttcct ttcctctgcc gtgctcagtc tg 42

<210> 3412

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3412

tggaggaagc agggcttcct ttcccgtgct cagtctg 37

<210> 3413

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3413

tggaggaagc agggcttccc gtgctcagtc tg 32

<210> 3414

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3414

tgtcctgcga cgccctctgg aggaagcgtg ctcagtctg 39

<210> 3415

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3415

tgtcctgcga cgccctctgg acgtgctcag tctg 34

<210> 3416

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3416

agtatcccgg tgcaggagct cgtgctcagt ctg 33

<210> 3417

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3417

tctgccatca tcgtgctcag tctg 24

<210> 3418

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3418

tctgccatca aagcgtgctc agtctg 26

<210> 3419

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3419

tctgccatcc gtgctcagtc tg 22

<210> 3420

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3420

tggaggaagc agggcttcct ttcctctgcc acgtgctcag tctg 44

<210> 3421

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3421

aacgaacacc tcagagtaat gactaagatg 30

<210> 3422

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3422

aacgaacacc tcagtacgta atgactaaga tg 32

<210> 3423

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3423

aacgaacacc caggtaatga ctaagatg 28

<210> 3424

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3424

aacgaacaca ggtaatgact aagatg 26

<210> 3425

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3425

ggaaaagcga tccagggtgc tgcagaaggg at 32

<210> 3426

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3426

ggaaaagcga tccaatcggt gctgcagaag ggat 34

<210> 3427

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3427

ggaaaagcga tcaggtgctg cagaagggat 30

<210> 3428

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3428

ggaaaagcga aggtgctgca gaagggat 28

<210> 3429

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3429

gtgatgggag caccttcttc ttctgctcgg a 31

<210> 3430

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3430

gtgatgggag cccttcgcat tcttctgctc gga 33

<210> 3431

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3431

gtgatgggag ccttcttctt ctgctcgga 29

<210> 3432

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3432

gtgatgggag ttcttcttct gctcgga 27

<210> 3433

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3433

tgtctgaagc catcggcttc ctcctgaaaa t 31

<210> 3434

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3434

tgtctgaagc catccatgct tcctcctgaa aat 33

<210> 3435

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3435

tgtctgaagc catgcttcct cctgaaaat 29

<210> 3436

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3436

tgtctgaagc cgcttcctcc tgaaaat 27

<210> 3437

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3437

aatgtgccat ctggagccgc tcatctggcc tgcaga 36

<210> 3438

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3438

aatgtgccat ctggagccct tgtcatctgg cctgcaga 38

<210> 3439

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3439

aatgtgccat ctggagcccc atctggcctg caga 34

<210> 3440

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3440

aatgtgccat ctggagccat ctggcctgca ga 32

<210> 3441

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3441

tcccgtcacc cgctgtttct ggcaccagg 29

<210> 3442

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3442

tcccgtcacc cctgtgattt ctggcaccag g 31

<210> 3443

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3443

tcccgtcacc ccgtttctgg caccagg 27

<210> 3444

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3444

tcccgtcacc gtttctggca ccagg 25

<210> 3445

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3445

tggaggaagc agggcttcct ttcctctgca tcaaagcgtg ctcagtctg 49

<210> 3446

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3446

tggaggaagc agggcttcct ttcctctgcc atgaaagcgt gctcagtctg 50

<210> 3447

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3447

tggaggaagc agggcttcct ttcctctgcg atccgtgctc agtctg 46

<210> 3448

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3448

tggaggaagc agggcttcct ttcctctgac atgacgtgct cagtctg 47

<210> 3449

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3449

aacgaacacc attggtaatg actaagatg 29

<210> 3450

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3450

aacgaacacg tcagtgtaat gactaagatg 30

<210> 3451

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3451

aacgaacaac tcgtaatgac taagatg 27

<210> 3452

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3452

aacgaacaca tcatgtaatg actaagatg 29

<210> 3453

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3453

ggaaaagcga tcggtagctg cagaagggat 30

<210> 3454

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3454

ggaaaagcga ttcaggtagc tgcagaaggg at 32

<210> 3455

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3455

ggaaaagcga tcaagtgctg cagaagggat 30

<210> 3456

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3456

gtcaaccagt atcccggtgc 20

<210> 3457

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3457

gtcaaccagt atcccggtgc 20

<210> 3458

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3458

gtcaaccatt aagcaaaaca t 21

<210> 3459

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3459

ggggtcccag gtgctgacgt 20

<210> 3460

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3460

gacatcgatg tcctccccat 20

<210> 3461

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3461

gatgaagcac tgtgggtacg a 21

<210> 3462

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3462

gatgtacaga gagcccaggg c 21

<210> 3463

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3463

gcccttcagc taaaataaag g 21

<210> 3464

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3464

ggcccagact gagcacgtga 20

<210> 3465

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3465

ggcccagact gagcacgtga 20

<210> 3466

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3466

ggcccagact gagcacgtga 20

<210> 3467

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3467

ggaatccctt ctgcagcacc 20

<210> 3468

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3468

ggaatccctt ctgcagcacc 20

<210> 3469

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3469

ggaatccctt ctgcagcacc 20

<210> 3470

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3470

gagtccgagc agaagaagaa 20

<210> 3471

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3471

gagtccgagc agaagaagaa 20

<210> 3472

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3472

gagtccgagc agaagaagaa 20

<210> 3473

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3473

gagtccgagc agaagaagaa 20

<210> 3474

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3474

gagtccgagc agaagaagaa 20

<210> 3475

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3475

gagtccgagc agaagaagaa 20

<210> 3476

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3476

gagtccgagc agaagaagaa 20

<210> 3477

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3477

ggcccagact gagcacgtga 20

<210> 3478

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3478

ggcccagact gagcacgtga 20

<210> 3479

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3479

tctgtcatca cgtgctcagt ctg 23

<210> 3480

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3480

tctgctatca cgtgctcagt ctg 23

<210> 3481

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3481

tctgccatta cgtgctcagt ctg 23

<210> 3482

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3482

ggaaaagtga tccaggtgct gcagaaggga t 31

<210> 3483

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3483

ggaaaagcga ttcaggtgct gcagaaggga t 31

<210> 3484

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3484

ggaaaagcga tctaggtgct gcagaaggga t 31

<210> 3485

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3485

gtgatgggag tccttcttct tctgctcgga 30

<210> 3486

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3486

gtgatgggag ctcttcttct tctgctcgga 30

<210> 3487

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3487

gtgatgggag cctttcttct tctgctcgga 30

<210> 3488

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3488

gtgatgggag ttcttcttct tctgctcgga 30

<210> 3489

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3489

gtgatgggag tctttcttct tctgctcgga 30

<210> 3490

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3490

gtgatgggag cttttcttct tctgctcgga 30

<210> 3491

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3491

gtgatgggag ttttttcttct tctgctcgga 30

<210> 3492

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3492

tctgccgtca cgtgctcagt ctg 23

<210> 3493

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3493

tctgccatcg cgtgctcagt ctg 23

<210> 3494

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3494

ggcccagact gagcacgtga 20

<210> 3495

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3495

ggcccagact gagcacgtga 20

<210> 3496

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3496

gagtccgagc agaagaagaa 20

<210> 3497

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3497

gagtccgagc agaagaagaa 20

<210> 3498

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3498

ggaatccctt ctgcagcacc 20

<210> 3499

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3499

ggaatccctt ctgcagcacc 20

<210> 3500

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3500

ggcccagact gagcacgtga 20

<210> 3501

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3501

ggcccagact gagcacgtga 20

<210> 3502

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3502

ggcccagact gagcacgtga 20

<210> 3503

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3503

ggcccagact gagcacgtga 20

<210> 3504

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3504

ggcccagact gagcacgtga 20

<210> 3505

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3505

ggcccagact gagcacgtga 20

<210> 3506

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3506

gcatggtgca cctgactcct g 21

<210> 3507

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3507

gcatggtgca cctgactcct g 21

<210> 3508

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3508

gcatggtgca cctgactcct g 21

<210> 3509

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3509

gcatggtgca cctgactcct g 21

<210> 3510

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3510

gcatggtgca cctgactcct g 21

<210> 3511

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3511

gcatggtgca cctgactcct g 21

<210> 3512

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3512

gtacctgaac cgtatatcct a 21

<210> 3513

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3513

gatccttcca gtcagggcca t 21

<210> 3514

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3514

gatccttcca gtcagggcca t 21

<210> 3515

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3515

gcagtggtgg ggggccttgg 20

<210> 3516

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3516

ggcccagact gagcacgtga 20

<210> 3517

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3517

ggcccagact gagcacgtga 20

<210> 3518

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3518

gtcatcttag tcattacctg 20

<210> 3519

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3519

gcatggtgca cctgactcct g 21

<210> 3520

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3520

ggcccagact gagcacgtga 20

<210> 3521

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3521

ggcccagact gagcacgtga 20

<210> 3522

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3522

tctgccatga cgtgctcagt ctg 23

<210> 3523

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3523

atgggagccc ttgttcttct gctcgg 26

<210> 3524

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3524

aaaagcgatc aaggtgctgc agaaggga 28

<210> 3525

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3525

tggaggaagc agggcttcct ttcctctgcc atcaatgatg gtgatgatgg tgcgtgctca 60

gtctg 65

<210> 3526

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3526

tctgcaatca cgtgctcagt ctg 23

<210> 3527

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3527

tctgccatca aagcgtgctc agtctg 26

<210> 3528

<211> 27

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 3528

agacttctcc acaggagtca ggtgcac 27

<210> 3529

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3529

agacttctcc tcaggagtca ggtgcac 27

<210> 3530

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3530

agacttctct tcaggagtca ggtgcac 27

<210> 3531

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3531

agtcagggcc ataggataga tatacggttc 30

<210> 3532

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3532

acctgaaccg tatatcctat ggccctgact g 31

<210> 3533

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3533

gtacctgaac cgtatatctt atggccctga ct 32

<210

> 3534

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3534

atgtagacgc caaggccccc cacc 24

<210> 3535

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3535

tctgccatcc cgtgctcagt ctg 23

<210> 3536

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3536

tctgccatca aagcgtgctc agtctg 26

<210> 3537

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3537

aacgaacacc tcacgtaatg actaagatg 29

<210> 3538

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3538

agacttctcc acaggagtca ggtgcac 27

<210> 3539

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3539

tggaggaagc agggcttcct ttcctctgcc atcaatgatg gtgatgatgg tgcgtgctca 60

gtctg 65

<210> 3540

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3540

tggaggaagc agggcttcct ttcctctgcc atcacttatc gtcgtcatcc ttgtaatccg 60

tgctcagtct g 71

<210> 3541

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3541

ggcccagact gagcacgtga 20

<210> 3542

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3542

ggcccagact gagcacgtga 20

<210> 3543

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3543

ggcccagact gagcacgtga 20

<210> 3544

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3544

ggcccagact gagcacgtga 20

<210> 3545

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3545

ggcccagact gagcacgtga 20

<210> 3546

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3546

gtcatcttag tcattacctg 20

<210> 3547

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3547

gagtccgagc agaagaagaa 20

<210> 3548

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3548

gtaacggcag acttctcctc 20

<210> 3549

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3549

tctgccatcc gtgctcagtc tg 22

<210> 3550

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3550

tctgccatca tcgtgctcag tctg 24

<210> 3551

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3551

tctgccatca aagcgtgctc agtctg 26

<210> 3552

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3552

tctgccatga cgtgctcagt ctg 23

<210> 3553

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3553

tggaggaagc agggcttcct ttcctctgcc atcacttatc gtcgtcatcc ttgtaatccg 60

tgctcagtct g 71

<210> 3554

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3554

aacgaacacc tcatgtaatg actaagatg 29

<210> 3555

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3555

atgggagccc ttgttcttct gctcgg 26

<210> 3556

<211> 26

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 3556

atctgactcc tgtggagaag tctgcc 26

<210> 3557

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3557

gatgtctgca ggccagatga 20

<210> 3558

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3558

gatgtctgca ggccagatga 20

<210> 3559

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3559

gatgtctgca ggccagatga 20

<210> 3560

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3560

gatgtctgca ggccagatga 20

<210> 3561

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3561

gatgtctgca ggccagatga 20

<210

> 3562

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3562

gatgtctgca ggccagatga 20

<210> 3563

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3563

gatgtctgca ggccagatga 20

<210> 3564

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3564

gatgtctgca ggccagatga 20

<210> 3565

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3565

gatgtctgca ggccagatga 20

<210> 3566

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3566

gatgtctgca ggccagatga 20

<210> 3567

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3567

gatgtctgca ggccagatga 20

<210> 3568

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3568

gatgtctgca ggccagatga 20

<210> 3569

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3569

gatgtctgca ggccagatga 20

<210> 3570

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3570

gatgtctgca ggccagatga 20

<210> 3571

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3571

gatgtctgca ggccagatga 20

<210> 3572

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3572

gatgtctgca ggccagatga 20

<210> 3573

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3573

gatgtctgca ggccagatga 20

<210> 3574

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3574

gatgtctgca ggccagatga 20

<210> 3575

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3575

gatgtctgca ggccagatga 20

<210> 3576

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3576

gatgtctgca ggccagatga 20

<210> 3577

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3577

gatgtctgca ggccagatga 20

<210> 3578

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3578

gatgtctgca ggccagatga 20

<210> 3579

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3579

gatgtctgca ggccagatga 20

<210> 3580

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3580

gatgtctgca ggccagatga 20

<210> 3581

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3581

gattcctggt gccagaaaca 20

<210> 3582

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3582

gattcctggt gccagaaaca 20

<210> 3583

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3583

gattcctggt gccagaaaca 20

<210> 3584

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3584

gattcctggt gccagaaaca 20

<210> 3585

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3585

gattcctggt gccagaaaca 20

<210> 3586

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3586

gattcctggt gccagaaaca 20

<210> 3587

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3587

gattcctggt gccagaaaca 20

<210> 3588

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3588

gattcctggt gccagaaaca 20

<210> 3589

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3589

gattcctggt gccagaaaca 20

<210> 3590

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3590

gattcctggt gccagaaaca 20

<210> 3591

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3591

gattcctggt gccagaaaca 20

<210> 3592

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3592

gattcctggt gccagaaaca 20

<210> 3593

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3593

gattcctggt gccagaaaca 20

<210> 3594

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3594

gattcctggt gccagaaaca 20

<210> 3595

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3595

gattcctggt gccagaaaca 20

<210> 3596

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3596

gattcctggt gccagaaaca 20

<210> 3597

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3597

gattcctggt gccagaaaca 20

<210> 3598

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3598

gattcctggt gccagaaaca 20

<210> 3599

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3599

gattcctggt gccagaaaca 20

<210> 3600

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3600

gattcctggt gccagaaaca 20

<210> 3601

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3601

gattcctggt gccagaaaca 20

<210> 3602

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3602

gattcctggt gccagaaaca 20

<210> 3603

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3603

gattcctggt gccagaaaca 20

<210> 3604

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3604

gattcctggt gccagaaaca 20

<210> 3605

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3605

gcattttcag gaggaagcga 20

<210> 3606

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3606

gcattttcag gaggaagcga 20

<210> 3607

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3607

gcattttcag gaggaagcga 20

<210> 3608

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3608

gcattttcag gaggaagcga 20

<210> 3609

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3609

gcattttcag gaggaagcga 20

<210> 3610

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3610

gcattttcag gaggaagcga 20

<210> 3611

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3611

gcattttcag gaggaagcga 20

<210> 3612

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3612

gcattttcag gaggaagcga 20

<210> 3613

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3613

gcattttcag gaggaagcga 20

<210> 3614

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3614

gcattttcag gaggaagcga 20

<210> 3615

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3615

gcattttcag gaggaagcga 20

<210> 3616

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3616

gcattttcag gaggaagcga 20

<210> 3617

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3617

gcattttcag gaggaagcga 20

<210> 3618

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3618

gcattttcag gaggaagcga 20

<210> 3619

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3619

gcattttcag gaggaagcga 20

<210> 3620

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3620

gcattttcag gaggaagcga 20

<210> 3621

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3621

gcattttcag gaggaagcga 20

<210> 3622

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3622

gcattttcag gaggaagcga 20

<210> 3623

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3623

gcattttcag gaggaagcga 20

<210> 3624

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3624

gcattttcag gaggaagcga 20

<210> 3625

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3625

gcattttcag gaggaagcga 20

<210> 3626

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3626

gcattttcag gaggaagcga 20

<210> 3627

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3627

gcattttcag gaggaagcga 20

<210> 3628

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3628

ccctctgaca atgtgccatc tggagcactc atctggcctg caga 44

<210> 3629

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3629

cctctgacaa tgtgccatct ggagcactca tctggcctgc aga 43

<210> 3630

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3630

ctctgacaat gtgccatctg gagcactcat ctggcctgca ga 42

<210> 3631

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3631

tctgacaatg tgccatctgg agcactcatc tggcctgcag a 41

<210> 3632

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3632

ctgacaatgt gccatctgga gcactcatct ggcctgcaga 40

<210> 3633

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3633

tgacaatgtg ccatctggag cactcatctg gcctgcaga 39

<210> 3634

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3634

gacaatgtgc catctggagc actcatctgg cctgcaga 38

<210> 3635

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3635

acaatgtgcc atctggagca ctcatctggc ctgcaga 37

<210> 3636

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3636

caatgtgcca tctggagcac tcatctggcc tgcaga 36

<210> 3637

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3637

aatgtgccat ctggagcact catctggcct gcaga 35

<210> 3638

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3638

atgtgccatc tggagcactc atctggcctg caga 34

<210> 3639

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3639

tgtgccatct ggagcactca tctggcctgc aga 33

<210> 3640

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3640

gtgccatctg gagcactcat ctggcctgca ga 32

<210> 3641

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3641

tgccatctgg agcactcatc tggcctgcag a 31

<210> 3642

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3642

gccatctgga gcactcatct ggcctgcaga 30

<210> 3643

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3643

ccatctggag cactcatctg gcctgcaga 29

<210> 3644

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3644

catctggagc actcatctgg cctgcaga 28

<210> 3645

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3645

atctggagca ctcatctggc ctgcaga 27

<210> 3646

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3646

tctggagcac tcatctggcc tgcaga 26

<210> 3647

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3647

ctggagcact catctggcct gcaga 25

<210> 3648

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3648

tggagcactc atctggcctg caga 24

<210> 3649

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3649

ggagcactca tctggcctgc aga 23

<210> 3650

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3650

gagcactcat ctggcctgca ga 22

<210> 3651

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3651

agcactcatc tggcctgcag a 21

<210> 3652

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3652

aggactagtt ctgccctccc gtcaccactg tttctggcac cagg 44

<210> 3653

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3653

ggactagttc tgccctcccg tcaccactgt ttctggcacc agg 43

<210> 3654

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3654

gactagttct gccctcccgt caccactgtt tctggcacca gg 42

<210> 3655

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3655

actagttctg ccctcccgtc accactgttt ctggcaccag g 41

<210> 3656

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3656

ctagttctgc cctcccgtca ccactgtttc tggcaccagg 40

<210> 3657

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3657

tagttctgcc ctcccgtcac cactgtttct ggcaccagg 39

<210> 3658

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3658

agttctgccc tcccgtcacc actgtttctg gcaccagg 38

<210> 3659

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3659

gttctgccct cccgtcacca ctgtttctgg caccagg 37

<210> 3660

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3660

ttctgccctc ccgtcaccac tgtttctggc accagg 36

<210> 3661

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3661

tctgccctcc cgtcaccact gtttctggca ccagg 35

<210> 3662

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3662

ctgccctccc gtcaccactg tttctggcac cagg 34

<210> 3663

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3663

tgccctcccg tcaccactgt ttctggcacc agg 33

<210> 3664

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3664

gccctcccgt caccactgtt tctggcacca gg 32

<210> 3665

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3665

ccctcccgtc accactgttt ctggcaccag g 31

<210> 3666

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3666

cctcccgtca ccactgtttc tggcaccagg 30

<210> 3667

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3667

ctcccgtcac cactgtttct ggcaccagg 29

<210> 3668

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3668

tcccgtcacc actgtttctg gcaccagg 28

<210> 3669

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3669

cccgtcacca ctgtttctgg caccagg 27

<210> 3670

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3670

ccgtcaccac tgtttctggc accagg 26

<210> 3671

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3671

cgtcaccact gtttctggca ccagg 25

<210> 3672

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3672

gtcaccactg tttctggcac cagg 24

<210> 3673

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3673

tcaccactgt ttctggcacc agg 23

<210> 3674

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3674

caccactgtt tctggcacca gg 22

<210> 3675

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3675

accactgttt ctggcaccag g 21

<210> 3676

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3676

aatgactcaa atatgctgtc tgaagcaatc gcttcctcct gaaaat 46

<210> 3677

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3677

atgactcaaa tatgctgtct gaagcaatcg cttcctcctg aaaat 45

<210> 3678

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3678

tgactcaaat atgctgtctg aagcaatcgc ttcctcctga aaat 44

<210> 3679

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3679

gactcaaata tgctgtctga agcaatcgct tcctcctgaa aat 43

<210> 3680

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3680

actcaaatat gctgtctgaa gcaatcgctt cctcctgaaa at 42

<210> 3681

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3681

ctcaaatatg ctgtctgaag caatcgcttc ctcctgaaaa t 41

<210> 3682

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3682

tcaaatatgc tgtctgaagc aatcgcttcc tcctgaaaat 40

<210> 3683

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3683

caaatatgct gtctgaagca atcgcttcct cctgaaaat 39

<210> 3684

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3684

aaatatgctg tctgaagcaa tcgcttcctc ctgaaaat 38

<210> 3685

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3685

aatatgctgt ctgaagcaat cgcttcctcc tgaaaat 37

<210> 3686

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3686

atatgctgtc tgaagcaatc gcttcctcct gaaaat 36

<210> 3687

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3687

tatgctgtct gaagcaatcg cttcctcctg aaaat 35

<210> 3688

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3688

atgctgtctg aagcaatcgc ttcctcctga aaat 34

<210> 3689

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3689

tgctgtctga agcaatcgct tcctcctgaa aat 33

<210> 3690

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3690

gctgtctgaa gcaatcgctt cctcctgaaa at 32

<210> 3691

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3691

ctgtctgaag caatcgcttc ctcctgaaaa t 31

<210> 3692

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3692

tgtctgaagc aatcgcttcc tcctgaaaat 30

<210> 3693

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3693

gtctgaagca atcgcttcct cctgaaaat 29

<210> 3694

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3694

tctgaagcaa tcgcttcctc ctgaaaat 28

<210> 3695

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3695

ctgaagcaat cgcttcctcc tgaaaat 27

<210> 3696

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3696

tgaagcaatc gcttcctcct gaaaat 26

<210> 3697

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3697

gaagcaatcg cttcctcctg aaaat 25

<210> 3698

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3698

aagcaatcgc ttcctcctga aaat 24

<210> 3699

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3699

gtaacggcag acttctccac 20

<210> 3700

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3700

gcatggtgca cctgactcct g 21

<210> 3701

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3701

gcatggtgca cctgactcct g 21

<210> 3702

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3702

gcatggtgca cctgactcct g 21

<210> 3703

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3703

gcatggtgca cctgactcct g 21

<210> 3704

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3704

gcatggtgca cctgactcct g 21

<210> 3705

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3705

gcatggtgca cctgactcct g 21

<210> 3706

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3706

gcatggtgca cctgactcct g 21

<210> 3707

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3707

gcatggtgca cctgactcct g 21

<210> 3708

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3708

gcatggtgca cctgactcct g 21

<210> 3709

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3709

gcatggtgca cctgactcct g 21

<210> 3710

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3710

gcatggtgca cctgactcct g 21

<210> 3711

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3711

gcatggtgca cctgactcct g 21

<210> 3712

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3712

gcatggtgca cctgactcct g 21

<210> 3713

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3713

atccttccag tcagggccat 20

<210> 3714

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3714

atccttccag tcagggccat 20

<210> 3715

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3715

atccttccag tcagggccat 20

<210> 3716

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3716

atccttccag tcagggccat 20

<210> 3717

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3717

atccttccag tcagggccat 20

<210> 3718

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3718

atccttccag tcagggccat 20

<210> 3719

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3719

atccttccag tcagggccat 20

<210> 3720

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3720

atccttccag tcagggccat 20

<210> 3721

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3721

atccttccag tcagggccat 20

<210> 3722

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3722

atccttccag tcagggccat 20

<210> 3723

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3723

atccttccag tcagggccat 20

<210> 3724

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3724

atccttccag tcagggccat 20

<210> 3725

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3725

atccttccag tcagggccat 20

<210> 3726

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3726

atccttccag tcagggccat 20

<210> 3727

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3727

atccttccag tcagggccat 20

<210> 3728

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3728

atccttccag tcagggccat 20

<210> 3729

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3729

atccttccag tcagggccat 20

<210> 3730

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3730

atccttccag tcagggccat 20

<210> 3731

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3731

atccttccag tcagggccat 20

<210> 3732

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3732

atccttccag tcagggccat 20

<210> 3733

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3733

atccttccag tcagggccat 20

<210> 3734

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3734

atccttccag tcagggccat 20

<210> 3735

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3735

atccttccag tcagggccat 20

<210> 3736

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3736

atccttccag tcagggccat 20

<210> 3737

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3737

atccttccag tcagggccat 20

<210> 3738

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3738

atccttccag tcagggccat 20

<210> 3739

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3739

atccttccag tcagggccat 20

<210> 3740

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3740

atccttccag tcagggccat 20

<210> 3741

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3741

atccttccag tcagggccat 20

<210> 3742

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3742

atccttccag tcagggccat 20

<210> 3743

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3743

atccttccag tcagggccat 20

<210> 3744

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3744

atccttccag tcagggccat 20

<210> 3745

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3745

atccttccag tcagggccat 20

<210> 3746

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3746

atccttccag tcagggccat 20

<210> 3747

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3747

atccttccag tcagggccat 20

<210> 3748

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3748

atccttccag tcagggccat 20

<210> 3749

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3749

atccttccag tcagggccat 20

<210> 3750

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3750

atccttccag tcagggccat 20

<210> 3751

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3751

atccttccag tcagggccat 20

<210> 3752

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3752

atccttccag tcagggccat 20

<210> 3753

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3753

atccttccag tcagggccat 20

<210> 3754

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3754

atccttccag tcagggccat 20

<210> 3755

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3755

agacttctcc tcaggagtca ggtgcac 27

<210> 3756

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3756

agacttctct tcaggagtca ggtgcac 27

<210> 3757

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3757

taacggcaga cttctcctca ggagtcaggt gcac 34

<210> 3758

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3758

acggcagact tctcctcagg agtcaggtgc ac 32

<210> 3759

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3759

ggcagacttc tcctcaggag tcaggtgcac 30

<210> 3760

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3760

gcagacttct cctcaggagt caggtgcac 29

<210> 3761

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3761

gacttctcct caggagtcag gtgcac 26

<210> 3762

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3762

acttctcctc aggagtcagg tgcac 25

<210> 3763

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3763

taacggcaga cttctcctca ggagtcaggt gca 33

<210> 3764

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3764

acggcagact tctcctcagg agtcaggtgc a 31

<210> 3765

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3765

ggcagacttc tcctcaggag tcaggtgca 29

<210> 3766

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3766

gcagacttct cctcaggagt caggtgca 28

<210> 3767

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3767

gacttctcct caggagtcag gtgca 25

<210> 3768

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3768

acttctcctc aggagtcagg tgca 24

<210> 3769

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3769

atatcttatg gccctgactg gaa 23

<210> 3770

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3770

tatatcttat ggccctgact ggaa 24

<210> 3771

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3771

gtatatctta tggccctgac tggaa 25

<210> 3772

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3772

accgtatatc ttatggccct gactggaa 28

<210> 3773

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3773

aaccgtatat cttatggccc tgactggaa 29

<210> 3774

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3774

gaaccgtata tcttatggcc ctgactggaa 30

<210> 3775

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3775

tgaaccgtat atcttatggc cctgactgga a 31

<210> 3776

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3776

atatcttatg gccctgact 19

<210> 3777

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3777

tatatcttat ggccctgact 20

<210> 3778

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3778

gtatatctta tggccctgac t 21

<210> 3779

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3779

accgtatatc ttatggccct gact 24

<210> 3780

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3780

aaccgtatat cttatggccc tgact 25

<210> 3781

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3781

gaaccgtata tcttatggcc ctgact 26

<210> 3782

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3782

tgaaccgtat atcttatggc cctgact 27

<210> 3783

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3783

tgaaccgtat atcttatggc cctgac 26

<210> 3784

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3784

tgaaccgtat atcttatggc cctgactg 28

<210> 3785

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3785

tgaaccgtat atcttatggc cctgactgg 29

<210> 3786

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3786

tgaaccgtat atcttatggc cctgactgga 30

<210> 3787

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3787

tgaaccgtat atcttatggc cctgactgga a 31

<210> 3788

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3788

tgaaccgtat atcttatggc cctgactgga ag 32

<210> 3789

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3789

tgaaccgtat atcttatggc cctgactgga agg 33

<210> 3790

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3790

acctgaaccg tatatcttat ggccctgact 30

<210> 3791

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3791

tacctgaacc gtatatctta tggccctgac t 31

<210> 3792

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3792

gtacctgaac cgtatatctt atggccctga ct 32

<210> 3793

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3793

ggtacctgaa ccgtatatct tatggccctg act 33

<210> 3794

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3794

tggtacctga accgtatatc ttatggccct gact 34

<210> 3795

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3795

acctgaaccg tatatcctat ggccctgact ggaa 34

<210> 3796

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3796

accgtatatc ctatggccct gactggaa 28

<210> 3797

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3797

acctgaaccg tatatcctat ggccctgact ggaagg 36

<210> 3798

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3798

acctgaaccg tatatcctat ggccctgact ggaag 35

<210> 3799

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3799

acctgaaccg tatatcctat ggccctgact gga 33

<210> 3800

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3800

acctgaaccg tatatcctat ggccctgact gg 32

<210> 3801

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3801

acctgaaccg tatatcctat ggccctgact g 31

<210> 3802

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3802

aaccgtatat cctatggccc tgactggaa 29

<210> 3803

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3803

tgaaccgtat atcctatggc cctgactgga a 31

<210> 3804

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3804

tacctgaacc gtatatccta tggccctgac tggaa 35

<210> 3805

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3805

tggtacctga accgtatatc ctatggccct gactggaa 38

<210> 3806

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3806

gtacctgaac cgtatatcct atggccctga ctggaa 36

<210> 3807

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3807

aaccgtatat cctatggccc tgactg 26

<210> 3808

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3808

tgaaccgtat atcctatggc cctgactg 28

<210> 3809

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3809

tacctgaacc gtatatccta tggccctgac tg 32

<210> 3810

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3810

tggtacctga accgtatatc ctatggccct gactg 35

<210> 3811

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3811

acactctttc cctacacgac gctcttccga tctnnnnatg tgggctgcct agaaagg 57

<210> 3812

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3812

tggagttcag acgtgtgctc ttccgatctc ccagccaaac ttgtcaacc 49

<210> 3813

<211> 57

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3813

acactctttc cctacacgac gctcttccga tctnnnnacg tctcatatgc cccttgg 57

<210> 3814

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3814

tggagttcag acgtgtgctc ttccgatcta cgtaggaatt ttggtgggac a 51

<210> 3815

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223>

n is a, c, g, or t

<400> 3815

acactctttc cctacacgac gctcttccga tctnnnngaa cccaggtagc cagagac 57

<210> 3816

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3816

tggagttcag acgtgtgctc ttccgatctt cctttcaacc cgaacggag 49

<210> 3817

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3817

acactctttc cctacacgac gctcttccga tctnnnncag ctcagcctga gtgttga 57

<210> 3818

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3818

tggagttcag acgtgtgctc ttccgatctc tcgtgggttt gtggttgc 48

<210> 3819

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3819

acactctttc cctacacgac gctcttccga tctnnnncat tgcagagagg cgtatca 57

<210> 3820

<211> 47

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 3820

tggagttcag acgtgtgctc ttccgatctg gggtcccagg tgctgac 47

<210> 3821

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3821

acactctttc cctacacgac gctcttccga tctnnnnagg gttggccaat ctactccc 58

<210> 3822

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3822

tggagttcag acgtgtgctc ttccgatctg tcttctctgt ctccacatgc c 51

<210> 3823

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3823

acactctttc cctacacgac gctcttccga tctnnnngtc agtggaacaa gccgagt 57

<210> 3824

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3824

tggagttcag acgtgtgctc ttccgatcta cttggttggg gtaacggtg 49

<210> 3825

<211> 57

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3825

acactctttc cctacacgac gctcttccga tctnnnncat acaggtgtgg cgagagg 57

<210> 3826

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3826

tggagttcag acgtgtgctc ttccgatctc cagcctcctt tggttagca 49

<210> 3827

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223>

n is a, c, g, or t

<400> 3827

acactctttc cctacacgac gctcttccga tctnnnntca caaacaagac agggaactg 59

<210> 3828

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3828

tggagttcag acgtgtgctc ttccgatcta gatgtagggc tagaggggtg 50

<210> 3829

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3829

acactctttc cctacacgac gctcttccga tctnnnnact tggtgccaaa ttcttctcc 59

<210> 3830

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3830

tggagttcag acgtgtgctc ttccgatcta aagagggaat gggctttgga 50

<210> 3831

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3831

acactctttc cctacacgac gctcttccga tctnnnncac aacagcttca tgtcagcc 58

<210> 3832

<211> 52

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 3832

tggagttcag acgtgtgctc ttccgatcta cgttaatgtt tcctgatggt cc 52

<210> 3833

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3833

acactctttc cctacacgac gctcttccga tctnnnntcc cctgttgacc tggagaa 57

<210> 3834

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3834

tggagttcag acgtgtgctc ttccgatctc actgtacttg ccctgacca 49

<210> 3835

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3835

acactctttc cctacacgac gctcttccga tctnnnnttg gtgttgacag ggagcaa 57

<210> 3836

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3836

tggagttcag acgtgtgctc ttccgatctc tgagatgtgg gcagaaggg 49

<210> 3837

<211> 57

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3837

acactctttc cctacacgac gctcttccga tctnnnntga gagggaacag aagggct 57

<210> 3838

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3838

tggagttcag acgtgtgctc ttccgatctg tccaaaggcc caagaacct 49

<210> 3839

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223>

n is a, c, g, or t

<400> 3839

acactctttc cctacacgac gctcttccga tctnnnntcc tagcactttg gaaggtcg 58

<210> 3840

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3840

tggagttcag acgtgtgctc ttccgatctg ctcatcttaa tctgctcagc c 51

<210> 3841

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3841

acactctttc cctacacgac gctcttccga tctnnnnggc atggcttctg agactca 57

<210> 3842

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3842

tggagttcag acgtgtgctc ttccgatctg tctcccttgc actccctgtc ttt 53

<210> 3843

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3843

acactctttc cctacacgac gctcttccga tctnnnnttt ggcaatggag gcattgg 57

<210> 3844

<211> 49

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 3844

tggagttcag acgtgtgctc ttccgatctg aagaggctgc ccatgagag 49

<210> 3845

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3845

acactctttc cctacacgac gctcttccga tctnnnnggt ctgaggctcg aatcctg 57

<210> 3846

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3846

tggagttcag acgtgtgctc ttccgatctc tgtggcctcc atatccctg 49

<210> 3847

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3847

acactctttc cctacacgac gctcttccga tctnnnnttt ccaccagaac tcagccc 57

<210> 3848

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3848

tggagttcag acgtgtgctc ttccgatctc ctcggttcct ccacaacac 49

<210> 3849

<211> 62

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3849

acactctttc cctacacgac gctcttccga tctnnnngtg gggagatttg catctgtgga 60

gg 62

<210> 3850

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3850

tggagttcag acgtgtgctc ttccgatctg cttttatacc atcttggggt tacag 55

<210> 3851

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3851

acactctttc cctacacgac gctcttccga tctnnnncaa tgtgcttcaa cccatcacgg 60

c 61

<210> 3852

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3852

tggagttcag acgtgtgctc ttccgatctc catgaatttg tgatggatgc agtctg 56

<210> 3853

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3853

acactctttc cctacacgac gctcttccga tctnnnngag aaggaggtgc aggagctaga 60

c 61

<210> 3854

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3854

tggagttcag acgtgtgctc ttccgatctc atcccgacct tcatccctcc tgg 53

<210> 3855

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3855

acactctttc cctacacgac gctcttccga tctnnnngta gttctgacat tcctcctgag 60

gg 62

<210> 3856

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3856

tggagttcag acgtgtgctc ttccgatctt caaacaaggt gcagatacag ca 52

<210> 3857

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3857

acactctttc cctacacgac gctcttccga tctnnnngcg ggcagtggcg tcttagtcg 59

<210> 3858

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3858

tggagttcag acgtgtgctc ttccgatctc cctgggtttg gttggctgct c 51

<210> 3859

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<

223> n is a, c, g, or t

<400> 3859

acactctttc cctacacgac gctcttccga tctnnnnctc cttgccgccc agccggtc 58

<210> 3860

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3860

tggagttcag acgtgtgctc ttccgatctc actggggaag aggcgaggac ac 52

<210> 3861

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3861

acactctttc cctacacgac gctcttccga tctnnnncca gtgtttccca tccccaacac 60

<210> 3862

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3862

tggagttcag acgtgtgctc ttccgatctg aatggatccc cccctagagc tc 52

<210> 3863

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 3863

acactctttc cctacacgac gctcttccga tctnnnncag gcccacaggt ccttctgga 59

<210> 3864

<211> 18

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 3864

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg

1 5 10 15

Lys Val

<210> 3865

<211> 236

<212> PRT

<213> Homo sapiens

<400> 3865

Met Ser Glu Ala Pro Arg Ala Glu Thr Phe Val Phe Leu Asp Leu Glu

1 5 10 15

Ala Thr Gly Leu Pro Ser Val Glu Pro Glu Ile Ala Glu Leu Ser Leu

20 25 30

Phe Ala Val His Arg Ser Ser Leu Glu Asn Pro Glu His Asp Glu Ser

35 40 45

Gly Ala Leu Val Leu Pro Arg Val Leu Asp Lys Leu Thr Leu Cys Met

50 55 60

Cys Pro Glu Arg Pro Phe Thr Ala Lys Ala Ser Glu Ile Thr Gly Leu

65 70 75 80

Ser Ser Glu Gly Leu Ala Arg Cys Arg Lys Ala Gly Phe Asp Gly Ala

85 90 95

Val Val Arg Thr Leu Gln Ala Phe Leu Ser Arg Gln Ala Gly Pro Ile

100 105 110

Cys Leu Val Ala His Asn Gly Phe Asp Tyr Asp Phe Pro Leu Leu Cys

115 120 125

Ala Glu Leu Arg Arg Leu Gly Ala Arg Leu Pro Arg Asp Thr Val Cys

130 135 140

Leu Asp Thr Leu Pro Ala Leu Arg Gly Leu Asp Arg Ala His Ser His

145 150 155 160

Gly Thr Arg Ala Arg Gly Arg Gln Gly Tyr Ser Leu Gly Ser Leu Phe

165 170 175

His Arg Tyr Phe Arg Ala Glu Pro Ser Ala Ala His Ser Ala Glu Gly

180 185 190

Asp Val His Thr Leu Leu Leu Ile Phe Leu His Arg Ala Ala Glu Leu
195 200 205
Leu Ala Trp Ala Asp Glu Gln Ala Arg Gly Trp Ala His Ile Glu Pro
210 215 220

Met Tyr Leu Pro Pro Asp Asp Pro Ser Leu Glu Ala
225 230 235
<210> 3866
<211> 236
<212> PRT
<213> Mus musculus
<400> 3866
Met Ser Glu Pro Pro Arg Ala Glu Thr Phe Val Phe Leu Asp Leu Glu
1 5 10 15
Ala Thr Gly Leu Pro Asn Met Asp Pro Glu Ile Ala Glu Ile Ser Leu
20 25 30
Phe Ala Val His Arg Ser Ser Leu Glu Asn Pro Glu Arg Asp Asp Ser
35 40 45

Gly Ser Leu Val Leu Pro Arg Val Leu Asp Lys Leu Thr Leu Cys Met
50 55 60
Cys Pro Glu Arg Pro Phe Thr Ala Lys Ala Ser Glu Ile Thr Gly Leu
65 70 75 80
Ser Ser Glu Ser Leu Met His Cys Gly Lys Ala Gly Phe Asn Gly Ala
85 90 95
Val Val Arg Thr Leu Gln Gly Phe Leu Ser Arg Gln Glu Gly Pro Ile
100 105 110

Cys Leu Val Ala His Asn Gly Phe Asp Tyr Asp Phe Pro Leu Leu Cys
115 120 125
Thr Glu Leu Gln Arg Leu Gly Ala His Leu Pro Gln Asp Thr Val Cys
130 135 140
Leu Asp Thr Leu Pro Ala Leu Arg Gly Leu Asp Arg Ala His Ser His
145 150 155 160
Gly Thr Arg Ala Gln Gly Arg Lys Ser Tyr Ser Leu Ala Ser Leu Phe

165 170 175

His Arg Tyr Phe Gln Ala Glu Pro Ser Ala Ala His Ser Ala Glu Gly
180 185 190

Asp Val His Thr Leu Leu Leu Ile Phe Leu His Arg Ala Pro Glu Leu
195 200 205

Leu Ala Trp Ala Asp Glu Gln Ala Arg Ser Trp Ala His Ile Glu Pro
210 215 220

Met Tyr Val Pro Pro Asp Gly Pro Ser Leu Glu Ala
225 230 235

<210> 3867

<211> 236

<212> PRT

<213> Rattus norvegicus

<400> 3867

Met Ser Glu Pro Leu Arg Ala Glu Thr Phe Val Phe Leu Asp Leu Glu
1 5 10 15

Ala Thr Gly Leu Pro Asn Met Asp Pro Glu Ile Ala Glu Ile Ser Leu
20 25 30

Phe Ala Val His Arg Ser Ser Leu Glu Asn Pro Glu Arg Asp Asp Ser
35 40 45

Gly Ser Leu Val Leu Pro Arg Val Leu Asp Lys Leu Thr Leu Cys Met

50 55 60

Cys Pro Glu Arg Pro Phe Thr Ala Lys Ala Ser Glu Ile Thr Gly Leu
65 70 75 80

Ser Ser Glu Gly Leu Met Asn Cys Arg Lys Ala Ala Phe Asn Asp Ala
85 90 95

Val Val Arg Thr Leu Gln Gly Phe Leu Ser Arg Gln Glu Gly Pro Ile
100 105 110

Cys Leu Val Ala His Asn Gly Phe Asp Tyr Asp Phe Pro Leu Leu Cys

115 120 125

Thr Glu Leu Gln Arg Leu Gly Ala His Leu Pro Arg Asp Thr Val Cys

130 135 140

Leu Asp Thr Leu Pro Ala Leu Arg Gly Leu Asp Arg Val His Ser His

145 150 155 160

Gly Thr Arg Ala Gln Gly Arg Lys Ser Tyr Ser Leu Ala Ser Leu Phe

165 170 175

His Arg Tyr Phe Gln Ala Glu Pro Ser Ala Ala His Ser Ala Glu Gly

180 185 190

Asp Val Asn Thr Leu Leu Leu Ile Phe Leu His Arg Ala Pro Glu Leu

195 200 205

Leu Ala Trp Ala Asp Glu Gln Ala Arg Ser Trp Ala His Ile Glu Pro

210 215 220

Met Tyr Val Pro Pro Asp Gly Pro Ser Leu Glu Ala

225 230 235

<210> 3868

<211> 803

<212> PRT

<213> Homo sapiens

<400> 3868

Met Gly Ile Gln Gly Leu Leu Gln Phe Ile Lys Glu Ala Ser Glu Pro

1 5 10 15

Ile His Val Arg Lys Tyr Lys Gly Gln Val Val Ala Val Asp Thr Tyr

20 25 30

Cys Trp Leu His Lys Gly Ala Ile Ala Cys Ala Glu Lys Leu Ala Lys

35 40 45

Gly Glu Pro Thr Asp Arg Tyr Val Gly Phe Cys Met Lys Phe Val Asn

50 55 60

Met Leu Leu Ser His Gly Ile Lys Pro Ile Leu Val Phe Asp Gly Cys

65 70 75 80

Thr Leu Pro Ser Lys Lys Glu Val Glu Arg Ser Arg Arg Glu Arg Arg

85 90 95

Gln Ala Asn Leu Leu Lys Gly Lys Gln Leu Leu Arg Glu Gly Lys Val

100 105 110

Ser Glu Ala Arg Glu Cys Phe Thr Arg Ser Ile Asn Ile Thr His Ala
115 120 125
Met Ala His Lys Val Ile Lys Ala Ala Arg Ser Gln Gly Val Asp Cys

130 135 140
Leu Val Ala Pro Tyr Glu Ala Asp Ala Gln Leu Ala Tyr Leu Asn Lys
145 150 155 160
Ala Gly Ile Val Gln Ala Ile Ile Thr Glu Asp Ser Asp Leu Leu Ala
165 170 175
Phe Gly Cys Lys Lys Val Ile Leu Lys Met Asp Gln Phe Gly Asn Gly
180 185 190
Leu Glu Ile Asp Gln Ala Arg Leu Gly Met Cys Arg Gln Leu Gly Asp

195 200 205
Val Phe Thr Glu Glu Lys Phe Arg Tyr Met Cys Ile Leu Ser Gly Cys
210 215 220
Asp Tyr Leu Ser Ser Leu Arg Gly Ile Gly Leu Ala Lys Ala Cys Lys
225 230 235 240
Val Leu Arg Leu Ala Asn Asn Pro Asp Ile Val Lys Val Ile Lys Lys
245 250 255
Ile Gly His Tyr Leu Lys Met Asn Ile Thr Val Pro Glu Asp Tyr Ile

260 265 270
Asn Gly Phe Ile Arg Ala Asn Asn Thr Phe Leu Tyr Gln Leu Val Phe
275 280 285
Asp Pro Ile Lys Arg Lys Leu Ile Pro Leu Asn Ala Tyr Glu Asp Asp
290 295 300
Val Asp Pro Glu Thr Leu Ser Tyr Ala Gly Gln Tyr Val Asp Asp Ser
305 310 315 320
Ile Ala Leu Gln Ile Ala Leu Gly Asn Lys Asp Ile Asn Thr Phe Glu

325 330 335
Gln Ile Asp Asp Tyr Asn Pro Asp Thr Ala Met Pro Ala His Ser Arg
340 345 350
Ser His Ser Trp Asp Asp Lys Thr Cys Gln Lys Ser Ala Asn Val Ser

355 360 365

Ser Ile Trp His Arg Asn Tyr Ser Pro Arg Pro Glu Ser Gly Thr Val

370 375 380

Ser Asp Ala Pro Gln Leu Lys Glu Asn Pro Ser Thr Val Gly Val Glu

385 390 395 400

Arg Val Ile Ser Thr Lys Gly Leu Asn Leu Pro Arg Lys Ser Ser Ile

405 410 415

Val Lys Arg Pro Arg Ser Ala Glu Leu Ser Glu Asp Asp Leu Leu Ser

420 425 430

Gln Tyr Ser Leu Ser Phe Thr Lys Lys Thr Lys Lys Asn Ser Ser Glu

435 440 445

Gly Asn Lys Ser Leu Ser Phe Ser Glu Val Phe Val Pro Asp Leu Val

450 455 460

Asn Gly Pro Thr Asn Lys Lys Ser Val Ser Thr Pro Pro Arg Thr Arg

465 470 475 480

Asn Lys Phe Ala Thr Phe Leu Gln Arg Lys Asn Glu Glu Ser Gly Ala

485 490 495

Val Val Val Pro Gly Thr Arg Ser Arg Phe Phe Cys Ser Ser Asp Ser

500 505 510

Thr Asp Cys Val Ser Asn Lys Val Ser Ile Gln Pro Leu Asp Glu Thr

515 520 525

Ala Val Thr Asp Lys Glu Asn Asn Leu His Glu Ser Glu Tyr Gly Asp

530 535 540

Gln Glu Gly Lys Arg Leu Val Asp Thr Asp Val Ala Arg Asn Ser Ser

545 550 555 560

Asp Asp Ile Pro Asn Asn His Ile Pro Gly Asp His Ile Pro Asp Lys

565 570 575

Ala Thr Val Phe Thr Asp Glu Glu Ser Tyr Ser Phe Glu Ser Ser Lys

580 585 590

Phe Thr Arg Thr Ile Ser Pro Pro Thr Leu Gly Thr Leu Arg Ser Cys

595 600 605

Phe Ser Trp Ser Gly Gly Leu Gly Asp Phe Ser Arg Thr Pro Ser Pro
610 615 620
Ser Pro Ser Thr Ala Leu Gln Gln Phe Arg Arg Lys Ser Asp Ser Pro
625 630 635 640
Thr Ser Leu Pro Glu Asn Asn Met Ser Asp Val Ser Gln Leu Lys Ser

645 650 655
Glu Glu Ser Ser Asp Asp Glu Ser His Pro Leu Arg Glu Glu Ala Cys
660 665 670
Ser Ser Gln Ser Gln Glu Ser Gly Glu Phe Ser Leu Gln Ser Ser Asn
675 680 685
Ala Ser Lys Leu Ser Gln Cys Ser Ser Lys Asp Ser Asp Ser Glu Glu
690 695 700
Ser Asp Cys Asn Ile Lys Leu Leu Asp Ser Gln Ser Asp Gln Thr Ser

705 710 715 720
Lys Leu Arg Leu Ser His Phe Ser Lys Lys Asp Thr Pro Leu Arg Asn
725 730 735
Lys Val Pro Gly Leu Tyr Lys Ser Ser Ser Ala Asp Ser Leu Ser Thr
740 745 750
Thr Lys Ile Lys Pro Leu Gly Pro Ala Arg Ala Ser Gly Leu Ser Lys
755 760 765
Lys Pro Ala Ser Ile Gln Lys Arg Lys His His Asn Ala Glu Asn Lys

770 775 780
Pro Gly Leu Gln Ile Lys Leu Asn Glu Leu Trp Lys Asn Phe Gly Phe
785 790 795 800
Lys Lys Phe

<210> 3869
<211> 846
<212> PRT
<213> Homo sapiens
<400> 3869
Met Gly Ile Gln Gly Leu Leu Gln Phe Ile Lys Glu Ala Ser Glu Pro

1 5 10 15

Ile His Val Arg Lys Tyr Lys Gly Gln Val Val Ala Val Asp Thr Tyr

20 25 30

Cys Trp Leu His Lys Gly Ala Ile Ala Cys Ala Glu Lys Leu Ala Lys

35 40 45

Gly Glu Pro Thr Asp Arg Tyr Val Gly Phe Cys Met Lys Phe Val Asn

50 55 60

Met Leu Leu Ser His Gly Ile Lys Pro Ile Leu Val Phe Asp Gly Cys

65 70 75 80

Thr Leu Pro Ser Lys Lys Glu Val Glu Arg Ser Arg Arg Glu Arg Arg

85 90 95

Gln Ala Asn Leu Leu Lys Gly Lys Gln Leu Leu Arg Glu Gly Lys Val

100 105 110

Ser Glu Ala Arg Glu Cys Phe Thr Arg Ser Ile Asn Ile Thr His Ala

115 120 125

Met Ala His Lys Val Ile Lys Ala Ala Arg Ser Gln Gly Val Asp Cys

130 135 140

Leu Val Ala Pro Tyr Glu Ala Asp Ala Gln Leu Ala Tyr Leu Asn Lys

145 150 155 160

Ala Gly Ile Val Gln Ala Ile Ile Thr Glu Asp Ser Asp Leu Leu Ala

165 170 175

Phe Gly Cys Lys Lys Val Ile Leu Lys Met Asp Gln Phe Gly Asn Gly

180 185 190

Leu Glu Ile Asp Gln Ala Arg Leu Gly Met Cys Arg Gln Leu Gly Asp

195 200 205

Val Phe Thr Glu Glu Lys Phe Arg Tyr Met Cys Ile Leu Ser Gly Cys

210 215 220

Asp Tyr Leu Ser Ser Leu Arg Gly Ile Gly Leu Ala Lys Ala Cys Lys

225 230 235 240

Val Leu Arg Leu Ala Asn Asn Pro Asp Ile Val Lys Val Ile Lys Lys

245 250 255

Ile Gly His Tyr Leu Lys Met Asn Ile Thr Val Pro Glu Asp Tyr Ile
260 265 270
Asn Gly Phe Ile Arg Ala Asn Asn Thr Phe Leu Tyr Gln Leu Val Phe
275 280 285

Asp Pro Ile Lys Arg Lys Leu Ile Pro Leu Asn Ala Tyr Glu Asp Asp
290 295 300
Val Asp Pro Glu Thr Leu Ser Tyr Ala Gly Gln Tyr Val Asp Asp Ser
305 310 315 320
Ile Ala Leu Gln Ile Ala Leu Gly Asn Lys Asp Ile Asn Thr Phe Glu
325 330 335
Gln Ile Asp Asp Tyr Asn Pro Asp Thr Ala Met Pro Ala His Ser Arg
340 345 350

Ser His Ser Trp Asp Asp Lys Thr Cys Gln Lys Ser Ala Asn Val Ser
355 360 365
Ser Ile Trp His Arg Asn Tyr Ser Pro Arg Pro Glu Ser Gly Thr Val
370 375 380
Ser Asp Ala Pro Gln Leu Lys Glu Asn Pro Ser Thr Val Gly Val Glu
385 390 395 400
Arg Val Ile Ser Thr Lys Gly Leu Asn Leu Pro Arg Lys Ser Ser Ile
405 410 415

Val Lys Arg Pro Arg Ser Ala Glu Leu Ser Glu Asp Asp Leu Leu Ser
420 425 430
Gln Tyr Ser Leu Ser Phe Thr Lys Lys Thr Lys Lys Asn Ser Ser Glu
435 440 445
Gly Asn Lys Ser Leu Ser Phe Ser Glu Val Phe Val Pro Asp Leu Val
450 455 460
Asn Gly Pro Thr Asn Lys Lys Ser Val Ser Thr Pro Pro Arg Thr Arg
465 470 475 480

Asn Lys Phe Ala Thr Phe Leu Gln Arg Lys Asn Glu Glu Ser Gly Ala
485 490 495
Val Val Val Pro Gly Thr Arg Ser Arg Phe Phe Cys Ser Ser Asp Ser

500 505 510

Thr Asp Cys Val Ser Asn Lys Val Ser Ile Gln Pro Leu Asp Glu Thr

515 520 525

Ala Val Thr Asp Lys Glu Asn Asn Leu His Glu Ser Glu Tyr Gly Asp

530 535 540

Gln Glu Gly Lys Arg Leu Val Asp Thr Asp Val Ala Arg Asn Ser Ser

545 550 555 560

Asp Asp Ile Pro Asn Asn His Ile Pro Gly Asp His Ile Pro Asp Lys

565 570 575

Ala Thr Val Phe Thr Asp Glu Glu Ser Tyr Ser Phe Glu Ser Ser Lys

580 585 590

Phe Thr Arg Thr Ile Ser Pro Pro Thr Leu Gly Thr Leu Arg Ser Cys

595 600 605

Phe Ser Trp Ser Gly Gly Leu Gly Asp Phe Ser Arg Thr Pro Ser Pro

610 615 620

Ser Pro Ser Thr Ala Leu Gln Gln Phe Arg Arg Lys Ser Asp Ser Pro

625 630 635 640

Thr Ser Leu Pro Glu Asn Asn Met Ser Asp Val Ser Gln Leu Lys Ser

645 650 655

Glu Glu Ser Ser Asp Asp Glu Ser His Pro Leu Arg Glu Glu Ala Cys

660 665 670

Ser Ser Gln Ser Gln Glu Ser Gly Glu Phe Ser Leu Gln Ser Ser Asn

675 680 685

Ala Ser Lys Leu Ser Gln Cys Ser Ser Lys Asp Ser Asp Ser Glu Glu

690 695 700

Ser Asp Cys Asn Ile Lys Leu Leu Asp Ser Gln Ser Asp Gln Thr Ser

705 710 715 720

Lys Leu Arg Leu Ser His Phe Ser Lys Lys Asp Thr Pro Leu Arg Asn

725 730 735

Lys Val Pro Gly Leu Tyr Lys Ser Ser Ser Ala Asp Ser Leu Ser Thr

740 745 750

Thr Lys Ile Lys Pro Leu Gly Pro Ala Arg Ala Ser Gly Leu Ser Lys
755 760 765

Lys Pro Ala Ser Ile Gln Lys Arg Lys His His Asn Ala Glu Asn Lys
770 775 780

Pro Gly Leu Gln Ile Lys Leu Asn Glu Leu Trp Lys Asn Phe Gly Phe
785 790 795 800

Lys Lys Asp Ser Glu Lys Leu Pro Pro Cys Lys Lys Pro Leu Ser Pro
805 810 815

Val Arg Asp Asn Ile Gln Leu Thr Pro Glu Ala Glu Glu Asp Ile Phe
820 825 830

Asn Lys Pro Glu Cys Gly Arg Val Gln Arg Ala Ile Phe Gln
835 840 845

<210> 3870

<211> 845

<212> PRT

<213> Homo sapiens

<400> 3870

Met Gly Ile Gln Gly Leu Leu Gln Phe Ile Lys Glu Ala Ser Glu Pro

1 5 10 15

Ile His Val Arg Lys Tyr Lys Gly Gln Val Val Ala Val Asp Thr Tyr
20 25 30

Cys Trp Leu His Lys Gly Ala Ile Ala Cys Ala Glu Lys Leu Ala Lys
35 40 45

Gly Glu Pro Thr Asp Arg Tyr Val Gly Phe Cys Met Lys Phe Val Asn
50 55 60

Met Leu Leu Ser His Gly Ile Lys Pro Ile Leu Val Phe Asp Gly Cys

65 70 75 80

Thr Leu Pro Ser Lys Lys Glu Val Glu Arg Ser Arg Arg Glu Arg Arg
85 90 95

Gln Ala Asn Leu Leu Lys Gly Lys Gln Leu Leu Arg Glu Gly Lys Val
100 105 110

Ser Glu Ala Arg Glu Cys Phe Thr Arg Ser Ile Asn Ile Thr His Ala

115 120 125

Met Ala His Lys Val Ile Lys Ala Ala Arg Ser Gln Gly Val Asp Cys

130 135 140

Leu Val Ala Pro Tyr Glu Ala Asp Ala Gln Leu Ala Tyr Leu Asn Lys

145 150 155 160

Ala Gly Ile Val Gln Ala Ile Ile Thr Glu Asp Ser Asp Leu Leu Ala

165 170 175

Phe Gly Cys Lys Lys Val Ile Leu Lys Met Asp Gln Phe Gly Asn Gly

180 185 190

Leu Glu Ile Asp Gln Ala Arg Leu Gly Met Cys Arg Gln Leu Gly Asp

195 200 205

Val Phe Thr Glu Glu Lys Phe Arg Tyr Met Cys Ile Leu Ser Gly Cys

210 215 220

Asp Tyr Leu Ser Ser Leu Arg Gly Ile Gly Leu Ala Lys Ala Cys Lys

225 230 235 240

Val Leu Arg Leu Ala Asn Asn Pro Asp Ile Val Lys Val Ile Lys Lys

245 250 255

Ile Gly His Tyr Leu Lys Met Asn Ile Thr Val Pro Glu Asp Tyr Ile

260 265 270

Asn Gly Phe Ile Arg Ala Asn Asn Thr Phe Leu Tyr Gln Leu Val Phe

275 280 285

Asp Pro Ile Lys Arg Lys Leu Ile Pro Leu Asn Ala Tyr Glu Asp Asp

290 295 300

Val Asp Pro Glu Thr Leu Ser Tyr Ala Gly Gln Tyr Val Asp Asp Ser

305 310 315 320

Ile Ala Leu Gln Ile Ala Leu Gly Asn Lys Asp Ile Asn Thr Phe Glu

325 330 335

Gln Ile Asp Asp Tyr Asn Pro Asp Thr Ala Met Pro Ala His Ser Arg

340 345 350

Ser His Ser Trp Asp Asp Lys Thr Cys Gln Lys Ser Ala Asn Val Ser

355 360 365

Ser Ile Trp His Arg Asn Tyr Ser Pro Arg Pro Glu Ser Gly Thr Val

370 375 380

Ser Asp Ala Pro Gln Leu Lys Glu Asn Pro Ser Thr Val Gly Val Glu

385 390 395 400

Arg Val Ile Ser Thr Lys Gly Leu Asn Leu Pro Arg Lys Ser Ser Ile

405 410 415

Val Lys Arg Pro Arg Ser Glu Leu Ser Glu Asp Asp Leu Leu Ser Gln

420 425 430

Tyr Ser Leu Ser Phe Thr Lys Lys Thr Lys Lys Asn Ser Ser Glu Gly

435 440 445

Asn Lys Ser Leu Ser Phe Ser Glu Val Phe Val Pro Asp Leu Val Asn

450 455 460

Gly Pro Thr Asn Lys Lys Ser Val Ser Thr Pro Pro Arg Thr Arg Asn

465 470 475 480

Lys Phe Ala Thr Phe Leu Gln Arg Lys Asn Glu Glu Ser Gly Ala Val

485 490 495

Val Val Pro Gly Thr Arg Ser Arg Phe Phe Cys Ser Ser Asp Ser Thr

500 505 510

Asp Cys Val Ser Asn Lys Val Ser Ile Gln Pro Leu Asp Glu Thr Ala

515 520 525

Val Thr Asp Lys Glu Asn Asn Leu His Glu Ser Glu Tyr Gly Asp Gln

530 535 540

Glu Gly Lys Arg Leu Val Asp Thr Asp Val Ala Arg Asn Ser Ser Asp

545 550 555 560

Asp Ile Pro Asn Asn His Ile Pro Gly Asp His Ile Pro Asp Lys Ala

565 570 575

Thr Val Phe Thr Asp Glu Glu Ser Tyr Ser Phe Glu Ser Ser Lys Phe

580 585 590

Thr Arg Thr Ile Ser Pro Pro Thr Leu Gly Thr Leu Arg Ser Cys Phe

595 600 605

Ser Trp Ser Gly Gly Leu Gly Asp Phe Ser Arg Thr Pro Ser Pro Ser

610 615 620

Pro Ser Thr Ala Leu Gln Gln Phe Arg Arg Lys Ser Asp Ser Pro Thr

625 630 635 640

Ser Leu Pro Glu Asn Asn Met Ser Asp Val Ser Gln Leu Lys Ser Glu

645 650 655

Glu Ser Ser Asp Asp Glu Ser His Pro Leu Arg Glu Glu Ala Cys Ser

660 665 670

Ser Gln Ser Gln Glu Ser Gly Glu Phe Ser Leu Gln Ser Ser Asn Ala

675 680 685

Ser Lys Leu Ser Gln Cys Ser Ser Lys Asp Ser Asp Ser Glu Glu Ser

690 695 700

Asp Cys Asn Ile Lys Leu Leu Asp Ser Gln Ser Asp Gln Thr Ser Lys

705 710 715 720

Leu Arg Leu Ser His Phe Ser Lys Lys Asp Thr Pro Leu Arg Asn Lys

725 730 735

Val Pro Gly Leu Tyr Lys Ser Ser Ser Ala Asp Ser Leu Ser Thr Thr

740 745 750

Lys Ile Lys Pro Leu Gly Pro Ala Arg Ala Ser Gly Leu Ser Lys Lys

755 760 765

Pro Ala Ser Ile Gln Lys Arg Lys His His Asn Ala Glu Asn Lys Pro

770 775 780

Gly Leu Gln Ile Lys Leu Asn Glu Leu Trp Lys Asn Phe Gly Phe Lys

785 790 795 800

Lys Asp Ser Glu Lys Leu Pro Pro Cys Lys Lys Pro Leu Ser Pro Val

805 810 815

Arg Asp Asn Ile Gln Leu Thr Pro Glu Ala Glu Glu Asp Ile Phe Asn

820 825 830

Lys Pro Glu Cys Gly Arg Val Gln Arg Ala Ile Phe Gln

835 840 845

<210> 3871

<211> 580

<212> PRT

<213> Homo sapiens

<400> 3871

Met Gly Thr Thr Gly Leu Glu Ser Leu Ser Leu Gly Asp Arg Gly Ala
1 5 10 15

Ala Pro Thr Val Thr Ser Ser Glu Arg Leu Val Pro Asp Pro Pro Asn
20 25 30

Asp Leu Arg Lys Glu Asp Val Ala Met Glu Leu Glu Arg Val Gly Glu
35 40 45

Asp Glu Glu Gln Met Met Ile Lys Arg Ser Ser Glu Cys Asn Pro Leu
50 55 60

Leu Gln Glu Pro Ile Ala Ser Ala Gln Phe Gly Ala Thr Ala Gly Thr
65 70 75 80

Glu Cys Arg Lys Ser Val Pro Cys Gly Trp Glu Arg Val Val Lys Gln
85 90 95

Arg Leu Phe Gly Lys Thr Ala Gly Arg Phe Asp Val Tyr Phe Ile Ser
100 105 110

Pro Gln Gly Leu Lys Phe Arg Ser Lys Ser Ser Leu Ala Asn Tyr Leu
115 120 125

His Lys Asn Gly Glu Thr Ser Leu Lys Pro Glu Asp Phe Asp Phe Thr
130 135 140

Val Leu Ser Lys Arg Gly Ile Lys Ser Arg Tyr Lys Asp Cys Ser Met
145 150 155 160

Ala Ala Leu Thr Ser His Leu Gln Asn Gln Ser Asn Asn Ser Asn Trp
165 170 175

Asn Leu Arg Thr Arg Ser Lys Cys Lys Lys Asp Val Phe Met Pro Pro
180 185 190

Ser Ser Ser Ser Glu Leu Gln Glu Ser Arg Gly Leu Ser Asn Phe Thr
195 200 205

Ser Thr His Leu Leu Leu Lys Glu Asp Glu Gly Val Asp Asp Val Asn
210 215 220

Phe Arg Lys Val Arg Lys Pro Lys Gly Lys Val Thr Ile Leu Lys Gly

225 230 235 240

Ile Pro Ile Lys Lys Thr Lys Lys Gly Cys Arg Lys Ser Cys Ser Gly
245 250 255
Phe Val Gln Ser Asp Ser Lys Arg Glu Ser Val Cys Asn Lys Ala Asp
260 265 270
Ala Glu Ser Glu Pro Val Ala Gln Lys Ser Gln Leu Asp Arg Thr Val
275 280 285
Cys Ile Ser Asp Ala Gly Ala Cys Gly Glu Thr Leu Ser Val Thr Ser
290 295 300

Glu Glu Asn Ser Leu Val Lys Lys Lys Glu Arg Ser Leu Ser Ser Gly
305 310 315 320
Ser Asn Phe Cys Ser Glu Gln Lys Thr Ser Gly Ile Ile Asn Lys Phe
325 330 335
Cys Ser Ala Lys Asp Ser Glu His Asn Glu Lys Tyr Glu Asp Thr Phe
340 345 350
Leu Glu Ser Glu Glu Ile Gly Thr Lys Val Glu Val Val Glu Arg Lys
355 360 365

Glu His Leu His Thr Asp Ile Leu Lys Arg Gly Ser Glu Met Asp Asn
370 375 380
Asn Cys Ser Pro Thr Arg Lys Asp Phe Thr Gly Glu Lys Ile Phe Gln
385 390 395 400
Glu Asp Thr Ile Pro Arg Thr Gln Ile Glu Arg Arg Lys Thr Ser Leu
405 410 415
Tyr Phe Ser Ser Lys Tyr Asn Lys Glu Ala Leu Ser Pro Pro Arg Arg
420 425 430

Lys Ala Phe Lys Lys Trp Thr Pro Pro Arg Ser Pro Phe Asn Leu Val
435 440 445
Gln Glu Thr Leu Phe His Asp Pro Trp Lys Leu Leu Ile Ala Thr Ile
450 455 460
Phe Leu Asn Arg Thr Ser Gly Lys Met Ala Ile Pro Val Leu Trp Lys
465 470 475 480

Phe Leu Glu Lys Tyr Pro Ser Ala Glu Val Ala Arg Thr Ala Asp Trp
485 490 495

Arg Asp Val Ser Glu Leu Leu Lys Pro Leu Gly Leu Tyr Asp Leu Arg
500 505 510

Ala Lys Thr Ile Val Lys Phe Ser Asp Glu Tyr Leu Thr Lys Gln Trp
515 520 525

Lys Tyr Pro Ile Glu Leu His Gly Ile Gly Lys Tyr Gly Asn Asp Ser
530 535 540

Tyr Arg Ile Phe Cys Val Asn Glu Trp Lys Gln Val His Pro Glu Asp
545 550 555 560

His Lys Leu Asn Lys Tyr His Asp Trp Leu Trp Glu Asn His Glu Lys
565 570 575

Leu Ser Leu Ser
580

<210> 3872

<211> 410

<212> PRT

<213> Homo sapiens

<400> 3872

Met Glu Ala Glu Asn Ala Gly Ser Tyr Ser Leu Gln Gln Ala Gln Ala
1 5 10 15

Phe Tyr Thr Phe Pro Phe Gln Gln Leu Met Ala Glu Ala Pro Asn Met
20 25 30

Ala Val Val Asn Glu Gln Gln Met Pro Glu Glu Val Pro Ala Pro Ala
35 40 45

Pro Ala Gln Glu Pro Val Gln Glu Ala Pro Lys Gly Arg Lys Arg Lys
50 55 60

Pro Arg Thr Thr Glu Pro Lys Gln Pro Val Glu Pro Lys Lys Pro Val
65 70 75 80

Glu Ser Lys Lys Ser Gly Lys Ser Ala Lys Ser Lys Glu Lys Gln Glu
85 90 95

Lys Ile Thr Asp Thr Phe Lys Val Lys Arg Lys Val Asp Arg Phe Asn

100 105 110

Gly Val Ser Glu Ala Glu Leu Leu Thr Lys Thr Leu Pro Asp Ile Leu

115 120 125

Thr Phe Asn Leu Asp Ile Val Ile Ile Gly Ile Asn Pro Gly Leu Met

130 135 140

Ala Ala Tyr Lys Gly His His Tyr Pro Gly Pro Gly Asn His Phe Trp

145 150 155 160

Lys Cys Leu Phe Met Ser Gly Leu Ser Glu Val Gln Leu Asn His Met

165 170 175

Asp Asp His Thr Leu Pro Gly Lys Tyr Gly Ile Gly Phe Thr Asn Met

180 185 190

Val Glu Arg Thr Thr Pro Gly Ser Lys Asp Leu Ser Ser Lys Glu Phe

195 200 205

Arg Glu Gly Gly Arg Ile Leu Val Gln Lys Leu Gln Lys Tyr Gln Pro

210 215 220

Arg Ile Ala Val Phe Asn Gly Lys Cys Ile Tyr Glu Ile Phe Ser Lys

225 230 235 240

Glu Val Phe Gly Val Lys Val Lys Asn Leu Glu Phe Gly Leu Gln Pro

245 250 255

His Lys Ile Pro Asp Thr Glu Thr Leu Cys Tyr Val Met Pro Ser Ser

260 265 270

Ser Ala Arg Cys Ala Gln Phe Pro Arg Ala Gln Asp Lys Val His Tyr

275 280 285

Tyr Ile Lys Leu Lys Asp Leu Arg Asp Gln Leu Lys Gly Ile Glu Arg

290 295 300

Asn Met Asp Val Gln Glu Val Gln Tyr Thr Phe Asp Leu Gln Leu Ala

305 310 315 320

Gln Glu Asp Ala Lys Lys Met Ala Val Lys Glu Glu Lys Tyr Asp Pro

325 330 335

Gly Tyr Glu Ala Ala Tyr Gly Gly Ala Tyr Gly Glu Asn Pro Cys Ser

340 345 350

Ser Glu Pro Cys Gly Phe Ser Ser Asn Gly Leu Ile Glu Ser Val Glu

355 360 365

Leu Arg Gly Glu Ser Ala Phe Ser Gly Ile Pro Asn Gly Gln Trp Met

370 375 380

Thr Gln Ser Phe Thr Asp Gln Ile Pro Ser Phe Ser Asn His Cys Gly

385 390 395 400

Thr Gln Glu Gln Glu Glu Glu Ser His Ala

405 410

<210> 3873

<211> 84

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400>

3873

Met Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu

1 5 10 15

Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val

20 25 30

Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp

35 40 45

Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu

50 55 60

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys

65 70 75 80

Ile Lys Met Leu

<210> 3874

<400> 3874

000

<210> 3875

<211> 1165

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3875

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn
20 25 30

Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys
35 40 45

Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn
50 55 60

Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr
65 70 75 80

Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg
85 90 95

Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp
100 105 110

Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp
115 120 125

Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val
130 135 140

Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu
145 150 155 160

Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu
165 170 175

Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu
180 185 190

Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln
195 200 205

Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val

210 215 220

Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu
225 230 235 240
Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe
245 250 255
Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser
260 265 270
Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr
275 280 285

Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr
290 295 300
Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu
305 310 315 320
Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser
325 330 335
Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu
340 345 350

Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile
355 360 365
Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly
370 375 380
Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys
385 390 395 400
Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu
405 410 415

Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile
420 425 430
His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr
435 440 445
Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe
450 455 460

Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe
465 470 475 480

Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe
485 490 495
Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg
500 505 510
Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys
515 520 525
His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys
530 535 540

Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly
545 550 555 560
Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys
565 570 575
Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys
580 585 590
Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser
595 600 605

Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe
610 615 620
Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr
625 630 635 640
Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr
645 650 655
Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg
660 665 670

Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile
675 680 685
Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp
690 695 700
Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu

705 710 715 720

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp

725 730 735

Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys

740 745 750

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val

755 760 765

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu

770 775 780

Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys

785 790 795 800

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu

805 810 815

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr

820 825 830

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

835 840 845

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe

850 855 860

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys

865 870 875 880

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys

885 890 895

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln

900 905 910

Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu

915 920 925

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln

930 935 940

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

945 950 955 960

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
965 970 975
Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
980 985 990

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
995 1000 1005
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1010 1015 1020
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1025 1030 1035
Met Ile Ala Lys Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val
1040 1045 1050

Glu Tyr Gly Leu Leu Pro Ile Gly Lys Ile Val Glu Lys Arg Ile
1055 1060 1065
Glu Cys Thr Val Tyr Ser Val Asp Asn Asn Gly Asn Ile Tyr Thr
1070 1075 1080
Gln Pro Val Ala Gln Trp His Asp Arg Gly Glu Gln Glu Val Phe
1085 1090 1095
Glu Tyr Cys Leu Glu Asp Gly Ser Leu Ile Arg Ala Thr Lys Asp
1100 1105 1110

His Lys Phe Met Thr Val Asp Gly Gln Met Leu Pro Ile Asp Glu
1115 1120 1125
Ile Phe Glu Arg Glu Leu Asp Leu Met Arg Val Asp Asn Leu Pro
1130 1135 1140
Asn Ser Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu
1145 1150 1155
Pro Lys Lys Lys Arg Lys Val
1160 1165
<210> 3876
<211> 1128
<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3876

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn
20 25 30

Val Tyr Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn
35 40 45

Gly Phe Ile Ala Ser Asn Cys Phe Asn Glu Ile Gly Lys Ala Thr Ala
50 55 60

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
65 70 75 80

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn
85 90 95

Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr
100 105 110

Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
115 120 125

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
130 135 140

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys
145 150 155 160

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val
165 170 175

Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu
180 185 190

Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro
195 200 205

Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu
210 215 220

Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
225 230 235 240
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu
245 250 255

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
260 265 270
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
275 280 285
Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
290 295 300
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
305 310 315 320

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala
325 330 335
Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala
340 345 350
Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
355 360 365
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
370 375 380

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly
385 390 395 400
Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
405 410 415
Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Thr
420 425 430
Leu Asn Ile Glu Asp Glu Tyr Arg Leu His Glu Thr Ser Lys Glu Pro
435 440 445

Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala Trp
450 455 460
Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu Ile

465 470 475 480

Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro

485 490 495

Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu

500 505 510

Leu Asp Gln Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro

515 520 525

Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val Gln

530 535 540

Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr Val

545 550 555 560

Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln Trp

565 570 575

Tyr Thr Val Leu Asp Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu His

580 585 590

Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu Met

595 600 605

Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe Lys

610 615 620

Asn Ser Pro Thr Leu Phe Asn Glu Ala Leu His Arg Asp Leu Ala Asp

625 630 635 640

Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp Asp

645 650 655

Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr Arg

660 665 670

Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala Lys

675 680 685

Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu Leu

690 695 700

Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val Met

705 710 715 720

Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly
725 730 735
Lys Ala Gly Phe Cys Arg Leu Phe Ile Pro Gly Phe Ala Glu Met Ala
740 745 750
Ala Pro Leu Tyr Pro Leu Thr Lys Pro Gly Thr Leu Phe Asn Trp Gly
755 760 765

Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr
770 775 780
Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu Phe
785 790 795 800
Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys Leu
805 810 815
Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp Pro
820 825 830

Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile Ala
835 840 845
Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu Val
850 855 860
Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro Asp
865 870 875 880
Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu Leu
885 890 895

Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro Ala
900 905 910
Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu Asp
915 920 925
Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln Pro
930 935 940
Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu Leu
945 950 955 960

Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr Glu

965 970 975

Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg Ala

980 985 990

Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys

995 1000 1005

Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His

1010 1015 1020

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Trp Leu Thr Ser Glu

1025 1030 1035

Gly Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys

1040 1045 1050

Ala Leu Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly

1055 1060 1065

His Gln Lys Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala

1070 1075 1080

Asp Gln Ala Ala Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr

1085 1090 1095

Ser Thr Leu Leu Ile Glu Asn Ser Ser Pro Ser Gly Gly Ser Lys

1100 1105 1110

Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys

1115 1120 1125

<210> 3877

<211> 1024

<212> PRT

<213> Streptococcus pyogenes

<400> 3877

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys

<210> 3878

<211> 1023

<212

> PRT

<213> Streptococcus pyogenes

<400> 3878

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

<210> 3879
<211> 344
<212> PRT
<213> Streptococcus pyogenes
<400> 3879
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1 5 10 15
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
20 25 30
Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
35 40 45

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
50 55 60
Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
65 70 75 80
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
85 90 95
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
100 105 110

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
115 120 125
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
130 135 140
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

145 150 155 160

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

165 170 175

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser

180 185 190

Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr

195 200 205

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

210 215 220

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

225 230 235 240

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

245 250 255

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

260 265 270

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

275 280 285

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

290 295 300

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp

305 310 315 320

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

325 330 335

Asp Leu Ser Gln Leu Gly Gly Asp

340

<210> 3880

<400> 3880

000

<210> 3881

<400> 3881

000

<210> 3882

<400> 3882

000

<210> 3883

<400> 3883

000

<210> 3884

<400> 3884

000

<210> 3885

<400> 3885

000

<210> 3886

<400> 3886

000

<210> 3887

<400> 3887

000

<210> 3888

<400> 3888

000

<210> 3889

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3889

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

1 5 10

<210> 3890

<400> 3890

000

<210> 3891

<400> 3891

000

<210> 3892

<400> 3892

000

<210> 3893

<400> 3893

000

<210> 3894

<400> 3894

000

<210> 3895

<400> 3895

000

<210> 3896

<400> 3896

000

<210> 3897

<400> 3897

000

<210> 3898

<400> 3898

000

<210> 3899

<400> 3899

000

<210> 3900

<400> 3900

000

<210> 3901

<400> 3901

000

<210> 3902

<400> 3902

000

<210> 3903

<400> 3903

000

<210> 3904

<400> 3904

000

<210> 3905

<400> 3905

000

<210> 3906

<400> 3906

000

<210> 3907

<400> 3907

000

<210> 3908

<400> 3908

000

<210> 3909

<400> 3909

000

<210> 3910

<400

> 3910

000

<210> 3911

<400> 3911

000

<210> 3912

<400> 3912

000

<210> 3913

<400> 3913

000

<210> 3914

<400> 3914

000

<210> 3915

<400> 3915

000

<210> 3916

<400> 3916

000

<210> 3917

<400> 3917

000

<210> 3918

<400> 3918

000

<210> 3919

<400> 3919

000

<210> 3920

<400> 3920

000

<210> 3921

<400> 3921

000

<210> 3922

<400> 3922

000

<210> 3923

<400> 3923

000

<210> 3924

<400> 3924

000

<210> 3925

<400> 3925

000

<210> 3926

<400> 3926

000

<210> 3927

<400> 3927

000

<210> 3928

<400> 3928

000

<210> 3929

<400> 3929

000

<210> 3930

<400> 3930

000

<210> 3931

<400> 3931

000

<210> 3932

<400> 3932

000

<210> 3933

<400> 3933

000

<210> 3934

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3934

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Phe

1 5 10

<210> 3935

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3935

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30

Arg Asn Gln Gly Gly Tyr
35

<210> 3936

<211> 16

<212> PRT

<213> Rattus norvegicus

<220><221> misc_feature

<222> (3)..(12)

<223> Xaa can be any naturally occurring amino acid

<400> 3936

Lys Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys Lys Lys Leu
1 5 10 15

<210> 3937

<211> 345

<212> PRT

<213> Homo sapiens

<400> 3937

Met Pro Ala Arg Ala Leu Leu Pro Arg Arg Met Gly His Arg Thr Leu
1 5 10 15

Ala Ser Thr Pro Ala Leu Trp Ala Ser Ile Pro Cys Pro Arg Ser Glu
20 25 30

Leu Arg Leu Asp Leu Val Leu Pro Ser Gly Gln Ser Phe Arg Trp Arg

35 40 45

Glu Gln Ser Pro Ala His Trp Ser Gly Val Leu Ala Asp Gln Val Trp

50 55 60

Thr Leu Thr Gln Thr Glu Glu Gln Leu His Cys Thr Val Tyr Arg Gly
65 70 75 80
Asp Lys Ser Gln Ala Ser Arg Pro Thr Pro Asp Glu Leu Glu Ala Val
85 90 95
Arg Lys Tyr Phe Gln Leu Asp Val Thr Leu Ala Gln Leu Tyr His His

100 105 110
Trp Gly Ser Val Asp Ser His Phe Gln Glu Val Ala Gln Lys Phe Gln
115 120 125
Gly Val Arg Leu Leu Arg Gln Asp Pro Ile Glu Cys Leu Phe Ser Phe
130 135 140
Ile Cys Ser Ser Asn Asn Asn Ile Ala Arg Ile Thr Gly Met Val Glu
145 150 155 160
Arg Leu Cys Gln Ala Phe Gly Pro Arg Leu Ile Gln Leu Asp Asp Val

165 170 175
Thr Tyr His Gly Phe Pro Ser Leu Gln Ala Leu Ala Gly Pro Glu Val
180 185 190
Glu Ala His Leu Arg Lys Leu Gly Leu Gly Tyr Arg Ala Arg Tyr Val
195 200 205
Ser Ala Ser Ala Arg Ala Ile Leu Glu Glu Gln Gly Gly Leu Ala Trp
210 215 220
Leu Gln Gln Leu Arg Glu Ser Ser Tyr Glu Glu Ala His Lys Ala Leu

225 230 235 240
Cys Ile Leu Pro Gly Val Gly Thr Lys Val Ala Asp Cys Ile Cys Leu
245 250 255
Met Ala Leu Asp Lys Pro Gln Ala Val Pro Val Asp Val His Met Trp
260 265 270
His Ile Ala Gln Arg Asp Tyr Ser Trp His Pro Thr Thr Ser Gln Ala
275 280 285
Lys Gly Pro Ser Pro Gln Thr Asn Lys Glu Leu Gly Asn Phe Phe Arg

290 295 300

Ser Leu Trp Gly Pro Tyr Ala Gly Trp Ala Gln Ala Val Leu Phe Ser
305 310 315 320
Ala Asp Leu Arg Gln Ser Arg His Ala Gln Glu Pro Pro Ala Lys Arg
325 330 335
Arg Lys Gly Ser Lys Gly Pro Glu Gly
340 345

<210> 3938
<211> 298
<212> PRT
<213> Homo sapiens
<400> 3938
Met Val Thr Pro Ala Leu Gln Met Lys Lys Pro Lys Gln Phe Cys Arg

1 5 10 15
Arg Met Gly Gln Lys Lys Gln Arg Pro Ala Arg Ala Gly Gln Pro His
20 25 30
Ser Ser Ser Asp Ala Ala Gln Ala Pro Ala Glu Gln Pro His Ser Ser
35 40 45
Ser Asp Ala Ala Gln Ala Pro Cys Pro Arg Glu Arg Cys Leu Gly Pro
50 55 60
Pro Thr Thr Pro Gly Pro Tyr Arg Ser Ile Tyr Phe Ser Ser Pro Lys

65 70 75 80
Gly His Leu Thr Arg Leu Gly Leu Glu Phe Phe Asp Gln Pro Ala Val
85 90 95
Pro Leu Ala Arg Ala Phe Leu Gly Gln Val Leu Val Arg Arg Leu Pro
100 105 110
Asn Gly Thr Glu Leu Arg Gly Arg Ile Val Glu Thr Glu Ala Tyr Leu
115 120 125
Gly Pro Glu Asp Glu Ala Ala His Ser Arg Gly Gly Arg Gln Thr Pro

130 135 140
Arg Asn Arg Gly Met Phe Met Lys Pro Gly Thr Leu Tyr Val Tyr Ile
145 150 155 160
Ile Tyr Gly Met Tyr Phe Cys Met Asn Ile Ser Ser Gln Gly Asp Gly

165 170 175

Ala Cys Val Leu Leu Arg Ala Leu Glu Pro Leu Glu Gly Leu Glu Thr

180 185 190

Met Arg Gln Leu Arg Ser Thr Leu Arg Lys Gly Thr Ala Ser Arg Val

195 200 205

Leu Lys Asp Arg Glu Leu Cys Ser Gly Pro Ser Lys Leu Cys Gln Ala

210 215 220

Leu Ala Ile Asn Lys Ser Phe Asp Gln Arg Asp Leu Ala Gln Asp Glu

225 230 235 240

Ala Val Trp Leu Glu Arg Gly Pro Leu Glu Pro Ser Glu Pro Ala Val

245 250 255

Val Ala Ala Ala Arg Val Gly Val Gly His Ala Gly Glu Trp Ala Arg

260 265 270

Lys Pro Leu Arg Phe Tyr Val Arg Gly Ser Pro Trp Val Ser Val Val

275 280 285

Asp Arg Val Ala Glu Gln Asp Thr Gln Ala

290 295

<210> 3939

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (5)..(6)

<223> may be non-consecutive

<400> 3939

gttttaaaac 10

<210> 3940

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3940

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95
Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175
His Asn Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3941

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3941

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Lys Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3942
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3942

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95
Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Lys Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3943

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3943

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Lys Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3944
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3944

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Ile Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240
<210> 3945
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3945
Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu His
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240
<210> 3946
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3946
Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Ile Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3947

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3947

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Gln Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3948

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3948

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Arg Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3949
<211> 240
<212> PRT
<213> Artificial Sequence

<220><223> Synthetic
<400> 3949

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95
Ser Gln Trp Asn Lys Leu Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3950

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3950

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Leu Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3951

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3951

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Val Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3952

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3952

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Val Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3953

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3953

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Gly
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240
<210> 3954
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3954

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95
Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln Arg Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240
<210> 3955

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3955

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Ser Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3956

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3956

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asn Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3957

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3957

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Pro Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3958

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3958

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Arg Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3959

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3959

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Thr Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3960

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3960

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95
Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Ser Tyr
115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240
<210> 3961
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3961

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Val Tyr
115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240
<210> 3962

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3962

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Val Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3963

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3963

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Gly Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240

<210> 3964
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3964
Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95
Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110
Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125
Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160
Val Tyr Tyr Arg Pro Met Asn Glu Tyr Ser Asn Gln Asn Asn Phe Val
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190
Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240
<210> 3965
<211> 240
<212> PRT
<213> Artificial Sequence

<220><223> Synthetic
<400> 3965
Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp
1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
65 70 75 80
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His
85 90 95
Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met
100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr
115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp
130 135 140
Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
145 150 155 160
Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Ser Gln Asn Asn Phe Val
165 170 175
His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr
180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg
195 200 205
Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala
210 215 220
Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val
225 230 235 240
<210> 3966
<211> 240
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 3966
Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15
Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn
20 25 30
Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg
35 40 45
Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly
50 55 60
Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Lys Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Pro Pro Val

225 230 235 240

<210> 3967

<211> 240

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3967

Met Ala Asn Leu Gly Cys Trp Met Leu Val Leu Phe Val Ala Thr Trp

1 5 10 15

Ser Asp Leu Gly Leu Cys Lys Lys Arg Pro Lys Pro Gly Gly Trp Asn

20 25 30

Thr Gly Gly Ser Arg Tyr Pro Gly Gln Gly Ser Pro Gly Gly Asn Arg

35 40 45

Tyr Pro Pro Gln Gly Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

50 55 60

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Pro His Gly Gly Gly

65 70 75 80

Trp Gly Gln Pro His Gly Gly Gly Trp Gly Gln Gly Gly Gly Thr His

85 90 95

Ser Gln Trp Asn Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met

100 105 110

Ala Gly Ala Ala Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr

115 120 125

Met Leu Gly Ser Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp

130 135 140

Tyr Glu Asp Arg Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln

145 150 155 160

Val Tyr Tyr Arg Pro Met Asp Glu Tyr Ser Asn Gln Asn Asn Phe Val

165 170 175

His Asp Cys Val Asn Ile Thr Ile Lys Gln His Thr Val Thr Thr Thr

180 185 190

Thr Lys Gly Glu Asn Phe Thr Glu Thr Asp Val Lys Met Met Glu Arg

195 200 205

Val Val Glu Gln Met Cys Ile Thr Gln Tyr Glu Arg Glu Ser Gln Ala

210 215 220

Tyr Tyr Gln Arg Gly Ser Ser Met Val Leu Phe Ser Ser Ser Pro Val

225 230 235 240

<210> 3968

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3968

Glu Ala Ala Ala Lys

1 5

<210> 3969

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3969

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys

1 5 10

<210> 3970

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3970

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys

1 5 10 15

<210> 3971

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3971

gattggcttt gatttcccta 20

<210> 3972

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3972

gtatatgatt gttactgaga 20

<210> 3973

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3973

gcagttgcct aagaactggt 20

<210> 3974

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3974

gaggagagta aagtgattgg 20

<210> 3975

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3975

gcagagccag gaacccctgt 20

<210> 3976

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3976

gggaaggggc aggagagcca 20

<210> 3977

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3977

gaatatgtcc cagatagcac 20

<210> 3978

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3978

ggggactctt taaggaaaga 20

<210> 3979

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3979

gagaaagaga aagggagtag 20

<210> 3980

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3980

gagtagaggc ggccacgacc 20

<210> 3981

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3981

gatcagtgaa acgcaccaga 20

<210> 3982

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3982

gcagctcagg ttctgggaga 20

<210> 3983

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3983

gtggccactg agaaccgggc 20

<210> 3984

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3984

gaatctgcct aacaggaggt 20

<210> 3985

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3985

gtcaccaatc ctgtccctag 20

<210> 3986

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3986

gctggccccc caccgcccca 20

<210> 3987

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3987

gacgtcacgg cgctgcccca 20

<210> 3988

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3988

ggtgtgctgg gcaggtcgcg 20

<210> 3989

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3989

gtccttggca agcccaggag 20

<210> 3990

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3990

gtgcgtccta ggtgttcacc 20

<210> 3991

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3991

gtcctggcag ggctgtggtg 20

<210> 3992

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3992

gtgacctgcc cggttctcag 20

<210> 3993

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3993

gagcttggca gggggtggga 20

<210> 3994

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3994

gagccagaga ggatcctggg 20

<210> 3995

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3995

gatggagcca gagaggatcc 20

<210> 3996

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3996

ggggccacta gggacaggat 20

<210> 3997

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3997

gtcccctcca ccccacagtg 20

<210> 3998

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3998

gggaccacct tatattccca 20

<210> 3999

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 3999

gacctgccca gcacaccctg 20

<210> 4000

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4000

gcgactcctg gaagtggcca 20

<210> 4001

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4001

ggacttccca gtgtgcatcg 20

<210> 4002

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4002

tttgcagttt atcaggatga 20

<210> 4003

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4003

ggttgagcag gtagatgtca 20

<210> 4004

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4004

ctgggcggca gcatagtgag 20

<210> 4005

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4005

gtcaagagtt gacacattgt 20

<210> 4006

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4006

cccaagtgat cacacttgtc 20

<210> 4007

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4007

ccacccaagt gatcacactt 20

<210> 4008

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4008

gggagagacg caaacacagc 20

<210> 4009

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4009

tctctgacct gtttttcctt 20

<210> 4010

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4010

tcttactgtc cccttctggg 20

<210> 4011

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4011

tctgggctca ctatgctgcc 20

<210> 4012

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4012

cccagtggga ctttggaaat 20

<210> 4013

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4013

gtgggacttt ggaaatacaa 20

<210> 4014

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4014

cctgacaatc gataggtacc 20

<210> 4015

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4015

ggaatcatct ttaccagatc 20

<210> 4016

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4016

gcagctctca ttttccatac 20

<210> 4017

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4017

tacagtcagt atcaattctg 20

<210> 4018

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4018

gactgaaact tcacagaata 20

<210> 4019

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4019

gatttatgag atcaacagca c 21

<210> 4020

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4020

gacaccaaaa actgccacag g 21

<210> 4021

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4021

gcactcattt cctccaagct c 21

<210> 4022

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4022

gtaggtacag gacagggcag 20

<210> 4023

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4023

gagataggta ggtacaggac a 21

<210> 4024

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4024

gtgaaaagat aggtaggtac 20

<210> 4025

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4025

gttatggttt actccatcta 20

<210> 4026

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4026

gttttggttt accctatcta 20

<210> 4027

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4027

gctgtggaac tgcaacacac t 21

<210> 4028

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4028

gtgccaccta acagtgagct g 21

<210> 4029

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4029

atgatcctga cgacggagac cgccgtcgtc gacaagcccg tgctcagtct g 51

<210> 4030

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4030

tcctgacgac ggagaccgcc gtcgtcgaca agcccgtgct cagtctg 47

<210> 4031

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4031

gacgacggag accgccgtcg tcgacaagcc cgtgctcagt ctg 43

<210> 4032

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4032

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatcc gggatactgg 50

<210> 4033

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4033

tgtcgacgac ggcggtctcc gtcgtcagga tcatccggga tactgg 46

<210> 4034

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4034

gacgacggcg gtctccgtcg tcaggatcat ccgggatact gg 42

<210> 4035

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4035

taccgtacac cactgagacc gcggtggttg accagacaaa cctcgtgctc agtctg 56

<210> 4036

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4036

gtacaccact gagaccgcgg tggttgacca gacaaacctc gtgctcagtc tg 52

<210> 4037

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4037

accactgaga ccgcggtggt tgaccagaca aacctcgtgc tcagtctg 48

<210> 4038

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4038

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctccggga tactgg 56

<210> 4039

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4039

ggtcaaccac cgcggtctca gtggtgtacg gtacaaacct ccgggatact gg 52

<210> 4040

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4040

aaccaccgcg gtctcagtgg tgtacggtac aaacctccgg gatactgg 48

<210> 4041

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4041

tggaggaagc agggcttcct ttcctctgcc atcacacctg cactcccgtg ctcagtctg 59

<210> 4042

<211> 83

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4042

tggaggaagc agggcttcct ttcctctgcc atcacccgtc tcctggggag atggtttcca 60

cctgcactcc cgtgctcagt ctg 83

<210> 4043

<211> 155

<212> DNA

<213> Artificial Sequence

<

220><223> Synthetic

<400> 4043

tggaggaagc agggcttcct ttcctctgcc atcaaaattt ctttccatct tcaagcatcc 60

cggtgtagtg caccacgcag gtctggccgc gcttggggaa ggtgcgcccg tctcctgggg 120

agatggtttc cacctgcact cccgtgctca gtctg 155

<210> 4044

<211> 368

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4044

tggaggaagc agggcttcct ttcctctgcc atcattccag ttttagaagc tccacatcga 60

agacgagagt ggcatgtggt gggatgatgc ctgggtgccc agtggcacca taggcataat 120

ctggagatat agtcagtttg gctctctgac ccacactcat ctgggcaacc ccttcttccc 180

agcctcggat cacctcctgc ttgcctagca taaacttaaa gggcttgttt ctgtcccggg 240

aggaatcaaa tttctttcca tcttcaagca tcccggtgta gtgcaccacg caggtctggc 300

cgcgcttggg gaaggtgcgc ccgtctcctg gggagatggt ttccacctgc actcccgtgc 360

tcagtctg 368

<210> 4045

<211> 84

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4045

accacgcagg tctggccgcg cttggggaag gtgcgcccgt ctcctgggga gatggtttcc 60

acctgcactc ccgtgctcag tctg 84

<210> 4046

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4046

accttcccca agcgcggcca gacctgcgtg gtgcactaca ccgggatgct tgaagatgga 60

aagaaatttc cgggatactg g 81

<210> 4047

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4047

tcctctgcca tcacgtgctc agtctg 26

<210> 4048

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4048

tgatggcaga ggaccgggat actgg 25

<210> 4049

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4049

tggaggaagc agggcttcct ttcctctgcc atcacgtgct cagtctg 47

<210> 4050

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4050

tgatggcaga ggaaaggaag ccctgcttcc tccaccggga tactgg 46

<210> 4051

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4051

tgcaggagct gcatcctctg ccatcacgtg ctcagtctg 39

<210> 4052

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4052

tgatggcaga ggatgcagct cctgcaccgg gatactgg 38

<210> 4053

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4053

gcccagccaa acttgtcaac cagtatcccg gcgtgctcag tctg 44

<210> 4054

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4054

gggtcaatcc ttggggccca gactgagcac gccgggatac tgg 43

<210> 4055

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4055

tcctctgcca tcacgtgctc agtctgttaa ataacgcggt tctatctagt tacgcgttaa 60

accaactaga a 71

<210> 4056

<211> 70

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 4056

tgatggcaga ggaccgggat actggaaaat atacgcggtt ctatctagtt acgcgttaaa 60

ccaactagaa 70

<210> 4057

<211> 92

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4057

tggaggaagc agggcttcct ttcctctgcc atcacgtgct cagtctgtta ataatcgcgg 60

ttctatctag ttacgcgtta aaccaactag aa 92

<210> 4058

<211> 91

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4058

tgatggcaga ggaaaggaag ccctgcttcc tccaccggga tactggaaaa aaaacgcggt 60

tctatctagt tacgcgttaa accaactaga a 91

<210> 4059

<211> 84

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4059

tgcaggagct gcatcctctg ccatcacgtg ctcagtctga taaataacgc ggttctatct 60

agttacgcgt taaaccaact agaa 84

<210> 4060

<211> 83

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4060

tgatggcaga ggatgcagct cctgcaccgg gatactggaa aaaaggcgcg gttctatcta 60

gttacgcgtt aaaccaacta gaa 83

<210> 4061

<211> 89

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4061

gcccagccaa acttgtcaac cagtatcccg gcgtgctcag tctgttaaat accgcggttc 60

tatctagtta cgcgttaaac caactagaa 89

<210> 4062

<211> 88

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4062

gggtcaatcc ttggggccca gactgagcac gccgggatac tggaaaataa tcgcggttct 60

atctagttac gcgttaaacc aactagaa 88

<210> 4063

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4063

ttgaatagga agtaaattaa tttgaagctg gaccctaggg aaatcaa 47

<210> 4064

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4064

ttgaatagga agtaaattaa tttgaagctg gaccctaggg aaatcaaag 49

<210> 4065

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4065

ttgaatagga agtaaattaa tttgaagctg gaccctaggg aaatcaaagc c 51

<210> 4066

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4066

tagggtccag cttcaaatta atttacttcc tattcaaagt tcttaggc 48

<210> 4067

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4067

tagggtccag cttcaaatta atttacttcc tattcaaagt tcttaggcaa 50

<210> 4068

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4068

atgcctggac aagtaactta agttaaataa gccttctcag taacaatcat 50

<210> 4069

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4069

agaaggctta tttaacttaa gttacttgtc caggcatagt tcttag 46

<210> 4070

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4070

tgctgagaga gaaacagttg cctaagaact cagtaacaat 40

<210> 4071

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4071

accacttcca caatgtatat gattgttact gagttcttag g 41

<210> 4072

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4072

accacttcca caatgtatat gattgttact gagttcttag gcaa 44

<210> 4073

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4073

atgatcctga cgacggagac cgccgtcgtc gacaagccca gtaacaatc 49

<210> 4074

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4074

atgatcctga cgacggagac cgccgtcgtc gacaagccca gtaacaatca ta 52

<210> 4075

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4075

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag ttcttaggc 49

<210> 4076

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4076

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag ttcttaggca ac 52

<210> 4077

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4077

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatat cactttactc 50

<210> 4078

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4078

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatat cactttactc tcc 53

<210> 4079

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4079

taccgtacac cactgagacc gcggtggttg accagacaaa cctggggttc ct 52

<210> 4080

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4080

taccgtacac cactgagacc gcggtggttg accagacaaa cctggggttc ctgg 54

<210> 4081

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4081

taccgtacac cactgagacc gcggtggttg accagacaaa cctggggttc ctggct 56

<210> 4082

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4082

taccgtacac cactgagacc gcggtggttg accagacaaa cctctctcct gc 52

<210> 4083

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4083

taccgtacac cactgagacc gcggtggttg accagacaaa cctctctcct gccc 54

<210> 4084

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4084

taccgtacac cactgagacc gcggtggttg accagacaaa cctctctcct gcccct 56

<210> 4085

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4085

taccgtacac cactgagacc gcggtggttg accagacaaa cctctatctg gg 52

<210> 4086

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4086

taccgtacac cactgagacc gcggtggttg accagacaaa cctctatctg ggac 54

<210> 4087

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4087

taccgtacac cactgagacc gcggtggttg accagacaaa cctctatctg ggacat 56

<210> 4088

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4088

taccgtacac cactgagacc gcggtggttg accagacaaa cctttcctta aa 52

<210> 4089

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4089

taccgtacac cactgagacc gcggtggttg accagacaaa cctttcctta aaga 54

<210> 4090

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4090

taccgtacac cactgagacc gcggtggttg accagacaaa cctttcctta aagagt 56

<210> 4091

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4091

taccgtacac cactgagacc gcggtggttg accagacaaa cctctccctt tc 52

<210> 4092

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4092

taccgtacac cactgagacc gcggtggttg accagacaaa cctctccctt tctc 54

<210> 4093

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4093

taccgtacac cactgagacc gcggtggttg accagacaaa cctctccctt tctctt 56

<210> 4094

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4094

taccgtacac cactgagacc gcggtggttg accagacaaa cctcgtggcc gc 52

<210> 4095

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4095

taccgtacac cactgagacc gcggtggttg accagacaaa cctcgtggcc gcct 54

<210> 4096

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4096

taccgtacac cactgagacc gcggtggttg accagacaaa cctcgtggcc gcctct 56

<210> 4097

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4097

taccgtacac cactgagacc gcggtggttg accagacaaa cctggtgcgt tt 52

<210> 4098

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4098

taccgtacac cactgagacc gcggtggttg accagacaaa cctggtgcgt ttca 54

<210> 4099

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4099

taccgtacac cactgagacc gcggtggttg accagacaaa cctggtgcgt ttcact 56

<210> 4100

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4100

taccgtacac cactgagacc gcggtggttg accagacaaa cctcccagaa cc 52

<210> 4101

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4101

taccgtacac cactgagacc gcggtggttg accagacaaa cctcccagaa cctg 54

<210> 4102

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4102

taccgtacac cactgagacc gcggtggttg accagacaaa cctcccagaa cctgag 56

<210> 4103

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4103

taccgtacac cactgagacc gcggtggttg accagacaaa cctcggttct ca 52

<210> 4104

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4104

taccgtacac cactgagacc gcggtggttg accagacaaa cctcggttct cagt 54

<210> 4105

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4105

taccgtacac cactgagacc gcggtggttg accagacaaa cctcggttct cagtgg 56

<210> 4106

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4106

taccgtacac cactgagacc gcggtggttg accagacaaa ccttcctgtt ag 52

<210> 4107

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4107

taccgtacac cactgagacc gcggtggttg accagacaaa ccttcctgtt aggc 54

<210> 4108

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4108

taccgtacac cactgagacc gcggtggttg accagacaaa ccttcctgtt aggcag 56

<210> 4109

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4109

taccgtacac cactgagacc gcggtggttg accagacaaa cctgggacag ga 52

<210> 4110

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4110

taccgtacac cactgagacc gcggtggttg accagacaaa cctgggacag gatt 54

<210> 4111

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4111

taccgtacac cactgagacc gcggtggttg accagacaaa cctgggacag gattgg 56

<210> 4112

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4112

taccgtacac cactgagacc gcggtggttg accagacaaa cctggcggtg gg 52

<210> 4113

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4113

taccgtacac cactgagacc gcggtggttg accagacaaa cctggcggtg gggg 54

<210> 4114

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4114

taccgtacac cactgagacc gcggtggttg accagacaaa cctggcggtg gggggc 56

<210> 4115

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4115

taccgtacac cactgagacc gcggtggttg accagacaaa cctggcagcg cc 52

<210> 4116

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4116

taccgtacac cactgagacc gcggtggttg accagacaaa cctggcagcg ccgt 54

<210> 4117

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4117

taccgtacac cactgagacc gcggtggttg accagacaaa cctggcagcg ccgtga 56

<210> 4118

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4118

taccgtacac cactgagacc gcggtggttg accagacaaa cctgacctgc cc 52

<210> 4119

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4119

taccgtacac cactgagacc gcggtggttg accagacaaa cctgacctgc ccag 54

<210> 4120

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4120

taccgtacac cactgagacc gcggtggttg accagacaaa cctgacctgc ccagca 56

<210> 4121

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4121

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctctgggc ttg 53

<210> 4122

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4122

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctctgggc ttgcc 55

<210> 4123

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4123

ctctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctctgggc ttgccaa 57

<210> 4124

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4124

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctgaacac cta 53

<210> 4125

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4125

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctgaacac ctagg 55

<210> 4126

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4126

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctgaacacc taggac 56

<210> 4127

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4127

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctcacagc cct 53

<210> 4128

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4128

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctcacagc cctgc 55

<210> 4129

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4129

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctcacagcc ctgcca 56

<210> 4130

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4130

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctagaacc ggg 53

<210> 4131

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4131

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctagaacc gggca 55

<210> 4132

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4132

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctagaaccg ggcagg 56

<210> 4133

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4133

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctcacccc ctg 53

<210> 4134

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4134

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctcacccc ctgcc 55

<210> 4135

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4135

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctcaccccc tgccaa 56

<210> 4136

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4136

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctaggatc ctc 53

<210> 4137

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4137

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctaggatc ctctc 55

<210> 4138

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4138

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctaggatcc tctctg 56

<210> 4139

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4139

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa accttcctct ctg 53

<210> 4140

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4140

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa accttcctct ctggc 55

<210> 4141

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4141

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa ccttcctctc tggctc 56

<210> 4142

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4142

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctctgtcc cta 53

<210> 4143

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4143

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctctgtcc ctagt 55

<210> 4144

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4144

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctctgtccc tagtgg 56

<210> 4145

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4145

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa accttgtggg gtg 53

<210> 4146

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4146

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa accttgtggg gtgga 55

<210> 4147

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4147

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa ccttgtgggg tggagg 56

<210> 4148

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4148

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctgaatat aag 53

<210> 4149

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4149

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctgaatat aaggt 55

<210> 4150

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4150

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctgaatata aggtgg 56

<210> 4151

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4151

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctggtgtg ctg 53

<210> 4152

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4152

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctggtgtg ctggg 55

<210> 4153

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4153

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctggtgtgc tgggca 56

<210> 4154

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4154

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctccactt cca 53

<210> 4155

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4155

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctccactt ccagg 55

<210> 4156

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4156

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctccacttc caggag 56

<210> 4157

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4157

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa accttgcaca ctg 53

<210> 4158

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4158

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa accttgcaca ctggg 55

<210> 4159

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4159

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa ccttgcacac tgggaa 56

<210> 4160

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4160

atgatcctga cgacggagac cgccgtcgtc gacaagccgc agtttatc 48

<210> 4161

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4161

atgatcctga cgacggagac cgccgtcgtc gacaagccgc agtttatcag 50

<210> 4162

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4162

atgatcctga cgacggagac cgccgtcgtc gacaagccgc agtttatcag g 51

<210> 4163

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4163

atgatcctga cgacggagac cgccgtcgtc gacaagcctg agcaggt 47

<210> 4164

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4164

atgatcctga cgacggagac cgccgtcgtc gacaagcctg agcaggtag 49

<210> 4165

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4165

atgatcctga cgacggagac cgccgtcgtc gacaagcctg agcaggtaga t 51

<210> 4166

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4166

atgatcctga cgacggagac cgccgtcgtc gacaagccgg cggcagc 47

<210> 4167

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4167

atgatcctga cgacggagac cgccgtcgtc gacaagccgg cggcagcat 49

<210> 4168

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4168

atgatcctga cgacggagac cgccgtcgtc gacaagccgg cggcagcata g 51

<210> 4169

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4169

atgatcctga cgacggagac cgccgtcgtc gacaagccaa gagttgac 48

<210> 4170

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4170

atgatcctga cgacggagac cgccgtcgtc gacaagccaa gagttgacac 50

<210> 4171

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4171

atgatcctga cgacggagac cgccgtcgtc gacaagccaa gagttgacac a 51

<210> 4172

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4172

atgatcctga cgacggagac cgccgtcgtc gacaagccaa gtgatcac 48

<210> 4173

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4173

atgatcctga cgacggagac cgccgtcgtc gacaagccaa gtgatcacac 50

<210> 4174

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4174

atgatcctga cgacggagac cgccgtcgtc gacaagccaa gtgatcacac tt 52

<210> 4175

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4175

atgatcctga cgacggagac cgccgtcgtc gacaagcccc caagtga 47

<210> 4176

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4176

atgatcctga cgacggagac cgccgtcgtc gacaagcccc caagtgatc 49

<210> 4177

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4177

atgatcctga cgacggagac cgccgtcgtc gacaagcccc caagtgatca c 51

<210> 4178

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4178

atgatcctga cgacggagac cgccgtcgtc gacaagccag agacgca 47

<210> 4179

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4179

atgatcctga cgacggagac cgccgtcgtc gacaagccag agacgcaaa 49

<210> 4180

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4180

atgatcctga cgacggagac cgccgtcgtc gacaagccag agacgcaaac a 51

<210> 4181

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4181

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatct gacctg 46

<210> 4182

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4182

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatct gacctgt 47

<210> 4183

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4183

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatct gacctgttt 49

<210> 4184

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4184

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatta ctgtccc 47

<210> 4185

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4185

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatta ctgtcccct 49

<210> 4186

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4186

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatta ctgtcccctt c 51

<210> 4187

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4187

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgg gctcact 47

<210> 4188

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4188

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgg gctcactat 49

<210> 4189

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4189

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgg gctcactatg c 51

<210> 4190

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4190

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag tgggact 47

<210> 4191

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4191

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag tgggactttg 50

<210> 4192

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4192

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag tgggactttg g 51

<210> 4193

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4193

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgg actttgg 47

<210> 4194

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4194

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgg actttggaa 49

<210> 4195

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4195

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgg actttggaaa t 51

<210> 4196

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4196

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatga caatcga 47

<210> 4197

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4197

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatga caatcgata 49

<210> 4198

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4198

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatga caatcgatag g 51

<210> 4199

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4199

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatat catcttta 48

<210> 4200

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4200

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatat catctttacc 50

<210> 4201

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4201

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatat catctttacc ag 52

<210> 4202

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4202

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgc tctcatt 47

<210> 4203

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4203

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgc tctcattttc 50

<210> 4204

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4204

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgc tctcattttc ca 52

<210> 4205

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4205

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag tcagtat 47

<210> 4206

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4206

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag tcagtatca 49

<210> 4207

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4207

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatag tcagtatcaa t 51

<210> 4208

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4208

tcctgacgac ggagaccgcc gtcgtcgaca agccagagac gcaaa 45

<210> 4209

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4209

tgtcgacgac ggcggtctcc gtcgtcagga tcatgctctc attttc 46

<210> 4210

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4210

acgacggaga ccgccgtcgt cgacaagcca gagacgcaaa 40

<210> 4211

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4211

acgacggcgg tctccgtcgt caggatcatg ctctcatttt c 41

<210> 4212

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4212

atgatcctga cgacggagtc cgccgtcgtc gacaagccag agacgcaaa 49

<210> 4213

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4213

ggcttgtcga cgacggcgga ctccgtcgtc aggatcatgc tctcattttc 50

<210> 4214

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4214

tcctgacgac ggagtccgcc gtcgtcgaca agccagagac gcaaa 45

<210> 4215

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4215

tgtcgacgac ggcggactcc gtcgtcagga tcatgctctc attttc 46

<210> 4216

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4216

acgacggagt ccgccgtcgt cgacaagcca gagacgcaaa 40

<210> 4217

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4217

acgacggcgg actccgtcgt caggatcatg ctctcatttt c 41

<210> 4218

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4218

gggtttgtac cgtacaccac tgagaccgcg gtggttgacc agacaaacca agagacgcaa 60

a 61

<210> 4219

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4219

tggtttgtct ggtcaaccac cgcggtctca gtggtgtacg gtacaaaccc gctctcattt 60

tc 62

<210> 4220

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4220

tgtaccgtac accactgaga ccgcggtggt tgaccagaca aaccaagaga cgcaaa 56

<210> 4221

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4221

tgtctggtca accaccgcgg tctcagtggt gtacggtaca aacccgctct cattttc 57

<210> 4222

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4222

cgtacaccac tgagaccgcg gtggttgacc agacaaacca agagacgcaa a 51

<210> 4223

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4223

ggtcaaccac cgcggtctca gtggtgtacg gtacaaaccc gctctcattt tc 52

<210> 4224

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4224

gggtttgtac cgtacaccac tgagtccgcg gtggttgacc agacaaacca agagacgcaa 60

a 61

<210> 4225

<211> 62

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 4225

tggtttgtct ggtcaaccac cgcggactca gtggtgtacg gtacaaaccc gctctcattt 60

tc 62

<210> 4226

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4226

tgtaccgtac accactgagt ccgcggtggt tgaccagaca aaccaagaga cgcaaa 56

<210> 4227

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4227

tgtctggtca accaccgcgg actcagtggt gtacggtaca aacccgctct cattttc 57

<210> 4228

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4228

cgtacaccac tgagtccgcg gtggttgacc agacaaacca agagacgcaa a 51

<210> 4229

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4229

ggtcaaccac cgcggactca gtggtgtacg gtacaaaccc gctctcattt tc 52

<210> 4230

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4230

aacttcacag aggcttgtcg acgacggcgg tctccgtcgt caggatcat 49

<210> 4231

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4231

gaaacttcac agaggcttgt cgacgacggc ggtctccgtc gtcaggatca t 51

<210> 4232

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4232

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatct gttgatctc 49

<210> 4233

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4233

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatct gttgatctca t 51

<210> 4234

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4234

taccgtacac cactgagacc gcggtggttg accagacaaa cctgtggcag tta 53

<210> 4235

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4235

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctgtggca gtta 54

<210> 4236

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4236

taccgtacac cactgagacc gcggtggttg accagacaaa cctgtggcag tttta 55

<210> 4237

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4237

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctgtggca gtttta 56

<210> 4238

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4238

taccgtacac cactgagacc gcggtggttg accagacaaa cctcttggag gaaa 54

<210> 4239

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4239

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctcttgga ggaaa 55

<210> 4240

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4240

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctccctgt cc 52

<210> 4241

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4241

taccgtacac cactgagacc gcggtggttg accagacaaa cctccctgtc c 51

<210> 4242

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4242

tctggtcaac caccgcggtc tcagtggtgt acggtacaaa cctcctgtac ct 52

<210> 4243

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4243

taccgtacac cactgagacc gcggtggttg accagacaaa cctcctgtac ct 52

<210> 4244

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4244

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctcctacc tatcta 56

<210> 4245

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4245

taccgtacac cactgagacc gcggtggttg accagacaaa cctcctacct atcta 55

<210> 4246

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4246

gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa acctatggag taaacct 57

<210> 4247

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4247

taccgtacac cactgagacc gcggtggttg accagacaaa cctatggagt aaacc 55

<210> 4248

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4248

atgatcctga cgacggagac cgccgtcgtc gacaagccat agggtaaacc a 51

<210> 4249

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4249

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatat agggtaaacc a 51

<210> 4250

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4250

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatgt gttgcagt 48

<210> 4251

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4251

atgatcctga cgacggagac cgccgtcgtc gacaagccgt gttgcagt 48

<210> 4252

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4252

ggcttgtcga cgacggcggt ctccgtcgtc aggatcatct cactgttagg t 51

<210> 4253

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4253

atgatcctga cgacggagac cgccgtcgtc gacaagccct cactgttagg t 51

<210> 4254

<211> 70

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(48)

<223> n is a, c, g, or t

<400> 4254

acactctttc cctacacgac gctcttccga tctnnnnnnn nnnnnnnntg ctggccagtt 60

tactaacaat 70

<210> 4255

<211> 74

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(48)

<223> n is a, c, g, or t

<400> 4255

acactctttc cctacacgac gctcttccga tctnnnnnnn nnnnnnnnca gaaagaagat 60

cttatcccat cttg 74

<210> 4256

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4256

tggagttcag acgtgtgctc ttccgatctg gctacttttg ttatttgcat t 51

<210> 4257

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4257

tggagttcag acgtgtgctc ttccgatctc cagagcaggg tcccgcttc 49

<210> 4258

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4258

acactctttc cctacacgac gctcttccga tctnnnnacg gggctcagtc tgaagag 57

<210> 4259

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4259

acactctttc cctacacgac gctcttccga tctnnnngcc aaggactcaa acccaga 57

<210> 4260

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4260

tggagttcag acgtgtgctc ttccgatctt ccgtgcgtca gttttacct 49

<210> 4261

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4261

acactctttc cctacacgac gctcttccga tctnnnnaac tgcttctcct cttgggaa 58

<210> 4262

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4262

tggagttcag acgtgtgctc ttccgatctt ccttgccaga acctctaagg t 51

<210> 4263

<211> 58

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4263

acactctttc cctacacgac gctcttccga tctnnnnatc ctctctggct ccatcgta 58

<210> 4264

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4264

tggagttcag acgtgtgctc ttccgatctt ccacttcagg acagcatgtt t 51

<210> 4265

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223>

n is a, c, g, or t

<400> 4265

acactctttc cctacacgac gctcttccga tctnnnncgc cgggaactgc cgctggc 57

<210> 4266

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4266

tggagttcag acgtgtgctc ttccgatctg aggaggccct catctggcg 49

<210> 4267

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4267

acactctttc cctacacgac gctcttccga tctnnnnaat caatgtgaag caaatcgcag 60

c 61

<210> 4268

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4268

tggagttcag acgtgtgctc ttccgatctt cgattgtcag gaggatgatg aa 52

<210> 4269

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4269

acactctttc cctacacgac gctcttccga tctnnnntcc ttcttactgt ccccttctgg 60

gc 62

<210> 4270

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4270

tggagttcag acgtgtgctc ttccgatctg caaacacagc atggacgac 49

<210> 4271

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4271

acactctttc cctacacgac gctcttccga tctnnnnaca atcgataggt acctggctgt 60

c 61

<210> 4272

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4272

tggagttcag acgtgtgctc ttccgatcta ccagccccaa gatgactat 49

<210> 4273

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4273

acactctttc cctacacgac gctcttccga tctnnnnctg aaaacctgag cttggagg 58

<210> 4274

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4274

tggagttcag acgtgtgctc ttccgatctg tctactccag cttaatggaa gtgg 54

<210> 4275

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4275

acactctttc cctacacgac gctcttccga tctnnnngag aagatgtgga aatgcctcac 60

<210> 4276

<211> 53

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 4276

tggagttcag acgtgtgctc ttccgatcta atcaacatga agggttgtgt tgt 53

<210> 4277

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4277

acactctttc cctacacgac gctcttccga tctnnnngtt cccacacatg cgttcctc 58

<210> 4278

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4278

tggagttcag acgtgtgctc ttccgatctg gcatgaaggg ttgtttttaa ttga 54

<210> 4279

<211> 68

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(48)

<223> n is a, c, g, or t

<400> 4279

acactctttc cctacacgac gctcttccga tctnnnnnnn nnnnnnnngc cctgtcctgt 60

acctacct 68

<210> 4280

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 4280

acactctttc cctacacgac gctcttccga tctnnnnttc agtttggggt atgtgcca 58

<210> 4281

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4281

tggagttcag acgtgtgctc ttccgatctc tcaaatttac ccgtggcagc 50

<210> 4282

<211> 5558

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4282

gatgccagct cattcctccc actcatgatc tatagatccc ccgggctgca ggaattctac 60

ccactctgtc gataccccac cgagacccca ttggggccaa tacgcccgcg tttcttcctt 120

ttccccaccc cacccccccaa gttcgggtga aggcccaggg ctcgcagcca acgtcggggc 180

ggcaagctta catcgagatc ccggcttgtc gacgacggcg gactccgtcg tcaggatcat 240

ccgtgagcaa gggcgaggag ctgttcaccg gggtggtgcc catcctggtc gagctggacg 300

gcgacgtaaa cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg 360

gcaagctgac cctgaagttc atctgcacca ccggcaagct gcccgtgccc tggcccaccc 420

tcgtgaccac cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc 480

agcacgactt cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct 540

tcaaggacga cggcaactac aagacccgcg ccgaggtgaa gttcgagggc gacaccctgg 600

tgaaccgcat cgagctgaag ggcatcgact tcaaggagga cggcaacatc ctggggcaca 660

agctggagta caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg 720

gcatcaaggt gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg 780

accactacca gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact 840

acctgagcac ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatggtcc 900

tgctggagtt cgtgaccgcc gccgggatca ctctcggcat ggacgagctg tacaagagcg 960

gcctgaggag cagagcccag gcgagcaaca gcgccgtgga cgccaccatg ggcgatcgcc 1020

cgggaattga ctagtgcggc cgcctaggat ccatgccgat agcgttggtt gagtggataa 1080

ccgtattacc gccaagctta tgcatgtgcc cgtcagtggg cagagcgcac atcgcccaca 1140

gtccccgaga agttgggggg aggggtcggc aattgaaccg gtgcctagag aaggtggcgc 1200

ggggtaaact gggaaagtga tgtcgtgtac tggctccgcc tttttcccga gggtggggga 1260

gaaccgtata taagtgcagt agtcgccgtg aacgttcttt ttcgcaacgg gtttgccgcc 1320

agaacacagg taagtgccgt gtgtggttcc cgcgggcctg gcctctttac gggttatggc 1380

ccttgcgtgc cttgaattac ttccacctgg ctgcagtacg tgattcttga tcccgagctt 1440
cgggttggaa gtgggtggga gagttcgagg ccttgcgctt aaggagcccc ttcgcctcgt 1500
gcttgagttg aggcctggcc tgggcgctgg ggccgccgcg tgcgaatctg gtggcacctt 1560
cgcgcctgtc tcgctgcttt cgataagtct ctagccattt aaaatttttg atgacctgct 1620
gcgacgcttt ttttctggca agatagtctt gtaaatgcgg gccaagatct gcacactggt 1680
atttcggttt ttggggccgc gggcggcgac ggggcccgtg cgtcccagcg cacatgttcg 1740
gcgaggcggg gcctgcgagc gcggccaccg agaatcggac gggggtagtc tcaagctggc 1800

cggcctgctc tggtgcctgg cctcgcgccg ccgtgtatcg ccccgccctg ggcggcaagg 1860
ctggcccggt cggcaccagt tgcgtgagcg gaaagatggc cgcttcccgg ccctgctgca 1920
gggagctcaa aatggaggac gcggcgctcg ggagagcggg cgggtgagtc acccacacaa 1980
aggaaaaggg cctttccgtc ctcagccgtc gcttcatgtg actccacgga gtaccgggcg 2040
ccgtccaggc acctcgatta gttctcgcgc ttttggagta cgtcgtcttt aggttggggg 2100
gaggggtttt atgcgatgga gtttccccac actgagtggg tggagactga agttaggcca 2160
gcttggcact tgatgtaatt ctccttggaa tttgcccttt ttgagtttgg atcttggttc 2220

attctcaagc ctcagacagt ggttcaaagt ttttttcttc catttcaggt gtcgtgagct 2280
agcccaccat gaccgagtac aagcccacgg tgcgcctcgc cacccgcgac gacgtccccc 2340
gggccgtacg caccctcgcc gccgcgttcg ccgactaccc cgccacgcgc cacaccgtcg 2400
acccggaccg ccacatcgag cgggtcaccg agctgcaaga actcttcctc acgcgcgtcg 2460
ggctcgacat cggcaaggtg tgggtcgcgg acgacggcgc cgcggtggcg gtctggacca 2520
cgccggagag cgtcgaagcg ggggcggtgt tcgccgagat cggcccgcgc atggccgagt 2580
tgagcggttc ccggctggcc gcgcagcaac agatggaagg cctcctggcg ccgcaccggc 2640

ccaaggagcc cgcgtggttc ctggccaccg tcggcgtctc gcccgaccac cagggcaagg 2700
gtctgggcag cgccgtcgtg ctccccggag tggaggcggc cgagcgcgcc ggggtgcccg 2760
ccttcctgga gacctccgcg ccccgcaacc tccccttcta cgagcggctc ggcttcaccg 2820
tcaccgccga cgtcgaggtg cccgaaggac cgcgcacctg gtgcatgacc cgcaagcccg 2880
gtgccggatc gggagagggc agaggaagtc tgctaacatg cggtgacgtc gaggagaatc 2940
ctggcccacc ggtcgccacc agcgagctga ttaaggagaa catgcacatg aagctgtaca 3000
tggagggcac cgtggacaac catcacttca agtgcacatc cgagggcgaa ggcaagccct 3060

acgagggcac ccagaccatg agaatcaagg tggtcgaggg cggccctctc cccttcgcct 3120

tcgacatcct ggctactagc ttcctctacg gcagcaagac cttcatcaac cacacccagg 3180

gcatccccga cttcttcaag cagtccttcc ctgagggctt cacatgggag agagtcacca 3240

catacgaaga cgggggcgtg ctgaccgcta cccaggacac cagcctccag gacggctgcc 3300

tcatctacaa cgtcaagatc agaggggtga acttcacatc caacggccct gtgatgcaga 3360

agaaaacact cggctgggag gccttcaccg agacgctgta ccccgctgac ggcggcctgg 3420

aaggcagaaa cgacatggcc ctgaagctcg tgggcgggag ccatctgatc gcaaacatca 3480

agaccacata tagatccaag aaacccgcta agaacctcaa gatgcctggc gtctactatg 3540

tggactacag actggaaaga atcaaggagg ccaacaacga gacctacgtc gagcagcacg 3600

aggtggcagt ggccagatac tgcgacctcc ctagcaaact ggggcacaag cttaattaag 3660

aattctctag aggatccaga catgataaga tacattgatg agtttggaca aaccacaact 3720

agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc tttatttgta 3780

accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt tatgtttcag 3840

gttcaggggg aggtgtggga ggtttttaa agcaagtaaa acctctacaa atgtggtatg 3900

gctgattatg atcctgcaag cctcgtcgcc gcggtttatt ctgttgacaa ttaatcatcg 3960

gcatagtata tcggcatagt ataatacgac aaggtgagga actaaaccat gggatcggcc 4020

attgaacaag atggattgca cgcaggttct ccggccgctt gggtggagag gctattcggc 4080

tatgactggg cacaacagac gatcggctgc tctgatgccg ccgtgttccg gctgtcagcg 4140

caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa tgaactgcag 4200

gacgaggcag cgcggctatc gtggctggcc acgacgggcg ttccttgcgc agctgtgctc 4260

gacgttgtca ctgaagcggg aagggactgg ctgctattgg gcgaagtgcc ggggcaggat 4320

ctcctgtcat ctcaccttgc tcctgccgag aaagtatcca tcatggctga tgcaatgcgg 4380

cggctgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa acatcgcatc 4440

gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct ggacgaagag 4500

catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcgcat gcccgacggc 4560

gaggatctcg tcgtgaccca tggcgatgcc tgcttgccga atatcatggt ggaaaatggc 4620

cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata 4680

gcgttggcta cccgtgatat tgctgaagag cttggcggcg aatgggctga ccgcttcctc 4740

gtgctttacg gtatcgccgc ccccgattcg cagcgcatcg ccttctatcg ccttcttgac 4800

gagttcttct gagcgggact ctggggttcg aataaagacc gaccaagcga cgtctgagag 4860

ctccctggcg aattcggtac caataaaaga gctttatttt catgatctgt gtgttggttt 4920

ttggccgcgt tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat 4980

cgacgctcaa gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc 5040

cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc 5100

gcctttctcc cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt 5160

tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac 5220

cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg 5280

ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca 5340

gagttcttga agtggtggcc taactacggc tacactagaa gaacagtatt tggtatctgc 5400

gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa 5460

accaccgctg gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa 5520

ggatctcaag aagatccttt gatcttttct agtgtgcg 5558

<210> 4283

<211> 5558

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4283

gatgccagct cattcctccc actcatgatc tatagatccc ccgggctgca ggaattctac 60

ccactctgtc gataccccac cgagacccca ttggggccaa tacgcccgcg tttcttcctt 120

ttccccaccc caccccccaa gttcgggtga aggcccaggg ctcgcagcca acgtcggggc 180

ggcaagctta catcgagatc ccggcttgtc gacgacggcg gtctccgtcg tcaggatcat 240

ccgtgagcaa gggcgaggag ctgttcaccg gggtggtgcc catcctggtc gagctggacg 300

gcgacgtaaa cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg 360

gcaagctgac cctgaagttc atctgcacca ccggcaagct gcccgtgccc tggcccaccc 420

tcgtgaccac cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc 480

agcacgactt cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct 540

tcaaggacga cggcaactac aagacccgcg ccgaggtgaa gttcgagggc gacaccctgg 600

tgaaccgcat cgagctgaag ggcatcgact tcaaggagga cggcaacatc ctggggcaca 660

agctggagta caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg 720

gcatcaaggt gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg 780

accactacca gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact 840

acctgagcac ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatggtcc 900

tgctggagtt cgtgaccgcc gccgggatca ctctcggcat ggacgagctg tacaagagcg 960

gcctgaggag cagagcccag gcgagcaaca gcgccgtgga cgccaccatg ggcgatcgcc 1020

cgggaattga ctagtgcggc cgcctaggat ccatgccgat agcgttggtt gagtggataa 1080

ccgtattacc gccaagctta tgcatgtgcc cgtcagtggg cagagcgcac atcgcccaca 1140

gtccccgaga agttgggggg aggggtcggc aattgaaccg gtgcctagag aaggtggcgc 1200

ggggtaaact gggaaagtga tgtcgtgtac tggctccgcc tttttcccga gggtggggga 1260

gaaccgtata taagtgcagt agtcgccgtg aacgttcttt ttcgcaacgg gtttgccgcc 1320

agaacacagg taagtgccgt gtgtggttcc cgcgggcctg gcctctttac gggttatggc 1380

ccttgcgtgc cttgaattac ttccacctgg ctgcagtacg tgattcttga tcccgagctt 1440

cgggttggaa gtgggtggga gagttcgagg ccttgcgctt aaggagcccc ttcgcctcgt 1500

gcttgagttg aggcctggcc tgggcgctgg ggccgccgcg tgcgaatctg gtggcacctt 1560

cgcgcctgtc tcgctgcttt cgataagtct ctagccattt aaaatttttg atgacctgct 1620

gcgacgcttt ttttctggca agatagtctt gtaaatgcgg gccaagatct gcacactggt 1680

atttcggttt ttggggccgc gggcggcgac ggggcccgtg cgtcccagcg cacatgttcg 1740

gcgaggcggg gcctgcgagc gcggccaccg agaatcggac gggggtagtc tcaagctggc 1800

cggcctgctc tggtgcctgg cctcgcgccg ccgtgtatcg ccccgccctg ggcggcaagg 1860

ctggcccggt cggcaccagt tgcgtgagcg gaaagatggc cgcttcccgg ccctgctgca 1920

gggagctcaa aatggaggac gcggcgctcg ggagagcggg cgggtgagtc acccacacaa 1980

aggaaaaggg cctttccgtc ctcagccgtc gcttcatgtg actccacgga gtaccgggcg 2040

ccgtccaggc acctcgatta gttctcgcgc ttttggagta cgtcgtcttt aggttggggg 2100

gaggggtttt atgcgatgga gtttccccac actgagtggg tggagactga agttaggcca 2160

gcttggcact tgatgtaatt ctccttggaa tttgcccttt ttgagtttgg atcttggttc 2220

attctcaagc ctcagacagt ggttcaaagt ttttttcttc catttcaggt gtcgtgagct 2280

agcccaccat gaccgagtac aagcccacgg tgcgcctcgc cacccgcgac gacgtccccc 2340

gggccgtacg cacctcgccc gccgcgttcg ccgactaccc cgccacgcgc cacaccgtcg 2400

acccggaccg ccacatcgag cgggtcaccg agctgcaaga actcttcctc acgcgcgtcg 2460

ggctcgacat cggcaaggtg tgggtcgcgg acgacggcgc cgcggtggcg gtctggacca 2520

cgccggagag cgtcgaagcg ggggcggtgt tcgccgagat cggcccgcgc atggccgagt 2580

tgagcggttc ccggctggcc gcgcagcaac agatggaagg cctcctggcg ccgcaccggc 2640

ccaaggagcc cgcgtggttc ctggccaccg tcggcgtctc gcccgaccac cagggcaagg 2700

gtctgggcag cgccgtcgtg ctccccggag tggaggcggc cgagcgcgcc ggggtgcccg 2760

ccttcctgga gacctccgcg ccccgcaacc tccccttcta cgagcggctc ggcttcaccg 2820

tcaccgccga cgtcgaggtg cccgaaggac cgcgcacctg gtgcatgacc cgcaagcccg 2880

gtgccggatc gggagagggc agaggaagtc tgctaacatg cggtgacgtc gaggagaatc 2940

ctggcccacc ggtcgccacc agcgagctga ttaaggagaa catgcacatg aagctgtaca 3000

tggagggcac cgtggacaac catcacttca agtgcacatc cgagggcgaa ggcaagccct 3060

acgagggcac ccagaccatg agaatcaagg tggtcgaggg cggccctctc cccttcgcct 3120

tcgacatcct ggctactagc ttcctctacg gcagcaagac cttcatcaac cacacccagg 3180

gcatccccga cttcttcaag cagtccttcc ctgagggctt cacatgggag agagtcacca 3240

catacgaaga cgggggcgtg ctgaccgcta cccaggacac cagcctccag gacggctgcc 3300

tcatctacaa cgtcaagatc agaggggtga acttcacatc caacggccct gtgatgcaga 3360

agaaaacact cggctgggag gccttcaccg agacgctgta ccccgctgac ggcggcctgg 3420

aaggcagaaa cgacatggcc ctgaagctcg tgggcgggag ccatctgatc gcaaacatca 3480

agaccacata tagatccaag aaacccgcta agaacctcaa gatgcctggc gtctactatg 3540

tggactacag actggaaaga atcaaggagg ccaacaacga gacctacgtc gagcagcacg 3600

aggtggcagt ggccagatac tgcgacctcc ctagcaaact ggggcacaag cttaattaag 3660

aattctctag aggatccaga catgataaga tacattgatg agtttggaca aaccacaact 3720

agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc tttatttgta 3780

accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt tatgtttcag 3840

gttcaggggg aggtgtggga ggtttttaa agcaagtaaa acctctacaa atgtggtatg 3900

gctgattatg atcctgcaag cctcgtcgcc gcggtttatt ctgttgacaa ttaatcatcg 3960

gcatagtata tcggcatagt ataatacgac aaggtgagga actaaaccat gggatcggcc 4020

attgaacaag atggattgca cgcaggttct ccggccgctt gggtggagag gctattcggc 4080

tatgactggg cacaacagac gatcggctgc tctgatgccg ccgtgttccg gctgtcagcg 4140

caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa tgaactgcag 4200

gacgaggcag cgcggctatc gtggctggcc acgacgggcg ttccttgcgc agctgtgctc 4260

gacgttgtca ctgaagcggg aagggactgg ctgctattgg gcgaagtgcc ggggcaggat 4320

ctcctgtcat ctcaccttgc tcctgccgag aaagtatcca tcatggctga tgcaatgcgg 4380

cggctgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa acatcgcatc 4440

gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct ggacgaagag 4500

catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcgcat gcccgacggc 4560

gaggatctcg tcgtgaccca tggcgatgcc tgcttgccga atatcatggt ggaaaatggc 4620
cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata 4680
gcgttggcta cccgtgatat tgctgaagag cttggcggcg aatgggctga ccgcttcctc 4740
gtgctttacg gtatcgccgc ccccgattcg cagcgcatcg ccttctatcg ccttcttgac 4800
gagttcttct gagcgggact ctggggttcg aataaagacc gaccaagcga cgtctgagag 4860
ctccctggcg aattcggtac caataaaaga gctttatttt catgatctgt gtgttggttt 4920
ttggccgcgt tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat 4980

cgacgctcaa gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc 5040
cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc 5100
gcctttctcc cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt 5160
tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac 5220
cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg 5280
ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca 5340
gagttcttga agtggtggcc taactacggc tacactagaa gaacagtatt tggtatctgc 5400

gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa 5460
accaccgctg gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa 5520
ggatctcaag aagatccttt gatcttttct agtgtgcg 5558
<210> 4284
<211> 5568
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 4284
gatgccagct cattcctccc actcatgatc tatagatccc ccgggctgca ggaattctac 60
ccactctgtc gataccccac cgagacccca ttggggccaa tacgcccgcg tttcttcctt 120
ttccccaccc cacccccaa gttcgggtga aggcccaggg ctcgcagcca acgtcggggc 180

ggcaagctta catcgagatc ccggtttgtc tggtcaacca ccgcggactc agtggtgtac 240
ggtacaaacc ccgtgagcaa gggcgaggag ctgttcaccg gggtggtgcc catcctggtc 300
gagctggacg gcgacgtaaa cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat 360
gccacctacg gcaagctgac cctgaagttc atctgcacca ccggcaagct gcccgtgccc 420

tggcccaccc tcgtgaccac cctgacctac ggcgtgcagt gcttcagccg ctaccccgac 480

cacatgaagc agcacgactt cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc 540

accatcttct tcaaggacga cggcaactac aagacccgcg ccgaggtgaa gttcgagggc 600

gacaccctgg tgaaccgcat cgagctgaag ggcatcgact tcaaggagga cggcaacatc 660

ctggggcaca agctggagta caactacaac agccacaacg tctatatcat ggccgacaag 720

cagaagaacg gcatcaaggt gaacttcaag atccgccaca acatcgagga cggcagcgtg 780

cagctcgccg accactacca gcagaacacc cccatcggcg acggccccgt gctgctgccc 840

gacaaccact acctgagcac ccagtccgcc ctgagcaaag accccaacga gaagcgcgat 900

cacatggtcc tgctggagtt cgtgaccgcc gccgggatca ctctcggcat ggacgagctg 960

tacaagagcg gcctgaggag cagagcccag gcgagcaaca gcgccgtgga cgccaccatg 1020

ggcgatcgcc cgggaattga ctagtgcggc cgcctaggat ccatgccgat agcgttggtt 1080

gagtggataa ccgtattacc gccaagctta tgcatgtgcc cgtcagtggg cagagcgcac 1140

atcgcccaca gtccccgaga agttgggggg aggggtcggc aattgaaccg gtgcctagag 1200

aaggtggcgc ggggtaaact gggaaagtga tgtcgtgtac tggctccgcc tttttcccga 1260

gggtggggga gaaccgtata taagtgcagt agtcgccgtg aacgttcttt ttcgcaacgg 1320

gtttgccgcc agaacacagg taagtgccgt gtgtggttcc cgcgggcctg gcctctttac 1380

gggttatggc ccttgcgtgc cttgaattac ttccacctgg ctgcagtacg tgattcttga 1440

tcccgagctt cgggttggaa gtgggtggga gagttcgagg ccttgcgctt aaggagcccc 1500

ttcgcctcgt gcttgagttg aggcctggcc tgggcgctgg ggccgccgcg tgcgaatctg 1560

gtggcacctt cgcgcctgtc tcgctgcttt cgataagtct ctagccattt aaaatttttg 1620

atgacctgct gcgacgcttt ttttctggca agatagtctt gtaaatgcgg gccaagatct 1680

gcacactggt atttcggttt ttggggccgc gggcggcgac ggggcccgtg cgtcccagcg 1740

cacatgttcg gcgaggcggg gcctgcgagc gcggccaccg agaatcggac gggggtagtc 1800

tcaagctggc cggcctgctc tggtgcctgg cctcgcgccg ccgtgtatcg ccccgccctg 1860

ggcggcaagg ctggcccggt cggcaccagt tgcgtgagcg gaaagatggc cgcttcccgg 1920

ccctgctgca gggagctcaa aatggaggac gcggcgctcg ggagagcggg cgggtgagtc 1980

acccacacaa aggaaaaggg cctttccgtc ctcagccgtc gcttcatgtg actccacgga 2040

gtaccgggcg ccgtccaggc acctcgatta gttctcgcgc ttttggagta cgtcgtcttt 2100

aggttggggg gaggggtttt atgcgatgga gtttccccac actgagtggg tggagactga 2160

agttaggcca gcttggcact tgatgtaatt ctccttggaa tttgcccttt ttgagtttgg 2220

atcttggttc attctcaagc ctcagacagt ggttcaaagt ttttttcttc catttcaggt 2280

gtcgtgagct agcccaccat gaccgagtac aagcccacgg tgcgcctcgc cacccgcgac 2340
gacgtccccc gggccgtacg caccctcgcc gccgcgttcg ccgactaccc cgccacgcgc 2400
cacaccgtcg acccggaccg ccacatcgag cgggtcaccg agctgcaaga actcttcctc 2460
acgcgcgtcg ggctcgacat cggcaaggtg tgggtcgcgg acgacggcgc cgcggtggcg 2520
gtctggacca cgccggagag cgtcgaagcg ggggcggtgt tcgccgagat cggcccgcgc 2580
atggccgagt tgagcggttc ccggctggcc gcgcagcaac agatggaagg cctcctggcg 2640
ccgcaccggc ccaaggagcc cgcgtggttc ctggccaccg tcggcgtctc gcccgaccac 2700

cagggcaagg gtctgggcag cgccgtcgtg ctccccggag tggaggcggc cgagcgcgcc 2760
ggggtgcccg ccttcctgga gacctccgcg ccccgcaacc tccccttcta cgagcggctc 2820
ggcttcaccg tcaccgccga cgtcgaggtg cccgaaggac cgcgcacctg gtgcatgacc 2880
cgcaagcccg gtgccggatc gggagagggc agaggaagtc tgctaacatg cggtgacgtc 2940
gaggagaatc ctggcccacc ggtcgccacc agcgagctga ttaaggagaa catgcacatg 3000
aagctgtaca tggagggcac cgtggacaac catcacttca agtgcacatc cgagggcgaa 3060
ggcaagccct acgagggcac ccagaccatg agaatcaagg tggtcgaggg cggccctctc 3120

cccttcgcct tcgacatcct ggctactagc ttcctctacg gcagcaagac cttcatcaac 3180
cacacccagg gcatccccga cttcttcaag cagtccttcc ctgagggctt cacatgggag 3240
agagtcacca catacgaaga cgggggcgtg ctgaccgcta cccaggacac cagcctccag 3300
gacggctgcc tcatctacaa cgtcaagatc agaggggtga acttcacatc caacggccct 3360
gtgatgcaga agaaaacact cggctgggag gccttcaccg agacgctgta ccccgctgac 3420
ggcggcctgg aaggcagaaa cgacatggcc ctgaagctcg tgggcgggag ccatctgatc 3480
gcaaacatca agaccacata tagatccaag aaacccgcta agaacctcaa gatgcctggc 3540

gtctactatg tggactacag actggaaaga atcaaggagg ccaacaacga gacctacgtc 3600
gagcagcacg aggtggcagt ggccagatac tgcgacctcc ctagcaaact ggggcacaag 3660
cttaattaag aattctctag aggatccaga catgataaga tacattgatg agtttggaca 3720
aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc 3780
tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt 3840
tatgtttcag gttcaggggg aggtgtggga ggtttttttaa agcaagtaaa acctctacaa 3900
atgtggtatg gctgattatg atcctgcaag cctcgtcgcc gcggtttatt ctgttgacaa 3960

ttaatcatcg gcatagtata tcggcatagt ataatacgac aaggtgagga actaaaccat 4020

gggatcggcc attgaacaag atggattgca cgcaggttct ccggccgctt gggtggagag 4080

gctattcggc tatgactggg cacaacagac gatcggctgc tctgatgccg ccgtgttccg 4140

gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa 4200

tgaactgcag gacgaggcag cgcggctatc gtggctggcc acgacgggcg ttccttgcgc 4260

agctgtgctc gacgttgtca ctgaagcggg aagggactgg ctgctattgg gcgaagtgcc 4320

ggggcaggat ctcctgtcat ctcaccttgc tcctgccgag aaagtatcca tcatggctga 4380

tgcaatgcgg cggctgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa 4440

acatcgcatc gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct 4500

ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcgcat 4560

gcccgacggc gaggatctcg tcgtgaccca tggcgatgcc tgcttgccga atatcatggt 4620

ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta 4680

tcaggacata gcgttggcta cccgtgatat tgctgaagag cttggcggcg aatgggctga 4740

ccgcttcctc gtgctttacg gtatcgccgc ccccgattcg cagcgcatcg ccttctatcg 4800

ccttcttgac gagttcttct gagcgggact ctggggttcg aataaagacc gaccaagcga 4860

cgtctgagag ctccctggcg aattcggtac caataaaaga gctttatttt catgatctgt 4920

gtgttggttt ttggccgcgt tgctggcgtt tttccatagg ctccgccccc ctgacgagca 4980

tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg acaggactat aaagatacca 5040

ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg 5100

atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcatagct cacgctgtag 5160

gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt 5220

tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca 5280

cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg 5340

cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa gaacagtatt 5400

tggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc 5460

cggcaaacaa accaccgctg gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg 5520

cagaaaaaaa ggatctcaag aagatccttt gatcttttct agtgtgcg 5568

<210> 4285

<211> 5568

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4285

gatgccagct cattcctccc actcatgatc tatagatccc ccgggctgca ggaattctac 60

ccactctgtc gataccccac cgagacccca ttggggccaa tacgcccgcg tttcttcctt 120

ttccccaccc cacccccccaa gttcgggtga aggcccaggg ctcgcagcca acgtcggggc 180

ggcaagctta catcgagatc ccggtttgtc tggtcaacca ccgcggtctc agtggtgtac 240

ggtacaaacc ccgtgagcaa gggcgaggag ctgttcaccg gggtggtgcc catcctggtc 300

gagctggacg gcgacgtaaa cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat 360

gccacctacg gcaagctgac cctgaagttc atctgcacca ccggcaagct gcccgtgccc 420

tggcccaccc tcgtgaccac cctgacctac ggcgtgcagt gcttcagccg ctaccccgac 480

cacatgaagc agcacgactt cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc 540

accatcttct tcaaggacga cggcaactac aagacccgcg ccgaggtgaa gttcgagggc 600

gacaccctgg tgaaccgcat cgagctgaag ggcatcgact tcaaggagga cggcaacatc 660

ctggggcaca agctggagta caactacaac agccacaacg tctatatcat ggccgacaag 720

cagaagaacg gcatcaaggt gaacttcaag atccgccaca acatcgagga cggcagcgtg 780

cagctcgccg accactacca gcagaacacc cccatcggcg acggccccgt gctgctgccc 840

gacaaccact acctgagcac ccagtccgcc ctgagcaaag accccaacga gaagcgcgat 900

cacatggtcc tgctggagtt cgtgaccgcc gccgggatca ctctcggcat ggacgagctg 960

tacaagagcg gcctgaggag cagagcccag gcgagcaaca gcgccgtgga cgccaccatg 1020

ggcgatcgcc cgggaattga ctagtgcggc cgcctaggat ccatgccgat agcgttggtt 1080

gagtggataa ccgtattacc gccaagctta tgcatgtgcc cgtcagtggg cagagcgcac 1140

atcgcccaca gtccccgaga agttgggggg aggggtcggc aattgaaccg gtgcctagag 1200

aaggtggcgc ggggtaaact gggaaagtga tgtcgtgtac tggctccgcc tttttcccga 1260

gggtggggga gaaccgtata taagtgcagt agtcgccgtg aacgttcttt ttcgcaacgg 1320

gtttgccgcc agaacacagg taagtgccgt gtgtggttcc cgcgggcctg gcctctttac 1380

gggttatggc ccttgcgtgc cttgaattac ttccacctgg ctgcagtacg tgattcttga 1440

tcccgagctt cgggttggaa gtgggtggga gagttcgagg ccttgcgctt aaggagcccc 1500

ttcgcctcgt gcttgagttg aggcctggcc tgggcgctgg ggccgccgcg tgcgaatctg 1560

gtggcacctt cgcgcctgtc tcgctgcttt cgataagtct ctagccattt aaaatttttg 1620

atgacctgct gcgacgcttt ttttctggca agatagtctt gtaaatgcgg gccaagatct 1680

gcacactggt atttcggttt ttggggccgc gggcggcgac ggggcccgtg cgtcccagcg 1740

cacatgttcg gcgaggcggg gcctgcgagc gcggccaccg agaatcggac gggggtagtc 1800

tcaagctggc cggcctgctc tggtgcctgg cctcgcgccg ccgtgtatcg ccccgccctg 1860

ggcggcaagg ctggcccggt cggcaccagt tgcgtgagcg gaaagatggc cgcttcccgg 1920

ccctgctgca gggagctcaa aatggaggac gcggcgctcg ggagagcggg cgggtgagtc 1980

acccacacaa aggaaaaggg cctttccgtc ctcagccgtc gcttcatgtg actccacgga 2040

gtaccgggcg ccgtccaggc acctcgatta gttctcgcgc ttttggagta cgtcgtcttt 2100

aggttggggg gaggggtttt atgcgatgga gtttccccac actgagtggg tggagactga 2160

agttaggcca gcttggcact tgatgtaatt ctccttggaa tttgcccttt ttgagtttgg 2220

atcttggttc attctcaagc ctcagacagt ggttcaaagt ttttttcttc catttcaggt 2280

gtcgtgagct agcccaccat gaccgagtac aagcccacgg tgcgcctcgc cacccgcgac 2340

gacgtccccc gggccgtacg caccctcgcc gccgcgttcg ccgactaccc cgccacgcgc 2400

cacaccgtcg acccggaccg ccacatcgag cgggtcaccg agctgcaaga actcttcctc 2460

acgcgcgtcg ggctcgacat cggcaaggtg tgggtcgcgg acgacggcgc cgcggtggcg 2520

gtctggacca cgccggagag cgtcgaagcg ggggcggtgt tcgccgagat cggcccgcgc 2580

atggccgagt tgagcggttc ccggctggcc gcgcagcaac agatggaagg cctcctggcg 2640

ccgcaccggc ccaaggagcc cgcgtggttc ctggccaccg tcggcgtctc gcccgaccac 2700

cagggcaagg gtctgggcag cgccgtcgtg ctccccggag tggaggcggc cgagcgcgcc 2760

ggggtgcccg ccttcctgga gacctccgcg ccccgcaacc tccccttcta cgagcggctc 2820

ggcttcaccg tcaccgccga cgtcgaggtg cccgaaggac cgcgcacctg gtgcatgacc 2880

cgcaagcccg gtgccggatc gggagagggc agaggaagtc tgctaacatg cggtgacgtc 2940

gaggagaatc ctggcccacc ggtcgccacc agcgagctga ttaaggagaa catgcacatg 3000

aagctgtaca tggagggcac cgtggacaac catcacttca agtgcacatc cgagggcgaa 3060

ggcaagccct acgagggcac ccagaccatg agaatcaagg tggtcgaggg cggccctctc 3120

cccttcgcct tcgacatcct ggctactagc ttcctctacg gcagcaagac cttcatcaac 3180

cacacccagg gcatccccga cttcttcaag cagtccttcc ctgagggctt cacatgggag 3240

agagtcacca catacgaaga cgggggcgtg ctgaccgcta cccaggacac cagcctccag 3300

gacggctgcc tcatctacaa cgtcaagatc agaggggtga acttcacatc caacggccct 3360

gtgatgcaga agaaaacact cggctgggag gccttcaccg agacgctgta ccccgctgac 3420

ggcggcctgg aaggcagaaa cgacatggcc ctgaagctcg tgggcgggag ccatctgatc 3480

gcaaacatca agaccacata tagatccaag aaacccgcta agaacctcaa gatgcctggc 3540

gtctactatg tggactacag actggaaaga atcaaggagg ccaacaacga gacctacgtc 3600

gagcagcacg aggtggcagt ggccagatac tgcgacctcc ctagcaaact ggggcacaag 3660

cttaattaag aattctctag aggatccaga catgataaga tacattgatg agtttggaca 3720

aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc 3780

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt 3840

tatgtttcag gttcaggggg aggtgtggga ggttttttaa agcaagtaaa acctctacaa 3900

atgtggtatg gctgattatg atcctgcaag cctcgtcgcc gcggtttatt ctgttgacaa 3960

ttaatcatcg gcatagtata tcggcatagt ataatacgac aaggtgagga actaaaccat 4020

gggatcggcc attgaacaag atggattgca cgcaggttct ccggccgctt gggtggagag 4080

gctattcggc tatgactggg cacaacagac gatcggctgc tctgatgccg ccgtgttccg 4140

gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa 4200

tgaactgcag gacgaggcag cgcggctatc gtggctggcc acgacgggcg ttccttgcgc 4260

agctgtgctc gacgttgtca ctgaagcggg aagggactgg ctgctattgg gcgaagtgcc 4320

ggggcaggat ctcctgtcat ctcaccttgc tcctgccgag aaagtatcca tcatggctga 4380

tgcaatgcgg cggctgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa 4440

acatcgcatc gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct 4500

ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcgcat 4560

gcccgacggc gaggatctcg tcgtgaccca tggcgatgcc tgcttgccga atatcatggt 4620

ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta 4680

tcaggacata gcgttggcta cccgtgatat tgctgaagag cttggcggcg aatgggctga 4740

ccgcttcctc gtgctttacg gtatcgccgc ccccgattcg cagcgcatcg ccttctatcg 4800

ccttcttgac gagttcttct gagcgggact ctggggttcg aataaagacc gaccaagcga 4860

cgtctgagag ctccctggcg aattcggtac caataaaaga gctttatttt catgatctgt 4920

gtgttggttt ttggccgcgt tgctggcgtt tttccatagg ctccgccccc ctgacgagca 4980

tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg acaggactat aaagatacca 5040

ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg 5100

atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcatagct cacgctgtag 5160

gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aacccccgt 5220

tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca 5280

cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg 5340

cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa gaacagtatt 5400

tggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc 5460

cggcaaacaa accaccgctg gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg 5520

cagaaaaaaa ggatctcaag aagatccttt gatcttttct agtgtgcg 5568

<210> 4286

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4286

tggggcccag actgagcacg aacaccatca tcaccatcat 40

<210> 4287

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4287

tggggcccag actgagcacg accaccatca tcaccatcat 40

<210> 4288

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4288

tggggcccag actgagcacg aactccatca tcaccatcat 40

<210> 4289

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4289

tggggcccag actgagcacg aacaccctca tcaccatcat 40

<210> 4290

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4290

tggggcccag actgagcacg cacaccatca tcaccatcat 40

<210> 4291

<211> 131

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4291

ggcccagact gagcacgtga gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgcaaat gatggtgatg atggtgttcg 120

tgctcagtct g 131

<210> 4292

<211> 130

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4292

gtcaaccagt atcccggtgc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgcaaca ccatcatcac catcatttcc 120

gggatactgg 130

<210> 4293

<211> 37

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4293

cacuggaucg aucgaugcug acuguagcua cagcuca 37

<210> 4294

<211> 37

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4294

gucagcaucg aucgauccag uguguagcua cagcuca 37

<210> 4295

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4295

tgagctgtag ctacagtcag catcgatcga tccagtg 37

<210> 4296

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4296

tgagctgtag ctacacactg gatcgatcga tgctgac 37

<210> 4297

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4297

acatcgatgt cgagtcactg gatcgatcga tgctgac 37

<210> 4298

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4298

gagcagagcc aggaacccca ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gccagcggca gttcccggcg 100

<210> 4299

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4299

gagcagagcc aggaacccca ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gcccgcggca gttcccggcg 100

<210> 4300

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4300

gagcagagcc aggaacccca ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gccagcggcc gttcccggcg 100

<210> 4301

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4301

gagcagagcc aggaacccca ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggtgg gccagcggca gttcccggcg 100

<210> 4302

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4302

gagcagagcc aggaacccct ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gccagcggca gttcccggcg 100

<210> 4303

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4303

gagcagagcc aggaacccca ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gccagcggca ggtcccggcg 100

<210> 4304

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4304

gagcagagcc aggaacccca ggtttgtctg gtcacccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gccagcggca gttcccggcg 100

<210> 4305

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4305

gagcagagcc aggaacccca ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gccagctgca gttcccggcg 100

<210> 4306

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4306

gagcagagcc aggaacccca ggtttgtctg gtcaaccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggcggggg gccagcggca gttcccggcg 100

<210> 4307

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4307

gagcagagcc aggaacccca ggtttgtctg gtcacccacc gcggtctcag tggtgtacgg 60

tacaaacctg gcggtggggg gcccgcggca gttcccggcg 100

<210> 4308

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4308

gggcttctgg gtttgagtcc ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4309

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4309

gggcttctgg gtttgagtcc ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgctg cacactggga agtccgcagc 120

<210> 4310

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4310

gggcttctgg gtttgagtcc ttggcaagcc cagggcttgt cgacgacggc ggcctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4311

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4311

gggcttctgg gtttgagtcc ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtcaggatca tgcacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4312

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4312

gggcttctgg gtttgcgtcc ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4313

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4313

gggcttctgg gtttgagtcc ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgatg cacactggga aggccgcagc 120

<210> 4314

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4314

gggcttctgg gtttgagtcc ttagcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4315

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4315

gggcttctgg gtttgagtcc ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtccggatca tccacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4316

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4316

gggcttctgg gtttgagtcc ttggcaagcc cagggcttgt cgacgactgc ggtctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4317

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4317

gggcttctgg gtttgagtcc ttcgcaagcc cagggcttgt cgacgacggc ggtctccgtc 60

gtcaggatca tccacttcca ggagtcgctg tgccccgatg cacactggga agtccgcagc 120

<210> 4318

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4318

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4319

<211> 112

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4319

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgggct tgtcgacgac 60

ggcggtctcc gtggtgtacg gtacaaacct tcctctctgg ctccatcgta ag 112

<210> 4320

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4320

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg ggtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4321

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4321

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg cgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4322

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4322

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt gcgtctggtc 60

aaccaccgcg gtctcagtgg tgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4323

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4323

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgggct tgtcgacgac 60

ggcggtctcc gtctggtcaa ccaccgcggt ctcagtggtg tacggtacaa accttcct 118

<210> 4324

<211> 116

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4324

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtacggtag caccgactcg gtgccacttt ttcaag 116

<210> 4325

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4325

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctgttc 60

aaccaccgcg gtctcagtgg tgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4326

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4326

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacggggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4327

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4327

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtacggcac aaaccttcct ctctggctcc atcgtaag 118

<210> 4328

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4328

tctgtcgata ccccccggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4329

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4329

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtacgggac aaaccttcct ctctggctcc atcgtaag 118

<210> 4330

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4330

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagcgg tgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4331

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4331

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtactgtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4332

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4332

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gcctcagtgg tgtacggtac aaaccttcct ctctggctcc atcgtaag 118

<210> 4333

<211> 118

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4333

tctgtcgata ccccacggtc aatccttggg gcccagactg agcacgaggt ttgtctggtc 60

aaccaccgcg gtctcagtgg tgtacgttac aaaccttcct ctctggctcc atcgtaag 118

<210> 4334

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4334

caaggacgcg actctcccga cgcggggagg ctattctgcc cataacacca tcatcaccat 60

catttgagaa aggctggcca gggtggcatc tgaatgggac 100

<210> 4335

<211> 78

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4335

caaggacgcg actctcccga caacaccatc atcaccatct gattttactt agaacctggg 60

tttcccctgg agaaaggc 78

<210> 4336

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4336

caaggacgcg actctcccga cgcggggagg ctattctgcc cataacacca tcatcaccat 60

catttgagaa aggctggcca gggcggcatc tgaatgggac 100

<210> 4337

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4337

gcctcagctc agctgcctcc ttggaggcag ctacctacct ctccagtcct gtcttcagag 60

aagagatgga agacacctgc gctattgaca cttttctcag 100

<210> 4338

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4338

gcctcagctc agctgcctcc ttggaggcag ctacctacct ctccagtcct gtc 53

<210> 4339

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4339

gcctcagctc agctgcctcc ttggaggcag ctacctacct ctccagtcct gtcctggggc 60

cattttctca g 71

<210> 4340

<211> 94

<212> DNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 4340

gcctcagctc agctgcctcc ttggaggcag ctacctacct ctccagtcct gtcctggggc 60

cagaagagat ggaagacact attgacactt ttct 94

<210> 4341

<211> 67

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4341

gcctcagctc agctgcctcc ttggaggcag ctacctacct ctccagtcct gtctccactt 60

ttctcag 67

<210> 4342

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4342

gaaacgcacc agacggccgc gtcagagcag ctcaggttct gggaggtttg tctggtcaac 60

caccgcggtc tcagtggtgt acggtacaaa cctgaacacc taggacgcac cattctcaca 120

<210> 4343

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4343

gaaacgcacc agacagccgc gtcagagcag ctcaggttct gggaggtttg tctggtcaac 60

caccgcggtc tcagtggtgt acggtacaaa cctgaacacc taggacgcac cattctcaca 120

<210> 4344

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4344

gaaacgcacc agacagccgc gtccgagcag ctcaggttct gggaggtttg tctggtcaac 60

caccgcggtc tcagtggtgt acggtacaaa cctgaacacc taggacgcac cattctcaca 120

<210> 4345

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4345

gaaacgcacc agacagccgc gtcagagcag ctcagtttct gggaggtttg tctggtcaac 60

caccgcggtc tcagtggtgt acggtacaaa cctgaacacc taggacgcac cattctcaca 120

<210> 4346

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4346

gaaacgcacc agacagccgc gtcagagcag ctcaggttct gggaggtttg tctggtcaac 60

caccgctgtc tcagtggtgt acggtacaaa cctgaacacc taggacgcac cattctcaca 120

<210> 4347

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4347

gaaacgcacc agacagccgc gtccgagcag ctcagtttct gggaggtttg tctggtcaac 60

caccgcggtc tcagtggtgt acggtacaaa cctgaacacc taggacgcac cattctcaca 120

<210> 4348

<211> 105

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4348

gaaacgcacc agacagccgc gtcagagcag ctcaggttct gggaggtttg tctggtcaac 60

caccgcggtc tcagtggtga acacctagga cgcaccattc tcaca 105

<210> 4349

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4349

caaggacgcg actctcccga cgcggggagg ctattctgcc cataacacca tcatcaccat 60

catttagaga cgcaaacaca gccaccaccc aagtgatcac 100

<210> 4350

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4350

cccaggtttg tctggtcaac caccgcggtc tcagtggtgt 40

<210> 4351

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4351

cccaggtttg accgcggtct cagtggtgt 29

<210> 4352

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4352

tccaggtttg tctggtcaac caccgcggtc tcagtggtgt 40

<210> 4353

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4353

ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc gtcaggatca tccacttcca 60

<210> 4354

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4354

ttggcaagcc cagggcttgt cgacgacggc ggtctccgtc gtcaggatca tgcacttcca 60

<210> 4355

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4355

tcatgcccct ggctctcctg cgacgacggc ggtctccgtc gtcaggatca tccacttcca 60

<210

> 4356

<211> 73

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4356

ggggcccaga ctgagcacgt gatggcagag gaaaggatat gtgcagctcc tgcaccggga 60

tactggttga caa 73

<210> 4357

<211> 73

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4357

ttgtcaacca gtatcccggt gcaggagctg cacatatcct ttcctctgcc atcacgtgct 60

cagtctgggc ccc 73

<210> 4358

<211> 76

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4358

ggggcccaga ctgagcacgg gcttgtcgac gacggcggtc tccgtcgtca ggatcatccg 60

ggatactggt tgacaa 76

<210> 4359

<211> 76

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 4359

ttgtcaacca gtatcccgga tgatcctgac gacggagacc gccgtcgtcg acaagcccgt 60

gctcagtctg ggcccc 76