(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년07월23일
(11) 등록번호 10-2687373
(24) 등록일자 2024년07월17일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) *A61K 38/00* (2006.01)
C12N 15/10 (2017.01) *C12N 9/78* (2006.01)
(52) CPC특허분류
C12N 9/22 (2013.01)
A61K 38/00 (2013.01)
(21) 출원번호 10-2019-7031141
(22) 출원일자(국제) 2018년03월23일
심사청구일자 2021년03월23일
(85) 번역문제출일자 2019년10월22일
(65) 공개번호 10-2019-0130613
(43) 공개일자 2019년11월22일
(86) 국제출원번호 PCT/US2018/024208
(87) 국제공개번호 WO 2018/176009
국제공개일자 2018년09월27일
(30) 우선권주장
62/475,830 2017년03월23일 미국(US)
(뒷면에 계속)

(56) 선행기술조사문헌
NATURE. VOL 533. pp420~435(2016.05.19.)*
Mol Cell. 60(3): 385-397(2015.11.05.)*
WO2017070632 A1
WO2018039438 A1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
프레지던트 앤드 펠로우즈 오브 하바드 칼리지
미합중국, 메사추세츠 02138, 캠브리지, 퀸시스트리트17
(72) 발명자
리우, 데이비드, 알.
미국 02420 매사추세츠주 렉싱턴 위트먼 서클 3
코모르, 알렉시스, 크리스틴
미국 92122 캘리포니아주 샌 디에고 주디셜 드라이브 9135 #3419
(뒷면에 계속)
(74) 대리인
양영준, 김영

전체 청구항 수 : 총 44 항 심사관 : 이형곤

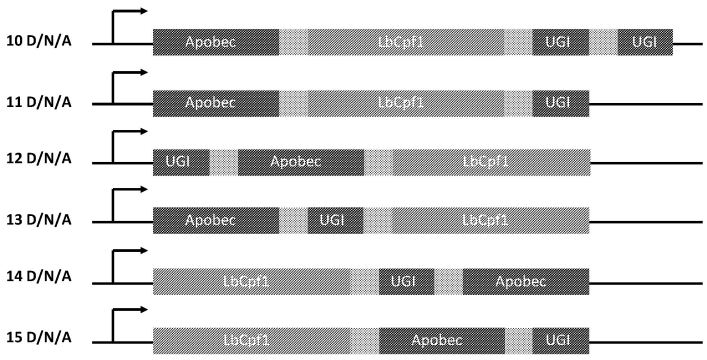
(54) 발명의 명칭 **핵산 프로그램가능한 DNA 결합 단백질을 포함하는 핵염기 편집제**

(57) 요 약

본 개시내용의 일부 측면은 세포 또는 대상체의 게놈 내에서, 예를 들어, 인간 게놈 내에서 단일 부위를 편집하는 것을 포함하는, 핵산의 표적화된 편집에 유용한 전략, 시스템, 시약, 방법, 및 키트를 제공한다. 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp), 예를 들어, Cpf1 또는 그의 변이체, 및 핵산 편집

(뒷면에 계속)

대 표 도



단백질 또는 단백질 도메인, 예를 들어, 데아미나제 도메인의 융합 단백질이 제공된다. 일부 실시양태에서, 표적화된 핵산 편집 방법이 제공된다. 일부 실시양태에서, 표적화된 핵산 편집 단백질, 예를 들어, napDNAbp (예를 들어, CasX, CasY, Cpf1, C2c1, C2c2, C2C3, 및 아르고노트) 및 핵산 편집 단백질 또는 도메인의 융합 단백질의 생성을 위한 시약 및 키트가 제공된다.

(52) CPC특허분류

C12N 15/102 (2013.01)

C12N 15/111 (2013.01)

C12N 15/62 (2013.01)

C12N 9/78 (2013.01)

C12N 2310/20 (2017.05)

(72) 발명자

첸, 리웨이

미국 02143 매사추세츠주 소머빌 에이번 스트리트 86 아파트먼트 2

리스, 홀리, 에이.

미국 02144 매사추세츠주 소머빌 호손 스트리트 34

(30) 우선권주장

62/490,587 2017년04월26일 미국(US)

62/511,934 2017년05월26일 미국(US)

62/551,951 2017년08월30일 미국(US)

명 세 서

청구범위

청구항 1

(i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp) 도메인; (ii) 시티딘 데아미나제 도메인; (iii) 제1 우라실 글리코실라제 억제제 (UGI) 도메인; 및 (iv) 제2 우라실 글리코실라제 억제제 (UGI) 도메인을 포함하는 융합 단백질이며, 여기서 napDNAbp 도메인이 CasX, CasY, Cpf1, C2c1, C2c2, C2c3 또는 아르고노트 (Ago) 단백질인 융합 단백질.

청구항 2

제1항에 있어서, 하기 구조:

NH_2-[시티딘 데아미나제 도메인]-[napDNAbp 도메인]-[제1 UGI 도메인]-[제2 UGI 도메인]-COOH;

NH_2-[제1 UGI 도메인]-[제2 UGI 도메인]-[시티딘 데아미나제 도메인]-[napDNAbp 도메인]-COOH;

NH_2-[napDNAbp 도메인]-[시티딘 데아미나제 도메인]-[제1 UGI 도메인]-[제2 UGI 도메인]-COOH;

NH_2-[제1 UGI 도메인]-[제2 UGI 도메인]-[napDNAbp 도메인]-[시티딘 데아미나제 도메인]-COOH;

NH_2-[시티딘 데아미나제 도메인]-[제1 UGI 도메인]-[제2 UGI 도메인]-[napDNAbp 도메인]-COOH; 또는

NH_2-[napDNAbp 도메인]-[제1 UGI 도메인]-[제2 UGI 도메인]-[시티딘 데아미나제 도메인]-COOH

를 포함하며, 여기서 각 경우의 "]-["는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 3

제1항 또는 제2항에 있어서, (v) 핵 국재화 서열 (NLS)을 추가로 포함하는 융합 단백질.

청구항 4

제3항에 있어서, NLS가 하기 아미노산 서열: PKKKRKV (서열식별번호: 614), MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (서열식별번호: 615), 또는 KRTADGSEFEPKKKRKV (서열식별번호: 740)를 포함하는 것인 융합 단백질.

청구항 5

제3항에 있어서, (1) 시티딘 데아미나제 도메인과 napDNAbp 도메인이 링커를 통해 연결되는 것, (2) 제1 UGI 도메인과 NLS가 링커를 통해 연결되는 것, (3) 제2 UGI 도메인과 NLS가 링커를 통해 연결되는 것, (4) napDNAbp 도메인과 제1 UGI 도메인이 링커를 통해 연결되는 것, (5) napDNAbp 도메인과 제2 UGI 도메인이 링커를 통해 연결되는 것, 및 (6) 제1 UGI 도메인과 제2 UGI 도메인이 링커를 통해 연결되는 것 중 적어도 하나인 융합 단백질.

청구항 6

제3항에 있어서, 하기 구조: NH_2-[시티딘 데아미나제 도메인]-[napDNAbp 도메인]-[제1 UGI 도메인]-[제2 UGI 도메인]-[NLS]-COOH를 포함하며, 여기서 각 경우의 "]-["는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 7

제3항에 있어서,

제2 UGI 도메인과 NLS가 아미노산 서열 SGGS (서열식별번호: 606)를 포함하는 링커를 통해 연결되거나;

napDNAbp 도메인과 제1 UGI 도메인이 아미노산 서열 SGGS (서열식별번호: 606)를 포함하는 링커를 통해 연결되거나;

napDNAbp 도메인과 제1 UGI 도메인이, 아미노산 서열 $SGGS(GGS)_n$ (서열식별번호: 612)를 포함하며 여기서 n은 2인 링커를 통해 연결되거나;

제1 UGI 도메인과 제2 UGI 도메인이, 아미노산 서열 $SGGS(GGS)_n$ (서열식별번호: 612)를 포함하며 여기서 n은 2인 링커를 통해 연결되는 것인

융합 단백질.

청구항 8

제7항에 있어서, 시티딘 데아미나제 도메인과 napDNAbp 도메인이 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605)를 포함하는 링커를 통해 연결되는 것인 융합 단백질.

청구항 9

제1항 또는 제2항에 있어서, 제1 UGI 도메인 및/또는 제2 UGI 도메인이 서열식별번호: 736의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 10

제1항에 있어서, napDNAbp 도메인이

(a) 서열식별번호: 31의 아미노산 서열을 포함하는 CasY 단백질이거나;

(b) 서열식별번호: 9-24 중 어느 하나의 아미노산 서열을 포함하는 Cpf1 단백질이거나;

(c) 서열식별번호: 26의 아미노산 서열을 포함하는 C2c1 단백질이거나;

(d) 서열식별번호: 27의 아미노산 서열을 포함하는 C2c2 단백질이거나;

(e) 서열식별번호: 28의 아미노산 서열을 포함하는 C2c3 단백질이거나;

(f) 서열식별번호: 25의 아미노산 서열을 포함하는 아르고노트 단백질이거나;

(g) 서열식별번호: 29 또는 30의 아미노산 서열을 포함하는 CasX 단백질인

융합 단백질.

청구항 11

제1항에 있어서, napDNAbp 도메인이 라크노스피라세아에(*Lachnospiraceae*) 종으로부터의 Cpf1 니카제 (LbCpf1) 또는 아시다미노코쿠스(*Acidaminococcus*) 종으로부터의 Cpf1 니카제 (AsCpf1)인 융합 단백질.

청구항 12

제1항에 있어서, 3, 4, 또는 5개의 UGI 도메인을 포함하는 융합 단백질.

청구항 13

제1항에 있어서, UGI 도메인이 서열식별번호: 736에 제시된 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 14

제2항에 있어서, 링커 중 어느 하나가 아미노산 서열 $SGGS(GGS)_n$ (서열식별번호: 612), $(GGGS)_n$ (서열식별번호: 613), $(GGGGS)_n$ (서열식별번호: 607), $(G)_n$ (서열식별번호: 608), $(EAAAK)_n$ (서열식별번호: 609), $(GGS)_n$ (서열식별번호: 610), $(SGGS)_n$ (서열식별번호: 606), SGSETPGTSESATPES (서열식별번호: 604), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605), $(XP)_n$ (서열식별번호: 611), 또는 그의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이고; X는 임의의 아미노산인 융합 단백질.

청구항 15

제1항 또는 제2항에 있어서, 시티딘 데아미나제 도메인이

(i) 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 데아미나제이거나;

(ii) 서열식별번호: 49-84 중 어느 하나의 아미노산 서열을 포함하거나;

(iii) 서열식별번호: 76의 W90Y, R126E, 및 R132E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 래트 APOBEC1 (rAPOBEC1) 데아미나제이거나;

(iv) 서열식별번호: 74의 W90Y, Q126E, 및 R132E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 인간 APOBEC1 (hAPOBEC1) 데아미나제이거나;

(v) 서열식별번호: 60의 W285Y, R320E, 및 R326E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 인간 APOBEC3G (hAPOBEC3G) 데아미나제이거나;

(vi) 활성화-유도된 데아미나제 (AID)이거나;

(vii) 페트로미존 마리누스(*Petromyzon marinus*)로부터의 시티딘 데아미나제 1 (pmCDA1)인

융합 단백질.

청구항 16

제1항 또는 제2항에 있어서, 시티딘 데아미나제 도메인이 APOBEC1 데아미나제, APOBEC2 데아미나제, APOBEC3A 데아미나제, APOBEC3B 데아미나제, APOBEC3C 데아미나제, APOBEC3D 데아미나제, APOBEC3F 데아미나제, APOBEC3G 데아미나제, 및 APOBEC3H 데아미나제로 이루어진 군으로부터 선택된 APOBEC 패밀리 데아미나제인 융합 단백질.

청구항 17

제10항에 있어서, 핵 국재화 서열 (NLS)을 추가로 포함하는 융합 단백질.

청구항 18

제1항 또는 제2항의 융합 단백질 및 상기 융합 단백질의 napDNAbp 도메인에 결합된 가이드 RNA를 포함하는 복합체.

청구항 19

제18항에 있어서, 가이드 RNA가 10 내지 100개 뉴클레오티드 길이이고, 표적 서열에 상보적인 적어도 10개의 인접 뉴클레오티드의 서열을 포함하는 것인 복합체.

청구항 20

제19항에 있어서, 표적 서열이 진핵생물의 게놈 내에 또는 포유동물의 게놈 내에 있는 것인 복합체.

청구항 21

제20항에 있어서, 포유동물이 인간, 마우스, 또는 래트인 복합체.

청구항 22

제1항 또는 제2항의 융합 단백질 및 서열식별번호: 618의 핵산 서열을 포함하는 가이드 RNA를 포함하는 복합체.

청구항 23

제1항 또는 제2항의 융합 단백질을 코딩하는 폴리뉴클레오티드.

청구항 24

제23항의 폴리뉴클레오티드를 포함하며, 상기 폴리뉴클레오티드의 발현을 유도하는 이종 프로모터를 포함하는

벡터.

청구항 25

제1항 또는 제2항의 융합 단백질, 또는

제1항 또는 제2항의 융합 단백질 및 상기 융합 단백질의 napDNAbp 도메인에 결합된 가이드 RNA를 포함하는 복합체

를 포함하는 단리된 세포.

청구항 26

제24항의 벡터를 포함하는 단리된 세포.

청구항 27

핵산 분자를 제1항 또는 제2항의 융합 단백질 및 가이드 RNA와 접촉시키는 단계를 포함하는 방법이며, 여기서 가이드 RNA는 표적 염기 쌍을 포함하는 유기체의 게놈에서의 표적 서열에 상보적인 적어도 10개의 인접 뉴클레오티드의 서열을 포함하고, 여기서 방법은 시험관내 또는 생체외, 또는 비-인간 동물의 생체내에서 수행되는 것인 방법.

청구항 28

제1항 또는 제2항의 융합 단백질, 또는

제1항 또는 제2항의 융합 단백질 및 상기 융합 단백질의 napDNAbp 도메인에 결합된 가이드 RNA를 포함하는 복합체

를 포함하는,

유전 질환; 신생물성 질환; 대사 질환; 리소솜 축적 질환; 자가면역 장애; 염증성 장애; 감염성 질환; 신경계 장애; 심혈관 장애; 증식성 장애; 호흡기 장애; 소화 장애; 내분비 장애; 대사 장애; 영양 장애; 비뇨기 장애; 심리적 장애; 피부 장애; 혈액 장애; 및 림프 장애로 이루어진 군으로부터 선택된 질환, 장애, 또는 상태를 갖거나 갖는 것으로 의심되는 대상체를 치료하는데 사용하기 위한 제약 조성물.

청구항 29

제24항의 벡터를 포함하는,

유전 질환; 신생물성 질환; 대사 질환; 리소솜 축적 질환; 자가면역 장애; 염증성 장애; 감염성 질환; 신경계 장애; 심혈관 장애; 증식성 장애; 호흡기 장애; 소화 장애; 내분비 장애; 대사 장애; 영양 장애; 비뇨기 장애; 심리적 장애; 피부 장애; 혈액 장애; 및 림프 장애로 이루어진 군으로부터 선택된 질환, 장애, 또는 상태를 갖거나 갖는 것으로 의심되는 대상체를 치료하는데 사용하기 위한 제약 조성물.

청구항 30

제25항의 단리된 세포를 포함하는,

유전 질환; 신생물성 질환; 대사 질환; 리소솜 축적 질환; 자가면역 장애; 염증성 장애; 감염성 질환; 신경계 장애; 심혈관 장애; 증식성 장애; 호흡기 장애; 소화 장애; 내분비 장애; 대사 장애; 영양 장애; 비뇨기 장애; 심리적 장애; 피부 장애; 혈액 장애; 및 림프 장애로 이루어진 군으로부터 선택된 질환, 장애, 또는 상태를 갖거나 갖는 것으로 의심되는 대상체를 치료하는데 사용하기 위한 제약 조성물.

청구항 31

제28항에 있어서, 질환, 장애, 또는 상태가 인간 대상체에 존재하는 것인, 질환, 장애 또는 상태를 갖거나 갖는 것으로 의심되는 대상체를 치료하는데 사용하기 위한 제약 조성물.

청구항 32

제31항에 있어서, 질환, 장애 또는 상태가 낭성 섬유증, 페닐케톤뇨, 표피박리성 과다각화증 (EHK), 샤르코-마

리-투스병 유형 4J, 신경모세포종 (NB), 폰 빌레브란트병 (vWD), 선천성 근긴장증, 유전성 신장 아밀로이드증, 확장성 심근병증 (DCM), 유전성 림프부종, 가족성 알츠하이머병, HIV, 프리온 질환, 만성 영아 신경학적 피부 관절 증후군 (CINCA), 데스민-관련 근병증 (DRM), 또는 돌연변이체 PI3KCA 단백질, 돌연변이체 CTNNB1 단백질, 돌연변이체 HRAS 단백질 또는 돌연변이체 p53 단백질과 연관된 신생물성 질환인 제약 조성물.

청구항 33

(i) 제1항 또는 제2항의 융합 단백질과 RNA를 수용액 중에서 복합체화함으로써 수용액 중 융합 단백질 및 RNA를 포함하는 복합체를 형성하는 단계; 및

(ii) (i)의 복합체를 양이온성 지질과 접촉시키는 단계

를 포함하는, 리보핵단백질 (RNP) 복합체를 생산하는 방법.

청구항 34

(i) 제1항 또는 제2항의 융합 단백질을 세포 내에서 발현시키는 단계로서, 여기서 융합 단백질은 친화성 태그를 추가로 포함하는 것인 단계;

(ii) (i)의 세포를 용해시킴으로써 용해물을 생성하는 단계; 및

(iii) 용해물을 친화성 크로마토그래피에 적용함으로써 정제된 융합 단백질을 포함하는 용리액을 생산하는 단계

를 포함하는, 융합 단백질을 정제하는 방법.

청구항 35

제1항 또는 제2항의 융합 단백질, 또는 제1항 또는 제2항의 융합 단백질을 포함하는 제약 조성물을 대상체의 내이(inner ear)로 전달하는 단계를 포함하며, 여기서 대상체는 비-인간 동물인 방법.

청구항 36

표적 핵염기 쌍을 포함하는 이중-가닥 DNA 서열의 표적 영역을, 핵염기 편집제 및 가이드 핵산을 포함하는 복합체와 접촉시키는 단계; 및 그에 의해

상기 표적 영역의 가닥 분리를 유도하는 단계;

표적 영역의 단일 가닥에서의 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계;

상기 표적 영역의 1개 이하의 가닥을 절단하는 단계; 및

제1 핵염기 염기에 상보적인 제3 핵염기를 제2 핵염기에 상보적인 제4 핵염기로 대체하는 단계

를 포함하는, 이중-가닥 DNA 서열의 핵염기 쌍을 편집하는 방법이며,

여기서 방법은 이중-가닥 DNA 서열에서 20% 미만의 indel 형성을 유발하고;

여기서 핵염기 편집제는 CasX, CasY, Cpf1, C2c1, C2c2, C2c3 또는 아르고노트 (Ago) 도메인을 포함하고;

여기서 방법은 시험관내 또는 생체외, 또는 비-인간 동물의 생체내에서 수행되는 것인 방법.

청구항 37

제36항에 있어서, 핵염기 편집제가 제1항 또는 제2항의 융합 단백질인 방법.

청구항 38

제36항에 있어서, 핵염기 편집제가 제1 UGI 도메인 및 제2 UGI 도메인을 포함하는 것인 방법.

청구항 39

제36항 또는 제38항에 있어서, 접촉시키는 단계가 이중-가닥 DNA 서열에서 19% 미만, 18% 미만, 17% 미만, 16% 미만, 15% 미만, 14% 미만, 13% 미만, 12% 미만, 11% 미만, 또는 10% 미만의 indel 형성을 유발하는 것인 방법.

청구항 40

제36항 또는 제38항에 있어서, 제2 핵염기를 제4 핵염기에 상보적인 제5 핵염기로 대체함으로써 의도된 편집된 염기 쌍을 생성하며, 의도된 편집된 염기 쌍의 생성 효율이 적어도 5%, 10%, 20%, 30%, 40%, 또는 50%인 방법.

청구항 41

제40항에 있어서, 의도된 편집된 염기 쌍 대 비의도된 편집된 염기 쌍의 비가 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 또는 8:1인 방법.

청구항 42

제36항 또는 제38항에 있어서, 표적 핵염기 쌍이 질환 또는 장애와 연관된 T에서 C로의 점 돌연변이를 포함하며, 돌연변이체 C 염기의 탈아미노화가 질환 또는 장애와 연관되지 않은 서열을 발생시키는 것인 방법.

청구항 43

제42항에 있어서,

표적 영역이 단백질을 코딩하고,

점 돌연변이가 코돈 내에 있고, 야생형 코돈에 비해 돌연변이체 코돈에 의해 코딩되는 아미노산의 변화를 발생시키는 것인 방법.

청구항 44

제42항에 있어서, 질환 또는 장애가 낭성 섬유증, 페닐케톤뇨, 표피박리성 과다각화증 (EHK), 샤르코-마리-투스병 유형 4J, 신경모세포종 (NB), 폰 빌레브란트병 (vWD), 선천성 근긴장증, 유전성 신장 아밀로이드증, 확장성 심근병증 (DCM), 유전성 림프부종, 가족성 알츠하이머병, HIV, 프리온 질환, 만성 영아 신경학적 피부 관절 증후군 (CINCA), 데스민-관련 근병증 (DRM), 또는 돌연변이체 PI3KCA 단백질, 돌연변이체 CTNNB1 단백질, 돌연변이체 HRAS 단백질 또는 돌연변이체 p53 단백질과 연관된 신생물성 질환인 방법.

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

청구항 101

삭제

청구항 102

삭제

청구항 103

삭제

청구항 104

삭제

청구항 105

삭제

청구항 106

삭제

청구항 107

삭제

청구항 108

삭제

청구항 109

삭제

청구항 110

삭제

청구항 111

삭제

청구항 112

삭제

청구항 113

삭제

청구항 114

삭제

청구항 115

삭제

청구항 116

삭제

청구항 117

삭제

청구항 118

삭제

청구항 119

삭제

청구항 120

삭제

청구항 121

삭제

청구항 122

삭제

청구항 123

삭제

청구항 124

삭제

청구항 125

삭제

청구항 126

삭제

청구항 127

삭제

청구항 128

삭제

청구항 129

삭제

청구항 130

삭제

청구항 131

삭제

청구항 132

삭제

청구항 133

삭제

청구항 134

삭제

청구항 135

삭제

청구항 136

삭제

청구항 137

삭제

청구항 138

삭제

청구항 139

삭제

청구항 140

삭제

청구항 141

삭제

청구항 142

삭제

청구항 143

삭제

청구항 144

삭제

청구항 145

삭제

청구항 146

삭제

청구항 147

삭제

청구항 148

삭제

청구항 149

삭제

청구항 150

삭제

청구항 151

삭제

청구항 152

삭제

청구항 153

삭제

청구항 154

삭제

청구항 155

삭제

청구항 156

삭제

청구항 157

삭제

청구항 158

삭제

청구항 159

삭제

청구항 160

삭제

청구항 161

삭제

청구항 162

삭제

청구항 163

삭제

청구항 164

삭제

청구항 165

삭제

청구항 166

삭제

청구항 167

삭제

청구항 168

삭제

청구항 169

삭제

청구항 170

삭제

청구항 171

삭제

청구항 172

삭제

청구항 173

삭제

청구항 174

삭제

청구항 175

삭제

청구항 176

삭제

청구항 177

삭제

청구항 178

삭제

청구항 179

삭제

청구항 180

삭제

청구항 181

삭제

청구항 182

삭제

청구항 183

삭제

청구항 184

삭제

청구항 185

삭제

청구항 186

삭제

청구항 187

삭제

청구항 188

삭제

청구항 189

삭제

청구항 190

삭제

청구항 191

삭제

청구항 192

삭제

청구항 193

삭제

청구항 194

삭제

청구항 195

삭제

청구항 196

삭제

청구항 197

삭제

청구항 198

삭제

청구항 199

삭제

청구항 200

삭제

청구항 201

삭제

청구항 202

삭제

청구항 203

삭제

청구항 204

삭제

청구항 205

삭제

청구항 206

삭제

청구항 207

삭제

발명의 설명

배 경 기 술

[0001] 핵산 서열의 표적화된 편집, 예를 들어, 게놈 DNA 내로의 특이적 변형의 표적화된 절단 또는 표적화된 도입은 유전자 기능의 연구를 위한 매우 유망한 접근법이며, 또한 인간 유전 질환에 대한 새로운 요법을 제공하는 잠재력을 갖는다.[1] 이상적인 핵산 편집 기술은 3가지 특징을 갖는다: (1) 목적하는 변형을 설치하는 높은 효율; (2) 최소의 오프-타겟(off-target) 활성; 및 (3) 주어진 핵산에서의 임의의 부위, 예를 들어, 인간 게놈 내의 임의의 부위를 정확하게 편집하도록 프로그램되는 능력.[2] 조작된 아연 핑거 뉴클레아제 (ZFN),[3] 전사 활성인자 유사 이펙터 뉴클레아제 (TALEN),[4] 및 가장 최근, RNA-가이드된 DNA 엔도뉴클레아제 Cas9[5]를 포함한 현행 게놈 조작 도구는 게놈에서 서열-특이적 DNA 절단을 실시한다. 이 프로그램가능한 절단은 비-상동 말단 연결 (NHEJ)을 통한 절단 부위에서의 DNA의 돌연변이, 또는 상동성-지정 복구 (HDR)를 통한 절단 부위 주위의 DNA의 대체를 발생시킬 수 있다.[6,7]

[0002] 현재의 기술에 대한 한 가지 결점은 NHEJ 및 HDR 둘 다가 전형적으로 보통의 유전자 편집 효율 뿐만 아니라, 목적하는 변경과 경쟁할 수 있는 원치 않는 유전자 변경을 발생시키는 확률적 과정라는 점이다.[8] 많은 유전 질환은 원칙적으로 게놈에서의 특이적 위치에서 특이적 뉴클레오티드 변화 (예를 들어, 질환과 연관된 유전자의 특이적 코돈에서의 C에서 T로의 변화)를 실시함으로써 치료될 수 있기 때문에,[9] 이러한 정확성 유전자 편집을 달성하기 위한 프로그램가능한 방식의 개발은 강한 새로운 연구 도구, 뿐만 아니라 유전자 편집-기반 인간 치료제에 대한 잠재적인 새로운 접근법 둘 다를 제시할 것이다.

발명의 내용

[0003] 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp), 예컨대 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) 시스템은 다양한 유기체 및 세포주에서 강건하고 일반적인 게놈 조작을 가능하게 하도록 변형된[11] 최근 발견된 원핵생물 적응 면역계이다.[10] CRISPR-Cas (CRISPR 연관된) 시스템은 염기-쌍형성을 통해 복합체를 표적 DNA 서열에 국재화하는 가이드로서 RNA 분자 (sgRNA)를 사용하는 단백질-RNA 복합체이다.[12] 천연 시스템에서, Cas 단백질은 그 후 표적화된 DNA 서열을 절단하는 엔도뉴클레아제로서 작용한다.[13] 표적 DNA 서열은 sgRNA에 양쪽 상보적이고, 또한 시스템이 기능하도록 하기 위해 상보적 영역의 3'-말단에 "프로토스페이서-인접 모티프" (PAM)를 함유해야 한다.[14]

[0004] 공지된 Cas 단백질 중에서, 에스. 피오게네스(*S. pyogenes*) Cas9는 게놈 조작을 위한 도구로서 주로 광범위하게 사용되었다.[15] 이 Cas9 단백질은 2개의 별개의 뉴클레아제 도메인을 함유하는 큰 다중-도메인 단백질이다. 점돌연변이는 Cas9 내로 도입되어 뉴클레아제 활성을 폐지할 수 있으며, 이는 sgRNA-프로그램된 방식으로 DNA에 결합하는 그의 능력을 여전히 보유하는 데드 Cas9 (dCas9)를 발생시킨다.[16] 원칙적으로, 또 다른 단백질 또는 도메인에 융합되는 경우, dCas9는 단순히 적절한 sgRNA와의 공동-발현에 의해 그 단백질 또는 도메인을 사실상 임의의 DNA 서열에 표적화할 수 있다.

[0005] 게놈 조작 목적을 위한 dCas9 복합체의 잠재력은 엄청나다. 단백질을 sgRNA에 의해 프로그램된 게놈에서의 특이적 부위로 가져오는 그의 고유한 능력은 이론적으로 데아미나제 (예를 들어, 시티딘 데아미나제), 전사 활성인자, 전사 리프레서, 히스톤-변형 단백질, 인테그라제, 및 레콤비나제를 포함한 뉴클레아제를 넘어 다양한 부위-특이적 게놈 조작 도구로 개발될 수 있다.[11] 이들 잠재적 용도 중 일부는 최근 RNA-가이드된 전사 활성인자,[17,18] 전사 리프레서,[16,19,20] 및 크로마틴 변형 효소[21]를 제공하기 위한 전사 활성인자와의 dCas9 융합물을 통해 실행되었다. 다양한 sgRNA와의 이들 융합물의 단순한 공동-발현은 표적 유전자의 특이적 발현을 발생시킨다. 이들 중대한 연구는 게놈의 정확한 조작을 위한 용이하게 프로그램가능한 서열-특이적 이펙터의 설계 및 구축을 위한 방식을 조성하였다.

[0006] 본 개시내용의 일부 측면은 링커에 의해 융합된 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp), 예를 들어 CasX, CasY, Cpf1, C2c1, C2c2, C2c3, 또는 아르고노트 단백질, 및 시티딘 데아미나제 도메인의 특정 구성이 표적 시티딘 잔기를 효율적으로 탈아미노화하는 데 유용하다는 인식에 기초한다. 본 개시내용의 다른 측면은 링커를 통해 napDNAbp의 N-말단에 융합된 시티딘 데아미나제 도메인을 갖는 핵염기 편집 융합 단백질이 이중 가닥 DNA 표적 분자에서 표적 핵산을 효율적으로 탈아미노화할 수 있었다는 인식에 관한 것이다. 예를 들어, 하기 실시예 3 및 4를 참조하며, 이는 본원에서 염기 편집제로도 지칭되는 융합 단백질이 다른 염기 편집제, 예컨대 UGI 도메인을 갖지 않는 염기 편집제보다 더 적은 indel을 생성하고, 더 효율적으로 표적 핵산을 탈아미노화한다는 것을 입증한다. 본 개시내용의 다른 측면은 링커를 통해 napDNAbp의 N-말단에 융합된 시티딘 데아미나제 도메인을 갖는 핵염기 편집 융합 단백질이, 융합 단백질이 1개 초과의 UGI 도메인으로 구성된 경우에 더 높은 효율 및 크게 개선된 생성물 순도를 갖는 염기 편집을 수행한다는 인식에 관한 것이다. 예를 들어, 실시예 17을 참조하며, 이는 2개의 UGI 도메인을 포함하는 융합 단백질 (예를 들어, 염기 편집제)이 다른 염기 편집제, 예컨대 1개의 UGI 도메인을 포함하는 것들보다 더 적은 indel을 생성하고 표적 핵산을 더 효율적으로 탈아미노화한다는 것을 입증한다.

[0007] 일부 실시양태에서, 융합 단백질은 (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp); (ii) 시티딘 데아미나제 도메인; 및 (iii) 우라실 글리코실라제 억제제 (UGI) 도메인을 포함하며, 여기서 napDNAbp는 CasX, CasY, Cpf1, C2c1, C2c2, C2c3 또는 아르고노트 단백질이다. 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백

질 (napDNAbp)은 CasX 단백질이다. 일부 실시양태에서, CasX 단백질은 서열식별번호: 29 또는 30과 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, CasX 단백질은 서열식별번호: 29 또는 30의 아미노산 서열을 포함한다.

[0008] 일부 실시양태에서, 융합 단백질은 (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp); (ii) 시티딘 데아미나제 도메인; (iii) 제1 우라실 글리코실라제 억제제 (UGI) 도메인; 및 (iv) 제2 우라실 글리코실라제 억제제 (UGI) 도메인을 포함하며, 여기서 napDNAbp는 Cas9, dCas9 또는 Cas9 니카제 단백질이다. 일부 실시양태에서, napDNAbp는 dCas9 단백질이다. 일부 실시양태에서, napDNAbp는 CasX, CasY, Cpf1, C2c1, C2c2, C2c3 또는 아르고노트 단백질이다. 일부 실시양태에서, dCas9 단백질은 에스. 피오게네스 dCas9 (SpCas9d)이다. 일부 실시양태에서, dCas9 단백질은 D10A 돌연변이를 보유하는 에스. 피오게네스 dCas9이다. 일부 실시양태에서, dCas9 단백질은 서열식별번호: 6 또는 7과 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, dCas9 단백질은 서열식별번호: 6 또는 7의 아미노산 서열을 포함한다. 일부 실시양태에서, dCas9 단백질은 에스. 아우레우스(*S. aureus*) dCas9 (SaCas9d)이다. 일부 실시양태에서, dCas9 단백질은 D10A 돌연변이를 보유하는 에스. 아우레우스 dCas9이다. 일부 실시양태에서, dCas9 단백질은 서열식별번호: 33-36과 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, dCas9 단백질은 서열식별번호: 33-36의 아미노산 서열을 포함한다.

[0009] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 CasY 단백질이다. 일부 실시양태에서, CasY 단백질은 서열식별번호: 31과 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, CasY 단백질은 서열식별번호: 31의 아미노산 서열을 포함한다.

[0010] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 Cpf1 또는 Cpf1 돌연변이체 단백질이다. 일부 실시양태에서, Cpf1 또는 Cpf1 돌연변이체 단백질은 서열식별번호: 9-24 중 어느 하나와 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cpf1 또는 Cpf1 돌연변이체 단백질은 서열식별번호: 9-24 중 어느 하나의 아미노산 서열을 포함한다.

[0011] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 C2c1 단백질이다. 일부 실시양태에서, C2c1 단백질은 서열식별번호: 26과 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, C2c1 단백질은 서열식별번호: 26의 아미노산 서열을 포함한다.

[0012] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 C2c2 단백질이다. 일부 실시양태에서, C2c2 단백질은 서열식별번호: 27과 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, C2c2 단백질은 서열식별번호: 27의 아미노산 서열을 포함한다.

[0013] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 C2c3 단백질이다. 일부 실시양태에서, C2c3 단백질은 서열식별번호: 28과 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, C2c3 단백질은 서열식별번호: 28의 아미노산 서열을 포함한다.

[0014] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 아르고노트 단백질이다. 일부 실시양태에서, 아르고노트 단백질은 서열식별번호: 25와 적어도 90% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 아르고노트 단백질은 서열식별번호: 25의 아미노산 서열을 포함한다.

[0015] 본 개시내용의 일부 측면은 본원에 제공된 융합 단백질이 큰 비율의 indel을 생성하지 않으면서 1개 이상의 돌연변이 (예를 들어, C에서 T로의 돌연변이)를 생성할 수 있다는 인식에 기초한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 (예를 들어, 염기 편집 단백질) 중 임의의 것은 10% 미만의 indel을 생성한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 (예를 들어, 염기 편집 단백질) 중 임의의 것은 10%, 9%, 8%, 7%, 6%, 5.5%, 5%, 4.5%, 4%, 3.5%, 3%, 2.5%, 2%, 1.5%, 1%, 0.5%, 또는 0.1% 미만의 indel을 생성한다.

[0016] 일부 실시양태에서, 융합 단백질은 napDNAbp 및 아포지단백질 B mRNA-편집 복합체 1 (APOBEC1) 데아미나제 도메인을 포함하며, 여기서 데아미나제 도메인은 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함하는 링커를 통해 napDNAbp 도메인의 N-말단에 융합된다. 일부 실시양태에서, napDNAbp는 본원에 제공된 napDNAbp 중 임의의 것의 아미노산 서열을 포함한다. 일부 실시양태에서, 데아미나제는 래트 APOBEC1 (서열식별번호: 76)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC1 (서열식별번호: 74)이다. 일부 실시양태에서, 데아미나제는 pmCDA1 (서열식별번호: 81)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G (서열식별번호: 60)이다. 일부 실시양태에서, 데아미나제는 (서열식별번호: 82-84) 중 어느 하나의 인간 APOBEC3G 변이체이다. 일부 실시양태에서, 융합 단백질은 napDNAbp 및 아포지단백질 B mRNA-편집 복합체 1 촉매 폴리펩티드-유사 3G

(APOBEC3G) 데아미나제 도메인을 포함하며, 여기서 데아미나제 도메인은 임의의 길이 또는 조성 (예를 들어, 아미노산 서열, 펩티드, 중합체, 또는 결합)의 링커를 통해 napDNAbp 도메인의 N-말단에 융합된다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605)를 포함한다.

[0017] 일부 실시양태에서, 융합 단백질은 napDNAbp 및 시티딘 데아미나제 1 (CDA1) 데아미나제 도메인을 포함하며, 여기서 데아미나제 도메인은 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함하는 링커를 통해 napDNAbp 도메인의 N-말단에 융합된다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605)를 포함한다. 일부 실시양태에서, napDNAbp는 본원에 제공된 napDNAbp 중 임의의 것의 아미노산 서열을 포함한다.

[0018] 일부 실시양태에서, 융합 단백질은 napDNAbp 및 활성화-유도된 시티딘 데아미나제 (AID) 데아미나제 도메인을 포함하며, 여기서 데아미나제 도메인은 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함하는 링커를 통해 napDNAbp 도메인의 N-말단에 융합된다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605)를 포함한다. 일부 실시양태에서, napDNAbp는 본원에 제공된 napDNAbp 중 임의의 것의 아미노산 서열을 포함한다.

[0019] 본 개시내용의 일부 측면은 링커에 의해 융합된 napDNAbp, 및 시티딘 데아미나제 도메인의 특정 구성이 표적 시티딘 잔기를 효율적으로 탈아미노화하는 데 유용하다는 인식에 기초한다. 본 개시내용의 다른 측면은 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함하는 링커를 통해 napDNAbp의 N-말단에 융합된 아포지단백질 B mRNA-편집 복합체 1 (APOBEC1) 데아미나제 도메인을 갖는 핵염기 편집 융합 단백질이 이중 가닥 DNA 표적 분자에서 표적 핵산을 효율적으로 탈아미노화할 수 있었다는 인식에 관한 것이다. 일부 실시양태에서, 융합 단백질은 napDNAbp 도메인 및 아포지단백질 B mRNA-편집 복합체 1 (APOBEC1) 데아미나제 도메인을 포함하며, 여기서, 데아미나제 도메인은 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함하는 링커를 통해 napDNAbp의 N-말단에 융합된다.

[0020] 본 개시내용의 일부 측면은 대상체의 게놈, 예를 들어, 인간의 게놈 내에서 단일 부위를 편집하는 것을 포함하는, 핵산의 표적화된 편집에 유용한 전략, 시스템, 시약, 방법, 및 키트를 제공한다. 일부 실시양태에서, napDNAbp (예를 들어, CasX, CasY, Cpf1, C2c1, C2c2, C2c3 또는 아르고노트 단백질) 및 데아미나제 또는 데아미나제 도메인의 융합 단백질이 제공된다. 일부 실시양태에서, 표적화된 핵산 편집 방법이 제공된다. 일부 실시양태에서, 표적화된 핵산 편집 단백질, 예를 들어, napDNAbp 및 데아미나제 또는 데아미나제 도메인의 융합 단백질의 생성을 위한 시약 및 키트가 제공된다.

[0021] 본 개시내용의 일부 측면은 제2 단백질 (예를 들어, 효소 도메인, 예컨대 시티딘 데아미나제 도메인)에 융합되고, 따라서 융합 단백질을 형성하는 본원에 제공된 바와 같은 napDNAbp를 포함하는 융합 단백질을 제공한다. 일부 실시양태에서, 제2 단백질은 효소 도메인, 또는 결합 도메인을 포함한다. 일부 실시양태에서, 효소 도메인은 뉴클레아제, 니카제, 레콤비나제, 데아미나제, 메틸트랜스퍼라제, 메틸라제, 아세틸라제, 아세틸트랜스퍼라제, 전사 활성인자, 또는 전사 리프레서 도메인이다. 일부 실시양태에서, 효소 도메인은 핵산 편집 도메인이다. 일부 실시양태에서, 핵산 편집 도메인은 데아미나제 도메인이다. 일부 실시양태에서, 데아미나제는 시토신 데아미나제 또는 시티딘 데아미나제이다. 일부 실시양태에서, 데아미나제는 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC1 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC2 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3A 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3B 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3C 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3D 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3E 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3F 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3G 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3H 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC4 데아미나제이다. 일부 실시양태에서, 데아미나제는 활성화-유도된 데아미나제 (AID)이다. 데아미나제는 임의의 적합한 유기체 (예를 들어, 인간 또는 래트)로부터의 것일 수 있다는 것이 인지되어야 한다. 일부 실시양태에서, 데아미나제는 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트, 또는 마우스로부터의 것이다. 일부 실시양태에서, 데아미나제는 래트 APOBEC1 (서열식별번호: 76)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC1 (서열식별번호: 74)이다. 일부 실시양태에서, 데아미나제는 pmCDA1이다.

[0022] 본 개시내용의 일부 측면은 (i) 서열식별번호: 32의 아미노산 서열을 포함하는 CasX, CasY, Cpf1, C2c1, C2c2,

C2c3 또는 아르고노트 단백질 도메인; 및 (ii) 아포지단백질 B mRNA-편집 복합체 1 (APOBEC1) 데아미나제 도메인을 포함하며, 데아미나제 도메인이 SGSETPGTSESATPES (서열식별번호: 604)의 아미노산 서열을 포함하는 링커를 통해 napDNAbp의 N-말단에 융합된 것인 융합 단백질을 제공한다. 일부 실시양태에서, 데아미나제는 래트 APOBEC1 (서열식별번호: 76)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC1 (서열식별번호: 74)이다. 일부 실시양태에서, 융합 단백질은 서열식별번호: 591의 아미노산 서열을 포함한다. 일부 실시양태에서, 융합 단백질은 서열식별번호: 5737의 아미노산 서열을 포함한다. 일부 실시양태에서, 데아미나제는 pmCDA1 (서열식별번호: 81)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G (서열식별번호: 60)이다. 일부 실시양태에서, 데아미나제는 서열식별번호: 82-84 중 어느 하나의 인간 APOBEC3G 변이체이다.

[0023] 본 개시내용의 다른 측면은 데아미나제 도메인, napDNAbp 도메인 및 우라실 글리코실라제 억제제 (UGI) 도메인을 포함하는 융합 단백질이 핵산 분자에서 표적 뉴클레오티드를 탈아미노화하기 위한 개선된 효율을 입증한다는 인식에 관한 것이다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, U:G 헤테로듀플렉스 DNA의 존재에 대한 세포의 DNA-복구 반응은 세포에서의 핵염기 편집 효율의 감소의 원인이 될 수 있다. 우라실 DNA 글리코실라제 (UDG)는 세포에서 DNA로부터 U의 제거를 촉매하며, 이는 가장 흔한 결과로서 U:G 쌍의 C:G 쌍으로의 복귀를 갖는 염기 절제 복구를 개시할 수 있다. 본원에 입증된 바와 같이, 우라실 DNA 글리코실라제 억제제 (UGI)는 인간 UDG 활성을 억제할 수 있다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 염기 절제 복구는 단일 가닥에 결합하고/거나, 편집된 염기를 차단하고/거나, UGI를 억제하고/거나, 염기 절제 복구를 억제하고/거나, 편집된 염기를 보호하고/거나, 비-편집단 가닥의 "고정"을 촉진시키는 등의 분자에 의해 억제될 수 있다. 따라서, 본 개시내용은 UGI 도메인에 융합된 napDNAbp-시티딘 데아미나제 도메인을 포함하는 융합 단백질을 고려한다.

[0024] 본 개시내용의 추가 측면은 데아미나제 도메인, napDNAbp 도메인, 및 1개 초과의 우라실 글리코실라제 억제제 (UGI) 도메인 (예를 들어, 1, 2, 3, 4, 5개, 또는 그 초과의 UGI 도메인)을 포함하는 융합 단백질이 핵산 분자에서의 표적 뉴클레오티드를 탈아미노화하기 위한 개선된 효율 및/또는 개선된 핵산 생성물 순도를 입증한다는 인식에 관한 것이다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 제2 UGI 도메인의 부가는 G:U 염기 편집 중간체에 대한 UDG의 접근을 실질적으로 감소시킴으로써 염기 편집의 효율을 개선시킬 수 있다.

[0025] 본 개시내용의 일부 측면은 본원에 제공된 염기 편집제 중 임의의 것이 유의한 비율의 indel을 생성하지 않고, 특이적 뉴클레오티드 염기를 변형할 수 있다는 인식에 기초한다. 본원에 사용된 "indel"은 핵산 내의 뉴클레오티드 염기의 삽입 또는 결실을 지칭한다. 이러한 삽입 또는 결실은 유전자의 코딩 영역 내의 프레임 시프트 돌연변이를 야기할 수 있다. 일부 실시양태에서, 핵산에서 삽입 또는 결실 (즉, indel)을 생성하지 않고, 핵산 내에서 특이적 뉴클레오티드를 효율적으로 변형시키는 (예를 들어 돌연변이시키거나 탈아미노화하는) 염기 편집제를 생성하는 것이 바람직하다. 특정 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 indel에 비해 보다 큰 비율의 의도된 변형 (예를 들어, 점 돌연변이 또는 탈아미노화)을 생성할 수 있다.

[0026] 특정 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 목적하는 돌연변이의 특정 백분율을 생성할 수 있다. 일부 실시양태에서, 목적하는 돌연변이는 C에서 T로의 돌연변이이다. 일부 실시양태에서, 목적하는 돌연변이는 C에서 A로의 돌연변이이다. 일부 실시양태에서, 목적하는 돌연변이는 C에서 G로의 돌연변이이다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 목적하는 돌연변이의 적어도 1%를 생성할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 목적하는 돌연변이의 적어도 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95% 또는 99%를 생성할 수 있다.

[0027] 본 개시내용의 일부 측면은 본원에 제공된 염기 편집제 중 임의의 것이 유의한 수의 비의도된 돌연변이, 예컨대 비의도된 점 돌연변이를 생성하지 않고, 핵산 (예를 들어 대상체의 게놈 내의 핵산)에서 의도된 돌연변이, 예컨대 점 돌연변이를 효율적으로 생성할 수 있다는 인식에 기초한다.

[0028] 일부 실시양태에서, 융합 단백질의 데아미나제 도메인은 napDNAbp 도메인의 N-말단에 융합된다. 일부 실시양태에서, UGI 도메인은 napDNAbp 도메인의 C-말단에 융합된다. 일부 실시양태에서, napDNAbp 및 핵산 편집 도메인은 링커를 통해 융합된다. 일부 실시양태에서, napDNAbp 도메인 및 UGI 도메인은 링커를 통해 융합된다. 일부 실시양태에서, 제2 UGI 도메인은 제1 UGI 도메인의 C-말단에 융합된다. 일부 실시양태에서, 제1 UGI 도메인 및 제2 UGI 도메인은 링커를 통해 융합된다.

[0029] 특정 실시양태에서, 링커는 본 발명의 펩티드 또는 펩티드 도메인 중 임의의 것을 연결하는 데 사용될 수 있다. 링커는 공유 결합만큼 단순할 수 있거나, 이는 다수의 원자 길이의 중합체 링커일 수 있다. 특정 실시양태에서, 링커는 폴리펩티드이거나, 아미노산에 기초한다. 다른 실시양태에서, 링커는 펩티드-유사가 아

니다. 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 아시클릭, 치환된 또는 비치환된, 분지된 또는 비분지된 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)에 기초한다. 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 고리에 기초한다. 링커는 펩티드에서 링커로 친핵체 (예를 들어, 티올, 아미노)의 부착을 용이하게 하는 관능화된 모이어티를 포함할 수 있다. 임의의 친전자체는 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.

[0030] 일부 실시양태에서, 링커는 아미노산 서열 $(GGGGS)_n$ (서열식별번호: 607), $(G)_n$ (서열식별번호: 608), $(EAAAK)_n$ (서열식별번호: 609), $(GGS)_n$ (서열식별번호: 610), $(SGGS)_n$ (서열식별번호: 606), SGSETPGTSESATPES (서열식별번호: 604), $(XP)_n$ (서열식별번호: 611), $SGGS(GGS)_n$ (서열식별번호: 612), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605), 또는 그의 임의의 조합을 포함하며, 여기서 n은 독립적으로 정수 1 내지 30이고, X는 임의의 아미노산이다. 일부 실시양태에서, 링커는 아미노산 서열 $(GGS)_n$ (서열식별번호: 610)을 포함하며, 여기서 n은 1, 3 또는 7이다. 일부 실시양태에서, 링커는 아미노산 서열 $SGGS(GGS)_n$ (서열식별번호: 612)을 포함하며, 여기서 n은 2이다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605)를 포함한다.

[0031] 일부 실시양태에서, 융합 단백질은 구조 [핵산 편집 도메인]-[임의적인 링커 서열]-[napDNAbp]-[임의적인 링커 서열]-[UGI]를 포함한다. 일부 실시양태에서, 융합 단백질은 구조 [핵산 편집 도메인]-[임의적인 링커 서열]-[UGI]-[임의적인 링커 서열]-[napDNAbp]; [UGI]-[임의적인 링커 서열]-[핵산 편집 도메인]-[임의적인 링커 서열]-[napDNAbp]; [UGI]-[임의적인 링커 서열]-[napDNAbp]-[임의적인 링커 서열]-[핵산 편집 도메인]; [napDNAbp]-[임의적인 링커 서열]-[UGI]-[임의적인 링커 서열]-[핵산 편집 도메인]; [napDNAbp]-[임의적인 링커 서열]-[핵산 편집 도메인]-[임의적인 링커 서열]-[UGI]; 또는 [핵산 편집 도메인]-[임의적인 링커 서열]-[napDNAbp]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]를 포함한다.

[0032] 일부 실시양태에서, 핵산 편집 도메인은 데아미나제를 포함한다. 일부 실시양태에서, 핵산 편집 도메인은 데아미나제를 포함한다. 일부 실시양태에서, 데아미나제는 시티딘 데아미나제이다. 일부 실시양태에서, 데아미나제는 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC1 데아미나제, APOBEC2 데아미나제, APOBEC3A 데아미나제, APOBEC3B 데아미나제, APOBEC3C 데아미나제, APOBEC3D 데아미나제, APOBEC3F 데아미나제, APOBEC3G 데아미나제, APOBEC3H 데아미나제 또는 APOBEC4 데아미나제이다. 일부 실시양태에서, 데아미나제는 활성화-유도된 데아미나제 (AID)이다. 일부 실시양태에서, 데아미나제는 시티딘 데아미나제 1 (CDA1)이다. 일부 실시양태에서, 데아미나제는 칠성장어 CDA1 (pmCDA1) 데아미나제이다.

[0033] 일부 실시양태에서, 데아미나제는 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스로부터의 것이다. 일부 실시양태에서, 데아미나제는 인간으로부터의 것이다. 일부 실시양태에서 데아미나제는 래트로부터의 것이다. 일부 실시양태에서, 데아미나제는 (서열식별번호: 76)에 제시된 아미노산 서열을 포함하는 래트 APOBEC1 데아미나제이다. 일부 실시양태에서, 데아미나제는 (서열식별번호: 74)에 제시된 아미노산 서열을 포함하는 인간 APOBEC1 데아미나제이다. 일부 실시양태에서, 데아미나제는 pmCDA1 (서열식별번호: 81)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G (서열식별번호: 60)이다. 일부 실시양태에서, 데아미나제는 (서열식별번호: 82-84) 중 어느 하나의 인간 APOBEC3G 변이체이다. 일부 실시양태에서, 데아미나제는 서열식별번호: 49-84에 제시된 아미노산 서열 중 어느 하나와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일하다.

[0034] 일부 실시양태에서, UGI 도메인은 서열식별번호: 134와 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, UGI 도메인은 서열식별번호: 134에 제시된 바와 같은 아미노산 서열을 포함한다.

[0035] 본 개시내용의 일부 측면은 본원에 제공된 바와 같은 napDNAbp 융합 단백질, 및 napDNAbp에 결합된 가이드 RNA를 포함하는 복합체를 제공한다.

[0036] 본 개시내용의 일부 측면은 본원에 제공된 napDNAbp, 융합 단백질 또는 복합체를 사용하는 방법을 제공한다. 예를 들어, 본 개시내용의 일부 측면은 DNA 분자를 (a) 본원에 제공된 바와 같은 napDNAbp 또는 융합 단백질과, 및 가이드 RNA와 (여기서 가이드 RNA는 약 15-100 뉴클레오티드 길이이고, 표적 서열에 상보적인 적어도 10개의 인접 뉴클레오티드를 포함함); (b) 또는 본원에 제공된 바와 같은 napDNAbp, napDNAbp 융합 단백질, 또는 gRNA를 갖는 napDNAbp 또는 napDNAbp 복합체와 접촉시키는 단계를 포함하는 방법을 제공한다.

[0037] 본 개시내용의 일부 측면은 (a) 본원에 제공된 바와 같은 napDNAbp 또는 napDNAbp 융합 단백질을 코딩하는 뉴클레오티드 서열; 및 (b) (a)의 서열의 발현을 유도하는 이종 프로모터를 포함하는 핵산 구축물을 포함하는 키트를 제공한다. 일부 실시양태에서, 키트는 가이드 RNA 백본을 코딩하는 발현 구축물을 추가로 포함하며, 여기서 구축물은 표적 서열과 동일하거나 그에 상보적인 핵산 서열의 가이드 RNA 백본 내로의 클로닝을 허용하도록 위치된 클로닝 부위를 포함한다.

[0038] 본 개시내용의 일부 측면은 본원에 제공된 바와 같은 융합 단백질의 napDNAbp를 코딩하는 폴리뉴클레오티드를 제공한다. 본 개시내용의 일부 측면은 이러한 폴리뉴클레오티드를 포함하는 벡터를 제공한다. 일부 실시양태에서, 벡터는 폴리뉴클레오티드의 발현을 유도하는 이종 프로모터를 포함한다.

[0039] 본 개시내용의 일부 측면은 본원에 제공된 바와 같은 napDNAbp 단백질, 융합 단백질, 핵산 분자 및/또는 벡터를 포함하는 세포를 제공한다.

[0040] Cas9 도메인 (예를 들어 Cas9, nCas9, 또는 dCas9)을 포함하는 본원에 제공된 융합 단백질 중 임의의 것은 본원에 제공된 napDNAbp 중 임의의 것, 예를 들어 CasX, CasY, Cpf1, C2c1, C2c2, C2c3 또는 아르고노트 단백질로 대체될 수 있다는 것이 인지되어야 한다.

[0041] 상기 리포터 시스템의 예시적 실시양태의 설명은 단지 예시적 목적으로 제공되며, 제한하는 것으로 비의도된다. 추가의 리포터 시스템, 예를 들어 상기 상세히 기재된 예시적 시스템의 변형이 또한 본 개시내용에 포괄된다.

[0042] 상기 요약은 비제한적 방식으로 본원에 개시된 기술의 실시양태, 이점, 특색 및 용도의 일부를 예시하는 것으로 의도된다. 본원에 개시된 기술의 다른 실시양태, 이점, 특색 및 용도는 상세한 설명, 도면, 실시예 및 청구범위로부터 명백할 것이다.

도면의 간단한 설명

[0043] 도 1은 단일 가닥 DNA 기질에 대한 데아미나제의 데아미나제 활성을 나타낸다. 무작위화된 PAM 서열 (NNN PAM)을 사용한 단일 가닥 DNA 기질을 음성 대조군으로서 사용하였다. 사용된 정규 PAM 서열은 (NGG PAM)을 포함한다.

도 2는 단일 가닥 DNA 기질에 대한 Cas9:데아미나제 융합 단백질의 활성을 나타낸다.

도 3은 Cas9:데아미나제:sgRNA 복합체에 의한 이중 가닥 DNA 기질 결합을 예시한다.

도 4는 이중 가닥 DNA 탈아미노화 검정을 예시한다.

도 5는 Cas9 융합물이 이중-가닥 DNA 표적 서열의 위치 3-11 (도 5에서의 개략도에 따라 넘버링됨)을 표적화할 수 있다는 것을 입증한다. 상부 겔: 1 μM rAPOBEC1-GGS-dCas9, 125 nM dsDNA, 1 당량 sgRNA. 중간 겔: 1 μM rAPOBEC1-$(GGS)_3$(서열식별번호: 610)-dCas9, 125 nM dsDNA, 1 당량 sgRNA. 하부 겔: 1.85 μM rAPOBEC1-XTEN-dCas9, 125 nM dsDNA, 1 당량 sgRNA.

도 6은 교정 가이드 RNA, 예를 들어, 교정 sgRNA가 데아미나제 활성에 요구된다는 것을 입증한다.

도 7은 데아미나제-dCas9:sgRNA 복합체에 의한 생체내 표적 서열의 표적 DNA 결합의 메카니즘을 예시한다.

도 8은 예시적인 질환-연관된 표적 서열의 성공적인 탈아미노화를 나타낸다.

도 9는 His6-rAPOBEC1-XTEN-dCas9를 사용한 시험관내 C→T 편집 효율을 나타낸다.

도 10은 HEK293T 세포에서의 C→T 편집 효율이 UGI와의 융합에 의해 크게 증진된다는 것을 나타낸다.

도 11a 내지 11c는 NBE1이 시험관내에서 특이적, 가이드 RNA-프로그램된 C에서 U로의 전환을 매개한다는 것을 나타낸다. 도 11a: 핵염기 편집 전략. 가이드 RNA에 의해 특정된 유전자좌에 표적 C를 갖는 DNA가 dCas9에 의해 결합되며, 이는 DNA 기질의 국소적 변성을 매개한다. 테더링된 APOBEC1 효소에 의한 시티딘 탈아미노화는 표적 C를 U로 전환시킨다. 생성된 G:U 헤테로듀플렉스는 DNA 복제 또는 복구 후에 A:T 염기 쌍으로 영구적으로 전환될 수 있다. U가 주형 DNA 가닥 중에 있는 경우, 이는 또한 전사 후에 G에서 A로의 돌연변이를 함유하는 RNA 전사체를 발생시킬 것이다. 도 11b: 대략 5개의 뉴클레오티드의 활성 윈도우를 나타내는 탈아미노화 검정. 37℃에서 2시간 동안 dsDNA 기질과 함께 NBE1-sgRNA 복합체의 인큐베이션 후, 5' 형광단-표지된 DNA를 단리하고, 37℃에서 1시간 동안 USER 효소 (우라실 DNA 글리코실라제 및 엔도뉴클레아제 VIII)와 함께 인큐베이션하여 임의의 우라실의 부위에서 DNA 절단을 유도하였다. 생성된 DNA를 변성 폴리아크릴아미드 겔 상에서 분할하고, 임의의 형광단-연결된 가닥을 가시화하였다. 각각의 레인은 프로토스페이서 내의 표적 C의 위치에 따라, 또는 표적 C가 존재하지 않는 경우 "-"로 표지되며, 이는 PAM으로부터 원위의 염기를 위치 1로서 카운팅한다. 도 11c: NBE1의 서열 특이성 및 sgRNA-의존성을 나타내는 데아미나제 검정. 위치 7에 표적 C를 갖는 DNA 기질을 도 11b에서와 같이 교정 sgRNA, 미스매칭된 sgRNA와 함께, 또는 sgRNA 없이 NBE1과 함께 인큐베이션하였다. C에서 U로의 편집은 미스매칭된 sgRNA에서 또는 sgRNA가 없는 것에서 관찰되지 않는다. 양성 대조군 샘플은 위치 7에 합성적으로 혼입된 U를 갖는 DNA 서열을 함유한다.

도 12a 내지 12b는 시험관내에서의 핵염기 편집 효율에 대한 서열 컨텍스트 및 표적 C 위치의 효과를 나타낸다. 도 12a: 시험관내에서의 편집 효율에 대한 표적 C 주위의 서열을 변화시키는 것의 효과. 프로토스페이서 서열 5'-TTATTTCGTGGATTTATTTA-3' (서열식별번호: 591)에서 C_7에 대한 표적화된 가닥의 80% (양쪽 가닥으로부터의 총 서열분석 판독물의 40%)의 탈아미노화 수율은 1.0으로서 정의되었으며, 위치 1-6 및 8-13에 모든 가능한 단일-염기 돌연변이를 함유하는 기질의 상대적 탈아미노화 효율이 나타나 있다. 값 및 오차 막대는 상이한 날에 수행된 2회 이상의 독립적인 생물학적 반복의 평균 및 표준 편차를 반영한다. 도 12b: 시험관내에서의 편집 효율에 대한 각각의 NC 모티프의 위치적 효과. 각각의 NC 표적 모티프는 우측에 제시된 서열에 지시된 바와 같이 프로토스페이서 내의 위치 1에서 8까지 다양화되었다 (PAM은 적색으로 나타냄, 프로토스페이서 플러스 프로토스페이서에 대해 5'의 1개의 염기가 또한 제시됨). NBE1과 함께 인큐베이션 후의 넘버링된 표적 C 위치의 각각에 T를 함유하는 총 서열 판독물의 백분율이 그래프에 나타나 있다. 시험관내에서 최대 가능한 탈아미노화 수율은 총 서열분석 판독물의 50% (표적화된 가닥의 100%)라는 것에 주목한다. 값 및 오차 막대는 상이한 날에 수행된 2회 또는 3회의 독립적인 생물학적 반복의 평균 및 표준 편차를 반영한다. 도 12b는 위에서 아래로 각각 서열식별번호: 619 내지 626을 나타낸다.

도 13a 내지 13c는 인간 세포에서의 핵염기 편집을 나타낸다. 도 13a: 핵염기 편집제에 의해 표적화된 6개의 포유동물 세포 게놈 유전자좌의 프로토스페이서 및 PAM 서열. 표적 C는 프로토스페이서 내의 그의 위치에 상응하는 아래첨자 수로 지시된다. 도 13a는 위에서 아래로 각각 서열식별번호: 127 내지 132를 나타낸다. 도 13b: HEK293T 세포를 NBE1, NBE2, 또는 NBE3 및 적절한 sgRNA를 발현하는 플라스미드로 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, 6개의 유전자좌에서 고처리량 DNA 서열분석에 의해 분석하였다. 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는, 세포의 C에서 T로의 전환 백분율이 모든 6개의 게놈 유전자좌에서 NBE1, NBE2, 및 NBE3에 대해, 및 6개의 부위 중 3개 (EMX1, HEK293 부위 3, 및 HEK293 부위 4)에서 공여자 HDR 주형을 갖는 야생형 Cas9에 대해 나타나 있다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 및 표준 편차를 반영한다. 도 13c: 모든 6개의 게놈 유전자좌에 대해 NBE2 및 NBE3으로의, 또는 6개의 부위 중 3개 (EMX1, HEK293 부위 3, 및 HEK293 부위 4)에서 야생형 Cas9 및 HDR에 대한 단일-가닥 DNA 주형으로의 HEK293T 세포의 처리 후의, 방법에 기재된 바와 같이 계산된 indel 형성의 빈도가 나타나 있다. 값은 상이한 날에 수행된 적어도 3회의 독립적인 생물학적 반복의 평균을 반영한다.

도 14a 내지 14c는 포유동물 세포에서의 3가지 질환-관련 돌연변이의 NBE2- 및 NBE3-매개된 교정을 나타낸다. 각각의 부위에 대해, 프로토스페이서의 서열은 돌연변이의 명칭의 우측에 지시되고, PAM 및 돌연변이의 원인이 되는 염기는 프로토스페이서 내의 그의 위치에 상응하는 아래첨자 수를 갖는 볼드체로 지시된다. 각각의 질환-연관된 대립유전자 위의 아미노산 서열은 적색으로의 핵염기 편집 후에 교정된 아미노산 서열과 함께 나타나 있다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 NBE2 또는 NBE3 및 적절한 sgRNA를 코딩하는 플라스미드로 뉴클레오펙션시켰다. 뉴클레오펙션 2일 후에, 게놈 DNA를 추출

하고, HTS에 의해 분석하여 병원성 돌연변이 교정을 평가하였다. 도 14a: 알츠하이머병-연관된 APOE4 대립유전자는 총 판독물의 11% (뉴클레오펙션된 성상세포의 44%)에서 NBE3에 의해 마우스 성상세포에서의 APOE3'로 전환된다. 2개의 인근의 C는 또한 T로 전환되지만, 생성된 단백질의 예측된 서열 (서열식별번호: 627)에 대한 변화는 없다. 도 14b: 암-연관된 p53 N239D 돌연변이는 돌연변이에 대해 이형접합성인 처리된 인간 림프종 세포의 11% (뉴클레오펙션된 세포의 12%)에서 NBE2에 의해 교정된다 (서열식별번호: 628). 도 14c: p53 Y163C 돌연변이는 뉴클레오펙션된 인간 유방암 세포의 7.6%에서 NBE3에 의해 교정된다 (서열식별번호: 629).

도 15a 내지 15d는 핵염기 편집에 대한 데아미나제-dCas9 링커 길이 및 조성의 효과를 나타낸다. GGS (도 15a), $(GGS)_3$ (서열식별번호: 610) (도 15b), XTEN (도 15c), 또는 $(GGS)_7$ (서열식별번호: 610)의 데아미나제-Cas9 링커를 갖는 핵염기 편집제의 탈아미노화 윈도우를 나타내는 겔-기반 데아미나제 검정 (도 15d). 37℃에서 2시간 동안 125 nM dsDNA 기질과 함께 1.85 μM 편집제-sgRNA 복합체의 인큐베이션 후, 염료-접합된 DNA를 단리하고, USER 효소 (우라실 DNA 글리코실라제 및 엔도뉴클레아제 VIII)와 함께 37℃에서 추가의 1시간 동안 인큐베이션하여 임의의 우라실의 부위에서 DNA 백본을 절단하였다. 생성된 DNA를 변성 폴리아크릴아미드 겔 상에서 분할하고, 염료-접합된 가닥을 영상화하였다. 각각의 레인은 프로토스페이서 내의 표적 C의 위치에 따라, 또는 표적 C가 존재하지 않는 경우 -로 넘버링된다. 8U는 위치 8에 합성적으로 혼입된 U를 갖는 양성 대조군 서열이다.

도 16a 내지 16b는 NBE1이 시험관내에서 질환-관련 돌연변이를 교정할 수 있다는 것을 나타낸다. 도 16a: 7가지 질환-관련 돌연변이의 프로토스페이서 및 PAM 서열. 각 경우에 질환-연관된 표적 C는 프로토스페이서 내의 그의 위치를 반영하는 아래첨자 수로 지시된다. 둘 다의 APOE4 SNP를 제외한 모든 돌연변이에 대해, 표적 C는 주형 (비-코딩) 가닥에 있다. 도 16a는 위에서 아래로 각각 서열식별번호: 631 내지 636을 나타낸다. 도 16b: NBE1와 함께 인큐베이션, DNA 단리, 및 USER 효소와 함께 인큐베이션하여 U를 함유하는 위치에서 DNA를 절단하기 전 (-) 및 후 (+)의 각각의 dsDNA 올리고뉴클레오티드를 나타내는 데아미나제 검정. USER 효소를 갖는 프로토스페이서 내의 다양한 위치에 U를 함유하는 합성 올리고뉴클레오티드의 인큐베이션으로부터의 양성 대조군 레인이 U의 위치를 지시하는 상응하는 수와 함께 나타나 있다.

도 17은 NBE1의 프로세싱성을 나타낸다. 8개의 연속적 C를 함유하는 60량체 DNA 올리고뉴클레오티드의 프로토스페이서 및 PAM이 상부에 나타나 있다. 올리고뉴클레오티드 (125 nM)를 NBE1 (2 μM)과 함께 37℃에서 2시간 동안 인큐베이션하였다. DNA를 단리하고, 고처리량 서열분석에 의해 분석하였다. 관찰된 가장 빈번한 9개의 서열에 대한 총 판독물의 퍼센트가 나타나 있다. 대다수의 편집된 가닥 (>93%)은 T로 전환된 1개 초과의 C를 갖는다. 이 도면은 서열식별번호: 309를 나타낸다.

도 18a 내지 18h는 UGI를 NBE1에 융합시켜 NBE2를 생성하는 것의 효과를 나타낸다. 도 18a: 핵염기 편집제로 표적화된 6개의 포유동물 세포 게놈 유전자좌의 프로토스페이서 및 PAM 서열. 편집가능한 C는 프로토스페이서 내의 그의 위치에 상응하는 표지로 지시된다. 도 18a는 위에서 아래로 각각 서열식별번호: 127 내지 132를 나타낸다. 도 18b 내지 18g: HEK293T 세포를 NBE1, NBE2, 또는 NBE1 및 UGI, 및 적절한 sgRNA를 발현하는 플라스미드로 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, 6개의 유전자좌에서 고처리량 DNA 서열분석에 의해 분석하였다. 지시된 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 모든 6개의 게놈 유전자좌에서 NBE1, NBE1 및 UGI, 및 NBE2에 대해 나타나 있다. 도 18h: NBE1, 별개의 플라스미드 상의 NBE1 플러스 UGI, NBE2, 및 비처리된 세포에 대한 관심 프로토스페이서 주위의 510개의 C에서의 C에서 T로의 돌연변이 비율이 나타나 있다. 데이터는 $1.5x10^6$개의 세포로부터의 3,000,000개의 DNA 서열분석 판독물의 결과를 나타낸다. 값은 상이한 날에 수행된 적어도 2회의 생물학적 실험의 평균을 반영한다.

도 19는 U2OS 및 HEK293T 세포에서의 NBE2의 핵염기 편집 효율을 나타낸다. NBE2에 의한 세포의 C에서 T로의 전환 백분율이 HEK293T 세포 및 U2OS 세포에서의 6개의 표적화된 게놈 유전자좌 각각에 대해 나타나 있다. HEK293T 세포를 리포펙타민 2000을 사용하여 형질감염시키고, U2OS 세포를 뉴클레오펙션시켰다. U2OS 뉴클레오펙션 효율은 74%였다. 플라스미드 전달 3일 후에, 게놈 DNA를 추출하고, HTS에 의해 6개의 게놈 유전자좌에서 핵염기 편집에 대해 분석하였다. 값 및 오차 막대는 상이한 날에 수행된 적어도 2회의 생물학적 실험의 평균 및 표준 편차를 반영한다.

도 20은 핵염기 편집이 다중 세포 분열에 걸쳐 지속된다는 것을 나타낸다. NBE2에 의한 세포의 C에서 T로의 전환 백분율이 세포를 계대하기 전 및 후의 HEK293T 세포에서의 2개의 게놈 유전자좌에서 제시된다. HEK293T 세포를 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 세포를 수거하고, 반으로 분할하였다.

하나의 반은 HTS 분석에 적용하고, 다른 반은 대략 5회의 세포 분열 동안 번식시킨 후, 수거하고, HTS 분석에 적용하였다.

도 21은 원칙적으로 핵염기 편집에 의해 교정될 수 있는 클린바(ClinVar)로부터의 유전적 변이체를 나타낸다. 인간 유전적 변이 및 그의 상응하는 표현형의 NCBI 클린바 데이터베이스[68]를 현행 핵염기 편집 기술에 의해 교정될 수 있는 유전 질환에 대해 조사하였다. 결과를 좌측에 열거된 연속적 제한을 시행함으로써 필터링하였다. x-축은 로그 스케일의 그 제한 및 모든 상기 제한들을 충족시키는 경우의 수를 나타낸다.

도 22는 6개의 게놈 유전자좌에서의 편집가능한 C의 시험관내 확인을 나타낸다. 6개의 상이한 게놈 부위를 매칭시키는 서열을 갖는 합성 80량체를 NBE1과 함께 인큐베이션한 후, HTS를 통해 핵염기 편집에 대해 분석하였다. 각각의 부위에 대해, 프로토스페이서의 서열은 부위의 명칭의 우측에 지시되고, PAM은 적색으로 강조된다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 DNA 서열분석 판독물의 백분율이 있다. 표적 C는 시험관내 전환 효율이 >10%인 경우 "편집가능한" 것으로 간주되었다. 최대 수율은 비-표적화된 가닥이 핵염기 편집에 대한 기질이 아니기 때문에 총 DNA 서열분석 판독물의 50%이다. 이 도면은 위에서 아래로 각각 서열식별번호: 127 내지 132를 나타낸다.

도 23은 EMX1 오프-타겟에서의 NBE1, NBE2, 및 NBE3의 활성을 나타낸다. HEK293T 세포를 NBE1, NBE2, 또는 NBE3 및 EMX1 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 EMX1 sgRNA에 대한 온-타겟(on-target) 유전자좌, 플러스 상부 10개의 공지된 Cas9 오프-타겟 유전자좌에서, GUIDE-seq 방법[55]을 사용하여 이전에 결정된 바와 같이 분석하였다. EMX1 오프-타겟 5개의 유전자좌는 증폭하지 않았으며, 나타내지 않는다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 NBE1, NBE2, 및 NBE3에 대해 나타나 있다. 맨 우측에는 각각의 서열에 대해 보고된 서열분석 판독물의 총 수가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 127, 및 637 내지 645를 나타낸다.

도 24는 FANCF 오프-타겟에서의 NBE1, NBE2, 및 NBE3의 활성을 나타낸다. HEK293T 세포를 NBE1, NBE2, 또는 NBE3 및 FANCF 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 FANCF sgRNA에 대한 온-타겟 유전자좌, 플러스 모든 공지된 Cas9 오프-타겟 유전자좌에서, GUIDE-seq 방법[55]을 사용하여 이전에 결정된 바와 같이 분석하였다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 NBE1, NBE2, 및 NBE3에 대해 나타나 있다. 맨 우측에는 각각의 서열에 대해 보고된 서열분석 판독물의 총 수가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 128 및 646 내지 653을 나타낸다.

도 25는 HEK293 부위 2 오프-타겟에서의 NBE1, NBE2, 및 NBE3의 활성을 나타낸다. HEK293T 세포를 NBE1, NBE2, 또는 NBE3 및 HEK293 부위 2 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 HEK293 부위 2 sgRNA에 대한 온-타겟 유전자좌, 플러스 모든 공지된 Cas9 오프-타겟 유전자좌에서, GUIDE-seq 방법[55]을 사용하여 이전에 결정된 바와 같이 분석하였다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 NBE1, NBE2, 및 NBE3에 대해 나타나 있다. 맨 우측에는 각각의 서열에 대해 보고된 서열분석 판독물의 총 수가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 129, 654, 및 655를 나타낸다.

도 26은 HEK293 부위 3 오프-타겟에서의 NBE1, NBE2, 및 NBE3의 활성을 나타낸다. HEK293T 세포를 NBE1, NBE2, 또는 NBE3 및 HEK293 부위 3 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 HEK293 부위 3 sgRNA에 대한 온-타겟 유전자좌, 플러스 모든 공지된 Cas9 오프-타겟 유전자좌에서, GUIDE-seq 방법[55]을 사용하여 이전에 결정된 바와 같이 분석하였다. 온-타겟 및 오프-타겟 프로토스페이서 및

프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 NBE1, NBE2, 및 NBE3에 대해 나타나 있다. 맨 우측에는 각각의 서열에 대해 보고된 서열분석 판독물의 총 수가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 130 및 656 내지 660을 나타낸다.

도 27은 HEK293 부위 4 오프-타겟에서의 NBE1, NBE2, 및 NBE3의 활성을 나타낸다. HEK293T 세포를 NBE1, NBE2, 또는 NBE3 및 HEK293 부위 4 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 HEK293 부위 4 sgRNA에 대한 온-타겟 유전자좌, 플러스 상부 10개의 공지된 Cas9 오프-타겟 유전자좌에서, GUIDE-seq 방법[55]을 사용하여 이전에 결정된 바와 같이 분석하였다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 NBE1, NBE2, 및 NBE3에 대해 나타나 있다. 맨 우측에는 각각의 서열에 대해 보고된 서열분석 판독물의 총 수가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 131 및 661 내지 670을 나타낸다.

도 28은 비-표적 C 돌연변이 비율을 나타낸다. 여기에는 시험된 6개의 온-타겟 및 34개의 오프-타겟 유전자좌 주위의 2,500개의 별개의 시토신에서의 C에서 T로의 돌연변이 비율이 나타나 있으며, 이는 대략 $1.8x10^6$개의 세포로부터 유래된 총 14,700,000개의 서열 판독물을 나타낸다.

도 29a 내지 29c는 인간 세포에서의 염기 편집을 나타낸다. 도 29a는 포유동물 세포에서의 가능한 염기 편집 결과물을 나타낸다. 초기 편집은 U:G 미스매치를 발생시켰다. 우라실 DNA 글리코실라제 (UDG)에 의한 U의 인식 및 절제는 염기 절제 복구 (BER)를 개시하였으며, 이는 C:G 출발 상태로의 복귀를 야기한다. BER은 BE2 및 BE3에 의해 저해되었으며, 이는 UDG를 억제하였다. U:G 미스매치는 또한 미스매치 복구 (MMR)에 의해 프로세싱되었으며, 이는 미스매치의 니킹된 가닥을 우선적으로 복구하였다. BE3은 G를 함유하는 비-편집된 가닥을 니킹하였으며, 이는 U:G 미스매치의 목적하는 U:A 또는 T:A 결과물로의 해결을 선호한다. 도 29b는 하기 실시예에서의 물질 및 방법에 기재된 바와 같이 처리된 HEK293T 세포를 나타낸다. 지시된 표적 위치에서 T를 갖는 총 DNA 서열분석 판독물의 백분율은 BE1, BE2, 또는 BE3으로의 처리, 또는 공여자 HDR 주형을 갖는 야생형 Cas9로의 처리에 대해 나타낸다. 도 29c는 도 29b에서의 처리 후의 indel 형성의 빈도를 나타낸다. 값은 도 34에 열거되어 있다. 도 29b 및 29c에 대해, 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 및 s.d.를 반영한다.

도 30a 내지 30b는 포유동물 세포에서의 2가지 질환-관련 돌연변이의 BE3-매개된 교정을 나타낸다. 프로토스페이서의 서열은 돌연변이의 우측에 나타나 있으며, PAM 및 표적 염기는 적색으로 프로토스페이서 내의 그의 위치를 지시하는 아래첨자 수와 함께 나타나 있다. 각각의 서열의 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 물질 및 방법에 기재된 바와 같이 처리하였다. 도 30a는 총 판독물의 74.9%에서 BE3에 의한 마우스 성상세포에서 APOE3r로 전환된 알츠하이머병-연관된 APOE4 대립유전자를 나타낸다. 2개의 인근의 C는 또한 T로 전환되었지만, 생성된 단백질의 예측된 서열에 대한 변화는 없었다. 야생형 Cas9 및 공여자 ssDNA로의 이들 세포의 동일한 처리는 26.1% indel 형성과 함께, 단지 0.3% 교정을 발생시킨다. 이 도면은 서열식별번호: 671 및 627을 나타낸다. 도 30b는 0.7% indel 형성을 갖는 뉴클레오펙션된 인간 유방암 세포의 7.6%에서 BE3에 의해 교정된 암 연관된 p53 Y163C 돌연변이를 나타낸다. 야생형 Cas9 및 공여자 ssDNA로의 이들 세포의 동일한 처리는 6.1% indel 형성과 함께, 돌연변이 교정을 발생시키지 않는다. 이 도면은 서열식별번호: 672 및 629를 나타낸다.

도 31은 HEK293 부위 2 오프-타겟에서의 BE1, BE2, 및 BE3의 활성을 나타낸다. HEK293T 세포를 BE1, BE2, 또는 BE3 및 HEK293 부위 2 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 HEK293 부위 2 sgRNA에 대한 온-타겟 유전자좌, 플러스 모든 공지된 Cas9 및 dCas9 오프-타겟 유전자좌에서, GUIDE-seq 방법을 사용하여 종(Joung) 및 동료들 (63), 및 크로마틴 면역침전 고처리량 서열분석 (ChIP-seq) 실험을 사용하여 아들리(Adli) 및 동료들 (18)에 의해 이전에 결정된 바와 같이 분석하였다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 BE1, BE2, 및 BE3에 대해 나타나 있다. 맨 우측에는 보고된 서열분석 판독물의 총 수, 및 각각의 서열에 대해 보고된 ChIP-seq 신호 강도가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 129, 654, 655 및 673

내지 677을 나타낸다.

도 32는 HEK293 부위 3 오프-타겟에서의 BE1, BE2, 및 BE3의 활성을 나타낸다. HEK293T 세포를 BE1, BE2, 또는 BE3 및 HEK293 부위 3 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 HEK293 부위 3 sgRNA에 대한 온-타겟 유전자좌, 플러스 모든 공지된 Cas9 오프-타겟 유전자좌 및 상부 5개의 공지된 dCas9 오프-타겟 유전자좌에서, GUIDE-seq 방법[54]을 사용하여, 및 크로마틴 면역침전 고처리량 서열분석 (ChIP-seq) 실험[61]을 사용하여 종 및 동료들에 의해 이전에 결정된 바와 같이 분석하였다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 BE1, BE2, 및 BE3에 대해 나타나 있다. 맨 우측에는 보고된 서열분석 판독물의 총 수, 및 각각의 서열에 대해 보고된 ChIP-seq 신호 강도가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 130, 656 내지 660 및 678-682를 나타낸다.

도 33은 HEK293 부위 4 오프-타겟에서의 BE1, BE2, 및 BE3의 활성을 나타낸다. HEK293T 세포를 BE1, BE2, 또는 BE3 및 HEK293 부위 4 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 고처리량 DNA 서열분석에 의해 HEK293 부위 4 sgRNA에 대한 온-타겟 유전자좌, 플러스 상부 10개의 공지된 Cas9 오프-타겟 유전자좌 및 상부 5개의 공지된 dCas9 오프-타겟 유전자좌에서, GUIDE-seq 방법[54]을 사용하여, 및 크로마틴 면역침전 고처리량 서열분석 (ChIP-seq) 실험[61]을 사용하여 이전에 결정된 바와 같이 분석하였다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다. 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 BE1, BE2, 및 BE3에 대해 나타나 있다. 맨 우측에는 보고된 서열분석 판독물의 총 수, 및 각각의 서열에 대해 보고된 ChIP-seq 신호 강도가 제시된다. 이 도면은 위에서 아래로 각각 서열식별번호: 131, 661 내지 670, 683 및 684를 나타낸다.

도 34는 마우스 성상세포에서 알츠하이머병-연관된 APOE4 대립유전자의 APOE3r로의 BE3-매개된 교정 후의 비-프로토스페이서 염기의 돌연변이 비율을 나타낸다. 도 30a 및 도 34b로부터의 프로토스페이서의 어느 한 측 상의 50개의 염기의 DNA 서열이 프로토스페이서에 대한 각각의 염기의 위치와 함께 나타나 있다. PAM에 대해 원위의 프로토스페이서의 측은 양수로 지정되는 반면, PAM과 함께 PAM을 포함하는 측은 음수로 지정된다. 각각의 서열의 아래에는 비처리된 세포에 대한, BE3 및 APOE4 C158R 돌연변이를 표적화하는 sgRNA로 처리된 세포에 대한, 또는 BE3 및 VEGFA 유전자좌를 표적화하는 sgRNA로 처리된 세포에 대한 상응하는 염기를 갖는 총 DNA 서열분석 판독물의 백분율이 있다. 어떠한 BE3-처리된 샘플도 비처리된 대조군의 것 초과의 돌연변이 비율을 발생시키지 않았다. 이 도면은 위에서 아래로 각각 서열식별번호: 685 내지 688을 나타낸다.

도 35는 HCC1954 인간 세포에서의 암-연관된 p53 Y163C 돌연변이의 BE3-매개된 교정 후의 비-프로토스페이서 염기의 돌연변이 비율을 나타낸다. 도 30b 및 도 39b로부터의 프로토스페이서의 어느 한 측 상의 50개의 염기의 DNA 서열이 프로토스페이서에 대한 각각의 염기의 위치와 함께 나타나 있다. PAM에 대해 원위의 프로토스페이서의 측은 양수로 지정되는 반면, PAM과 함께 PAM을 포함하는 측은 음수로 지정된다. 각각의 서열의 아래에는 비처리된 세포에 대한, BE3 및 TP53 Y163C 돌연변이를 표적화하는 sgRNA로 처리된 세포에 대한, 또는 BE3 및 VEGFA 유전자좌를 표적화하는 sgRNA로 처리된 세포에 대한 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 어떠한 BE3-처리된 샘플도 비처리된 대조군의 것 초과의 돌연변이 비율을 발생시키지 않았다. 이 도면은 위에서 아래로 각각 서열식별번호: 689 내지 692를 나타낸다.

도 36a 내지 36f는 염기 편집에 대한 데아미나제, 링커 길이, 및 링커 조성의 효과를 나타낸다. 도 36a는 ssDNA에 대한 rAPOBEC1, pmCDA1, hAID, hAPOBEC3G, rAPOBEC1-GGS-dCas9, rAPOBEC1-$(GGS)_3$(서열식별번호: 610)-dCas9, 및 dCas9-$(GGS)_3$(서열식별번호: 610)-rAPOBEC1의 활성을 나타내는 겔-기반 데아미나제 검정을 나타낸다. 효소를 포유동물 세포 용해물-유래된 시험관내 전사-번역 시스템에서 발현시키고, 1.8 μM 염료-접합된 ssDNA 및 USER 효소 (우라실 DNA 글리코실라제 및 엔도뉴클레아제 VIII)와 함께 37℃에서 2시간 동안 인큐베이션하였다. 생성된 DNA를 변성 폴리아크릴아미드 겔 상에서 분할하고, 영상화하였다. 양성 대조군은 표적 C와 동일한 위치에 합성적으로 혼입된 U를 갖는 서열이다. 도 36b는 도 36c 내지 36f에 사용된 발현되고 정제된 단백질의

쿠마시-염색된 변성 PAGE 겔을 나타낸다. 도 36c 내지 36f는 GGS (도 36c), $(GGS)_3$ (서열식별번호: 610) (도 36d), XTEN (도 36e), 또는 $(GGS)_7$ (서열식별번호: 610) (도 36f)의 데아미나제-Cas9 링커를 갖는 염기 편집제의 탈아미노화 윈도우를 나타내는 겔-기반 데아미나제 검정을 나타낸다. 125 nM dsDNA 기질과 함께 sgRNA와 복합체화된 1.85 μM 데아미나제-dCas9 융합물을 37℃에서 2시간 동안 인큐베이션한 후, 염료-접합된 DNA를 단리하고, USER 효소와 함께 37℃에서 1시간 동안 인큐베이션하여 임의의 우라실의 부위에서 DNA 백본을 절단하였다. 생성된 DNA를 변성 폴리아크릴아미드 겔 상에서 분해하고, 염료-접합된 가닥을 영상화하였다. 각각의 레인은 프로토스페이서 내의 표적 C의 위치에 따라, 또는 표적 C가 존재하지 않는 경우 -로 넘버링된다. 8U는 위치 8에 합성적으로 혼입된 U를 갖는 양성 대조군 서열이다.

도 37a 내지 37c는 BE1 염기 편집 효율이 포유동물 세포에서 극적으로 감소되었음을 나타낸다. 도 37a: 염기 편집제에 의해 표적화된 6개의 포유동물 세포 게놈 유전자좌의 프로토스페이서 및 PAM 서열. 표적 C는 프로토스페이서 내의 그의 위치에 상응하는 아래첨자 수와 함께 적색으로 지시된다. 도 37b는 6개의 상이한 게놈 부위를 매칭시키는 서열을 갖는 합성 80량체를 BE1과 함께 인큐베이션한 후, HTS에 의해 염기 편집에 대해 분석하였음을 나타낸다. 각각의 부위에 대해, PAM과 함께 프로토스페이서의 서열은 부위의 명칭의 우측에 지시된다. 각각의 서열의 아래에는 상응하는 염기를 갖는 총 DNA 서열분석 판독물의 백분율이 있다. 본 발명자들은 시험관내 전환 효율이 >10%인 경우 표적 C를 "편집가능한" 것으로 간주하였다. 최대 수율은 비-표적화된 가닥이 BE1에 의해 영향을 받지 않기 때문에 총 DNA 서열분석 판독물의 50%이다. 값은 단일 실험으로부터 나타나 있다. 도 37c는 HEK293T 세포를 BE1 및 적절한 sgRNA를 발현하는 플라스미드로 형질감염시켰음을 나타낸다. 형질감염 3일 후에, 게놈 DNA를 추출하고, 고처리량 DNA 서열분석에 의해 6개의 유전자좌에서 분석하였다. 지시된 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 BE1에 대해 모든 6개의 게놈 유전자좌에서 나타나 있다. HEK293T 세포로부터의 모든 데이터의 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 및 표준 편차를 반영한다. 도 37a는 위에서 아래로 각각 서열식별번호: 127 내지 132를 나타낸다. 도 37b는 위에서 아래로 각각 서열식별번호: 127 내지 132를 나타낸다.

도 38은 염기 편집이 다중 세포 분열에 걸쳐 지속된다는 것을 나타낸다. BE2 및 BE3에 의한 세포의 C에서 T로의 전환 백분율이 세포를 계대하기 전 및 후에 HEK293T 세포에서 HEK293 부위 3 및 4에 대해 나타나 있다. HEK293T 세포를 BE2 또는 BE3 및 HEK293 부위 3 또는 4를 표적화하는 sgRNA를 발현하는 플라스미드로 뉴클레오펙션시켰다. 뉴클레오펙션 3일 후에, 세포를 수거하고, 반으로 분할하였다. 하나의 반은 HTS 분석으로 처리하고, 다른 반은 대략 5회의 세포 분열 동안 번식시킨 후, 수거하고, HTS 분석에 적용하였다. 값 및 오차 막대는 적어도 2회의 생물학적 실험의 평균 및 표준 편차를 반영한다.

도 39a 내지 39c는 비-표적 C/G 돌연변이 비율을 나타낸다. 여기에는 시험된 6개의 온-타겟 및 34개의 오프-타겟 유전자좌 주위의 2,500개의 별개의 시토신 및 구아닌에서의 C에서 T로의 및 G에서 A로의 돌연변이 비율이 나타나 있으며, 이는 대략 $1.8x10^6$개의 세포로부터 유래된 총 14,700,000개의 서열 판독물을 나타낸다. 도 39a 및 39b는 BE1, BE2, 및 BE3에 의한 세포의 비-표적 C에서 T로의 및 G에서 A로의 전환 백분율이 모든 2,500개의 시토신/구아닌에 대한 프로토스페이서에 대한 그의 위치에 대해 개별적으로 플롯팅된다는 것을 나타낸다. PAM에 대해 원위의 프로토스페이서의 측은 양수로 지정되는 반면, PAM을 포함하는 측은 음수로 지정된다. 도 39c는 BE1, BE2, 및 BE3에 의한 평균 비-표적 세포의 C에서 T로의 및 G에서 A로의 전환 백분율, 뿐만 아니라 최고 및 최저 개별 전환 백분율이 나타나 있다는 것을 나타낸다.

도 40a 내지 40b는 포유동물 세포에서의 2가지 질환-관련 돌연변이의 BE3-매개된 교정의 추가적인 데이터 세트를 나타낸다. 각각의 부위에 대해, 프로토스페이서의 서열은 돌연변이의 명칭의 우측에 지시되고, PAM 및 돌연변이의 원인이 되는 염기는 프로토스페이서 내의 그의 위치에 상응하는 아래첨자 수와 함께 적색 볼드체로 지시된다. 각각의 질환-연관된 대립유전자 위의 아미노산 서열이 염기 편집 후의 교정된 아미노산 서열과 함께 나타나 있다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 BE3 및 적절한 sgRNA를 코딩하는 플라스미드로 뉴클레오펙션시켰다. 뉴클레오펙션 2일 후에, 게놈 DNA를 뉴클레오펙션된 세포로부터 추출하고, HTS에 의해 분석하여 병원성 돌연변이 교정을 평가하였다. 도 40a는 알츠하이머병-연관된 APOE4 대립유전자가 단지 교정 sgRNA로 처리된 경우에만 총 판독물의 58.3%에서 BE3에 의해 마우스 성상세포에서 APOE3r로 전환된다는 것을 나타낸다. 2개의 인근의 C는 또한 T로 전환되지만, 생성된 단백질의 예측된 서열에 대한 변화는 없다. 야생형 Cas9 및 공여자 ssDNA로의 이들 세포의 동일한 처리는 26.7% indel 형성과 함께, 0.2% 교정을 발생시킨다. 도 40b는 암-연관된 p53 Y163C 돌연변이가 단지 교정 sgRNA로 처리된 경우에만

뉴클레오펙션된 인간 유방암 세포의 3.3%에서 BE3에 의해 교정된다는 것을 나타낸다. 야생형 Cas9 및 공여자 ssDNA로의 이들 세포의 동일한 처리는 8.0% indel 형성과 함께, 검출가능한 돌연변이 교정을 발생시키지 않는다. 도 40a 내지 40b는 서열식별번호: 671, 627, 672 및 629를 나타낸다.

도 41은 단일-가닥 DNA (ssDNA) 기질에 대한 다양한 데아미나제의 활성을 시험하는 데 사용될 수 있는 예시적인 USER (우라실-특이적 절제 시약) 효소-기반 검정의 개략적 표현을 나타낸다.

도 42는 pmCDA-nCas9-UGI-NLS 구축물 및 염기 편집제 (rAPOBEC1) 및 음성 대조군 (비처리)에 비한 HeK-3 부위에서의 그의 활성의 개략도이다. 이 도면은 서열식별번호: 693을 나타낸다.

도 43은 pmCDA1-XTEN-nCas9-UGI-NLS 구축물 및 염기 편집제 (rAPOBEC1) 및 음성 대조군 (비처리)에 비한 HeK-3 부위에서의 그의 활성의 개략도이다. 이 도면은 서열식별번호: 694를 나타낸다.

도 44는 시티딘 데아미나제 (CDA) 또는 APOBEC를 사용하여 T로 전환된 표적 C를 갖는 총 서열분석 판독물의 퍼센트를 나타낸다.

도 45는 데아미나제 (CDA) 또는 APOBEC를 사용하여 A로 전환된 표적 C를 갖는 총 서열분석 판독물의 퍼센트를 나타낸다.

도 46은 데아미나제 (CDA) 또는 APOBEC를 사용하여 G로 전환된 표적 C를 갖는 총 서열분석 판독물의 퍼센트를 나타낸다.

도 47은 huAPOBEC3G-XTEN-nCas9-UGI-NLS 구축물, 및 돌연변이된 형태 (huAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS, 염기 편집제 (rAPOBEC1) 및 음성 대조군 (비처리)에 비한 HeK-2 부위에서의 그의 활성의 개략도이다. 이 도면은 서열식별번호: 695를 나타낸다.

도 48은 실시예 7의 선택 검정에 사용된 LacZ 구축물의 개략도를 나타낸다.

도 49은 상이한 플라스미드 및 구축물로부터의 복귀 데이터를 나타낸다.

도 50은 lacZ 복귀 및 복귀된 클론의 정제의 확인을 나타낸다.

도 51은 실시예 7에 사용된 탈아미노화 선택 플라스미드를 나타내는 개략도이다.

도 52는 클로람페니콜 복귀 검정 (pmCDA1 융합)의 결과를 나타낸다.

도 53a 내지 53b는 2개의 구축물의 DNA 교정 유도를 입증하였다.

도 54는 클로람페니콜 복귀 검정 (huAPOBEC3G 융합)의 결과를 나타낸다.

도 55는 EMX1 오프-타겟에서의 BE3 및 HF-BE3의 활성을 나타낸다. 서열은 위에서 아래로 서열식별번호: 127 및 637-645에 상응한다.

도 56은 BE3 및 HF-BE3의 온-타겟 염기 편집 효율을 나타낸다.

도 57은 돌연변이가 다양한 정도로 시티딘 탈아미노화에 영향을 미친다는 것을 입증하는 그래프이다. 각각이 촉매작용을 약간 손상시키는 돌연변이의 조합은 다른 것에 비해 한 위치에서 선택적 탈아미노화를 허용한다. FANCF 부위는 $GGAATC_6C_7C_8TTC_{11}TGCAGCACCTGG$ (서열식별번호: 128)였다.

도 58은 차세대 염기 편집제를 나타내는 개략도이다.

도 59는 Cas9 변이체로부터 제조된 새로운 염기 편집제를 예시하는 개략도이다.

도 60은 상이한 NGA PAM 부위의 염기-편집된 백분율을 나타낸다.

도 61은 NGCG PAM EMX (VRER BE3) 및 $C_1TC_3C_4C_5ATC_8AC_{10}ATCAACCGGT$ (서열식별번호: 696) 스페이서를 사용한 시티딘의 염기-편집된 백분율을 나타낸다.

도 62는 상이한 NNGRRT PAM 부위로부터 생성되는 염기화된-편집된 백분율을 나타낸다.

도 63은 상이한 NNHRRT PAM 부위로부터 생성되는 염기화된-편집된 백분율을 나타낸다.

도 64a 내지 64c는 Cpf1 BE2를 사용하여 상이한 TTTN PAM 부위로부터 생성되는 염기-편집된 백분율을 나타낸다.

사용된 스페이서는 $TTTCCTC_3C_4C_5C_6C_7C_8C_9AC_{11}$AGGTAGAACAT (도 64a, 서열식별번호: 697), $TTTCC_1C_2TC_4TGTC_8C_9AC_{11}$ACCCTCATCCTG (도 64b, 서열식별번호: 698), 및 $TTTCC_1C_2C_3AGTC_7C_8TC_{10}C_{11}AC_{13}AC_{15}C_{16}C_{17}$TGAAAC (도 64c, 서열식별번호: 699)였다.

도 65는 시티딘 데아미나제 점 돌연변이유발의 동역학 조정을 통해 달성된 바와 같은 선택적 탈아미노화를 나타내는 개략도이다.

도 66은 스페이서에서 다중 시티딘으로의 세포 배양에서 프로빙된 탈아미노화 윈도우에 대한 다양한 돌연변이의 효과를 나타내는 그래프이다. 사용된 스페이서는 $TGC_3C_4C_5C_6TC_8C_9C_{10}TC_{12}C_{13}C_{14}$TGGCCC (서열식별번호: 700)였다.

도 67은 스페이서에서 다중 시티딘으로의 세포 배양에서 프로빙된 탈아미노화 윈도우에 대한 다양한 돌연변이의 효과를 나타내는 그래프이다. 사용된 스페이서는 $AGAGC_5C_6C_7C_8C_9C_{10}C_{11}TC_{13}$AAAGAGA (서열식별번호: 701)였다.

도 68은 제한된 수의 시티딘을 갖는 FANCF 부위에 대한 다양한 돌연변이의 효과를 나타내는 그래프이다. 사용된 스페이서는 $GGAATC_6C_7C_8TTC_{11}$TGCAGCACCTGG (서열식별번호: 128)였다. 삼중 돌연변이체 (W90Y, R126E, R132E)는 제6 위치에서 시티딘을 우선적으로 편집한다는 것에 주의한다.

도 69는 제한된 수의 시티딘을 갖는 HEK3 부위에 대한 다양한 돌연변이의 효과를 나타내는 그래프이다. 사용된 스페이서는 $GGCC_4C_5$AGACTGAGCACGTGATGG (서열식별번호: 702)였다. 이중 및 삼중 돌연변이체는 제4 위치에서의 시티딘에 비해 제5 위치에서의 시티딘을 우선적을 편집한다는 것에 주의한다.

도 70은 제한된 수의 시티딘을 갖는 EMX1 부위에 대한 다양한 돌연변이의 효과를 나타내는 그래프이다. 사용된 스페이서는 $GAGTC_5C_6$GAGCAGAAGAAGAAGGG (서열식별번호: 703)였다. 삼중 돌연변이체는 제6이 아니라 단지 제5 위치에서의 시티딘을 편집한다는 것에 주의한다.

도 71은 제한된 수의 시티딘을 갖는 HEK2 부위에 대한 다양한 돌연변이의 효과를 나타내는 그래프이다. 사용된 스페이서는 $GAAC_4AC_6$AAAGCATAGACTGCGGG (서열식별번호: 704)였다.

도 72는 불멸화 성상세포에서의 BE3 및 돌연변이 W90Y R132E를 포함하는 BE3의 온-타겟 염기 편집 효율을 나타낸다.

도 73은 3가지 Cpf1 융합 구축물의 개략도를 나타낸다.

도 74는 BE3 및 HF-BE3 (EMX1, FANCF, 및 RNF2)의 플라스미드 전달의 비교를 나타낸다.

도 75는 BE3 및 HF-BE3 (HEK3 및 HEK 4)의 플라스미드 전달의 비교를 나타낸다.

도 76은 모든 10개의 부위에서의 EMX-1의 오프-타겟 편집을 나타낸다. 이 도면은 서열식별번호: 127 및 637-645를 나타낸다.

도 77은 GAGTCCGAGCAGAAGAAGAAG (서열식별번호: 705) 스페이서를 사용한 HEK 세포에의 데아미나제 단백질 리포펙션을 나타낸다. EMX-1 온-타겟 및 EMX-1 오프 타겟 부위 2를 조사하였다.

도 78은 GGAATCCCTTCTGCAGCACCTGG (서열식별번호: 706) 스페이서를 사용한 HEK 세포에의 데아미나제 단백질 리포펙션을 나타낸다. FANCF 온 타겟 및 FANCF 오프 타겟 부위 1을 조사하였다.

도 79는 GGCCCAGACTGAGCACGTGA (서열식별번호: 707) 스페이서를 사용한 HEK 세포에의 데아미나제 단백질 리포펙션을 나타낸다. HEK-3 온 타겟 부위를 조사하였다.

도 80은 GGCACTGCGGCTGGAGGTGGGGG (서열식별번호: 708) 스페이서를 사용한 HEK 세포에의 데아미나제 단백질 리포펙션을 나타낸다. HEK-4 온 타겟, 오프 타겟 부위 1, 부위 3, 및 부위 4.

도 81은 sgHR_13 (GTCAGGTCGAGGGTTCTGTC (서열식별번호: 709) 스페이서; C8 표적: 정지를 위해 G51), sgHR_14 (GGGCCGCAGTATCCTCACTC (서열식별번호: 710) 스페이서; C7 표적; C7 표적: 정지를 위해 Q68), 및 sgHR_15 (CCGCCAGTCCCAGTACGGGA (서열식별번호: 711) 스페이서; C10 및 C11은 정지를 위한 표적: W239 또는 W237임)에 대한 sgRNA 활성에 대한 시험관내 검정의 결과를 나타낸다.

도 82는 sgHR_17 (CAACCACTGCTCAAAGATGC (서열식별번호: 712) 스페이서; C4 및 C5는 정지를 위한 표적: W410임), 및 sgHR_16 (CTTCCAGGATGAGAACACAG (서열식별번호: 713) 스페이서; C4 및 C5는 정지를 위한 표적:

W273임)에 대한 시험관내 검정의 결과를 나타낸다.

도 83은 제브라피시 배아에서의 sgHR_13과 복합체화된 BE3 단백질의 직접 주사를 나타낸다.

도 84는 제브라피시 배아에서의 sgHR_16과 복합체화된 BE3 단백질의 직접 주사를 나타낸다.

도 85는 제브라피시 배아에서의 sgHR_17과 복합체화된 BE3 단백질의 직접 주사를 나타낸다.

도 86은 시토신에서 티민으로의 변화를 생성할 수 있는 염기 편집제를 사용하여 생성될 수 있는 예시적인 핵산 변화를 나타낸다.

도 87은 염기 편집제 (예를 들어, BE3)가 하나의 APOE 이소형 (예를 들어, APOE4)을 알츠하이머병의 감소된 위험과 연관된 또 다른 APOE 이소형 (예를 들어, APOE3r)으로 편집하는 데 사용될 수 있는 방법을 입증하는 아포지단백질 E (APOE) 이소형의 예시를 나타낸다.

도 88은 마우스 성상세포에서의 APOE4에서 APOE3r로의 염기 편집을 나타낸다. 이 도면은 서열식별번호: 671 및 627을 나타낸다.

도 89는 아르기닌 잔기 37에서의 단백질의 조기 말단절단을 유발하는 PRNP의 염기 편집을 나타낸다. 이 도면은 서열식별번호: 557 및 714를 나타낸다.

도 90은 (UGI가 억제하는) UDG를 녹아웃시키는 것이 C에서 T로의 염기 편집의 효율의 청결도를 극적으로 개선시킨다는 것을 나타낸다.

도 91은 니카제를 갖지만 UGI를 갖지 않는 염기 편집제의 사용이 결과물의 혼합물을 매우 높은 indel 비율로 야기한다는 것을 나타낸다.

도 92a 내지 92g는 SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, 및 VRER-BE3이 인간 세포에서 비-NGG PAM을 함유하는 표적 부위에서의 효율적인 염기 편집을 매개한다는 것을 나타낸다. 도 92a는 에스. 피오게네스 및 에스. 아우레우스 Cas9를 사용한 염기 편집제 아키텍처를 나타낸다. 도 92b는 대안의 또는 완화된 PAM 요건을 갖는 최근 특징화된 Cas9 변이체를 나타낸다. 도 92c 및 92d는 실시예 12에 기재된 바와 같이 제시된 염기 편집제 변이체로 처리된 HEK293T 세포를 나타낸다. 지시된 표적 위치에 T로 전환된 C를 갖는 (형질감염된 세포에 대해 풍부화가 없는) 총 DNA 서열분석 판독물의 백분율이 나타나 있다. 시험된 각각의 표적의 PAM 서열은 X-축 아래에 나타나 있다. 차트는 NNGRRT PAM을 갖는 게놈 유전자좌에서의 SaBE3 및 SaKKH-BE3 (도 92c), NNNRRT PAM을 갖는 게놈 유전자좌에서의 SaBE3 및 SaKKH-BE3 (도 92d), NGAG PAM (도 92e), 및 NGAH PAM (도 92f)을 갖는 게놈 유전자좌에서의 VQR-BE3 및 EQR-BE3, 및 NGCG PAM을 갖는 게놈 유전자좌에서의 VRER-BE3 (도 92g)에 대한 결과를 나타낸다. 값 및 오차 막대는 적어도 2회의 생물학적 반복의 평균 및 표준 편차를 반영한다.

도 93a 내지 93c는 시티딘 데아미나제 도메인에서의 돌연변이를 갖는 염기 편집제가 좁혀진 편집 윈도우를 나타낸다는 것을 입증한다. 도 93a 내지 93c는 돌연변이체 염기 편집제 및 적절한 sgRNA를 발현하는 플라스미드로 형질감염된 HEK293T 세포를 나타낸다. 형질감염 3일 후에, 게놈 DNA를 추출하고, 고처리량 DNA 서열분석에 의해 지시된 유전자좌에서 분석하였다. 지시된 표적 위치에 T로 변화된 C를 갖는 (형질감염된 세포에 대한 풍부화가 없는) 총 DNA 서열분석 판독물의 백분율이 EMX1 부위 (서열식별번호: 721), HEK293 부위 3 (서열식별번호: 719), FANCF 부위 (서열식별번호: 722), HEK293 부위 2 (서열식별번호: 720), 부위 A (서열식별번호: 715), 및 부위 B (서열식별번호: 718) 유전자좌에 대해 나타나 있다. 도 93a는 염기 편집 윈도우를 좁히는 특정 시티딘 데아미나제 돌연변이를 예시한다. 추가의 돌연변이의 특징화에 대해서는 도 98을 참조한다. 도 93b는 게놈 유전자좌 상의 편집 윈도우 폭을 실시하는 시티딘 데아미나제 돌연변이의 효과를 나타낸다. 유익한 돌연변이를 조합하는 것은 편집 윈도우를 좁히는 것에 대해 상가 효과를 갖는다. 도 93c는 YE1-BE3, YE2-BE3, EE-BE3, 및 YEE-BE3이 염기 편집의 생성물 분포를 실시하여, BE3과 대조적으로 우세하게 단독으로-변형된 생성물을 생성한다는 것을 나타낸다. 값 및 오차 막대는 적어도 2회의 생물학적 반복의 평균 및 표준 편차를 반영한다.

도 94a 및 94b는 원칙적으로 이 연구에서 개발된 염기 편집제에 의해 교정될 수 있는 클린바로부터의 유전적 변이체를 나타낸다. 인간 유전적 변이 및 그의 상응하는 표현형의 NCBI 클린바 데이터베이스를 이론적으로 염기 편집에 의해 교정될 수 있는 유전 질환에 대해 조사하였다. 도 94a는 변경된 PAM 특이성을 갖는 염기 편집제의 사용을 통한 클린바 데이터베이스에서의 모든 병원성 T→C 돌연변이 중에서 염기 편집 표적화 범주의 개선을 입증한다. 백색 분획은 BE3, 또는 본 작업에서 개발된 5개의 변형된-PAM 염기 편집제와 함께의 BE3 중 어느 하나의 PAM 요건에 기초하여 접근가능한 병원성 T→C 돌연변이의 비율을 나타낸다. 도 94b는 좁혀진 활성 윈도우를 갖는 염기 편집제의 사용을 통한 클린바 데이터베이스에서의 모든 병원성 T→C 돌연변이 중에서 염기 편집 표적

화 범주의 개선을 나타낸다. BE3은 도 93a 내지 93c에 나타낸 것과 대등한 효율로 위치 4-8에서 C를 편집하는 것으로 가정하였다. YEE-BE3은 그의 활성 윈도우 내에서 C5>C6>C7>다른 것들 선호도로 편집하는 것으로 가정하였다. 백색 분획은 다른 C와 대등한 편집을 갖지 않는 편집된 BE3일 수 있는 (좌측), 또는 다른 C와 대등한 편집을 갖지 않는 편집된 BE3 또는 YEE-BE3일 수 있는 (우측) 병원성 T→C 돌연변이의 비율을 나타낸다.

도 95a 내지 95b는 염기 편집 윈도우 폭에 대한 말단절단된 가이드 RNA의 효과를 나타낸다. HEK293T 세포를 상이한 5' 말단절단 길이의 BE3 및 sgRNA를 발현하는 플라스미드로 형질감염시켰다. 처리된 세포를 실시예에 기재된 바와 같이 분석하였다. 도 95a는 EMX1 게놈 유전자좌 내의 부위에서의, 프로토스페이서 및 PAM 서열 (상부, 서열식별번호: 715) 및 지시된 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율을 나타낸다. 이 부위에서, 염기 편집 윈도우는 17-nt 말단절단된 gRNA의 사용을 통해 변경되었다. 도 95b는 HEK 부위 3 및 부위 4 게놈 유전자좌 내의 부위에서의, 프로토스페이서 및 PAM 서열 (상부, 서열식별번호: 715 및 716) 및 지시된 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율을 나타낸다. 이들 부위에서, 염기 편집 윈도우의 변화는 관찰되지 않았지만, sgRNA가 말단절단되는 경우 모든 기질 염기에 대한 편집 효율의 선형 감소가 주목되었다.

도 96은 염기 편집 윈도우 폭에 대한 APOBEC1-Cas9 링커 길이의 효과를 나타낸다. HEK293T 세포를 XTEN, GGS, $(GGS)_3$ (서열식별번호: 610), $(GGS)_5$ (서열식별번호: 610), 또는 $(GGS)_7$ (서열식별번호: 610) 및 sgRNA의 rAPOBEC1-Cas9 링커를 갖는 염기 편집제를 발현하는 플라스미드로 형질감염시켰다. 처리된 세포를 실시예에 기재된 바와 같이 분석하였다. 지시된 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 상이한 링커를 갖는 다양한 염기 편집제에 대해 나타나 있다.

도 97a 내지 97c는 염기 편집 윈도우 폭에 대한 rAPOBEC 돌연변이의 효과를 나타낸다. 도 97c는 부위 A 또는 부위 B 중 어느 하나를 표적화하는 sgRNA 및 지시된 BE3 점 돌연변이체를 발현하는 플라스미드로 형질감염된 HEK293T 세포를 나타낸다. 처리된 세포를 실시예에 기재된 바와 같이 분석하였다. 프로토스페이서에서의 및 프로토스페이서의 3개의 염기 쌍 내의 모든 C가 제시되며, 세포의 C에서 T로의 전환 백분율이 나타나 있다. 그 내에서 편집 효율이 반수-최대 값을 초과하는 뉴클레오티드의 계산된 수로서 정의되는 '편집 윈도우 폭'이 모든 시험된 돌연변이체에 대해 제시된다.

도 98은 포유동물 세포에서의 염기 편집의 생성물 분포에 대한 APOBEC1 돌연변이의 효과를 나타낸다. HEK293T 세포를 BE3 또는 그의 돌연변이체 및 적절한 sgRNA를 발현하는 플라스미드로 형질감염시켰다. 처리된 세포를 실시예에 기재된 바와 같이 분석하였다. 지시된 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 나타나 있다 (좌측). C에서 T로의 전환을 함유하는 총 서열분석 판독물의 퍼센트는 우측에 나타나 있다. BE3 점 돌연변이체는 단지 1개의 표적 시티딘을 갖는 부위인 HEK 부위 4에서 염기 편집 효율에 유의하게 영향을 미치지 않는다.

도 99는 BE3 및 HF-BE3에서의 온-타겟 편집 혈장 전달의 비교를 나타낸다.

도 100은 BE3의 단백질 및 혈장 전달에서의 온-타겟 편집의 비교를 나타낸다.

도 101은 HF-BE3의 단백질 및 혈장 전달에서의 온-타겟 편집의 비교를 나타낸다.

도 102는 리포펙션 및 HF 돌연변이를 설치하는 것 둘 다가 오프-타겟 탈아미노화 사건을 감소시킨다는 것을 나타낸다. 다이아몬드는 오프 타겟이 검출되지 않았음을 지시하며, 특이성 비는 100으로 설정되었다.

도 103은 프로토스페이서 ($NNNNTC_2TC_4TC_6TC_8TC_{10}TC_{12}TC_{14}TC_{16}TC_{18}TC_{20}NGG$, 서열식별번호: 723)에서의 짝수 위치에 놓인 C를 사용한 합성 기질 상의 시험관내 C에서 T로의 편집을 나타낸다.

도 104는 프로토스페이서 ($NNNNTC_2TC_4TC_6TC_8TC_{10}TC_{12}TC_{14}TC_{16}TC_{18}TC_{20}NGG$, 서열식별번호: 723)에서의 홀수 위치에 놓인 C를 사용한 합성 기질 상의 시험관내 C에서 T로의 편집을 나타낸다.

도 105는 플라스미드 대 단백질 전달을 갖는 염기 편집의 특이성 비를 나타내는 2개의 그래프를 포함한다.

도 106a 내지 106b는 비-NGG PAM 부위에 대한 BE3 활성을 나타낸다. HEK293T 세포를 BE3 및 적절한 sgRNA를 발현하는 플라스미드로 형질감염시켰다. 처리된 세포를 실시예에 기재된 바와 같이 분석하였다. 도 106a는 부위에 대한 BE3 활성이 SaBE3 또는 SaKKH-BE3에 의해 효율적으로 표적화될 수 있다는 것을 나타낸다. BE3은 NAG PAM에 대해 낮지만 유의한 활성을 나타낸다. 이 도면은 서열식별번호: 728 및 729를 나타낸다. 도 106b는 BE3이 VQR-BE3 또는 VRER-BE3과는 대조적으로 NGA 또는 NGCG PAM을 갖는 부위에서 유의하게 감소된 편집을 갖는다

는 것을 나타낸다. 이 도면은 서열식별번호: 730 및 731을 나타낸다.

도 107a 내지 107b는 VQR-BE3 및 SaKKH-BE3에 대한 APOBEC1 돌연변이의 효과를 나타낸다. HEK293T 세포를 VQR-BE3, SaKKH-BE3 또는 그의 돌연변이체 및 적절한 sgRNA를 발현하는 플라스미드로 형질감염시켰다. 처리된 세포를 하기 실시예에 기재된 바와 같이 분석하였다. 지시된 표적 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 나타나 있다. 도 107a는 윈도우-조정 돌연변이가 VQR-BE3에 적용되어 NGA PAM에 의해 표적화가능한 부위에서 선택적 염기 편집을 가능하게 할 수 있다는 것을 나타낸다. 이 도면은 서열식별번호: 732 및 733을 나타낸다. 도 107b는 SaKKH-BE3에 적용되는 경우, 돌연변이가 표적 윈도우 내의 염기 선택성을 부여하지 않고 염기 편집 효율의 전반적 감소를 유발한다는 것을 나타낸다. 이 도면은 서열식별번호: 728 및 734를 나타낸다.

도 108은 뉴클레오티드 편집의 개략적 표현을 나타낸다. 하기 약어가 사용된다: (MMR) - 미스매치 복구, (BE3 니카제) - Cas9 니카제 도메인을 포함하는 염기 편집제 3을 지칭함, (UGI) - 우라실 글리코실라제 억제제, UDG - 우라실 DNA 글리코실라제, (APOBEC) - APOBEC 시티딘 데아미나제를 지칭함.

도 109는 예시적인 염기 편집 구축물의 개략적 표현을 나타낸다. 염기 편집 구축물의 구조적 배열은 BE3, BE4-pmCDA1, BE4-hAID, BE4-3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE4-XTEN, BE4-32aa, BE4-2xUGI 및 BE4에 대해 나타나 있다. 링커는 회색으로 나타나 있다 (XTEN, SGGS (서열식별번호: 606), $(GGS)_3$ (서열식별번호: 610), 및 32aa). 데아미나제가 제시된다 (rAPOBEC1, pmCDA1, hAID, 및 hAPOBEC3G). 우라실 DNA 글리코실라제 억제제 (UGI)가 제시된다. 단일-가닥 DNA 결합 단백질 (SSB)은 자주색으로 제시된다. Cas9 니카제인 dCas9 (A840H)는 적색으로 제시된다. 도 109에는 또한 하기 표적 서열이 나타나 있다: EMX1, FANCF, HEK2, HEK3, HEK4, 및 RNF2. 아미노산 서열은 위에서 아래로 서열식별번호: 127-132로 지시된다. PAM 서열은 마지막 3개의 뉴클레오티드이다. 표적 시토신 (C)은 넘버링되고 적색으로 지시된다.

도 110은 EMX1 서열, $GAGTC_5CGAGCAGAAGAAGAAGGG$ (서열식별번호: 127)의 표적화된 시토신 (C_5) 상에서 지시된 염기 편집 구축물 (BE3, pmCDA1 hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, 및 BE4-2xUGI)에 대한 염기 편집 결과를 보여준다. 돌연변이된 표적화된 시토신(C_5)의 총 백분율은 "C_5" 하에 각각의 염기 편집 구축물에 대해 지시된다. indel의 총 백분율은 "indel" 하에 각각의 염기 편집 구축물에 대해 지시된다. 아데닌 (A), 구아닌 (G) 또는 티민 (T)으로 돌연변이된 돌연변이된 시토신의 비율이 파이 차트에서 각각의 염기 편집 구축물에 대해 지시된다.

도 111은 FANCF 서열, $GGAATCCC_8TTCTGCAGCACCTGG$ (서열식별번호: 128)의 표적화된 시토신 (C_8) 상에 표시된 염기 편집 구축물 (BE3, pmCDA1 hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, 및 BE4-2xUGI)에 대한 염기 편집 결과를 보여준다. 돌연변이된 표적화된 시토신 (C_8)의 총 백분율은 "C_8" 하에 각각의 염기 편집 구축물에 대해 지시된다. indel의 총 백분율은 "indel" 하에 각각의 염기 편집 구축물에 대해 지시된다. 아데닌(A), 구아닌(G) 또는 티민(T)으로 돌연변이된 돌연변이된 시토신의 비율이 파이 차트에서 각각의 염기 편집 구축물에 대해 지시된다.

도 112는 HEK2 서열 $GAACAC_6AAAGCATAGACTGCGGG$ (서열식별번호: 129)의 표적화된 시토신 (C_6) 상에서 지시된 염기 편집 구축물 (BE3, pmCDA1 hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, 및 BE4-2xUGI)에 대한 염기 편집 결과를 보여준다. 돌연변이된 표적화된 시토신 (C_6)의 총 백분율은 "C_6" 하에 각각의 염기 편집 구축물에 대해 지시된다. indel의 총 백분율은 "indel" 하에 각각의 염기 편집 구축물에 대해 지시된다. 아데닌(A), 구아닌(G) 또는 티민(T)으로 돌연변이된 돌연변이된 시토신의 비율이 파이 차트에서 각각의 염기 편집 구축물에 대해 지시된다.

도 113은 HEK3 서열 $GGCCC_5AGACTGAGCACGTGATGG$ (서열식별번호: 130)의 표적화된 시토신 (C_5) 상에서 지시된 염기 편집 구축물 (BE3, pmCDA1 hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, 및 BE4-2xUGI)에 대한 염기 편집 결과를 보여준다. 돌연변이된 표적화된 시토신 (C_5)의 총 백분율은 "C_5" 하에 각각의 염기 편집 구축물에 대해 지시된다. indel의 총 백분율은 "indel" 하에 각각의 염기 편집 구축물에 대해 지시된다. 아데닌(A), 구아닌(G) 또는 티민(T)으로 돌연변이된 돌연변이된 시토신의 비율을 파이 차트에서 각각의 염기 편집 구축물에 대해 지시된다.

도 114는 HEK4 서열 GGCAC$_5$TGCGGCTGGAGGTCCGGG (서열식별번호: 131)의 표적화된 시토신 (C_5) 상에서 지시된 염기 편집 구축물 (BE3, pmCDA1 hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, 및 BE4-2xUGI)에 대한 염기 편집 결과를 보여준다. 돌연변이된 표적화된 시토신 (C_5)의 총 백분율은 "C_5" 하에 각각의 염기 편집 구축물에 대해 지시된다. indel의 총 백분율은 "indel" 하에 각각의 염기 편집 구축물에 대해 지시된다. 아데닌(A), 구아닌(G) 또는 티민(T)으로 돌연변이된 돌연변이된 시토신의 비율을 파이 차트에서 각각의 염기 편집 구축물에 대해 지시된다.

도 115는 RNF2 서열 GTCATC$_6$TTAGTCATTACCTGAGG (서열식별번호: 132)의 표적화된 시토신 (C_6) 상에서 지시된 염기 편집 구축물 (BE3, pmCDA1 hAID, hAPOBEC3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE-XTEN, BE4-32aa, 및 BE4-2xUGI)에 대한 염기 편집 결과를 보여준다. 돌연변이된 표적화된 시토신 (C_6)의 총 백분율은 "C_6" 하에 각각의 염기 편집 구축물에 대해 지시된다. indel의 총 백분율은 "indel" 하에 각각의 염기 편집 구축물에 대해 지시된다. 아데닌(A), 구아닌(G) 또는 티민(T)으로 돌연변이된 돌연변이된 시토신의 비율을 파이 차트에서 각각의 염기 편집 구축물에 대해 지시된다.

도 116은 표적 가닥을 니킹하는 Cpf1 돌연변이체에 대해 시험하기 위해 사용된 예시적인 형광 표지된 (Cy3 표지된) DNA 구축물을 보여준다. 1의 DNA 구축물에서, 비-표적 가닥 (상부 가닥) 및 표적 가닥 (하부 가닥)은 둘 다 형광 표지된다. 2의 DNA 구축물에서, 비-표적 가닥 (상부 가닥)은 형광 표지되고, 표적 가닥 (하부 가닥)은 형광 표지되지 않는다. 3의 DNA 구축물에서, 비-표적 가닥 (상부 가닥)은 형광 표지되지 않고, 표적 가닥 (하부 가닥)은 형광 표지된다.

도 117은 30분 (30 min) 또는 2시간 초과 (2h+)의 반응 시간에 걸쳐 도 116에 제시된 DNA 구축물의 표적 및 비-표적 가닥을 절단하는 다양한 Cpf1 구축물 (예를 들어, R836A, R1138A, 야생형)의 능력을 입증하는 데이터를 보여준다.

도 118은 아키텍처를 갖는 염기 편집제, APOBEC-AsCpf1(R912A)-UGI가 표적 서열의 제7에서 제11 염기의 윈도우를 갖는 C 잔기 (예를 들어, 표적 서열 FANCF1, FANCF2, HEK3-3 및 HEK3-4의 것)를 편집할 수 있다는 것을 입증하는 데이터를 보여준다. BG는 배경 돌연변이 수준 (비처리)을 나타낸다. AsCpf1은 AsCpf1 단독 처리군 (대조군)을 나타내고, APOBEC-AsCpf1 (R912A)-UGI는 표적 가닥을 우선적으로 절단하는 Cpf1을 함유하는 염기 편집제를 나타내고, APOBEC-AsCpf1 (R1225A)-UGI는 비-표적 가닥을 절단하는 Cpf1을 함유하는 자기-무력화 염기 편집제를 나타낸다. FANCF1, FANCF2, HEK3-3, 및 HEK3-4의 표적 서열은 하기와 같다:

FANCF1 GCGGATGTTCCAATCAGTACGCA (SEQ ID NO: 724)

FANCF2 CGAGCTTCTGGCGGTCTCAAGCA (SEQ ID NO: 725)

HEK3-3 TGCTTCTCCAGCCCTGGCCTGG (SEQ ID NO: 726)

HEK3-4 AGACTGAGCACGTGATGGCAGAG (SEQ ID NO: 727)

도 119는 Cpf1 단백질 (예를 들어, AsCpf1 또는 LbCpf1)을 포함하는 염기 편집제의 개략적 표현을 나타낸다. 상이한 링커 서열 (예를 들어, XTEN, GGS, $(GGS)_3$ (서열식별번호: 610), $(GGS)_5$ (서열식별번호: 610), 및 $(GGS)_7$ (서열식별번호: 610))을 "링커"로 표지된 부분에 대해 시험하였고, 그 결과는 도 120에 제시된다.

도 120은 HEK3 부위 TGCTTCTC$_8$CAGCCCTGGCCTGG (서열식별번호: 592)의 C_8 잔기를 편집하는 도 119에 제시된 구축물의 능력을 입증하는 데이터를 보여준다. APOBEC 도메인을 Cpf1 도메인 (예를 들어, LbCpf1(R836A) 또는 AsCpf1(R912A))에 연결하는 상이한 링커 서열을 시험하였다. 시험된 예시적인 링커는 XTEN, GGS, $(GGS)_3$ (서열식별번호: 610), $(GGS)_5$ (서열식별번호: 610), 및 $(GGS)_7$ (서열식별번호: 610)을 포함한다.

도 121은, HEK3의 잔기 TGCTTCTC$_8$C$_9$AGCCCTGGCCTGG (서열식별번호: 592)의 C_8 및 C_9 편집을 위한, LbCpf1 도메인을 갖는 도 119에 제시된 구축물의 능력을 입증하는 데이터를 보여준다. APOBEC 도메인을 LbCpf1 도메인에 연결하는 통합 생물정보학 VU의 센터에 의해 유지된 데이터베이스로부터의 상이한 링커 서열을 시험하였다. 시험된 예시적인 링커는 1au7, 1c1k, 1c20, 1ee8, 1flz, 1ign, 1jmc, 1sfe, 2ezx 및 2reb를 포함한다.

도 122는 N 및 C 말단이 표시된 AsCpf1의 구조의 개략적 표현을 나타낸다.

도 123은 N 및 C 말단이 표시된 SpCas9의 구조의 개략적 표현을 나타낸다.

도 124는 AsCpf1의 개략적 표현을 나타내고, 여기서 적색 원은 편집 윈도우가 있다고 예측되는 면적을 나타낸다. 정사각형은 APOBEC 활성을 방해할 수 있는 나선 영역을 나타낸다.

도 125a 및 125b는 고충실도 염기 편집제 (HF-BE3)의 조작 및 시험관내 특징화를 보여준다. 도 125a는 HF-BE3의 개략적 표현을 보여준다. BE3에 도입되어 HF-BE3을 생성하는 점 돌연변이가 제시된다. 상기 표현은 PDB 구조 4UN3 (Cas9), 4ROV (시티딘 데아미나제) 및 1UGI (우라실 DNA 글리코실라제 억제제)를 사용하였다. 도 125b는 'TC' 반복 프로토스페이서를 함유하는 합성 기질의 시험관내 탈아미노화를 보여준다. 값 및 오차 막대는 상이한 날에 수행된 2개의 독립적인 반복의 평균 및 범위를 반영한다.

도 126a 내지 126c는 염기 편집제 단백질의 정제를 보여준다. 도 126a는 염기 편집제 발현을 위한 최적 이. 콜라이 균주의 선택을 보여준다. 18℃에서 16시간 동안 IPTG-유도된 단백질 발현 후에, 조 세포 용해물을 단백질 함량에 대해 분석하였다. BL21 스타 (DE3) (써모 피셔(Thermo Fisher)) 세포는 BE3 및 HF-BE3 둘 다의 가장 유망한 발현후 수준을 보여주었고, 염기 편집제의 발현에 사용되었다. 도 126b는 발현된 염기 편집제 단백질의 정제를 보여준다. 염기 편집제의 C-말단 상에 His6 태그를 위치시키는 것은 BE3 및 HF-BE3 둘 다에 대한 말단 절단 생성물의 생산을 유도한다 (레인 1 및 2). 예상외로, 이러한 말단절단 생성물은 His6 태그를 단백질의 N-말단에 위치시킴으로써 제거하였다 (레인 3-6). Cas9 발현에 최적인 것 (OD600 = 0.4)보다 늦게 OD600 = 0.7 (레인 4-5)의 세포 밀도로 염기 편집제의 발현을 유도하는 것은, 염기 편집 단백질의 수율을 개선시킨다. 수동 HisPur 수지 칼럼을 사용하여 정제를 수행한 후, 양이온 교환 FPLC (Akta)를 수행하였다. 도 126c는 정제된 BE3 및 HF-BE3을 보여준다. 상이한 농도의 정제된 BE3 및 HF-BE3을 열 및 LDS를 사용하여 변성시키고, 폴리아크릴아미드 겔 상에 로딩하였다. 단백질 샘플은 본 연구에 사용된 단백질을 대표한다. 도 126a 내지 126c의 겔은 BOLT 비스-트리스 플러스 4-12% 폴리아크릴아미드 (써모 피셔)이다. 전기영동 및 염색을 방법에 기재된 바와 같이 수행하였다.

도 127a 내지 127d는 인간 세포에서의 고충실도 염기 편집제 (HF-BE3)의 활성을 보여준다. 도 127a 내지 127c는 비-반복적 게놈 유전자좌 EMX1 (도 127a), FANCF (도 127b), 및 HEK293 부위 3 (도 127c)을 표적화하는 sgRNA로 처리된 HEK293T 세포로부터의 게놈 DNA의 고처리량 서열분석을 사용하여 BE3 및 HF-BE3의 플라스미드 형질감염과 연관된 온- 및 오프-타겟 편집을 검정하였다는 것을 보여준다. 각각의 sgRNA와 연관된 온- 및 오프-타겟 유전자좌는 수직선에 의해 분리된다. 도 127d는 VEGFA 부위 2를 표적화하는 고도로 반복되는 sgRNA와 연관된 온- 및 오프-타겟 편집을 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다. 도 127a 내지 127c의 경우, 별은 처리된 샘플과 비처리된 대조군 사이의 비교에 기초한 유의한 편집을 나타낸다. * $p \leq 0.05$, ** $p \leq 0.01$ 및 *** $p \leq 0.001$ (스튜던트 양측 t-검정). 도 127d의 경우, 모든 처리된 샘플이 대조군에 비해 유의한 편집을 나타내었기 때문에 별표는 제시되지 않았다. 개별 p-값이 표 16에 열거되어 있다.

도 128a 내지 128c는 배양된 인간 세포에서 온-타겟 및 오프-타겟 염기 편집의 효율에 대한 BE3 단백질 또는 플라스미드의 투여량의 효과를 보여준다. 도 128a는 각각의 sgRNA에 대한 활성 윈도우에서의 모든 편집된 시토신에 걸쳐 각각의 4개의 게놈 유전자좌에서의 온-타겟 편집 효율을 평균내었음을 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3개의 독립적인 생물학적 반복의 평균 ± S.E.M을 반영한다. 도 128b 및 128c는 HEK293T 세포에서의 BE3 플라스미드 적정 (도 128b) 또는 BE3 단백질 적정 (도 128c)으로부터 발생한 EMX1 부위에서의 온- 및 오프-타겟 편집을 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 129a 내지 129b는 BE3 및 HF-BE3의 플라스미드 및 단백질 전달에 대한 온-표적:오프-타겟 염기 편집 빈도 비를 보여준다. 활성 윈도우 내의 특정한 시토신에서의 온-타겟 편집 백분율을 표시된 오프-타겟 유전자좌에 대한 상응하는 시토신에서의 오프-타겟 편집 백분율로 나눔으로써 염기 편집 온-타겟:오프-타겟 특이성 비를 계산하였다 (방법 참조). 오프-타겟 편집이 검출 한계값 미만인 경우 (서열분석 판독물의 0.025%), 본 발명자들은 오프-타겟 편집을 검출 한계 (0.025%)로 설정하고, 이 상한으로 온-타겟 편집 백분율을 나누었다. ◆로 표시되는 이들 경우에, 제시된 특이성 비는 하한을 나타낸다. 특이성 비가 비-반복적 sgRNA FANCF, HEK293 부위 3, 및 FANCF (도 129a) 및 고도로 반복적인 sgRNA VEGFA 부위 2 (도 129b)에 대해 제시된다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 130a 내지 130d는 배양된 인간 세포로의 염기 편집제의 단백질 전달을 보여준다. 도 130a 내지 130d는 EMX1 (도 130a), FANCF (도 130b), HEK293 부위 3 (도 130c) 및 VEGFA 부위 2 (도 130d)를 표적화하는 sgRNA와 복합

체화된 염기 편집제의 RNP 전달과 연관된 온- 및 오프-타겟 편집을 보여준다. 오프-타겟 염기 편집은 비-반복적 sgRNA에 대한 모든 서열분석된 유전자좌에서 검출불가능하였다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다. 별은 처리된 샘플과 비처리 대조군 사이의 비교에 기초한 유의한 편집을 나타낸다. * $p \leq 0.05$, ** $p \leq 0.01$ 및 *** $p \leq 0.001$ (스튜던트 양측 t-검정).

도 131a 내지 131c는 게놈 유전자좌에서의 염기 편집과 연관된 indel 형성을 나타낸다. 도 131a는 VEGFA 부위 2, EMX1, FANCF, 및 HEK293 부위 3 sgRNA에 대한 온-타겟 유전자좌에서의 indel 빈도를 보여준다. 도 131b는 염기 편집:indel 형성의 비를 보여준다. 다이아몬드 (◆)는 어떤 indel도 검출되지 않았음을 나타낸다 (처리된 샘플 및 비처리된 대조군에서 indel 빈도의 유의한 차이가 없음). 도 131c는 도 131a에서 조사된 온-타겟 부위와 연관된 오프-타겟 유전자좌에서 관찰된 indel을 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 132a 내지 132d는 BE3의 RNP 전달을 사용하여 제브라피시 배아에서의 및 살아있는 마우스의 내이에서의 DNA-무함유 생체내 염기 편집을 보여준다. 도 132a는 표시된 sgRNA와 복합체화된 BE3의 주사 4일 후 수거된 제브라피쉬에서의 온-타겟 게놈 편집을 보여준다. 값 및 오차 막대는 평균 ± s.d.의 3개의 주사된 및 3개의 대조군 제브라피쉬를 반영한다. 대조군은 비관련된 sgRNA와 복합체화된 BE3을 주사하였다. 도 132b는 양이온성 지질 나노입자 내로 캡슐화된 BE3:sgRNA 복합체의 생체내 주사를 보여주는 개략도를 보여준다. 도 132c는 VEGFA 부위 2 게놈 유전자좌에서의 염기 편집제 윈도우 내의 시토신 잔기의 염기 편집을 보여준다. 도 132d는 VEGFA 부위 2 표적 유전자좌의 염기 편집 윈도우 내의 각각의 시토신에서의 온-타겟 편집을 보여준다. 도 132d (도 132c 및 132d)는 값 및 오차 막대가 VEGFA 부위 2를 표적화하는 sgRNA로 주사된 3마리의 마우스, 3마리의 주사되지 않은 마우스 및 비관련 sgRNA로 주사된 1마리의 마우스의 평균 ± S.E.M.을 반영한다.

도 133a 내지 133e는 뮤린 NIH/3T3 세포에서의 온- 및 오프-타겟 염기 편집을 보여준다. 도 133a는 'VEGFA 부위 2' sgRNA와 연관된 온-타겟 염기 편집을 보여준다 (서열에 대해서는 도 132e 참조). 음성 대조군은 BE3을 코딩하지만 sgRNA를 코딩하지는 않는 플라스미드로 처리된 세포에 상응한다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다. 도 133b 내지 133e는 이러한 부위와 연관된 오프-타겟 편집이 이러한 sgRNA (도 132e에 제시된 서열)에 대한 상위 4개의 예측된 오프-타겟 유전자좌에서의 고처리량 DNA 서열분석을 사용하여 측정되었음을 보여준다. 도 133b는 오프-타겟 2를 보여주고, 도 133c는 오프-타겟 1을 보여주고, 도 133d는 오프-타겟 3을 보여주고, 도 133e는 오프-타겟 4를 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 134a 내지 134b는 생체내-편집된 뮤린 조직으로부터의 오프-타겟 염기 편집 및 온-타겟 indel 분석을 보여준다. 도 134a는 VEGFA 부위 2와 연관된 오프-타겟 유전자좌의 염기 편집 윈도우 내의 각각의 시토신에 대해 플롯팅된 편집을 보여준다. 도 134b는 온-타겟 염기 편집제 유전자좌에서의 indel 비율을 보여준다. 값 및 오차 막대는 3개의 주사된 및 3개의 대조군 마우스의 평균 ± S.E.M을 반영한다.

도 135a 내지 135c는 UNG 녹아웃의 염기 편집 생성물 순도에 대한 효과를 보여준다. 도 135a는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3으로 처리된 HAP1 (UNG^+) 및 HAP1 UNG^- 세포를 보여준다. 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포가 제시된다. 도 135b는 시험된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 135a에서 분석된 표적 C는 적색으로 제시된다. 도 135c는 HAP1 세포 또는 HAP1 UNG^- 세포에서 BE3으로 처리한 후의 indel 형성의 빈도를 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 136a 내지 136d는 생성물 순도에 대한 다중-C 염기 편집의 효과를 보여준다. 도 136a는 비처리된, BE3-처리된, 및 AID-BE3-처리된 인간 HEK293T 세포의 대표적인 고처리량 서열분석 데이터를 보여준다. 프로토스페이서의 서열은 상부에 제시되고, PAM 및 표적 C는 적색으로 제시되고, 아래첨자처리된 숫자는 프로토스페이서 내에서의 그의 위치를 나타낸다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 비-T 염기가 아니라 T로 명확하게 편집된 표적 C의 상대적인 백분율은 1개의 C만을 편집한 BE3-처리된 세포에서보다, 이 유전자에서 3개의 C를 편집한 AID-BE3-처리된 세포에서 훨씬 더 높다. 도 136b는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3, CDA1-BE3, 및 AID-BE3으로 처리된 HEK293T 세포를 보여준다. 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포가 제시된다. 도 136c는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 136b에서 분석된 표적 C는 적색으로 제시된다. 도 136d는 도 136a에 제시된 처리 후의 indel 형성의 빈도를 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3

회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 137a 내지 137c는 BE3의 아키텍처 변화의 C→T 편집 효율 및 생성물 순도에 대한 효과를 보여준다. 도 137a는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 137c에서의 표적 C는 자주색 및 적색으로 제시되고, 도 137b에서의 표적 C는 적색으로 제시된다. 도 137b는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3, SSB-BE3, N-UGI-BE3, 및 BE3-2xUGI로 처리된 HEK293T 세포를 보여준다. 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포가 BE3, N-UGI-BE3 및 BE3-2xUGI에 대해 제시된다. 도 137c는 C→T 염기 편집 효율을 나타낸다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 138a 내지 138d는 C→T 편집 효율 및 생성물 순도에 대한 BE3에서의 링커 길이 변동의 효과를 보여준다. 도 138a는 BE3, BE3C, BE3D, 및 BE3E의 아키텍처를 보여주고, 도 138b는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 138c에서의 표적 C는 자주색 및 적석으로 제시되고, 도 138d에서의 표적 C는 적색으로 제시된다. 도 138c는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3, BE3C, BE3D, 또는 BE3E로 처리된 HEK293T 세포를 보여준다. C→T 염기 편집 효율이 제시된다. 도 138d는 BE3, BE3C, BE3D 및 BE3E에 대한 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포를 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 139a 내지 139d는 BE4가 BE3과 비교하여 염기 편집 효율 및 생성물 순도를 증가시킨다는 것을 보여준다. 도 139a는 BE3, BE4, 및 표적-AID의 아키텍처를 나타낸다. 도 139b는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 139c에서의 표적 C는 자주색 및 적색으로 제시되고, 도 139d에서의 표적 C는 적색으로 제시된다. 도 139c는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3, BE4, 또는 표적-AID로 처리된 HEK293T 세포를 보여준다. C→T 염기 편집 효율이 제시된다. 도 139d는 BE3 및 BE4에 대한 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의의 생성물 분포를 보여준다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 140a 내지 140c는 CDA1-BE3 및 AID-BE3이 BE3에 비해 더 효율적으로 표적 G에 이은 C를 편집한다는 것을 보여준다. 도 140a는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, BE3, CDA1-BE3, 및 AID-BE3에 의해 편집된 표적 C는 적색으로 제시되고, 오직 CDA1-BE3 및 AID-BE3에 의해 편집된 표적 C (G에 이어짐)는 자주색으로 제시된다. 도 140b는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3, CDA1-BE3, AID-BE3, 또는 APOBEC3G-BE3으로 처리된 HEK293T 세포를 보여준다. C→T 염기 편집 효율이 제시된다. 도 140c는 HEK2 유전자좌를 표적화하는 BE3, CDA1-BE3, 또는 AID-BE3으로 처리되고 프로토스페이서의 서열에 따라 비닝되고 분석된 HEK293T 세포로부터의 개별 DNA 서열분석 판독물을 보여주고, 이는 CDA1-BE3 및 AID-BE3에 의한 명확한 C에서 T로의 편집을 갖는 서열분석 판독물의 > 85%가 둘 다 T로 편집된 C를 갖는다는 것을 밝혀냈다 (도 140c).

도 141a 내지 141c는 다수의 편집가능한 C를 갖는 부위에서의 불균일한 편집이 보다 낮은 생성물 순도를 발생시킨다는 것을 보여준다. 도 141a는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 141c에서의 표적 C는 자주색 및 적색으로 제시되고, 도 141b에서의 표적 C는 적색으로 제시된다. 도 141b 및 141c는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3으로 처리된 HEK293T 세포를 보여준다. 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포가 제시된다. C에서 비-T로의 편집은 편집 효율이 동일한 유전자좌 내의 2개의 C에 대해 동일하지 않은 경우에 더 빈번하다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 142a 내지 142d는 다수의 C의 염기 편집이 보다 높은 염기 편집 생성물 순도를 발생시킨다는 것을 보여준다. 도 142a는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 142b에서 조사된 표적 C는 적색으로 제시된다. 도 142b는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3 또는 BE3B (UGI가 결여됨)로 처리된 HEK293T 세포를 보여준다. 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포가 제시된다. 도 142c는 HEK2 유전자좌를 표적화하는 BE3 또는 BE3B (UGI가 결여됨)로 처리된 HEK293T 세포로부터의 HTS 판독물이 위치 6의 1차 표적 C의 정체에 따라 비닝된 것을 나타낸다. 이어서, 생성된 판독물을 위치 4에서 2차 표적 C에서의 염기의 정체에 대해 분석하였다. 상기 판독물에서 단일 편집 사건만이 있는 경우, C_6은 비-T로 부정확하게 편집될 가능성이 더 크다. 도 142d는 HEK4 유전자좌 (오직 단일 편집가능한 C를 갖는 부위)를 표적화하는 BE3 또는 BE3B로 처리된 세포에서, C_5에서 A, G, 및 T를 갖는 편집된 판독물의 분포를 보여주고, 이는 단일 G:U 미스매치가 UNG-개시된 염기 절제 복구를 통해 프로세싱되어 생성물의 혼합물을 제공

한다는 것을 예시한다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 143은 다수의 C의 염기 편집이 HEK3 및 RNF2 유전자좌에서 보다 높은 염기 편집 생성물 순도를 발생시키는 것을 보여준다. HEK3 및 RNF2 유전자좌를 표적화하는 BE3 또는 BE3B (UGI 없음)로 처리된 HEK293T 세포로부터의 DNA 서열분석 판독물을 1차 표적 C 위치 (적색)에서의 염기의 정체에 따라 분리하였다. 이어서, 서열분석 판독물의 4개 군을 2차 표적 C 위치 (자주색)에서의 염기의 정체에 대해 조사하였다. BE3의 경우, 1차 표적 C (적색)가 부정확하게 G로 편집될 경우, 2차 표적 C는 C로 유지될 가능성이 더 크다. 반대로, 1차 표적 C (적색)가 T로 전환되는 경우, 2차 표적 C는 또한 동일한 서열분석 판독물에서 T로 편집될 가능성이 더 크다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 144a 내지 144c는 BE4가 BE3보다 더 낮은 indel 빈도를 유도하고, 표적-AID는 CDA1-BE3과 유사한 생성물 순도를 나타낸다. 도 144a는 실시예 17의 물질 및 방법에 기재된 바와 같은 BE3, BE4, 또는 표적-AID로 처리된 HEK293T 세포를 보여준다. indel 형성의 빈도 (실시예 17의 물질 및 방법 참조)가 제시된다. 도 144b는 실시예 17의 물질 및 방법에 기재된 바와 같은 CDA1-BE3 또는 표적-AID로 처리된 HEK293T 세포를 보여준다. 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포가 제시된다. 도 144c는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 144b에서 조사된 표적 C는 적색으로 제시된다. 값 및 오차 막대는 상이한 날에 수행된 3개의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 145a 내지 145c는 SaBE4가 SaBE3과 비교하여 증가된 염기 편집 수율 및 생성물 순도를 나타낸다는 것을 보여준다. 도 145a는 실시예 17의 물질 및 방법에 기재된 바와 같은 SaBE3 및 SaBE4로 처리된 HEK293T 세포를 보여준다. 나타난 표적 위치에서 T를 갖는 총 DNA 서열분석 판독물의 백분율이 제시된다. 도 145b는 연구된 게놈 유전자좌의 프로토스페이서 및 PAM 서열을 보여주고, 도 145a에서의 표적 C는 자주색 및 적색으로 제시되고, 도 145c에서 연구된 표적 C는 적색으로 제시된다. 도 145c는 편집된 DNA 서열분석 판독물 (표적 C가 돌연변이된 판독물) 중에서의 생성물 분포를 나타낸다. 값 및 오차 막대는 상이한 날에 수행된 3회의 독립적인 생물학적 반복의 평균 ± S.D.를 반영한다.

도 146은 EMX1 유전자좌에서의 BE3, CDA1-BE3, AID-BE3, 또는 APOBEC3G-BE3으로의 처리로부터의 염기 편집 결과를 보여준다. 프로토스페이서의 서열은 상단에 제시되고, PAM 및 표적 염기는 적색으로 표시되고, 아래첨자 수는 프로토스페이서 내에서의 그의 위치를 나타낸다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 실시예 17의 물질 및 방법에 기재된 바와 같이 처리하였다. 제시된 값은 하나의 대표적인 실험으로부터의 것이다.

도 147은 FANCF 유전자좌에서의 BE3, CDA1-BE3, AID-BE3, 또는 APOBEC3G-BE3으로의 처리로부터의 염기 편집 결과를 보여준다. 프로토스페이서의 서열은 상단에 제시되고, PAM 및 표적 염기는 적색으로 표시되고, 아래첨자 표시된 숫자는 프로토스페이서 내에서의 그의 위치를 나타낸다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 실시예 17의 물질 및 방법에 기재된 바와 같이 처리하였다. 제시된 값은 하나의 대표적인 실험으로부터의 것이다.

도 148은 HEK2 유전자좌에서의 BE3, CDA1-BE3, AID-BE3, 또는 APOBEC3G-BE3으로의 처리로부터의 염기 편집 결과를 보여준다. 프로토스페이서의 서열은 상단에 제시되고, PAM 및 표적 염기는 적색으로 표시되고, 아래첨자 표시된 숫자는 프로토스페이서 내에서의 그의 위치를 나타낸다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 실시예 17의 물질 및 방법에 기재된 바와 같이 처리하였다. 제시된 값은 하나의 대표적인 실험으로부터의 것이다.

도 149는 HEK3 유전자좌에서의 BE3, CDA1-BE3, AID-BE3, 또는 APOBEC3G-BE3으로의 처리로부터의 염기 편집 결과를 보여준다. 프로토스페이서의 서열은 상단에 제시되고, PAM 및 표적 염기는 적색으로 표시되고, 아래첨자 표시된 숫자는 프로토스페이서 내에서의 그의 위치를 나타낸다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 실시예 17의 물질 및 방법에 기재된 바와 같이 처리하였다. 제시된 값은 하나의 대표적인 실험으로부터의 것이다.

도 150은 HEK4 유전자좌에서의 BE3, CDA1-BE3, AID-BE3, 또는 APOBEC3G-BE3으로의 처리로부터의 염기 편집 결과를 보여준다. 프로토스페이서의 서열은 상단에 제시되고, PAM 및 표적 염기는 적색으로 표시되고, 아래첨자 표시된 숫자는 프로토스페이서 내에서의 그의 위치를 나타낸다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 실시예 17의 물질 및 방법에 기재된 바와 같이 처리하였다. 제

시된 값은 하나의 대표적인 실험으로부터의 것이다.

도 151은 RNF2 유전자좌에서의 BE3, CDA1-BE3, AID-BE3, 또는 APOBEC3G-BE3으로의 처리로부터의 염기 편집 결과를 보여준다. 프로토스페이서의 서열은 상단에 제시되고, PAM 및 표적 염기는 적색으로 표시되고, 아래첨자 표시된 숫자는 프로토스페이서 내에서의 그의 위치를 나타낸다. 각각의 서열 아래에는 상응하는 염기를 갖는 총 서열분석 판독물의 백분율이 있다. 세포를 실시예 17의 물질 및 방법에 기재된 바와 같이 처리하였다. 제시된 값은 하나의 대표적인 실험으로부터의 것이다.

도 152는 LBCpf1 융합 구축물의 개략도를 보여준다. 구축물 10은 [Apobec]-[LbCpf1]-[UGI]-[UGI]의 도메인 배열을 갖고; 구축물 11은 [Apobec]-[LbCpf1]-[UGI]의 도메인 배열을 갖고; 구축물 12는 [UGI]-[Apobec]-[LbCpf1]의 도메인 배열을 갖고; 구축물 13은 [Apobec]-[UGI]-[LbCpf1]의 도메인 배열을 갖고; 구축물 14는 [LbCpf1]-[UGI]-[Apobec]의 도메인 배열을 갖고; 구축물 15는 [LbCpf1]-[Apobec]-[UGI]의 도메인 배열을 갖는다. 각각의 구축물에 대해, 3개의 상이한 LbCpf1 단백질 (D/N/A, 이는 뉴클레아제 데드 LbCpf1 (D); LbCpf1 니카제 (N) 및 뉴클레아제 활성 LbCpf1 (A)을 지칭함)을 사용하였다.

도 153은 19개 뉴클레오티드 길이의 가이드, 즉 EMX19: TACTTTGTCCTCCGGTTCT (서열식별번호: 744)를 사용한 EMX 표적 TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC12C$_{13}$GGTTC$_{18}$TG (서열식별번호: 738) 내 6개 C 잔기의 C에서 T로의 편집의 백분율을 보여준다. 편집을 도 152에 제시된 몇몇 구축물에 대해 시험하였다.

도 154는 18개 뉴클레오티드 길이의 가이드, 즉 EMX18: TACTTTGTCCTCCGGTTC (서열식별번호: 745)를 사용한 EMX 표적 TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC12C$_{13}$GGTTC$_{18}$TG (서열식별번호: 738) 내 6개 C 잔기의 C에서 T로의 편집의 백분율을 보여준다. 편집을 도 152에 제시된 몇몇 구축물에 대해 시험하였다.

도 155는 17개 뉴클레오티드 길이의 가이드, 즉 EMX17: TACTTTGTCCTCCGGTT (서열식별번호: 746)를 사용한 EMX 표적 TTTGTAC$_3$TTTGTC$_9$C$_{10}$TC12C$_{13}$GGTTC$_{18}$TG (서열식별번호: 738) 내 6개 C 잔기의 C에서 T로의 편집의 백분율을 보여준다. 편집을 도 152에 제시된 몇몇 구축물에 대해 시험하였다.

도 156은 23개 뉴클레오티드의 길이의 가이드, 즉, Hek2_23: CAGCCCGCTGGCCCTGTAAAGGA (서열식별번호: 747)를 사용한 HEK2 표적 TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA (서열식별번호: 739) 내 8개 C 잔기의 C에서 T로의 편집의 백분율을 보여준다. 편집을 도 152에 제시된 몇몇 구축물에 대해 시험하였다.

도 157은 20개 뉴클레오티드 길이의 가이드, 즉, Hek2_20: CAGCCCGCTGGCCCTGTAAA (서열식별번호: 748)를 사용한 HEK2 표적 TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA (서열식별번호: 739) 내 8개 C 잔기의 C에서 T로의 편집의 백분율을 보여준다. 편집을 도 152에 제시된 몇몇 구축물에 대해 시험하였다.

도 158은 19개 뉴클레오티드 길이의 가이드, 즉, Hek2_19: CAGCCCGCTGGCCCTGTAA (서열식별번호: 749)를 사용한 HEK2 표적 TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA (서열식별번호: 739) 내 8개 C 잔기의 C에서 T로의 편집의 백분율을 보여준다. 편집을 도 152에 제시된 몇몇 구축물에 대해 시험하였다.

도 159는 18개 뉴클레오티드의 길이의 가이드, 즉, Hek2_18: CAGCCCGCTGGCCCTGTA (서열식별번호: 750)를 사용한 HEK2 표적 TTTCC$_1$AGC$_4$C$_5$C$_6$GC$_8$TGGC$_{12}$C$_{13}$C$_{14}$TGTAAA (서열식별번호: 739) 내 8개 C 잔기의 C에서 T로의 편집의 백분율을 보여준다. 편집을 도 152에 제시된 몇몇 구축물에 대해 시험하였다.

도 160은 (indel 카운트에 기초한 조정 후의) 편집 백분율 값, 및 도 153에 도시된 실험에 대한 indel의 백분율을 나타낸다.

도 161은 (indel 카운트에 기초한 조정 후의) 편집 백분율 값, 및 도 154에 도시된 실험에 대한 indel의 백분율을 나타낸다.

도 162는 (indel 카운트에 기초한 조정 후의) 편집 백분율 값, 및 도 155에 도시된 실험에 대한 indel의 백분율을 나타낸다.

도 163은 (indel 카운트에 기초한 조정 후의) 편집 백분율 값, 및 도 156에 도시된 실험에 대한 indel의 백분율을 나타낸다.

도 164는 (indel 카운트에 기초한 조정 후의) 편집 백분율 값, 및 도 157에 도시된 실험에 대한 indel의 백분율을 나타낸다.

도 165는 (indel 카운트에 기초한 조정 후의) 편집 백분율 값, 및 도 158에 도시된 실험에 대한 indel의 백분율을 나타낸다.

도 166은 (indel 카운트에 기초한 조정 후의) 편집 백분율 값, 및 도 159에 도시된 실험에 대한 indel의 백분율을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0044] 정의

[0045] 본원 및 청구범위에 사용된 단수 형태는 문맥이 달리 명백하게 나타내지 않는 한 단수 및 복수 지시대상을 포함한다. 따라서, 예를 들어, "작용제"에 대한 언급은 단일 작용제 및 복수의 이러한 작용제를 포함한다.

[0046] 용어 "핵산 프로그램가능한 DNA 결합 단백질" 또는 "napDNAbp"는 핵산 (예를 들어, DNA 또는 RNA), 예컨대 napDNAbp를 특정 핵산 서열로 가이드하는 가이드 핵산 (예를 들어, gRNA)과 예를 들어 표적 핵산 서열에 혼성화함으로써 회합되는 단백질을 지칭한다. 예를 들어, Cas9 단백질은 Cas9 단백질을 가이드 RNA에 상보적인 특정 DNA 서열로 가이드하는 가이드 RNA와 회합될 수 있다. 일부 실시양태에서, napDNAbp는 부류 2 미생물 CRISPR-Cas 이펙터이다. 일부 실시양태에서, napDNAbp는 Cas9 도메인, 예를 들어 뉴클레아제 활성 Cas9, Cas9 니카제 (nCas9) 또는 뉴클레아제 불활성 Cas9 (dCas9)이다. 핵산 프로그램가능한 DNA 결합 단백질의 예는, 비제한적으로, Cas9 (예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, C2c1, C2c2, C2C3 및 아르고노트를 포함한다. 그러나, 핵산 프로그램가능한 DNA 결합 단백질은 또한 RNA에 결합하는 핵산 프로그램가능한 단백질을 포함한다는 것이 인지되어야 한다. 예를 들어, napDNAbp는 napDNAbp를 RNA로 가이드하는 핵산과 회합될 수 있다. 다른 핵산 프로그램가능한 DNA 결합 단백질은 이들이 본 개시내용에 구체적으로 기재되어 있지 않을 수 있더라도 또한 본 개시내용의 범주 내에 있다.

[0047] 일부 실시양태에서, napDNAbp는 "RNA-프로그램가능한 뉴클레아제" 또는 "RNA-가이드된 뉴클레아제"이다. 용어는 본원에서 상호교환가능하게 사용되며, 절단을 위한 표적이 아닌 1개 이상의 RNA(들)와 복합체를 형성하는 (예를 들어, 그와 결합하거나 회합하는) 뉴클레아제를 지칭한다. 일부 실시양태에서, RNA-프로그램가능한 뉴클레아제는, RNA와의 복합체로 존재하는 경우, 뉴클레아제:RNA 복합체로 지칭될 수 있다. 전형적으로, 결합된 RNA(들)는 가이드 RNA (gRNA)로 지칭된다. gRNA는 2개 이상의 RNA의 복합체로서, 또는 단일 RNA 분자로서 존재할 수 있다. 단일 RNA 분자로서 존재하는 gRNA는 단일-가이드 RNA (sgRNA)로 지칭될 수 있지만, "gRNA"는 또한 단일 분자로서 또는 2개 이상의 분자의 복합체로서 존재하는 가이드 RNA를 지칭하는 데 사용된다. 전형적으로, 단일 RNA 종으로서 존재하는 gRNA는 2개의 도메인을 포함한다: (1) 표적 핵산에 대한 상동성을 공유하는 (즉, 표적에 대한 Cas9 복합체의 결합을 지시하는) 도메인; 및 (2) Cas9 단백질에 결합하는 도메인. 일부 실시양태에서, 도메인 (2)는 tracrRNA로 공지된 서열에 상응하며, 스템-루프 구조를 포함한다. 일부 실시양태에서, 도메인 (2)는 문헌 [Jinek et al., Science 337:816-821 (2012)] (이의 전체 내용은 본원에 참조로 포함됨)에 제공된 바와 같은 tracrRNA와 동일하거나 상동이다. gRNA의 다른 예 (예를 들어, 도메인 2를 포함하는 것)는 2013년 9월 6일에 출원된 발명의 명칭 "스위칭가능한 Cas9 뉴클레아제 및 그의 용도"인 미국 특허 가출원 U.S.S.N. 61/874,682, 및 2013년 9월 6일에 출원된 발명의 명칭 "기능적 뉴클레아제를 위한 전달 시스템"인 미국 특허 가출원 U.S.S.N. 61/874,746에서 찾아볼 수 있으며, 각각의 전체 내용은 그 전문이 본원에 참조로 포함된다. 일부 실시양태에서, gRNA는 도메인 (1) 및 (2) 중 2개 이상을 포함하며, "연장된 gRNA"로 지칭될 수 있다. 예를 들어, 연장된 gRNA는, 본원에 기재된 바와 같이, 2개 이상의 Cas9 단백질과 결합될 것이고, 2개 이상의 별개의 영역에서 표적 핵산과 결합될 것이다. gRNA는 표적 부위에 상보적인 뉴클레오티드 서열을 포함하며, 이는 상기 표적 부위에 대한 뉴클레아제/RNA 복합체의 결합을 매개하여, 뉴클레아제:RNA 복합체의 서열 특이성을 제공한다. 일부 실시양태에서, RNA-프로그램가능한 뉴클레아제는 스트렙토코쿠스 피오게네스(*Streptococcus pyogenes*)로부터의 (CRISPR-회합된 시스템) Cas9 엔도뉴클레아제, 예를 들어, Cas9 (Csn1)이다 (예를 들어, 각각의 전체 내용이 본원에 참조로 포함된 문헌 ["Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti 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).

[0048] 일부 실시양태에서, 본원에 제공된 sgRNA 중 임의의 것은 napDNAbp에 결합하는 서열, 예를 들어, sgRNA 백본 서열을 포함한다. 예를 들어 sgRNA는 문헌 [Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, and Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337, 816-812; Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. Nature Methods, 10, 957-963; Li JF, Norville JE, Aach J, McCromack M, Zhang D, Bush J, Church GM, and Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nature Biotech, 31, 688-691; Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JRJ, Joung JK (2013) Efficient in vivo genome editing using RNA-guided nucleases. Nat Biotechnol, 31, 227-229; Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRIPSR/Cas systems. Science, 339, 819-823; Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol, 31, 230-232; Jinek MJ, East A, Cheng A, Lin S, Ma E, Doudna J (2013) RNA-programmed genome editing in human cells. eLIFE, 2:e00471; DiCarlo JE, Norville JE, Mali P, Rios, Aach J, Church GM (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucl Acids Res, 41, 4336-4343; Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, and Barrangou R (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. Mol Cell, 56, 333-339]에 기재되어 있으며, 이들 각각의 내용은 본원에 참조로 포함된다. 일부 실시양태에서, 본원에 제공된 gRNA (예를 들어, sgRNA) 중 임의의 것은 GTAATTTCTACTAAGTGTAGAT (서열식별번호: 741)의 핵산 서열 (여기서 서열식별번호: 741의 각각의 T는 우라실 (U)이며, 즉, GUAAUUUCUACUAAGUGUAGAU임), 또는 서열 GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU-3' (서열식별번호: 618)를 포함한다.

[0049] RNA-프로그램가능한 뉴클레아제 (예를 들어, Cas9)는 DNA 절단 부위를 표적화하기 위해 RNA:DNA 혼성화를 사용하기 때문에, 이들 단백질은 원칙적으로 가이드 RNA에 의해 특정된 임의의 서열을 표적화할 수 있다. 부위-특이적 절단을 위해 (예를 들어, 게놈을 변형시키기 위해) RNA-프로그램가능한 뉴클레아제, 예컨대 Cas9를 사용하는 방법이 관련 기술분야에 공지되어 있다 (예를 들어, 각각의 전체 내용이 본원에 참조로 포함된 문헌 [Cong, L. et al., Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali, P. et al., RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang, W.Y. et al., Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek, M. et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo, J.E. et al., Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Research (2013); Jiang, W. et al., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology 31, 233-239 (2013)] 참조).

[0050] 용어 "Cas9" 또는 "Cas9 뉴클레아제"는 Cas9 단백질, 또는 그의 단편 (예를 들어, Cas9의 활성, 불활성, 또는 부분적 활성 DNA 절단 도메인, 및/또는 Cas9의 gRNA 결합 도메인을 포함하는 단백질)을 포함하는 RNA-가이드된 뉴클레아제를 지칭한다. Cas9 뉴클레아제는 또한 때때로 casn1 뉴클레아제 또는 CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부)-연관된 뉴클레아제로 지칭된다. CRISPR은 이동성 유전 요소 (바이러스, 수송가능한 요소 및 접합성 플라스미드)에 대한 보호를 제공하는 적응 면역계이다. CRISPR 클러스터는 스페이서, 선행 이동성 요소에 상보적인 서열, 및 표적 침입 핵산을 함유한다. CRISPR 클러스터는 전사되고, CRISPR RNA (crRNA)로 프로세싱된다. 유형 II CRISPR 시스템에서, 프리-crRNA의 교정 프로세싱은 트랜스-코딩된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질을 요구한다. tracrRNA는 프리-crRNA의 리보뉴클레아제 3-보조된 프로세싱을 위한 가이드로서 작용한다. 이어서, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 원형 dsDNA 표적을 내핵분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 내핵분해적으로 절단된 후, 외핵분해적으로 3'-5' 트리밍된다. 사실상, DNA-결합 및 절단은 전형적으로 단백질 및 둘 다의 RNA를 요구한다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 간단히 "gNRA")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, 그의 전체 내용은 본원에 참조로 포함된다. Cas9는 자기 대 비-자기를 구별하는 것을 돕기 위한 CRISPR 반복 서열 (PAM 또는 프로토스페이서 인접 모티프)에서의 짧은 모티프를 인식한다. Cas9 뉴클레아제 서열 및 구조는 관련 기술분야의 통상

의 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 오르톨로그는 에스. 피오게네스 및 에스. 써모필루스(*S. thermophilus*)를 포함하나 이에 제한되지 않는 다양한 종에서 기재되었다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이며, 이러한 Cas9 뉴클레아제 및 서열은 문헌 [Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737] (그의 전체 내용은 본원에 참조로 포함됨)에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다. 일부 실시양태에서, Cas9 뉴클레아제는 불활성 (예를 들어, 불활성화된) DNA 절단 도메인을 가지며, 즉, Cas9는 니카제이다.

[0051] 뉴클레아제-불활성화된 Cas9 단백질은 상호교환가능하게 (뉴클레아제-"데드" Cas9에 대해) "dCas9" 단백질로 지칭될 수 있다. 불활성 DNA 절단 도메인을 갖는 Cas9 단백질 (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 2개의 Cas9 도메인 중 1개를 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 Cas9와 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 99.9% 동일하다. 일부 실시양태에서, Cas9 변이체는 야생형 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 단편이 야생형 Cas9의 상응하는 단편과 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 99.9% 동일하도록 Cas9의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함한다. 일부 실시양태에서, 단편은 상응하는 야생형 Cas9의 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% 동일하거나, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%이다.

[0052] 일부 실시양태에서, 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, 1250, 또는 적어도 1300개 아미노산 길이이다. 일부 실시양태에서, 야생형 Cas9는 스트렙토코쿠스 피오게네스로부터의 Cas9 (NCBI 참조 서열: NC_017053.1, 서열식별번호: 1 (뉴클레오티드); 서열식별번호: 2 (아미노산))에 상응한다.

[0053]

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGG
CGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAA
TACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGT
GGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACA
CGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGA
AAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGAC
AAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATC
ATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGA
TAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTA
TTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGA
TTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGA
ATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGG
CAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAA
TTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAG
GCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGAC
TCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGA
AGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAA
TTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACA
ACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGA
CAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCT
TGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTG
CATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGT
TGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATA
AAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTT
ACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACA
AATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAAT
GTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGC
GCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAG
AAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGG
GGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGA
TGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTG
ATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAAT
CAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACA
TTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATG
AACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACT
GTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCG
TTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGC

GAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTC
TTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTAT
TATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTT
TAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCA
ATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACG
TTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAA
CGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGA
GGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCG
CCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATAC
GATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAAT
TAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAAT
TACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAA
GAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATG
TTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATA
TTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATG
GAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGT
CTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAA
GTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCA
ATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATC
CAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTT
GCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTA
GGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAG
AAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATA
TAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAA
TTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTT
AGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACA
ATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGT
GAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGC
ATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCAT
TTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAAC
AATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCC
ATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGT
GACTGA (SEQ ID NO:1)

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKR
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
[0054]

NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT
STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:2)
[0055]

[0056] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0057] 일부 실시양태에서, 야생형 Cas9는 서열식별번호:3 (뉴클레오티드) 및/또는 서열식별번호: 4 (아미노산)에 상응하거나, 이를 포함한다:

[0058]

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGC
TGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAAC
ACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGG
CGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACG
TCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAA
GTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAA
GAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCAT
GAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATA
AAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGG
CACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTT
CATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCA
AGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGC
TAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAA
CCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAG
CTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAA
TCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACC
TTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAG
GCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGA
CACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATT
CTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAA
GAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAG
AGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGA
CAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGA
AGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAA
TCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGG
TTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGG
AAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTT
TGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAG
TATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGC
GTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATT
CAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAA
AATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGT
CACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGA
TAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTG
AAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACG
ATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTC
GCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGAT
TTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATG
ACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGA

CTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGC
ATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACA
AACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGG
GGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAA
CTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACG
AGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA
ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCT
TTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCG
AGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTA
TTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTA
ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAAC
GTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTC
CCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGT
AATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATA
AAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGT
CGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTAT
GGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAG
ATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAA
GACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACC
AATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTG
AGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGA
CCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCAT
CGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACA
GTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAAC
TGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGA
AAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGA
TCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAA
CGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGT
CTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCA
CCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCG
ACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGC
CAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGT
GAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGC
CGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAG
GAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTC
GGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAG
TCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTA
CAAGGATGACGATGACAAGGCTGCAGGA (SEQ ID NO:3)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
[0059]

LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:4)
[0060]

[0061] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0062] 일부 실시양태에서, 야생형 Cas9는 스트렙토코쿠스 피오게네스로부터의 Cas9 (NCBI 참조 서열: NC_002737.2, 서

열식별번호: 5 (뉴클레오티드); 및 유니프롯 참조 서열: Q99ZW2, 서열식별번호: 6 (아미노산))에 상응한다.

[0063]

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGG
CGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAA
TACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGT
GGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACA
CGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGA
AAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGAC
AAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATC
ATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGA
TAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTA
TTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGC
AAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGA
TTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGA
ATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGG
CAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAA
TTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATT
TATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAG
GCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGAC
TCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGA
AGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAA
TTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACA
ACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGA
CAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCT
TGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTG
CATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGT
TGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATA
AAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTT

ACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAAC
CAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACA
AATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAAT
GTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGT
ACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAG
AAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGG
GAGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGA
TGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTG
ATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAAT
CAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACA
TTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATG
AACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACT
GTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATA
TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTC
GCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGAT
TCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCT
ATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCG
TTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATT
CAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAA
CGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTA
AACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTG
GAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAAC
TCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAA
TACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTA
AATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAAC
AATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGAT
TAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATG
ATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAA
ATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAA
ATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAA
TTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCC
CCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGA
GTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGG
GATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGT
GGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTT
ACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTT
TTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTA
AATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGG
AGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTA
TATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAA
AACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAAT
CAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTA
GTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTA
[0064] TTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATA

CAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCT
TATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAG
GAGGTGACTGA (SEQ ID NO: 5)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 6)

[0065]

[0066] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0067] 일부 실시양태에서, Cas9는 코리네박테리움 울세란스(*Corynebacterium ulcerans*) (NCBI Ref: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아(*Corynebacterium diphtheria*) (NCBI Ref: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라(*Spiroplasma syrphidicola*) (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아(*Prevotella intermedia*) (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨스(*Spiroplasma taiwanense*) (NCBI Ref: NC_021846.1); 스트렙토코쿠스 이니아에(*Streptococcus iniae*) (NCBI Ref: NC_021314.1); 벨리엘라 발티카(*Belliella baltica*) (NCBI Ref: NC_018010.1); 시크로플렉수스 토르퀴스 I(*Psychroflexus torquisI*) (NCBI Ref: NC_018721.1); 스트렙토코쿠스 써모필루스(*Streptococcus thermophilus*) (NCBI Ref: YP_820832.1), 리스테리아 인노쿠아(*Listeria innocua*) (NCBI Ref: NP_472073.1), 캄필로박터 제주니(*Campylobacter jejuni*) (NCBI Ref: YP_002344900.1) 또는 네이세리아 메닝기티디스(*Neisseria meningitidis*) (NCBI Ref: YP_002342100.1)로부터의 Cas9 또는 실시예 5에 열거된 유기체 중 임의의 것으로부터의 Cas9를 지칭한다.

[0068] 일부 실시양태에서, dCas9는 Cas9 뉴클레아제 활성을 불활성화시키는 1개 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 상응하거나, 부분적으로 또는 전체적으로 이를 포함한다. 예를 들어, 일부 실시양태에서, dCas9 도메인은 D10A 및/또는 H840A 돌연변이를 포함한다.

[0069] dCas9 (D10A 및 H840A):

MDKK**YSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS GET**AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYAD LFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQL IHDDSLTFKEDIQKAQVSGQG**DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH** KPENIVIEMA**RENQTTQK**GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG**GLSELDKAGFIK** RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 7)

[0070]

[0071] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인).

[0072] 일부 실시양태에서, Cas9 도메인은 D10A 돌연변이를 포함하는 반면, 위치 840에서의 잔기는 서열식별번호: 6에 제공된 아미노산 서열에, 또는 또 다른 Cas9 도메인, 예컨대 본원에 제공된 Cas9 단백질 중 임의의 것에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 위치에 히스티딘을 보유한다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 촉매 잔기 H840의 존재는 Cas9의 활성을 복원하여 표적화된 C와 대향하는 G를 함유하는 비-편집된 (예를 들어, 비-탈아미노화된) 가닥을 절단한다. H840의 (예를 들어, A840으로부터의) 복원은 C를 함유하는 표적 가닥의 절단을 발생시키지 않는다. 이러한 Cas9 변이체는 gRNA-한정된 표적 서열에 기초하여 특이적 위치에서 단일-가닥 DNA 파괴 (닉)를 생성할 수 있으며, 이는 비-편집된 가닥의 복구를 발생시키고, 궁극적으로 비-편집된 가닥 상의 G에서 A로의 변화를 발생시킨다. 이 과정의 개략적 표현은 도 108에 나타나 있다. 간략하게, C-G 염기 쌍의 C를 데아미나제, 예를 들어, APOBEC 데아미나제에 의해 U로 탈아미노화할 수 있다. G를 갖는 비-편집된 가닥을 니킹하는 것은 미스매치 복구 메카니즘을 통한 G의 제거를 용이하게 한다. UGI는 UDG를 억제하고, 이는 U의 제거를 방지한다.

[0073] 다른 실시양태에서, 예를 들어, 뉴클레아제 불활성화된 Cas9 (dCas9)를 발생시키는, D10A 및 H840A 이외의 돌연변이를 갖는 dCas9 변이체가 제공된다. 이러한 돌연변이는, 예로서, D10 및 H820에 다른 아미노산 치환, 또는 Cas9의 뉴클레아제 도메인 내에 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인에서의 치환)을 포함한다. 일부 실시양태에서, 서열식별번호: 6과 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, 적어도 약 99.9% 동일한 dCas9의 변이체 또는 상동체 (예를 들어, 서열식별번호: 6의 변이체)가 제공된다. 일부 실시양태에서, 서열식별번호: 6보다 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 이상 더 짧거나, 더 긴 아미노산 서열을 갖는 dCas9의 변이체 (예를 들어, 서열식별번호: 6의 변이체)가 제공된다.

[0074] 일부 실시양태에서, 본원에 제공된 바와 같은 Cas9 융합 단백질은 Cas9 단백질의 전장 아미노산 서열, 예를 들어, 본원에 제공된 Cas9 서열 중 하나를 포함한다. 그러나, 다른 실시양태에서, 본원에 제공된 바와 같은 융합 단백질은 전장 Cas9 서열을 포함하지 않지만, 단지 그의 단편을 포함한다. 예를 들어, 일부 실시양태에서, 본원에 제공된 Cas9 융합 단백질은 Cas9 단편을 포함하고, 단편은 crRNA 및 tracrRNA 또는 sgRNA에 결합하지만,

예를 들어, 그것이 단지 뉴클레아제 도메인의 말단절단된 형태만을 포함하거나, 뉴클레아제 도메인을 전혀 포함하지 않는다는 점에서 기능적 뉴클레아제 도메인을 포함하지 않는다. 적합한 Cas9 도메인 및 Cas9 단편의 예시적인 아미노산 서열은 본원에 제공되며, Cas9 도메인 및 단편의 추가의 적합한 서열은 관련 기술분야의 통상의 기술자에게 명백할 것이다.

[0075] 일부 실시양태에서, Cas9는 코리네박테리움 울세란스 (NCBI Ref: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아 (NCBI Ref: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라 (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아 (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨스 (NCBI Ref: NC_021846.1); 스트렙토코쿠스 이니아에 (NCBI Ref: NC_021314.1); 벨리엘라 발티카 (NCBI Ref: NC_018010.1); 시크로플렉수스 토르퀴스I (NCBI Ref: NC_018721.1); 스트렙토코쿠스 써모필루스 (NCBI Ref: YP_820832.1); 리스테리아 인노쿠아 (NCBI Ref: NP_472073.1); 캄필로박터 제주니 (NCBI Ref: YP_002344900.1); 또는 네이세리아 메닝기티디스 (NCBI Ref: YP_002342100.1)로부터의 Cas9를 지칭한다.

[0076] 본원에 사용된 용어 "데아미나제" 또는 "데아미나제 도메인"은 탈아미노화 반응을 촉매하는 단백질 또는 효소를 지칭한다. 일부 실시양태에서, 데아미나제 또는 데아미나제 도메인은 각각 시티딘 또는 데옥시시티딘의 우리딘 또는 데옥시우리딘으로의 가수분해적 탈아미노화를 촉매하는 시티딘 데아미나제이다. 일부 실시양태에서, 데아미나제 또는 데아미나제 도메인은 시토신의 우라실로의 가수분해적 탈아미노화를 촉매하는 시티딘 데아미나제 도메인이다. 일부 실시양태에서, 데아미나제 또는 데아미나제 도메인은 유기체, 예컨대 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트, 또는 마우스로부터의 자연-발생 데아미나제이다. 일부 실시양태에서, 데아미나제 또는 데아미나제 도메인은 자연에서 발생하지 않는 유기체로부터의 자연-발생 데아미나제의 변이체이다. 예를 들어, 일부 실시양태에서, 데아미나제 또는 데아미나제 도메인은 유기체로부터의 자연-발생 데아미나제와 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% 동일하다.

[0077] 본원에 사용된 용어 "유효량"은 목적하는 생물학적 반응을 유발하는 데 충분한 생물학적 활성제의 양을 지칭한다. 예를 들어, 일부 실시양태에서, 뉴클레아제의 유효량은 뉴클레아제에 의해 특이적으로 결합되고 절단되는 표적 부위의 절단을 유도하는 데 충분한 뉴클레아제의 양을 지칭할 수 있다. 일부 실시양태에서, 본원에 제공된 융합 단백질의, 예를 들어, 뉴클레아제-불활성 Cas9 도메인 및 핵산 편집 도메인 (예를 들어, 데아미나제 도메인)을 포함하는 융합 단백질의 유효량은 융합 단백질에 의해 특이적으로 결합되고 편집되는 표적 부위의 편집을 유도하는 데 충분한 융합 단백질의 양을 지칭할 수 있다. 통상의 기술자에 의해 인지될 바와 같이, 작용제, 예를 들어, 융합 단백질, 뉴클레아제, 데아미나제, 레콤비나제, 혼성 단백질, 단백질 이량체, 단백질 (또는 단백질 이량체) 및 폴리뉴클레오티드의 복합체, 또는 폴리뉴클레오티드의 유효량은 예를 들어, 목적하는 생물학적 반응, 예를 들어, 편집될 특이적 대립유전자, 게놈, 또는 표적 부위, 표적화되는 세포 또는 조직, 및 사용되는 작용제와 같은 다양한 인자에 의존할 수 있다.

[0078] 본원에 사용된 용어 "링커"는 2개의 분자 또는 모이어티, 예를 들어, 융합 단백질의 2개의 도메인, 예컨대, 예를 들어, 뉴클레아제-불활성 Cas9 도메인 및 핵산 편집 도메인 (예를 들어, 데아미나제 도메인)을 연결하는 화학적 기 또는 분자를 지칭한다. 링커는 예를 들어 임의의 길이 및 조성의 아미노산 서열, 펩티드, 또는 중합체일 수 있다. 일부 실시양태에서, 링커는 Cas9 뉴클레아제 도메인, 및 핵산 편집 단백질의 촉매 도메인을 포함한 RNA-프로그램가능한 뉴클레아제의 gRNA 결합 도메인을 연결한다. 일부 실시양태에서, 링커는 dCas9 및 핵산 편집 단백질을 연결한다. 전형적으로, 링커는 2개의 기, 분자, 또는 다른 모이어티 사이에 위치되거나, 그에 의해 플랭킹되고, 공유 결합을 통해 각각의 것에 대해 연결되며, 따라서 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 1 내지 100개 아미노산 길이, 예를 들어, 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, 30 내지 35, 35 내지 40, 40 내지 45, 45 내지 50, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 90 내지 100, 100 내지 150, 또는 150 내지 200개 아미노산 길이이다. 보다 긴 또는 보다 짧은 링커가 또한 고려된다.

[0079] 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 서열 내의 잔기의 위치에 의해 이어지는 원래 잔기를 확인함으로써, 및 새롭게 치환된 잔기의 확인에 의해 본원에 기재된다. 본원에 제공된 아미노산 치환 (돌연변이)을 생성하는 다양한 방법은 관련 기술분야에 널리 공지되어 있으며, 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.

[0080] 본원에 사용된 용어 "핵산" 및 "핵산 분자"는 핵염기 및 산성 모이어티, 예를 들어, 뉴클레오시드, 뉴클레오티드, 또는 뉴클레오티드의 중합체를 포함하는 화합물을 지칭한다. 전형적으로, 중합체 핵산, 예를 들어, 3개 이상의 뉴클레오티드를 포함하는 핵산 분자는 인접한 뉴클레오티드가 포스포디에스테르 연결을 통해 서로 연결된 선형 분자이다. 일부 실시양태에서, "핵산"은 개별 핵산 잔기 (예를 들어 뉴클레오티드 및/또는 뉴클레오시드)를 지칭한다. 일부 실시양태에서, "핵산"은 3개 이상의 개별 뉴클레오티드 잔기를 포함하는 올리고뉴클레오티드 쇄를 지칭한다. 본원에 사용된 용어 "올리고뉴클레오티드" 및 "폴리뉴클레오티드"는 상호교환가능하게 뉴클레오티드 (예를 들어, 일련의 적어도 3개의 뉴클레오티드)의 중합체를 지칭한다. 일부 실시양태에서, "핵산"은 RNA 뿐만 아니라 단일 및/또는 이중-가닥 DNA를 포괄한다. 핵산은 예를 들어, 게놈, 전사체, mRNA, tRNA, rRNA, siRNA, snRNA, 플라스미드, 코스미드, 염색체, 크로마티드 또는 다른 자연 발생 핵산 분자의 맥락에서 자연 발생일 수 있다. 한편, 핵산 분자는 비-자연 발생 분자, 예를 들어, 재조합 DNA 또는 RNA, 인공 염색체, 조작된 게놈, 또는 그의 단편, 또는 합성 DNA, RNA, DNA/RNA 하이브리드이거나, 비-자연 발생 뉴클레오티드 또는 뉴클레오시드를 포함할 수 있다. 게다가, 용어 "핵산", "DNA", "RNA", 및/또는 유사한 용어는 핵산 유사체, 예를 들어, 포스포디에스테르 백본 이외의 것을 갖는 유사체를 포함한다. 핵산은 천연 공급원으로부터 정제되고, 재조합 발현 시스템을 사용하여 생성되고, 임의로 정제되고, 화학적으로 합성되는 등을 할 수 있다. 적절한 경우, 예를 들어, 화학적으로 합성된 분자의 경우, 핵산은 뉴클레오시드 유사체, 예컨대 화학적으로 변형된 염기 또는 당, 및 백본 변형을 갖는 유사체를 포함할 수 있다. 핵산 서열은 달리 나타내지 않는 한 5'에서 3' 방향으로 제시된다. 일부 실시양태에서, 핵산은 천연 뉴클레오시드 (예를 들어 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, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-아이오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노아데노신, 7-데아자아데노신, 7-데아자구아노신, 8-옥소아데노신, 8-옥소구아노신, O(6)-메틸구아닌, 및 2-티오시티딘); 화학적으로 변형된 염기; 생물학적으로 변형된 염기 (예를 들어, 메틸화된 염기); 개재된 염기; 변형된 당 (예를 들어, 2'-플루오로리보스, 리보스, 2'-데옥시리보스, 아라비노스, 및 헥소스); 및/또는 변형된 포스페이트 기 (예를 들어, 포스포로티오에이트 및 5'-N-포스포르아미다이트 연결)이거나, 이를 포함한다.

[0081] 본원에 사용된 용어 "핵산 편집 도메인"은 핵산 (예를 들어, DNA 또는 RNA)에 대해 1개 이상의 변형 (예를 들어, 시티딘 잔기의 탈아미노화)을 생성할 수 있는 단백질 또는 효소를 지칭한다. 예시적인 핵산 편집 도메인은 데아미나제, 뉴클레아제, 니카제, 레콤비나제, 메틸트랜스퍼라제, 메틸라제, 아세틸라제, 아세틸트랜스퍼라제, 전사 활성인자, 또는 전사 리프레서 도메인을 포함하나 이에 제한되지는 않는다. 일부 실시양태에서, 핵산 편집 도메인은 데아미나제 (예를 들어, 시티딘 데아미나제, 예컨대 APOBEC 또는 AID 데아미나제)이다.

[0082] 본원에 사용된 용어 "증식성 질환"은 세포 또는 세포 집단이 비정상적으로 상승된 증식 속도를 나타낸다는 점에서 세포 또는 조직 항상성이 교란된 임의의 질환을 지칭한다. 증식성 질환은 과다증식성 질환, 예컨대 전-신생물성 과다형성 상태 및 신생물성 질환을 포함한다. 신생물성 질환은 세포의 비정상적 증식을 특징으로 하며, 양성 및 악성 신생물 둘 다를 포함한다. 악성 신생물은 또한 암으로 지칭된다.

[0083] 용어 "단백질", "펩티드", 및 "폴리펩티드"는 본원에서 상호교환가능하게 사용되며, 펩티드 (아미드) 결합에 의해 함께 연결된 아미노산 잔기의 중합체를 지칭한다. 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 단백질의 집합을 지칭할 수 있다. 단백질, 펩티드, 또는 폴리펩티드에서의 아미노산 중 1개 이상은 예를 들어, 화학 물질, 예컨대 탄수화물 기, 히드록실 기, 포스페이트 기, 파르네실 기, 이소파르네실 기, 지방산 기, 접합용 링커의 첨가, 관능화, 또는 다른 변형 등에 의해 변형될 수 있다. 단백질, 펩티드, 또는 폴리펩티드는 또한 단일 분자일 수 있거나, 다중-분자 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 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 핵산 편집 단백질의 상이한 도메인, 예를 들어, 핵산 결합 도메인 (예를 들어, 표적 부위에의 단백질의 결

합을 지정하는 Cas9의 gRNA 결합 도메인) 및 핵산 절단 도메인 또는 촉매 도메인을 포함할 수 있다. 일부 실시양태에서, 단백질은 단백질성 부분, 예를 들어, 핵산 결합 도메인을 구성하는 아미노산 서열, 및 유기 화합물, 예를 들어, 핵산 절단제로서 작용할 수 있는 화합물을 포함한다. 일부 실시양태에서, 단백질은 핵산, 예를 들어, RNA와의 복합체에 있거나, 그와 회합된다. 본원에 제공된 단백질 중 임의의 것은 관련 기술분야에 공지된 임의의 방법에 의해 생성될 수 있다. 예를 들어, 본원에 제공된 단백질은 펩티드 링커를 포함하는 융합 단백질에 특히 적합한 재조합 단백질 발현 및 정제를 통해 생성될 수 있다. 재조합 단백질 발현 및 정제를 위한 방법은 널리 공지되어 있으며, 그의 전체 내용이 본원에 참조로 포함되는 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 것들을 포함한다.

[0084] 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, 곤충류, 파리, 또는 선충류이다. 일부 실시양태에서, 대상체는 연구 동물이다. 일부 실시양태에서, 대상체는 유전자 조작된, 예를 들어, 유전자 조작된 비-인간 대상체이다. 대상체는 양쪽 성별의 및 임의의 발달 단계의 것일 수 있다.

[0085] 용어 "표적 부위"는 데아미나제 또는 데아미나제를 포함하는 융합 단백질 (예를 들어, 본원에 제공된 dCas9-데아미나제 융합 단백질)에 의해 탈아미노화되는 핵산 분자 내의 서열을 지칭한다.

[0086] 용어 "치료", "치료하다", 및 "치료하는"은 본원에 기재된 바와 같은 질환 또는 장애, 또는 그의 1종 이상의 증상을 역전시키거나, 경감시키거나, 그의 발병을 지연시키거나, 그의 진행을 억제하는 것을 목표로 하는 임상적 개입을 지칭한다. 본원에 사용된 용어 "치료", "치료하다", 및 "치료하는"은 본원에 기재된 바와 같은 질환 또는 장애, 또는 그의 1종 이상의 증상을 역전시키거나, 경감시키거나, 그의 발병을 지연시키거나, 그의 진행을 억제하는 것을 목표로 하는 임상적 개입을 지칭한다. 일부 실시양태에서, 치료는 1종 이상의 증상이 발달된 후에 및/또는 질환이 진단된 후에 투여될 수 있다. 다른 실시양태에서, 치료는 증상의 부재 하에서, 예를 들어, 증상의 발병을 예방하거나 지연시키거나, 또는 질환의 발병 또는 진행을 억제하기 위해 투여될 수 있다. 예를 들어, 치료는 증상의 발병 전에 (예를 들어, 증상의 병력의 관점에서 및/또는 유전적 또는 다른 감수성 인자의 관점에서) 감수성 개체에게 투여될 수 있다. 치료는 또한 증상이 해결된 후에, 예를 들어, 그의 재발을 예방하거나 지연시키기 위해 계속될 수 있다.

[0087] 단백질 또는 핵산과 관련하여 본원에 사용된 용어 "재조합"은 자연에서 발생하지 않지만, 인간 조작의 생성물인 단백질 또는 핵산을 지칭한다. 예를 들어, 일부 실시양태에서, 재조합 단백질 또는 핵산 분자는 임의의 자연 발생 서열과 비교하여 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 또는 적어도 7개의 돌연변이를 포함하는 아미노산 또는 뉴클레오티드 서열을 포함한다.

[0088] 본원에 사용된 용어 "제약 조성물"은 질환 또는 장애의 치료와 관련하여 대상체에게 투여될 수 있는 조성물을 지칭한다. 일부 실시양태에서, 제약 조성물은 활성 성분, 예를 들어 뉴클레아제 또는 뉴클레아제를 코딩하는 핵산, 및 제약상 허용되는 부형제를 포함한다.

[0089] 본원에 사용된 용어 "염기 편집제 (BE)" 또는 "핵염기 편집제 (NBE)"는 핵산 서열 (예를 들어, DNA 또는 RNA) 내의 염기 (예를 들어, A, T, C, G 또는 U)에 변형을 가할 수 있는 폴리펩티드를 포함하는 작용제를 지칭한다. 일부 실시양태에서, 염기 편집제는 핵산 내의 염기를 탈아미노화할 수 있다. 일부 실시양태에서, 염기 편집제는 DNA 분자 내의 염기를 탈아미노화할 수 있다. 일부 실시양태에서, 염기 편집제는 DNA 중 시토신 (C)을 탈아미노화할 수 있다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제 도메인에 융합된 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)을 포함하는 융합 단백질이다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 Cas9 (예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, C2c1, C2c2, C2c3, 또는 아르고노트 단백질을 포함한다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 Cas9 니카제 (nCas9)이다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 뉴클레아제-불활성 Cas9 (dCas9)이다. 일부 실시양태에서, 염기 편집제는 염기 절제 복구의 억제제, 예를 들어 UGI 도메인에 융합된다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 CasX 단백질을 포함한다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 CasY 단백질을 포함한다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 Cpf1 단백질을 포함한다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 C2c1 단백질을 포함한다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 C2c2 단백질을 포함한다. 일부 실

시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 C2c3 단백질을 포함한다. 일부 실시양태에서, 염기 편집제는 시티딘 데아미나제에 융합된 아르고노트 단백질을 포함한다.

[0090] 본원에 사용된 용어 "우라실 글리코실라제 억제제" 또는 "UGI"는 우라실-DNA 글리코실라제 염기-절제 복구 효소를 억제할 수 있는 단백질을 지칭한다.

[0091] 본원에 사용된 용어 "Cas9 니카제"는 듀플렉스 핵산 분자 (예를 들어, 듀플렉스 DNA 분자)의 단지 1개의 가닥만을 절단할 수 있는 Cas9 단백질을 지칭한다. 일부 실시양태에서, Cas9 니카제는 D10A 돌연변이를 포함하고, 서열식별번호: 6의 위치 H840에 히스티딘, 또는 또 다른 Cas9 도메인, 예컨대 본원에 제공된 Cas9 단백질 중 임의의 것에서의 상응하는 돌연변이를 갖는다. 예를 들어, Cas9 니카제는 서열식별번호: 8에 제시된 바와 같은 아미노산 서열을 포함할 수 있다. 이러한 Cas9 니카제는 활성 HNH 뉴클레아제 도메인을 가지며, DNA의 비-표적화된 가닥, 즉, gRNA에 의해 결합된 가닥을 절단할 수 있다. 또한, 이러한 Cas9 니카제는 불활성 RuvC 뉴클레아제 도메인을 가지며, DNA의 표적화된 가닥, 즉, 염기 편집을 목적하는 하는 가닥을 절단할 수 없다.

[0092] 예시적인 Cas9 니카제 (클로닝 벡터 pPlatTET-gRNA2; 수탁 번호 BAV54124).

[0093]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA

AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 8)

[0094]

[0095] 발명의 특정 실시양태의 상세한 설명

[0096] 본 개시내용의 일부 측면은 뉴클레오티드 서열에 결합할 수 있는 도메인 (예를 들어, Cas9, 또는 Cpf1 단백질) 및 효소 도메인, 예를 들어, DNA-편집 도메인, 예컨대, 예를 들어, 데아미나제 도메인을 포함하는 융합 단백질을 제공한다. 데아미나제에 의한 핵염기의 탈아미노화는 각각의 잔기에서 점 돌연변이를 발생시킬 수 있으며, 이는 본원에서 핵산 편집으로 지칭된다. 따라서, Cas9 변이체 또는 도메인 및 DNA 편집 도메인을 포함하는 융합 단백질은 핵산 서열의 표적화된 편집에 사용될 수 있다. 이러한 융합 단백질은 시험관내에서 DNA의 표적화된 편집에, 예를 들어, 돌연변이체 세포 또는 동물의 생성에; 표적화된 돌연변이의 도입에, 예를 들어, 생체외에서의 세포에서의, 예를 들어, 동일한 또는 또 다른 대상체 내로 후속적으로 재-도입된 대상체로부터 얻어진 세포에서의 유전적 결함의 교정에; 및 표적화된 돌연변이의 도입, 예를 들어 대상체에서 질환-연관된 유전자에

서의 유전적 결함의 교정 또는 탈활성화 돌연변이의 도입에 유용하다. 전형적으로, 본원에 기재된 융합 단백질의 Cas9 도메인은 임의의 뉴클레아제 활성을 갖지 않지만, 대신 Cas9 단편 또는 dCas9 단백질 또는 도메인이다. 본 발명의 다른 측면은 (i) 핵산 서열에 결합할 수 있는 도메인 (예를 들어, Cas9 또는 Cpf1 단백질); (ii) 효소 도메인, 예를 들어, DNA-편집 도메인 (예를 들어, 데아미나제 도메인); 및 (iii) 1개 이상의 우라실 글리코실라제 억제제 (UGI) 도메인을 포함하는 융합 단백질을 제공한다. 적어도 1개의 UGI 도메인의 존재는 UGI 도메인이 없는 융합 단백질과 비교하여 염기 편집 효율을 증가시킨다. 2개의 UGI 도메인을 포함하는 융합 단백질은 1개의 UGI 도메인을 갖거나 UGI 도메인을 갖지 않는 융합 단백질과 비교하여 염기 편집 효율 및 생성물 순도를 추가로 증가시킨다. 본원에 기재된 바와 같은 Cas9 융합 단백질의 사용 방법이 또한 제공된다.

[0097] 핵산 프로그램가능한 DNA 결합 단백질

[0098] 본 개시내용의 일부 측면은 단백질, 예컨대 염기 편집제를 특정 핵산 (예를 들어, DNA 또는 RNA) 서열에 가이드하는 데 사용될 수 있는 핵산 프로그램가능한 DNA 결합 단백질을 제공한다. 본원에 제공된 융합 단백질 (예를 들어, 염기 편집제) 중 임의의 것은 임의의 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)을 포함할 수 있다는 것이 인지되어야 한다. 예를 들어, Cas9 도메인을 포함하는 본원에 기재된 융합 단백질 중 임의의 것은 Cas9 도메인 대신 또 다른 napDNAbp, 예컨대 CasX, CasY, Cpf1, C2c1, C2c2, C2c3, 및 아르고노트를 사용할 수 있다. 핵산 프로그램가능한 DNA 결합 단백질은, 비제한적으로, Cas9 (예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, C2c1, C2c2, C2C3, 및 아르고노트를 포함한다. Cas9와 상이한 PAM 특이성을 갖는 핵산 프로그램가능한 DNA 결합 단백질의 한 예는 프레보텔라 및 프란시셀라 1로부터의 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (Cpf1)이다. Cas9와 유사하게, Cpf1도 또한 부류 2 CRISPR 이펙터이다. Cpf1은 Cas9와 구분되는 특색을 가지면서 강건한 DNA 간섭을 매개한다. Cpf1은 tracrRNA가 결여된 단일 RNA-가이드 엔도뉴클레아제이고, 이는 T-풍부 프로토스페이서-인접 모티프 (TTN, TTTN 또는 YTN)를 사용한다. 더욱이, Cpf1은 엇갈린 DNA 이중-가닥 파괴를 통해 DNA를 절단한다. 16종의 Cpf1-패밀리 단백질 중에서, 아시다미노코쿠스 및 라크노스피라세아에로부터의 2종의 효소는 인간 세포에서 효율적인 게놈-편집 활성을 갖는 것으로 밝혀졌다. Cpf1 단백질은 관련 기술분야에 공지되어 있고, 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 이전에 기재되어 있으며; 이들 전체 내용은 본원에 참조로 포함된다.

[0099] 가이드 뉴클레오티드 서열-프로그램가능한 DNA 결합 단백질 도메인으로서 사용될 수 있는 뉴클레아제-불활성 Cpf1 (dCpf1) 변이체가 또한 본 발명의 조성물 및 방법에 유용하다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레아제 도메인을 갖지만, HNH 엔도뉴클레아제 도메인은 갖지 않고, Cpf1의 N-말단은 Cas9의 알파-나선 인식 로브를 갖지 않는다. Cpf1의 RuvC-유사 도메인은 양쪽 DNA 가닥의 절단을 담당하고, RuvC-유사 도메인의 불활성화는 Cpf1 뉴클레아제 활성을 불활성화한다는 것이 문헌 [Zetsche et al., Cell, 163, 759-771, 2015 (이는 본원에 참조로 포함됨)]에 제시되었다. 예를 들어, 프란시셀라 노비시다(*Francisella novicida*) Cpf1 (서열식별번호: 15)에서 D917A, E1006A, 또는 D1255A에 상응하는 돌연변이는 Cpf1 뉴클레아제 활성을 불활성화한다. 일부 실시양태에서, 데드 Cpf1 (dCpf1)은 서열식별번호: 9에서 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, 또는 D917A/E1006A/D1255A에 상응하는 돌연변이를 포함한다. Cpf1의 RuvC 도메인을 불활성화하는 임의의 돌연변이, 예를 들어 치환 돌연변이, 결실 또는 삽입이 본 개시내용에 따라 사용될 수 있다는 것이 이해되어야 한다.

[0100] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 Cpf1 단백질이다. 일부 실시양태에서, Cpf1 단백질은 Cpf1 니카제 (nCpf1)이다. 일부 실시양태에서, Cpf1 단백질은 뉴클레아제 불활성 Cpf1 (dCpf1)이다. 일부 실시양태에서, Cpf1, nCpf1, 또는 dCpf1은 서열식별번호: 9-24 중 어느 하나와 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, dCpf1은 서열식별번호: 9-16 중 어느 하나와 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 포함하고, 서열식별번호: 9에서 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A 또는 D917A/ E1006A/D1255A에 상응하는 돌연변이를 포함한다. 일부 실시양태에서, dCpf1 단백질은 서열식별번호: 9-16 중 어느 하나의 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 Cpf1도 또한 본 개시내용에 따라 사용될 수 있다는 것이 인지되어야 한다.

[0101] 야생형 프란시셀라 노비시다 Cpf1 (서열식별번호: 9) (D917, E1006, 및 D1255는 볼드체 및 밑줄표시됨)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK

[0102]

RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 9)

[0103]

[0104] 프란시셀라 노비시다 Cpf1 D917A (서열식별번호: 10) (A917, E1006, 및 D1255는 볼드체 및 밑줄표시됨)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF

[0105]

ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 10)

[0106]

[0107] 프란시셀라 노비시다 Cpf1 E1006A (서열식별번호: 11) (D917, A1006, 및 D1255는 볼드체 및 밑줄표시됨)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 11)

[0108]

[0109] 프란시셀라 노비시다 Cpf1 D1255A (서열식별번호: 12) (D917, E1006, 및 A1255는 볼드체 및 밑줄표시됨)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 12)

[0110]

[0111] 프란시셀라 노비시다 Cpf1 D917A/E1006A (서열식별번호: 13) (A917, A1006, 및 D1255는 볼드체 및 밑줄표시됨)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY

[0112]

FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 13)

[0113]

[0114] 프란시셀라 노비시다 Cpf1 D917A/D1255A (서열식별번호: 14) (A917, E1006, 및 A1255는 볼드체 및 밑줄표시됨)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
[0115]
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 14)

[0116]

[0117] 프란시셀라 노비시다 Cpf1 E1006A/D1255A (서열식별번호: 15) (D917, A1006, 및 A1255는 볼드체 및 밑줄표시됨)

[0118] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 15)

[0119]

[0120] 프란시셀라 노비시다 Cpf1 D917A/E1006A/D1255A (서열식별번호: 16) (A917, A1006, 및 A1255는 볼드체 및 밑줄표시됨)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQ
FFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKF
KNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTY
FKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD
LAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSF
YEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIG
TAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQC
RFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQ
TNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYS
DEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKEN
KGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTF
ENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYK
LNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHC
PITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFN
IIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY
QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDK
ICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPT
KELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLI
SPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVI
KNEEYFEFVQNRNN (SEQ ID NO: 16)

[0121]

[0122] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질은 아시다미노코쿠스(*Acidaminococcus*) 종 (AsCpf1)으로부터의 Cpf1 단백질이다. 아시다미노코쿠스 종으로부터의 Cpf1 단백질이 이전에 기재되었고, 통상의 기술자에게 명백할 것이다. 예시적인 아시다미노코쿠스 Cpf1 단백질 (AsCpf1)은, 비제한적으로, 본원에 제공된 AsCpf1 단백질 중 임의의 것을 포함한다.

[0123] 야생형 AsCpf1- 잔기 R912는 볼드체로 밑줄표시되어 제시되고, 잔기 661-667은 이탤릭체로 밑줄표시되어 제시된다.

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD
KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYA*KKTGDQK*GYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLG
EYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWT
GLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLY
QELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAA
NSPSKFNQRVNAYLKEHPETPIIGIDRGE**R**NLIYITVIDSTGKILEQRSLNTIQQFDYQKK
LDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFK
SKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGT
QSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILH
FKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYR
DLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATG
EDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKL
QNGISNQDWLAYIQELRN (SEQ ID NO: 17)

[0124]

[0125] AsCpf1(R912A)- 잔기 A912는 볼드체로 밑줄표시되어 제시되고, 잔기 661-667은 이탤릭체로 밑줄표시되어 제시된다.

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPII
DRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIG
RTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFS
GFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGI
FVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDET
AHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALF
NELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSL
KHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQL
DSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVE
KFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYD
LNNPEKEPKKFQTAYA*KKTGDQK*GYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSS
QYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNL
HTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKT
PIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPIT
LNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGE**A**NLIYITVIDSTGKILEQRSLNTIQQ
FDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLE
NLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFT
SFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDV
KTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIEN
HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQM
RNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHL
KESKDLKLQNGISNQDWLAYIQELRN (SEQ ID NO: 19)

[0126]

[0127] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질은 라크노스피라세아에(*Lachnospiraceae*) 종 (LbCpf1)으로부터의 Cpf1 단백질이다. 라크노스피라세아에 종으로부터의 Cpf1 단백질이 이전에 기재되었고, 통상의 기술자에게 명백할 것이다. 예시적인 라크노스피라세아에 Cpf1 단백질 (LbCpf1)은, 비제한적으로, 본원에 제공된 LbCpf1 단백질 중 임의의 것을 포함한다.

[0128] 일부 실시양태에서, LbCpf1은 니카제이다. 일부 실시양태에서, LbCpf1 니카제는 서열식별번호: 18과 관련하여 R836X 돌연변이체를 포함하며, 여기서 X는 R을 제외한 임의의 아미노산이다. 일부 실시양태에서, LbCpf1 니카제는 서열식별번호: 18과 관련하여 R836A 돌연변이체를 포함한다. 일부 실시양태에서, LbCpf1은 뉴클레아제 불활성 LbCpf1 (dLbCpf1)이다. 일부 실시양태에서, dLbCpf1은 서열식별번호: 18과 관련하여 D832X 돌연변이체를 포함하며, 여기서 X는 D를 제외한 임의의 아미노산이다. 일부 실시양태에서, dLbCpf1은 서열식별번호: 18과 관련하여 D832A 돌연변이체를 포함한다. 추가의 dCpf1 단백질은 관련 기술분야, 예를 들어, 문헌 [Li et al. "Base editing with a Cpf1-cytidine deaminase fusion" Nature Biotechnology; March 2018 DOI: 10.1038/nbt.4102]에 기재되어 있으며; 그의 전체 내용은 본원에 참조로 포함된다. 일부 실시양태에서, dCpf1은 문헌 [Li et al.]에 기재된 Cpf1의 점 돌연변이 D832A, E1006A, D1125A 중 1, 2 또는 3개를 포함한다.

[0129] 야생형 LbCpf1 - 잔기 R836 및 R1138은 볼드체로 밑줄표시되어 제시된다.

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT
SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGE**R**NLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS
LLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQM**R**NSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH (SEQ ID NO: 18)

[0130]

[0131] LbCpf1 (R836A)- 잔기 A836은 볼드체로 밑줄표시되어 제시된다.

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT
SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGE**A**NLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS
LLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH (SEQ ID NO: 20)

[0132]

[0133] LbCpf1 (R1138A)- 잔기 A1138은 볼드체로 밑줄표시되어 제시된다.

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRY
YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYK
SLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCIN
ENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV
YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYT
SDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEK
LKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKA
FFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQF
MGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKL
LPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS
ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKL
YMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEEL
VVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTE
VRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHS
LLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFK
NSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMST
QNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD
YKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGIN
YQQGDIRALLCEQSDKAFYSSFMALMSLMLQM**A**NSITGRTDVDFLISPVKNSDGIFYD
SRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLE
YAQTSVKH (SEQ ID NO: 21)

[0134]

[0135] 일부 실시양태에서, Cpf1 단백질은 손상된 Cpf1 단백질이다. 본원에 사용된 "손상된 Cpf1" 단백질은 야생형 Cpf1 단백질과 비교하여 감소된 뉴클레아제 활성을 갖는 Cpf1 단백질이다. 일부 실시양태에서, 손상된 Cpf1 단백질은 비-표적 가닥보다 표적 가닥을 더 효율적으로 우선적으로 절단한다. 예를 들어, Cpf1 단백질은 편집될 뉴클레오티드가 상주하는 듀플렉싱된 핵산 분자의 가닥을 우선적으로 절단한다. 일부 실시양태에서, 손상된 Cpf1 단백질은 표적 가닥보다 비-표적 가닥을 더 효율적으로 우선적으로 절단한다. 예를 들어, Cpf1 단백질은 편집될 뉴클레오티드가 상주하지 않는 듀플렉싱된 핵산 분자의 가닥을 우선적으로 절단한다. 일부 실시양태에서, 손상된 Cpf1 단백질은 그것이 비-표적 가닥을 절단하는 것보다 표적 가닥을 적어도 5% 더 효율적으로 우선적으로 절단한다. 일부 실시양태에서, 손상된 Cpf1 단백질은 그것이 비-표적 가닥을 절단하는 것보다 표적 가닥을 적어도 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, 90%, 또는 적어도 100% 더 효율적으로 우선적으로 절단한다.

[0136] 일부 실시양태에서, 손상된 Cpf1 단백질은 비-자연 발생 Cpf1 단백질이다. 일부 실시양태에서, 손상된 Cpf1 단백질은 야생형 Cpf1 단백질에 비해 1개 이상의 돌연변이를 포함한다. 일부 실시양태에서, 손상된 Cpf1 단백질은 야생형 Cpf1 단백질에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개의 돌연변이를 포함한다. 일부 실시양태에서, 손상된 Cpf1 단백질은 서열식별번호: 18에, 또는 또 다른 Cpf1 단백질 내의 상응하는 아미노산에 제시된 바와 같은 R836A 돌연변이를 포함한다. 서열식별번호: 18의 R836A에 대한 상동 잔기 (예를 들어, 상응하는 아미노산)를 포함하는 Cpf1이 또한 유사한 결과를 달성하기 위해 돌연변이될 수 있다는 것이 인지되어야 한다. 일부 실시양태에서, 손상된 Cpf1 단백질은 서열식별번호: 18에, 또는 또 다른 Cpf1 단백질 내의 상응하는 아미노산에 제시된 바와 같은 R1138A 돌연변이를 포함한다. 일부 실시양태에서, 손상된 Cpf1 단백질은 서열식별번호: 17에, 또는 또 다른 Cpf1 단백질 내의 상응하는 아미노산에 제시된 바와 같은 R912A 돌연변이를 포함한다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 서열식별번호: 18 (LbCpf1)의 잔기 R836 및 서열식별번호: 17 (AsCpf1)의 잔기 R912는 상응하는 (예를 들어, 상동) 잔기의 예이다. 예를 들어, 서열식별번호: 17 및 18 사이의 정렬 부분은 R912 및 R836이 상응하는 잔기임을 보여준다.

```
AsCpf1      YQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQ--
LbCpf1      KCPKN-IFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINN
                *   *:* .*.. **.. :  :**********:**.*:*..*:*:** *** *
```

[0137]

[0138] 일부 실시양태에서, 본원에 제공된 Cpf1 단백질 중 임의의 것은 1개 이상의 아미노산 결실을 포함한다. 일부 실시양태에서, 본원에 제공된 Cpf1 단백질 중 임의의 것은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개의 아미노산 결실을 포함한다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, Cpf1에 융합된 데아미나제 (예를 들어, APOBEC)의 기능을 폐쇄할 수 있는, AsCpf1 (서열식별번호: 17)의 잔기 661-667을 포함하는 Cpf1 내의 나선 영역이 존재한다. 이러한 영역은 아미노산 서열 KKTGDQK를 포함한다. 따라서, 본 개시내용의 측면은 Cpf1에서의 이러한 나선 영역을 방해하는 돌연변이 (예를 들어, 결실)를 포함하는 Cpf1 단백질을 제공한다. 일부 실시양태에서, Cpf1 단백질은 서열식별번호: 17에서 하기 잔기 중 1개 이상의 결실, 또는 또 다른 Cpf1 단백질에서 1개 이상의 상응하는 결실을 포함한다: K661, K662, T663, G664, D665, Q666 및 K667. 일부 실시양태에서, Cpf1 단백질은 서열식별번호: 17에서 T663 및 D665 결실, 또는 또 다른 Cpf1 단백질에서 상응하는 결실을 포함한다. 일부 실시양태에서, Cpf1 단백질은 서열식별번호: 17에서 K662, T663, D665, 및 Q666 결실, 또는 또 다른 Cpf1 단백질에서 상응하는 결실을 포함한다. 일부 실시양태에서, Cpf1 단백질은 서열식별번호: 17에서 K661, K662, T663, D665, Q666 및 K667 결실, 또는 또 다른 Cpf1 단백질에서 상응하는 결실을 포함한다.

[0139] AsCpf1 (T663 및 D665 결실됨)

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD

[0140]

KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYAKKGQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEY
YAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGL
FSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQE
LYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANS
PSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD
NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSK
RTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQS
GFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFK
MNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDL
YPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGED
YINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQN
GISNQDWLAYIQELRN (SEQ ID NO: 22)

[0141]

[0142] AsCpf1 (K662, T663, D665, 및 Q666 결실됨)

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD
KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYAKGKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYA
ELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFS
PENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELY
DYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPS
KFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKR
TGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQS
GFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFK
MNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDL
YPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGED
YINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQN
GISNQDWLAYIQELRN (SEQ ID NO: 23)

[0143]

[0144] AsCpf1 (K661, K662, T663, D665, Q666, 및 K667 결실됨)

TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTD
AINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK
NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEE
VFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLP
HRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTMLKKQEEKEILKSQLDSLLGLY
HLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNF
QMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFD
KMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK
EPKKFQTAYAGGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAEL
NPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPE
NLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYD
YVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKF
NQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNRE
KERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTG
IAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFL
FYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMN
RNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYP
ANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYI
NSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGIS
NQDWLAYIQELRN (SEQ ID NO: 24)

[0145]

[0146] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 표적 서열에서 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-가이드 엔도뉴클레아제이다. NgAgo는 ~24개 뉴클레오티드 (gDNA) 길이의 5'-인산화된 ssDNA에 결합하여 그것을 그의 표적 부위로 가이드하고, gDNA 부위에서 DNA 이중-가닥 파괴를 만든다. 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 용도는 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 있으며, 이들 각각은 본원에 참조로 포함된다. 나트로노박테리움 그레고리 아르고노트의 서열은 서열식별번호: 25에 제공된다.

[0147] 일부 실시양태에서, napDNAbp는 아르고노트 단백질이다. 일부 실시양태에서, napDNAbp는 자연 발생 아르고노트 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 아르고노트 단백질이다. 일부 실시양태에서, napDNAbp는 서열식별번호: 25 중 어느 하나와 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 서열식별번호: 25 중 어느 하나의 아미노산 서열을 포함한다.

[0148] 야생형 나트로노박테리움 그레고리 아르고노트 (서열식별번호: 25)

MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLW
KNTTPKDVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDED
ETFAGGEPLDHHLDDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLTATD
GAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLDCGVRV
EADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDEFDL
HERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIV
WGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRRQGHGD
DAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAFAERLDPVRL
NGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEV
AVVLPEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDP
SEVDAAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRFRDAKIFYT
RNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATAVY
KDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDL
DPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFG
APEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSHIQVHNSTARLPITTAYA
DQASTHATKGYLVQTGAFESNVGFL (SEQ ID NO: 25)

[0149]

[0150] 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 있으며, 이는 본원에 참조로 포함된다. 일부 실시양태에서, napDNAbp는 마리니토가 피에조필라(*Marinitoga piezophila*) 아르고노트 (MpAgo) 단백질이다. CRISPR-회합된 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. 이러한 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, 예를 들어 데아미나제 (예를 들어, 시티딘 데아미나제)를 표적 핵산 (예를 들어, ssRNA)으로 가이드하기 위해, 본원에 기재된 융합 단백질 (예를 들어, 염기 편집제) 중 임의의 것에서 사용될 수 있다는 것이 인지되어야 한다.

[0151] 일부 실시양태에서, 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 미생물 CRISPR-Cas 시스템의 단일 이펙터이다. 미생물 CRISPR-Cas 시스템의 단일 이펙터는, 비제한적으로, Cas9, Cpf1, C2c1, C2c2, 및 C2c3을 포함한다. 전형적으로, 미생물 CRISPR-Cas 시스템은 부류 1 및 부류 2 시스템으로 나뉜다. 부류 1 시스템은 멀티 서브유닛 이펙터 복합체를 갖는 반면에, 부류 2 시스템은 단일 단백질 이펙터를 갖는다. Cas9 및 Cpf1은 부류 2 이펙터이다. Cas9 및 Cpf1 이외에, 3종의 별개의 부류 2 CRISPR-Cas 시스템 (C2c1, C2c2, 및 C2c3)은 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. 상기 시스템 중 2종인 C2c1 및 C2c3의 이펙터는 Cpf1과 관련된 RuvC-유사 엔도뉴클레아제 도메인을 함유한다. 제3 시스템인 C2c2는 2개의 예측된 HEPN RNase 도메인을 갖는 이펙터를 함유한다. 성숙 CRISPR RNA의 생산은, C2c1에 의한 CRISPR RNA의 생산과 달리 tracrRNA-비의존성이다. C2c1은 DNA 절단을 위해 CRISPR RNA 및 tracrRNA 둘 다에 의존한다. 박테리아 C2c2는 그의 RNA-활성화된 단일-가닥 RNA 분해 활성과 구분되는, CRISPR RNA 성숙을 위한 고유한 RNase 활성을 보유하는 것으로 밝혀졌다. 이들 RNase 기능은 서로 상이하고, Cpf1의 CRISPR RNA-프로세싱 거동과 상이하다. 예를 들어, 문헌 [East-Seletsky, et al., "Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection", Nature, 2016 Oct 13;538(7624):270-273]을 참조하고, 그의 전체 내용은 본원에 참조로 포함된다. 렙토트리키아 샤히이(*Leptotrichia shahii*)의 C2c2의 시험관내 생화학적 분석은 C2c2가 단일 CRISPR RNA에 의해 가이드되고 상보적

프로토스페이서를 보유하는 ssRNA 표적을 절단하도록 프로그램화될 수 있다는 것을 밝혀냈다. 2개의 보존된 HEPN 도메인 내의 촉매 잔기는 절단을 매개한다. 촉매 잔기 내의 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.

[0152] 알리시클로바실루스 아시도테레스트리스(*Alicyclobaccillus acidoterrastris*) C2c1 (AacC2c1)의 결정 구조는 키메라 단일-분자 가이드 RNA (sgRNA)와의 복합체로 보고되었다. 예를 들어, 본원에 참조로 포함된 문헌 [Liu et al., "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", Mol. Cell, 2017 Jan 19;65(2):310-322]을 참조한다. 또한 3원 복합체로서 표적 DNA에 결합된 알리시클로바실루스 아시도테레스트리스 C2c1에 대한 결정 구조가 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 위치된 채로 포획되었고, 이때 C2c1-매개 절단은 표적 DNA의 엇갈린 7개-뉴클레오티드 파괴를 일으켰다. C2c1 3원 복합체와 이전에 확인된 Cas9 및 Cpf1 대응부 사이의 구조적 비교는 CRISPR-Cas9 시스템에 의해 사용된 메카니즘의 다양성을 입증한다.

[0153] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 C2c1, C2c2 또는 C2c3 단백질이다. 일부 실시양태에서, napDNAbp는 C2c1 단백질이다. 일부 실시양태에서, napDNAbp는 C2c2 단백질이다. 일부 실시양태에서, napDNAbp는 C2c3 단백질이다. 일부 실시양태에서, napDNAbp는 자연 발생 C2c1, C2c2 또는 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 C2c1, C2c2 또는 C2c3 단백질이다. 일부 실시양태에서, napDNAbp는 서열식별번호: 26-28 중 어느 하나와 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 서열식별번호: 26-28 중 어느 하나의 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 C2c1, C2c2, 또는 C2c3이 또한 본 개시내용에 따라 사용될 수 있다는 것이 인지되어야 한다.

[0154] C2c1 (uniprot.org/uniprot/T0D7A2#)

[0155] sp|T0D7A2|C2C1_ALIAG CRISPR-회합된 엔도뉴클레아제 C2c1 OS=알리시클로바실루스 아시도테레스트리스 (균주 ATCC 49025 / DSM 3922 / CIP 106132 / NCIMB 13137 / GD3B) GN=c2c1 PE=1 SV=1

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGD
GEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGA
KGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETR
KSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQA
IERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASP
GLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGS
HDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWT
RFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDN
LLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSV
RVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSG
LRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLK
LPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVD
AANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDW
RKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGS
RFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEE
LSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDART
GAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPF
SAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTA
DSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEADEAREKSVVLM
RDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENTGDI (SEQ ID
NO: 26)

[0156]

[0157] C2c2 (uniprot.org/uniprot/P0DOC6)

[0158] >sp|P0DOC6|C2C2_LEPSD CRISPR-회합된 엔도뉴클레아제 C2c2 OS=렙토트리키아 샤히이 (균주 DSM 19757 / CCUG 47503 / CIP 107916 / JCM 16776 / LB37) GN=c2c2 PE=1 SV=1

MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKID
NNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIE
AYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDC
SIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEEHLREKLLKDD
KIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVE
DIADFVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKK
DKIVKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNF
DSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKIL
KRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFASTNMELN
KIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTK
IGTNERNRILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNII
TKINDIKISEENNNDIKYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKE
LYKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIENYYKNAQISASKGNNKAIKK
YQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTIVI
NDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECIT
ENWNLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLE
KKLEKIVIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKSKIL
CRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLFLNIGNPN
FDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEKYI
KKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLIDINWKL
AIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDE
ESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFYIVRNPFADYSIAEQIDRVSNLLS
YSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSVLELES
YNSDYIKNLIIELLTKIENTNDTL (SEQ ID NO: 27)

[0159]

[0160] C2c3, >CEPX01008730.1 해양 메타게놈 게놈 어셈블리 TARA_037_MES_0.1-0.22, 콘티그 TARA_037_MES_0.1-0.22_스캐폴드22115_1, 전체 게놈 샷건 서열로부터 번역됨.

MRSNYHGGRNARQWRKQISGLARRTKETVFTYKFPLETDAAEIDFDKAVQ
TYGIAEGVGHGSLIGLVCAFHLSGFRLFSKAGEAMAFRNRSRYPTDAFAEKLSAIMGI
QLPTLSPEGLDLIFQSPPRSRDGIAPVWSENEVRNRLYTNWTGRGPANKPDEHLLEIAG
EIAKQVFPKFGGWDDLASDPDKALAAADKYFQSQGDFPSIASLPAAIMLSPANSTVDF
EGDYIAIDPAAETLLHQAVSRCAARLGRERPDLDQNKGPFVSSLQDALVSSQNNGLS
WLFGVGFQHWKEKSPKELIDEYKVPADQHGAVTQVKSFVDAIPLNPLFDTTHYGEFR
ASVAGKVRSWVANYWKRLLDLKSLLATTEFTLPESISDPKAVSLFSGLLVDPQGLKK
VADSLPARLVSAEEAIDRLMGVGIPTAADIAQVERVADEIGAFIGQVQQFNNQVKQKL
ENLQDADDEEFLKGLKIELPSGDKEPPAINRISGGAPDAAAEISELEEKLQRLLDARSE
HFQTISEWAEENAVTLDPIAAMVELERLRLAERGATGDPEEYALRLLLQRIGRLANRV
SPVSAGSIRELLKPVFMEEREFNLFFHNRLGSLYRSPYSTSRHQPFSIDVGKAKAIDWIA
GLDQISSDIEKALSGAGEALGDQLRDWINLAGFAISQRLRGLPDTVPNALAQVRCPDD
VRIPPLLAMLLEEDDIARDVCLKAFNLYVSAINGCLFGALREGFIVRTRFQRIGTDQIH
YVPKDKAWEYPDRLNTAKGPINAAVSSDWIEKDGAVIKPVETVRNLSSTGFAGAGVS
EYLVQAPHDWYTPLDLRDVAHLVTGLPVEKNITKLKRLTNRTAFRMVGASSFKTHLD
SVLLSDKIKLGDFTIIIDQHYRQSVTYGGKVKISYEPERLQVEAAVPVVDTRDRTVPEP
DTLFDHIVAIDLGERSVGFAVFDIKSCLRTGEVKPIHDNNGNPVVGTVAVPSIRRLMK
AVRSHRRRRQPNQKVNQTYSTALQNYRENVIGDVCNRIDTLMERYNAFPVLEFQIKN
FQAGAKQLEIVYGS (SEQ ID NO: 28)

[0161]

[0162] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 단세포 원핵 미생물의 역 및 계를 구성하는 고세균 (예를 들어 나노고세균)으로부터의 Cas9이다. 일부 실시양태에서, napDNAbp는 예를 들어 본원에 참조로 포함된 문헌 [Burstein et al., "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21]에 기재되어 있는 CasX 또는 CasY이다. 게놈-분해 메타게놈학을 사용하여, 생명체의 고세균 역에서 최초 보고된 Cas9를 포함한, 수많은 CRISPR-Cas 시스템이 확인되었다. 이러한 분기 Cas9 단백질은 나노고세균에서 활성 CRISPR-Cas 시스템의 일부로서 발견되었다. 박테리아에서, 2종의 이전에 공지되지 않은 시스템인 CRISPR-CasX 및 CRISPR-CasY가 발견되었으며, 이는 이제껏 발견된 시스템 중 가장 소형이다. 일부 실시양태에서, Cas9는 CasX 또는 CasX의 변이체를 지칭한다. 일부 실시양태에서, Cas9는 CasY 또는 CasY의 변이체를 지칭한다. 다른 RNA-가이드 DNA 결합 단백질은 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)로서 사용될 수 있고 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다.

[0163] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 CasX 또는 CasY 단백질이다. 일부 실시양태에서, napDNAbp는 CasX 단백질이다. 일부 실시양태에서, napDNAbp는 CasY 단백질이다. 일부 실시양태에서, napDNAbp는 자연 발생 CasX 또는 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 CasX 또는 CasY 단백질이다. 일부 실시양태에서, napDNAbp는 서열식별번호: 29-31 중 어느 하나와 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 서열식별번호: 29-31 중 어느 하나의 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 CasX 및 CasY가 또한 본 개시내용에 따라 사용될 수 있다는 것이 인지되어야 한다.

[0164] CasX (uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53)

[0165] >tr|F0NN87|F0NN87_SULIH CRISPR-회합된 Casx 단백질 OS=술폴로부스 이슬란디쿠스 (균주 HVE10/4) GN=SiH_0402 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVK
ECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVS
EGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRI
YTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANY
IYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG (SEQ ID NO: 29)

[0166]

[0167] >tr|F0NH53|F0NH53_SULIR CRISPR 회합된 단백질, Casx OS=술폴로부스 이슬란디쿠스 (균주 REY15A) GN=SiRe_0771 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVK
ECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKV
SEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVS
IYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLAN
YIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG (SEQ ID NO: 30)

[0168]

[0169] CasY (ncbi.nlm.nih.gov/protein/APG80656.1)

[0170] >APG80656.1 CRISPR-회합된 단백질 CasY [배양되지 않은 파르쿠박테리아 군 박테리아]

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIV
SAINDDYVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLK
NVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFK
AEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNP
LNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFS
NFLGEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKF
DNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTK
EEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIK
ERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKK
YKNAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDK
WKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALA
RELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTVK
DFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELL
YIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYE
LTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQF
LEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTI
FPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKL
DQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKV
YATLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVD
ETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMRGNS
CLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI
(SEQ ID NO: 31)

[0171]

[0172] 핵염기 편집제의 Cas9 도메인

[0173] 비제한적, 예시적인 Cas9 도메인이 본원에 제공된다. Cas9 도메인은 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, 또는 Cas9 니카제일 수 있다. 일부 실시양태에서, Cas9 도메인은 뉴클레아제 활성 도메인이다. 예를 들어, Cas9 도메인은 듀플렉싱된 핵산의 양쪽 가닥 (예를 들어, 듀플렉싱된 DNA 분자의 양쪽 가닥)을 절단하는 Cas9 도메인일 수 있다. 일부 실시양태에서, Cas9 도메인은 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나를 포함한다. 일부 실시양태에서 Cas9 도메인은 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나와 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% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나와 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 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나와 비교하여 적어도 10, 적어도 15, 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 150, 적어도 200, 적어도 250, 적어도 300, 적어도 350, 적어도 400, 적어도 500, 적어도 600, 적어도 700, 적어도 800, 적어도 900, 적어도 1000, 적어도 1100, 또는 적어도 1200개의 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0174] 일부 실시양태에서, Cas9 도메인은 뉴클레아제-불활성 Cas9 도메인 (dCas9)이다. 예를 들어, dCas9 도메인은 듀플렉싱된 핵산 분자의 어느 하나의 가닥을 절단하지 않고 듀플렉싱된 핵산 분자에 (예를 들어, gRNA 분자를 통해) 결합할 수 있다. 일부 실시양태에서, 뉴클레아제-불활성 dCas9 도메인은 서열식별번호: 6에 제시된 아미노산 서열의 D10X 돌연변이 및 H840X 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산 변화이다.

일부 실시양태에서, 뉴클레아제-불활성 dCas9 도메인은 서열식별번호: 6에 제시된 아미노산 서열의 D10A 돌연변이 및 H840A 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 한 예로서, 뉴클레아제-불활성 Cas9 도메인은 서열식별번호: 32에 제시된 아미노산 서열을 포함한다 (클로닝 벡터 pPlatTET-gRNA2, 수탁 번호 BAV54124).

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN
PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 32; 예를 들어, 전체 내용이 본원에 참조로 포함된 문헌 [Qi et al., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013; 152(5):1173-83] 참조).

[0175]

[0176] 추가의 적합한 뉴클레아제-불활성 dCas9 도메인은 본 개시내용 및 관련 기술분야의 지식에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이며, 이는 본 개시내용의 범주 내에 있다. 이러한 추가의 예시적인 적합한 뉴클레아제-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하나 이에 제한되지는 않는다 (예를 들어, 전체 내용이 본원에 참조로 포함된 문헌 [Prashant et al., CAS9 transcriptional activators for target specificity screening 및 paired nickases for cooperative genome engineering. Nature Biotechnology. 2013; 31(9): 833-838] 참조). 일부 실시양태에서 dCas9 도메인은 본원에 제공된 dCas9 도메인 중 어느 하나와 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% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나와 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 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나와 비교하여 적어도 10, 적어도 15, 적어도 20, 적어도 30, at leat 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 150, 적어도 200, 적어도 250, 적어도 300, 적어도 350, 적어도 400, 적어도 500, 적어도 600, 적어도 700, 적어도 800, 적어도 900, 적어도 1000, 적어도 1100, 또는 적어도 1200개의 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0177] 일부 실시양태에서, Cas9 도메인은 Cas9 니카제이다. Cas9 니카제는 듀플렉싱된 핵산 분자 (예를 들어, 듀플렉싱된 DNA 분자)의 단지 1개의 가닥만을 절단할 수 있는 Cas9 단백질일 수 있다. 일부 실시양태에서, Cas9 니카제는 듀플렉스 핵산 분자의 표적 가닥을 절단하며, 이는 Cas9 니카제가 Cas9에 결합된 gRNA (예를 들어, sgRNA)에 염기 쌍형성된 (그에 상보적인) 가닥을 절단한다는 것을 의미한다. 일부 실시양태에서, Cas9 니카제는 D10A 돌연변이를 포함하며, 서열식별번호: 6의 위치 840에 히스티딘, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 돌연변이를 갖는다. 예를 들어, Cas9 니카제는 서열식별번호: 8에 제시된 바와 같은 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, Cas9 니카제는 듀플렉스 핵산 분자의 비-표적, 비-염기-편집된 가닥을 절단하며, 이는 Cas9 니카제가 Cas9에 결합된 gRNA (예를 들어, sgRNA)에 염기 쌍형성되지 않은 가닥을 절단한다는 것을 의미한다. 일부 실시양태에서, Cas9 니카제는 H840A 돌연변이를 포함하며, 서열식별번호: 6의 위치 10에 아스파르트산 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 갖는다. 일부 실시양태에서, Cas9 니카제는 본원에 제공된 Cas9 니카제 중 어느 하나와 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% 동일한 아미노산 서열을 포함한다. 추가의 적합한 Cas9 니카제는 본 개시내용 및 관련 기술분야의 지식에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이며, 본 개시내용의 범위 내에 있다.

[0178] 감소된 PAM 배타성을 갖는 Cas9 도메인

[0179] 본 개시내용의 일부 측면은 상이한 PAM 특이성을 갖는 Cas9 도메인을 제공한다. 전형적으로, Cas9 단백질, 예컨대 에스. 피오게네스로부터의 Cas9 (spCas9)는 특정한 핵산 영역에 결합하는 정규 NGG PAM 서열을 요구한다. 이는 게놈 내에서 목적하는 염기를 편집하는 능력을 제한할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집 융합 단백질은 예를 들어 표적 염기가 PAM의 대략 15 염기 상류인 4개의 염기 영역 (예를 들어, "탈아미노화 윈도우") 내에 놓이는 정확한 위치에 놓일 필요가 있을 수 있다. 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 참조하며, 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 있으며, 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 있으며; 각각의 전체 내용은 본원에 참조로 포함된다.

[0180] 일부 실시양태에서, Cas9 도메인은 스타필로코쿠스 아우레우스로부터의 Cas9 도메인 (SaCas9)이다. 일부 실시양태에서, SaCas9 도메인은 뉴클레아제 활성 SaCas9, 뉴클레아제 불활성 SaCas9 (SaCas9d), 또는 SaCas9 니카제 (SaCas9n)이다. 일부 실시양태에서, SaCas9는 아미노산 서열 서열식별번호: 33을 포함한다. 일부 실시양태에서, SaCas9는 서열식별번호: 33의 N579X 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 N을 제외한 임의의 아미노산이다. 일부 실시양태에서, SaCas9는 서열식별번호: 33의 N579A 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 비-정규 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시양태에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 NNGRRT PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 실시양태에서, SaCas9 도메인은 서열식별번호: 33의 E781X, N967X, 및 R1014X 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SaCas9 도메인은 서열식별번호: 33의 E781K, N967K, 및 R1014H 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 1개 이상의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SaCas9 도메인은 서열식별번호: 33의 E781K, N967K, 또는 R1014H 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다.

[0181] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 33-36 중 어느 하

나와 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% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 33-36 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 33-36 중 어느 하나의 아미노산 서열로 이루어진다.

[0182] 예시적인 SaCas9 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 33)

[0183]

[0184] 밑줄표시되어 있고, 볼드체인 서열식별번호: 33의 잔기 N579는 돌연변이되어 (예를 들어, A579로) SaCas9 니카제를 생성할 수 있다.

[0185] 예시적인 SaCas9d 서열

KRNYILGL**A**IGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 34)

[0186]

[0187] 밑줄표시되어 있고, 볼드체인 서열식별번호: 34의 잔기 D10은 돌연변이되어 (예를 들어, A10으로) 뉴클레아제 불활성 SaCas9d를 생성할 수 있다.

[0188] 예시적인 SaCas9n 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 35).

[0189]

[0190]

[0191] 서열식별번호: 33의 N579로부터 돌연변이되어 SaCas9 니카제를 생성할 수 있는 서열식별번호: 35의 잔기 A579는

밑줄표시되어 있고, 볼드체이다.

[0192] 예시적인 SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRG
VHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSD
YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEM
LMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFK
QKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLD
QIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHT
NDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLP
NDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPF
QYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD
TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAED
ALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHI
KDFKDYKYSHRVDKKPNR*K*LINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLI
NKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPV
IKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDV
IKKENYYEVNSKCYEEAKKLKKISNQAEFIASFY*K*NDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPP*H*IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKK
G (SEQ ID NO: 36).

[0193]

[0194] 서열식별번호: 36의 N579로부터 돌연변이되어 SaCas9 니카제를 생성할 수 있는 서열식별번호: 36의 잔기 A579는 밑줄표시되어 있고, 볼드체이다. 서열식별번호: 36의 E781, N967, 및 R1014로부터 돌연변이되어 SaKKH Cas9를 생성할 수 있는 서열식별번호: 36의 잔기 K781, K967, 및 H1014는 밑줄표시되어 있고, 이탤릭체이다.

[0195] 일부 실시양태에서, Cas9 도메인은 스트렙토코쿠스 피오게네스로부터의 Cas9 도메인 (SpCas9)이다. 일부 실시양태에서, SpCas9 도메인은 뉴클레아제 활성 SpCas9, 뉴클레아제 불활성 SpCas9 (SpCas9d), 또는 SpCas9 니카제 (SpCas9n)이다. 일부 실시양태에서, SpCas9는 아미노산 서열 서열식별번호: 37을 포함한다. 일부 실시양태에서, SpCas9는 서열식별번호: 37의 D9X 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 D를 제외한 임의의 아미노산이다. 일부 실시양태에서, SpCas9는 서열식별번호: 37의 D9A 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인, SpCas9d 도메인, 또는 SpCas9n 도메인은 비-정규 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시양태에서, SpCas9 도메인, SpCas9d 도메인, 또는 SpCas9n 도메인은 NGG, NGA, 또는 NGCG PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134X, R1334X, 및 T1336X 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134E, R1334Q, 및 T1336R 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134E, R1334Q, 및 T1336R 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134X, R1334X, 및 T1336X 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134V, R1334Q, 및 T1336R 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134V, R1334Q, 및 T1336R 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부

실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134X, G1217X, R1334X, 및 T1336X 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134V, G1217R, R1334Q, 및 T1336R 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 37의 D1134V, G1217R, R1334Q, 및 T1336R 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다.

[0196] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 37-41 중 어느 하나와 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% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 37-41 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 37-41 중 어느 하나의 아미노산 서열로 이루어진다.

[0197] 예시적인 SpCas9

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH

[0198]

VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 37)

[0199]

[0200] 예시적인 SpCas9n

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 38)

[0201]

[0202] 예시적인 SpEQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**
Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 39)

[0203]

[0204] 서열식별번호: 39의 D1134, R1334, 및 T1336로부터 돌연변이되어 SpEQR Cas9를 생성할 수 있는 서열식별번호: 39의 잔기 E1134, Q1334, 및 R1336은 밑줄표시되어 있고, 볼드체이다.

[0205] 예시적인 SpVQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**
Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 40)

[0206]

[0207] 서열식별번호: 40의 D1134, R1334, 및 T1336으로부터 돌연변이되어 SpVQR Cas9를 생성할 수 있는 서열식별번호: 40의 잔기 V1134, Q1334, 및 R1336은 밑줄표시되어 있고, 볼드체이다.

[0208] 예시적인 SpVRER Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN
GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED

[0209]

RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKS
VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA**R**
ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**
Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 41)

[0210]

[0211] 서열식별번호: 41의 D1134, G1217, R1334, 및 T1336으로부터 돌연변이되어 SpVRER Cas9를 생성할 수 있는 서열식별번호: 41의 잔기 V1134, R1217, Q1334, 및 R1336은 밑줄표시되어 있고, 볼드체이다.

[0212] 하기는 비-정규 (예를 들어, 비-NGG) PAM 서열을 갖는 핵산 서열에 결합할 수 있는 예시적인 융합 단백질 (예를 들어, 염기 편집 단백질)이다:

[0213] 예시적인 SaBE3 (rAPOBEC1-XTEN-SaCas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPES<u>KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR
GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAAL
LHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRG
SINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWK
DIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKE
IIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINL
ILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVIN
AIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIE
KIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEA**S**
KKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKD
FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFI
TPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKY
SKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF</u>

[0214]

<u>VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIG
VNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKS
KKHPQIIKKG</u>SGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO:
42)

[0215]

[0216] 예시적인 SaKKH-BE3 (rAPOBEC1-XTEN-SaCas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR
GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAAL
LHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRG
SINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWK
DIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKE
IIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINL
ILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVIN
AIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIE
KIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEAS
KKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKD
FINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFI
TPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKD
NDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKY
SKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKF
VTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIG
VNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKS
KKHPQIIKKGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO:
43)

[0217]

[0218] 예시적인 EQR-BE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH
FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF

[0219]

MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSPKKKRKV (SEQ ID NO: 44)

[0220]

[0221] VQR-BE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH
FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSPKKKRKV (SEQ ID NO: 45)

[0222]

[0223] VRER-BE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE
DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH
FLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRG
KSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK
RMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSPKKKRKV (SEQ ID NO: 46)

[0224]

[0225] 고 충실도 염기 편집제

[0226] 본 개시내용의 일부 측면은 고 충실도를 갖는 Cas9 도메인을 포함하는 Cas9 융합 단백질 (예를 들어, 본원에 제공된 융합 단백질 중 임의의 것)을 제공한다. 본 개시내용의 추가의 측면은 야생형 Cas9 도메인과 비교하여 Cas9 도메인 및 DNA의 당-포스페이트 백본 사이의 감소된 정전기적 상호작용을 갖는 Cas9 도메인을 포함하는 Cas9 융합 단백질 (예를 들어, 본원에 제공된 융합 단백질 중 임의의 것)을 제공한다. 일부 실시양태에서, Cas9 도메인 (예를 들어, 야생형 Cas9 도메인)은 Cas9 도메인 및 DNA의 당-포스페이트 백본 사이의 회합을 감소시키는 1개 이상의 돌연변이를 포함한다. 일부 실시양태에서, 본원에 제공된 Cas9 융합 단백질 중 임의의 것은 서열식별번호: 6에 제공된 아미노산 서열의 N497X, R661X, Q695X, 및/또는 Q926X 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 본원에 제공된 Cas9 융합 단백질 중 임의의 것은 서열식별번호: 6에 제공된 아미노산 서열의 N497A, R661A, Q695A, 및/또는 Q926A 돌연변이 중 1개 이상, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, Cas9 도메인은 서열식별번호: 6에 제공된 아미노산 서열의 D10A 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, (예를 들어, 본원에 제공된 융합 단백질 중 임의의 것의) Cas9 도메인은 서열식별번호: 47에 제시된 바와 같은 아미노산 서열을 포함한다. 일부 실시양태에서, 융합 단백질은 서열식별번호: 48에 제시된 바와 같은 아미노산 서열을 포함한다. 고 충실도를 갖는 Cas9 도메인은 관련 기술분야에 공지되어 있으며, 통상의 기술자에게 명백할 것이다. 예를 들어, 고 충실도를 갖는 Cas9 도메인은 문헌 [Kleinstiver, B.P., et al. "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature 529, 490-495 (2016); 및 Slaymaker, I.M., et al. "Rationally engineered Cas9 nucleases with improved specificity." Science 351, 84-88 (2015)]에 기재되어 있으며; 각각의 전체 내용은 본원에 참조로 포함된다.

[0227] 본원에 제공된 염기 편집제, 예를 들어 염기 편집제 2 (BE2) 또는 염기 편집제 3 (BE3)은 본원에 기재된 바와 같은 Cas9 도메인을 변형시켜 고 충실도 염기 편집제, 예를 들어 고 충실도 염기 편집제 2 (HF-BE2) 또는 고 충실도 염기 편집제 3 (HF-BE3)을 생성함으로써 고 충실도 염기 편집제로 전환될 수 있다는 것이 인지되어야 한다. 일부 실시양태에서, 염기 편집제 2 (BE2)는 데아미나제 도메인, dCas9, 및 UGI 도메인을 포함한다. 일부 실시양태에서, 염기 편집제 3 (BE3)은 데아미나제 도메인, nCas9 도메인 및 UGI 도메인을 포함한다.

[0228] 서열식별번호: 6의 Cas9에 대한 돌연변이가 볼드체 및 밑줄표시되어 있는 Cas9 도메인

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG

[0229]

LSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ
KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 47)

[0230]

[0231] HF-BE3

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFT
TERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQI
MTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK
DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD (SEQ ID NO: 48)

[0232]

[0233] Cas9 융합 단백질

[0234] 본원에 개시된 Cas9 도메인 (예를 들어, 뉴클레아제 활성 Cas9 단백질, 뉴클레아제-불활성 dCas9 단백질, 또는 Cas9 니카제 단백질) 중 임의의 것은 제2 단백질에 융합될 수 있으며, 따라서 본원에 제공된 융합 단백질은 본원에 제공된 Cas9 도메인 및 제2 단백질, 또는 "융합 파트너"를 포함한다. 일부 실시양태에서, 제2 단백질은 Cas9 도메인의 N-말단에 융합된다. 그러나, 다른 실시양태에서, 제2 단백질은 Cas9 도메인의 C-말단에 융합된다. 일부 실시양태에서, Cas9 도메인에 융합되는 제2 단백질은 핵산 편집 도메인이다. 일부 실시양태에서, Cas9 도메인 및 핵산 편집 도메인은 링커를 통해 융합되는 반면, 다른 실시양태에서, Cas9 도메인 및 핵산 편집 도메인은 서로에게 직접적으로 융합된다. 일부 실시양태에서, Cas9 도메인 및 핵산 편집 도메인은 임의의 길이 또는 조성의 링커를 통해 융합된다. 예를 들어, 링커는 임의의 길이 및 조성의 결합, 1개 이상의 아미노산, 펩티드, 또는 중합체일 수 있다. 일부 실시양태에서, 링커는 $(GGGS)_n$ (서열식별번호: 613), $(GGGGS)_n$ (서열식별번호: 607), $(G)_n$ (서열식별번호: 608), $(EAAAK)_n$ (서열식별번호: 609), $(GGS)_n$ (서열식별번호: 610), $(SGGS)_n$ (서열식별번호: 606), SGSETPGTSESATPES (서열식별번호: 604), $SGGS(GGS)_n$ (서열식별번호: 612), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605), 또는 $(XP)_n$ (서열식별번호: 611) 모티프, 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이고, X는 임의의 아미노산이다. 일부 실시양태에서, 링커는 $(GGS)_n$ 모티프를 포함하며, 여기서 n은 1, 3, 또는 7이다. 일부 실시양태에서, 링커는 $(GGS)_n$ (서열식별번호: 610) 모티프를 포함하며, 여기서 n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15이다. 일부 실시양태에서, 링커는 아미노산 서열 $SGGS(GGS)_n$ (서열식별번호: 612)을 포함하며, 여기서 n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10이다. 일부 실시양태에서, 링커는 아미노산 서열 $SGGS(GGS)_n$ (서열식별번호: 612)을 포함하며, 여기서 n은 2이다. 일부 실시양태에서, 링커는 SGSETPGTSESATPES의 아미노산 서열 (서열식별번호: 604) (실시예에서 XTEN 링커로도 지칭됨)을 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605) (실시예에서 32개 아미노산 링커로도 지칭됨)를 포함한다. 링커의 길이는 실시예에 예시된 바와 같이 편집되는 염기에 영향을 미칠 수 있다. 예를 들어, 3-아미노산 길이의 링커 (예를 들어, $(GGS)_1$)는 PAM 서열에 대해 2 내지 5, 2 내지 4, 2 내지 3, 3 내지

4개 염기 편집 윈도우를 줄 수 있는 반면, 9-아미노산 링커 (예를 들어, $(GGS)_3$ (서열식별번호: 610))는 PAM 서열에 대해 2 내지 6, 2 내지 5, 2 내지 4, 2 내지 3, 3 내지 6, 3 내지 5, 3 내지 4, 4 내지 6, 4 내지 5, 5 내지 6개 염기 편집 윈도우를 줄 수 있다. 16-아미노산 링커 (예를 들어, XTEN 링커)는 예외적으로 강한 활성으로 PAM 서열에 대해 2 내지 7, 2 내지 6, 2 내지 5, 2 내지 4, 2 내지 3, 3 내지 7, 3 내지 6, 3 내지 5, 3 내지 4, 4 내지 7, 4 내지 6, 4 내지 5, 5 내지 7, 5 내지 6, 6 내지 7개 염기 윈도우를 줄 수 있으며, 21-아미노산 링커 (예를 들어, $(GGS)_7$ (서열식별번호: 610))는 PAM 서열에 대해 3 내지 8, 3 내지 7, 3 내지 6, 3 내지 5, 3 내지 4, 4 내지 8, 4 내지 7, 4 내지 6, 4 내지 5, 5 내지 8, 5 내지 7, 5 내지 6, 6 내지 8, 6 내지 7, 7 내지 8개 염기 편집 윈도우를 줄 수 있다. 링커 길이를 다양화하는 것은 본 개시내용의 dCas9 융합 단백질이 PAM 서열과 상이한 거리의 핵염기를 편집하는 것을 허용한다는 신규한 발견은, PAM 서열이 유전자에서 교정되는 질환-유발 돌연변이에 대한 거리를 다양화하는 것일 수 있기 때문에 상당한 임상적 중요성을 부여한다. 여기서 실시예에 기재된 바와 같은 링커 길이는 제한적인 것을 의미하지 않는 것으로 이해되어야 한다.

[0235] 일부 실시양태에서, 제2 단백질은 효소 도메인을 포함한다. 일부 실시양태에서, 효소 도메인은 핵산 편집 도메인이다. 이러한 핵산 편집 도메인은 비제한적으로 뉴클레아제, 니카제, 레콤비나제, 데아미나제, 메틸트랜스퍼라제, 메틸라제, 아세틸라제, 또는 아세틸트랜스퍼라제일 수 있다. 본 개시내용에 따라 사용될 수 있는 비제한적 예시적인 결합 도메인은 전사 활성인자 도메인 및 전사 리프레서 도메인을 포함한다.

[0236] 데아미나제 도메인

[0237] 일부 실시양태에서, 제2 단백질은 핵산 편집 도메인을 포함한다. 일부 실시양태에서, 핵산 편집 도메인은 C에서 U로의 염기 변화를 촉매할 수 있다. 일부 실시양태에서, 핵산 편집 도메인은 데아미나제 도메인이다. 일부 실시양태에서, 데아미나제는 시티딘 데아미나제 또는 시티딘 데아미나제이다. 일부 실시양태에서, 데아미나제는 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC1 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC2 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3A 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3B 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3C 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3D 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3E 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3F 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3G 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3H 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC4 데아미나제이다. 일부 실시양태에서, 데아미나제는 활성화-유도된 데아미나제 (AID)이다. 일부 실시양태에서, 데아미나제는 척추동물 데아미나제이다. 일부 실시양태에서, 데아미나제는 무척추동물 데아미나제이다. 일부 실시양태에서, 데아미나제는 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트, 또는 마우스 데아미나제이다. 일부 실시양태에서, 데아미나제는 인간 데아미나제이다. 일부 실시양태에서, 데아미나제는 래트 데아미나제, 예를 들어, rAPOBEC1이다. 일부 실시양태에서, 데아미나제는 활성화-유도된 시티딘 데아미나제 (AID)이다. 일부 실시양태에서, 데아미나제는 시티딘 데아미나제 1 (CDA1)이다. 일부 실시양태에서, 데아미나제는 페트로미존 마리누스 (*Petromyzon marinus*) 시티딘 데아미나제 1 (pmCDA1)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G (서열식별번호: 60)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G의 단편 (서열식별번호: 83)이다. 일부 실시양태에서, 데아미나제는 D316R_D317R 돌연변이를 포함하는 인간 APOBEC3G 변이체 (서열식별번호: 82)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G 및 서열식별번호: 60에서의 D316R_D317R 돌연변이에 상응하는 돌연변이를 포함하는 단편 (서열식별번호: 84)이다.

[0238] 일부 실시양태에서, 핵산 편집 도메인은 서열식별번호: 49-84 중 어느 하나의 데아미나제 도메인과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일하다. 일부 실시양태에서, 핵산 편집 도메인은 서열식별번호: 49-84 중 어느 하나의 아미노산 서열을 포함한다.

[0239] 염기 편집제의 편집 윈도우를 조정하는 데아미나제 도메인

[0240] 본 개시내용의 일부 측면은 예를 들어 데아미나제 도메인에서의 점 돌연변이를 생성함으로써 본원에 제공된 융합 단백질 중 임의의 것의 데아미나제 도메인 촉매 활성을 조정하는 것이 융합 단백질 (예를 들어, 염기 편집제)의 프로세싱성에 영향을 미친다는 인식에 기초한다. 예를 들어, 염기 편집 융합 단백질 내의 데아미나제 도메인의 촉매 활성을 감소시키지만 제거하지는 않는 돌연변이는 데아미나제 도메인이 표적 잔기에 인접한 잔기의 탈아미노화를 촉매함으로써, 탈아미노화 윈도우를 좁힐 가능성을 적게 할 수 있다. 탈아미노화 윈도우

를 좁히는 능력은 특이적 표적 잔기의 인접한 잔기의 원치 않는 탈아미노화를 방지할 수 있으며, 이는 오프-타겟 효과를 감소시키거나 방지할 수 있다.

[0241] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 촉매 데아미나제 활성을 감소시킨 데아미나제 도메인 (예를 들어, 시티딘 데아미나제 도메인)을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 적절한 대조군과 비교하여 촉매 데아미나제 활성을 감소시킨 데아미나제 도메인 (예를 들어, 시티딘 데아미나제 도메인)을 포함한다. 예를 들어, 적절한 대조군은 1개 이상의 돌연변이를 데아미나제 내로 도입하기 전의 데아미나제의 데아미나제 활성일 수 있다. 다른 실시양태에서, 적절한 대조군은 야생형 데아미나제일 수 있다. 일부 실시양태에서, 적절한 대조군은 야생형 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 데아미나제이다. 일부 실시양태에서, 적절한 대조군은 APOBEC1 데아미나제, APOBEC2 데아미나제, APOBEC3A 데아미나제, APOBEC3B 데아미나제, APOBEC3C 데아미나제, APOBEC3D 데아미나제, APOBEC3F 데아미나제, APOBEC3G 데아미나제, 또는 APOBEC3H 데아미나제이다. 일부 실시양태에서, 적절한 대조군은 활성화 유도된 데아미나제 (AID)이다. 일부 실시양태에서, 적절한 대조군은 페트로미존 마리누스로부터의 시티딘 데아미나제 1 (pmCDA1)이다. 일부 실시양태에서, 데아미나제 도메인은 적절한 대조군과 비교하여 적어도 1%, 적어도 5%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 또는 적어도 95% 더 적은 촉매 데아미나제 활성을 갖는 데아미나제 도메인일 수 있다.

[0242] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 H121X, H122X, R126X, R126X, R118X, W90X, W90X, 및 R132X로 이루어진 군으로부터 선택되는 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 H121R, H122R, R126A, R126E, R118A, W90A, W90Y, 및 R132E로 이루어진 군으로부터 선택되는 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다.

[0243] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 D316X, D317X, R320X, R320X, R313X, W285X, W285X, R326X로 이루어진 군으로부터 선택되는 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 D316R, D317R, R320A, R320E, R313A, W285A, W285Y, R326E로 이루어진 군으로부터 선택되는 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다.

[0244] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 H121R 및 H122R 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 R126A 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 R126E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 R118A 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 W90A 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 W90Y 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 R132E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 W90Y 및 R126E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 R126E 및 R132E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 W90Y 및 R132E 돌연변이,

또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 76)의 W90Y, R126E, 및 R132E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다.

[0245] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 D316R 및 D317R 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 R320A 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 R320E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 R313A 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 W285A 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 W285Y 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 R326E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 W285Y 및 R320E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 R320E 및 R326E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 W285Y 및 R326E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 60)의 W285Y, R320E, 및 R326E 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다.

[0246] 본 개시내용의 일부 측면은 (i) 뉴클레아제-불활성 Cas9 도메인; 및 (ii) 핵산 편집 도메인을 포함하는 융합 단백질을 제공한다. 일부 실시양태에서, 뉴클레아제-불활성 Cas9 도메인 (dCas9)은 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에 의해 제공된 바와 같은 Cas9의 아미노산 서열과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함하고, Cas9의 뉴클레아제 활성을 불활성화시키는 돌연변이를 포함한다. Cas9의 뉴클레아제 도메인을 불활성이 되게 하는 돌연변이는 관련 기술분야에 널리 공지되어 있다. 예를 들어, Cas9의 DNA 절단 도메인은 2가지 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, 본 개시내용의 dCas9는 서열식별번호: 6에 제공된 아미노산 서열의 D10A 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, 본 개시내용의 dCas9는 서열식별번호: 6에 제공된 아미노산 서열의 H840A 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, 본 개시내용의 dCas9는 서열식별번호: 6에 제공된 아미노산 서열의 D10A 및 H840A 돌연변이 둘 다, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, Cas9는 서열식별번호: 6에 제공된 아미노산 서열의 위치 840에 히스티딘 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 추가로 포함한다. 촉매 잔기 H840의 존재는 Cas9의 활성을 복원하여 표적화된 C에 대향하는 G를 함유하는 비-편집된 가닥을 절단한다. H840의 복원은 C를 함유하는 표적 가닥의 절단을 발생시키지 않는다. 일부 실시양태에서, dCas9는 서열식별번호: 32의 아미노산 서열

을 포함한다. Cas9의 뉴클레아제 도메인을 불활성화시키는 다른 돌연변이는 또한 본 개시내용의 dCas9에 포함될 수 있다는 것이 이해되어야 한다.

[0247] 본원에 개시된 돌연변이를 포함하는 Cas9 또는 dCas9 도메인은 전장 Cas9, 또는 그의 단편일 수 있다. 일부 실시양태에서, Cas9, 또는 그의 단편을 포함하는 단백질은 "Cas9 변이체"로 지칭된다. Cas9 변이체는 Cas9, 또는 그의 단편과 상동성을 공유한다. 예를 들어 Cas9 변이체는 야생형 Cas9와 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 99.9% 동일하다. 일부 실시양태에서, Cas9 변이체는 단편이 야생형 Cas9, 예를 들어, 서열식별번호: 6의 아미노산 서열을 포함하는 Cas9의 상응하는 단편과 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 99.9% 동일하도록 Cas9의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함한다.

[0248] 본 개시내용의 Cas9 융합 단백질 중 임의의 것은 핵산 편집 도메인 (예를 들어, 핵산을 변형시킬 수 있는 효소, 예컨대 데아미나제)을 추가로 포함할 수 있다. 일부 실시양태에서, 핵산 편집 도메인은 DNA-편집 도메인이다. 일부 실시양태에서, 핵산 편집 도메인은 데아미나제 활성을 갖는다. 일부 실시양태에서, 핵산 편집 도메인은 데아미나제 또는 데아미나제 도메인을 포함하거나, 이로 이루어진다. 일부 실시양태에서, 데아미나제는 시티딘 데아미나제이다. 일부 실시양태에서, 데아미나제는 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC1 패밀리 데아미나제이다. 일부 실시양태에서, 데아미나제는 활성화-유도된 시티딘 데아미나제 (AID)이다. 일부 핵산 편집 도메인 뿐만 아니라 이러한 도메인을 포함하는 Cas9 융합 단백질은 본원에 상세히 기재되어 있다. 추가의 적합한 핵산 편집 도메인은 본 개시내용 및 관련 기술분야의 지식에 기초하여 통상의 기술자에게 명백할 것이다.

[0249] 본 개시내용의 일부 측면은 핵산 편집 도메인에 융합된 Cas9 도메인을 포함하며, 핵산 편집 도메인이 Cas9 도메인의 N-말단에 융합된 것인 융합 단백질을 제공한다. 일부 실시양태에서, Cas9 도메인 및 핵산 편집-편집 도메인은 링커를 통해 융합된다. 일부 실시양태에서, 링커는 $(GGGS)_n$ (서열식별번호: 613), $(GGGGS)_n$ (서열식별번호: 607), $(G)_n$ (서열식별번호: 608), $(EAAAK)_n$ (서열식별번호: 609), $(GGS)_n$ (서열식별번호: 610), $(SGGS)_n$ (서열식별번호: 606), SGSETPGTSESATPES (서열식별번호: 604) 모티프 (예를 들어, 문헌 [Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82] 참조; 전체 내용은 본원에 참조로 포함됨), 또는 $(XP)_n$ (서열식별번호: 611) 모티프, 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이다. 일부 실시양태에서, n은 독립적으로 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, 또는 30이거나, 1개 초과의 링커 또는 1개 초과의 링커 모티프가 존재하는 경우, 그의 임의의 조합이다. 일부 실시양태에서, 링커는 $(GGS)_n$ (서열식별번호: 610) 모티프를 포함하며, 여기서 n은 1,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15이다. 일부 실시양태에서, 링커는 $(GGS)_n$ (서열식별번호: 610) 모티프를 포함하며, 여기서 n은 1, 3, 또는 7이다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함한다. 추가의 적합한 링커 모티프 및 링커 배위는 관련 기술분야의 통상의 기술자에게 명백할 것이다. 일부 실시양태에서, 적합한 링커 모티프 및 배위는 문헌 [Chen et al., Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013; 65(10):1357-69]에 기재된 것들을 포함하며, 그의 전체 내용은 본원에 참조로 포함된다. 추가의 적합한 링커 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다. 일부 실시양태에서, 본원에 제공된 예시적인 Cas9 융합 단백질의 일반적 아키텍처는 구조:

[0250] $[NH_2]$-[핵산 편집 도메인]-[Cas9]-[COOH] 또는

[0251] $[NH_2]$-[핵산 편집 도메인]-[링커]-[Cas9]-[COOH]

[0252] 를 포함하며, 여기서 NH_2는 융합 단백질의 N-말단이고, COOH는 융합 단백질의 C-말단이다.

[0253] 본 개시내용의 융합 단백질은 하나 이상의 추가의 특색을 포함할 수 있다. 예를 들어, 일부 실시양태에서, 융합 단백질은 핵 국재화 서열 (NLS)을 포함한다. 일부 실시양태에서, 융합 단백질의 NLS는 핵산 편집 도메인 및

Cas9 도메인 사이에 국재화된다. 일부 실시양태에서, 융합 단백질의 NLS는 Cas9 도메인의 C-말단에 국재화된다.

[0254] 존재할 수 있는 다른 예시적인 특색은 국재화 서열, 예컨대 세포질 국재화 서열, 방출 서열, 예컨대 핵 방출 서열, 또는 다른 국재화 서열, 뿐만 아니라 융합 단백질의 가용화, 정제, 또는 검출에 유용한 서열 태그이다. 본원에 제공된 적합한 단백질 태그는 비오틴 카르복실라제 운반체 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.

[0255] 일부 실시양태에서, 핵산 편집 도메인은 데아미나제이다. 예를 들어, 일부 실시양태에서, 데아미나제 도메인을 갖는 예시적인 Cas9 융합 단백질의 일반적 아키텍처는 구조:

[0256] $[NH_2]$-[NLS]-[데아미나제]-[Cas9]-[COOH],

[0257] $[NH_2]$-[Cas9]-[데아미나제]-[COOH],

[0258] $[NH_2]$-[데아미나제]-[Cas9]-[COOH], 또는

[0259] $[NH_2]$-[데아미나제]-[Cas9]-[NLS]-[COOH]

[0260] 를 포함하며; 여기서 NLS는 핵 국재화 서열이고, NH_2는 융합 단백질의 N-말단이고, COOH는 융합 단백질의 C-말단이다. 핵 국재화 서열은 관련 기술분야에 공지되어 있으며, 통상의 기술자에게 명백할 것이다. 예를 들어, NLS 서열은 플랭크(Plank) 등, PCT/EP2000/011690에 기재되어 있으며, 그의 내용은 예시적인 핵 국재화 서열의 그의 개시내용에 대해 본원에 참조로 포함된다. 일부 실시양태에서, NLS는 아미노산 서열 PKKKRKV (서열식별번호: 614) 또는 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (서열식별번호: 615)를 포함한다. 일부 실시양태에서, 링커는 Cas9 및 데아미나제 사이에 삽입된다. 일부 실시양태에서, NLS는 Cas9 도메인의 C-말단에 위치한다. 일부 실시양태에서, NLS는 Cas9 도메인의 N-말단에 위치한다. 일부 실시양태에서, NLS는 데아미나제 및 Cas9 도메인 사이에 위치한다. 일부 실시양태에서, NLS는 데아미나제 도메인의 N-말단에 위치한다. 일부 실시양태에서, NLS는 데아미나제 도메인의 C-말단에 위치한다.

[0261] 핵산 편집 도메인의 한 예시적인 적합한 유형은 예를 들어, APOBEC 패밀리의 시티딘 데아미나제이다. 시티딘 데아미나제 효소의 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리는 제어되고 유익한 방식으로 돌연변이유발을 개시하는 기능을 하는 11개의 단백질을 포괄한다.[29] 한 패밀리 구성원인 활성화-유도된 시티딘 데아미나제 (AID)는 ssDNA에서의 시토신을 전사-의존성, 가닥-편형된 방식으로 우라실로 전환시킴으로써 항체의 성숙을 담당한다.[30] 아포지단백질 B 편집 복합체 3 (APOBEC3) 효소는 역전사된 바이러스 ssDNA에서의 시토신의 탈아미노화를 통해 특정 HIV-1 균주에 대해 인간 세포에 보호를 제공한다.[31] 이들 단백질 모두는 촉매 활성을 위해 Zn^{2+}-배위 모티프 (His-X-Glu-X_{23-26}-Pro-Cys-X_{2-4}-Cys; 서열식별번호: 616) 및 결합된 물 분자를 요구한다. Glu 잔기는 탈아미노화 반응에서 친핵성 공격을 위해 물 분자를 수산화아연으로 활성화시키는 작용을 한다. 각각의 패밀리 구성원은 그 자신의 특정 "핫스팟"에서 우선적으로 탈아미노화하며, 이는 hAID에 대해 WRC (W는 A 또는 T이고, R은 A 또는 G임), 내지 hAPOBEC3F에 대해 TTC의 범위이다.[32] APOBEC3G의 촉매 도메인의 최근의 결정 구조는 6개의 α-나선에 의해 플랭킹된 5-가닥 β-시트 코어로 구성된 2차 구조를 밝혀내었으며, 이는 전체 패밀리에 걸쳐 보존되는 것으로 여겨진다.[33] 활성 중심 루프는 ssDNA 결합 및 "핫스팟" 동일성을 측정하는 것 둘 다를 담당하는 것으로 나타났다.[34] 이들 효소의 과다발현은 게놈 불안정성 및 암과 관련되었으며, 따라서, 서열-특이적 표적화의 중요성을 강조한다.[35]

[0262] 본 개시내용의 일부 측면은 시티딘 데아미나제 효소, 예컨대 APOBEC 효소의 활성이 게놈 DNA에서의 특이적 부위에 지정될 수 있다는 인식에 관한 것이다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, Cas9를 인식제로서 사용하는 것의 이점은 (1) Cas9의 서열 특이성이 sgRNA 서열을 간단히 변화시킴으로써 용이하게 변경될

수 있고; (2) Cas9가 dsDNA를 변성시킴으로써 그의 표적 서열에 결합하고, 이는 단일-가닥인 DNA의 신장, 및 따라서 데아미나제에 대한 생존성 기질을 발생시킨다는 것을 포함한다. 다른 촉매 도메인, 또는 다른 데아미나제로부터의 촉매 도메인은 또한 Cas9와 융합 단백질을 생성하는 데 사용될 수 있으며, 본 개시내용은 이와 관련하여 제한되지 않는 것으로 이해되어야 한다.

[0263] 본 개시내용의 일부 측면은 Cas9:데아미나제 융합 단백질이 도 3에서의 넘버링 스킴에 따른 위치 3 내지 11에서 뉴클레오티드를 효율적으로 탈아미노화할 수 있다는 인식에 기초한다. Cas9:데아미나제 융합 단백질에 의해 표적화될 수 있는 뉴클레오티드에 관하여 본원에 제공된 결과의 관점에서, 관련 기술분야의 통상의 기술자는 융합 단백질을 탈아미노화되는 뉴클레오티드를 포함하는 표적 서열에 표적화하기 위한 적합한 가이드 RNA를 설계할 수 있을 것이다.

[0264] 일부 실시양태에서, 데아미나제 도메인 및 Cas9 도메인은 링커를 통해 서로에게 융합된다. 특정 용도를 위해 데아미나제 활성을 위한 최적 길이를 달성하기 위해, 데아미나제 도메인 (예를 들어, AID) 및 Cas9 도메인 사이의 다양한 링커 길이 및 유연성이 사용될 수 있다 (예를 들어, 형태 $(GGGGS)_n$ (서열식별번호: 607), $(GGS)_n$ (서열식별번호: 610), 및 $(G)_n$ (서열식별번호: 608)의 매우 유연한 링커 내지 형태 $(EAAAK)_n$ (서열식별번호: 609), $(SGGS)_n$ (서열식별번호: 606), SGSETPGTSESATPES (서열식별번호: 604) (예를 들어, 문헌 [Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82] 참조; 전체 내용은 본원에 참조로 포함됨) 및 $(XP)_n$ (서열식별번호: 611)의 보다 단단한 링커)[36]. 일부 실시양태에서, 링커는 $(GGS)_n$ (서열식별번호: 610) 모티프를 포함하며, 여기서 n은 1, 3, 또는 7이다. 일부 실시양태에서, 링커는 (SGSETPGTSESATPES (서열식별번호: 604) 모티프를 포함한다.

[0265] 본 개시내용의 측면에 따른 Cas9 도메인에 융합될 수 있는 일부 예시적인 적합한 핵산 편집 도메인, 예를 들어, 데아미나제 및 데아미나제 도메인이 하기 제공된다. 일부 실시양태에서, 각각의 서열의 활성 도메인, 예를 들어, 국재화 신호가 없는 도메인 (핵 방출 신호, 세포질 국재화 신호가 없는 핵 국재화 서열)이 사용될 수 있다는 것이 이해되어야 한다.

[0266] 인간 AID:

[0267] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS
VRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 49)

[0268] (밑줄: 핵 국재화 서열; 이중 밑줄: 핵 방출 신호)

[0269] 마우스 AID:

[0270] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENS
VRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 51)

[0271] (밑줄: 핵 국재화 서열; 이중 밑줄: 핵 방출 신호)

[0272] 개 AID:

[0273] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENS
VRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 52)

[0274] (밑줄: 핵 국재화 서열; 이중 밑줄: 핵 방출 신호)

[0275] 소 AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCH
VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARL
YFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS
[0276] VRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 53)

[0277] (밑줄: 핵 국재화 서열; 이중 밑줄: 핵 방출 신호)

[0278] 래트 AID:

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRW
LRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLD
FGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGN
PNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSV
[0279] RLSRRLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 54)

[0280] (밑줄: 핵 국재화 서열; 이중 밑줄: 핵 방출 신호)

[0281] 마우스 APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVS
LHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLAT
HHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGR
RFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMD
PLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEI*
*LFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQ
KGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKES
[0282] WGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 55)

[0283] (이탤릭체: 핵산 편집 도메인)

[0284] 래트 APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLAT
HHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGR
RFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVH
LLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEI*
*LFLDKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQ
KGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKES
[0285] WGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 56)

[0286] (이탤릭체: 핵산 편집 도메인)

[0287] 레수스 마카크 APOBEC-3G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY
*HPEM*RFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVAR
LYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNN
LPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWV
PLNQHRGFLRNQAPNIHGFPKGR*HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSC*AQ
EMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTF
VDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 57)

[0288]

[0289] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국재화 신호)

[0290] 침팬지 APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVT
LTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF
EPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHND
TWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPC
FSC*AQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHC
WDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 58)

[0291]

[0292] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국재화 신호)

[0293] 녹색 원숭이 APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKL
YPEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTL
TIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPF
KPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSH
NDTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSP
CFSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFE
YCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 59)

[0294]

[0295] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국재화 신호)

[0296] 인간 APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVT
LTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF
EPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHN
DTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSP
CFSC*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHC
WDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 60)

[0297]

[0298] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국재화 신호)

[0299] 인간 APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQ
VYSQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTL
TISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWY
KFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKH
HSPVSWKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGE
VAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENF
VYNDDEPFKPWKGLKYNFLFLDSKLQEILE (SEQ ID NO: 61)

[0300]

[0301] (이탤릭체: 핵산 편집 도메인)

[0302] 인간 APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG
QVYFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVT
LTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPW
YKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTW
VLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSW
GC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYC
WDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 62)

[0303]

[0304] (이탤릭체: 핵산 편집 도메인)

[0305] 래트 APOBEC-3B:

MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNN
FLCYEVNGMDCALPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTW
YMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVA
AMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVFYLQ
FNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMR
SMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFYWRKKFQKGLCT
LWRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL
(SEQ ID NO: 63)

[0306]

[0307] 소 APOBEC-3B:

DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLF
KQQFGNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAEIRFIDKINSL
DLNPSQSYKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQD
LQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILTAPI
(SEQ ID NO: 64)

[0308]

[0309] 침팬지 APOBEC-3B:

MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFR
GQMYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHP
NVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPF
MPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERLDN
GTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISW
SPCFSWGCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTY
DEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPP
QSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSF
HSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG (SEQ ID NO: 65)

[0310]

[0311] 인간 APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR
NQVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVN
LTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWK
GLKTNFRLLKRRLRESLQ (SEQ ID NO: 66)

[0312]

[0313] (이탤릭체: 핵산 편집 도메인)

[0314] 고릴라 APOBEC3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR
NQVDSETH*CHAERCFLSWFCDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVN
LTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPW
KGLKYNFRFLKRRLQEILE (SEQ ID NO: 67)

[0315]

[0316] (이탤릭체: 핵산 편집 도메인)

[0317] 인간 APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHN
QAKNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQE
NTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCP
FQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 68)

[0318]

[0319] (이탤릭체: 핵산 편집 도메인)

[0320] 레수스 마카크 APOBEC-3A:

MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRG
FLCNKAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVR
VFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDR
QGRPFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 69)

[0321]

[0322] (이탤릭체: 핵산 편집 도메인)

[0323] 소 APOBEC-3A:

MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*H*
*AELYFLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHN
RFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCT
ELQAILKTQQN (SEQ ID NO: 70)

[0324]

[0325] (이탤릭체: 핵산 편집 도메인)

[0326] 인간 APOBEC-3H:

[0327] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEIC FINEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCK PQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIK RRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 71)

[0328] (이탤릭체: 핵산 편집 도메인)

[0329] 레수스 마카크 APOBEC-3H:

[0330] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEI RFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHW RPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAI KRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR (SEQ ID NO: 72)

[0331] 인간 APOBEC-3D:

[0332] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG PVLPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPC*VVK VTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWE NFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRN ESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVT WYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKI MGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (SEQ ID NO: 73)

[0333] (이탤릭체: 핵산 편집 도메인)

[0334] 인간 APOBEC-1:

[0335] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNT TNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVA RLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPP LWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPS VAWR (SEQ ID NO: 74)

[0336] 마우스 APOBEC-1:

[0337] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNT SNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIAR LYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWV KLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK (SEQ ID NO: 75)

[0338] 래트 APOBEC-1:

[0339] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 76)

[0340] 인간 APOBEC-2:

MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRN
VE
YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRY
NVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCK
LRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK (SEQ ID
NO: 77)

[0341]

[0342] 마우스 APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRN
VEYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPAL
KYNVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEA
GCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK (SEQ ID
NO: 78)

[0343]

[0344] 래트 APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRN
VEYSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPAL
KYNVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEA
GCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK (SEQ
ID NO: 79)

[0345]

[0346] 소 APOBEC-2:

MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRN
VE
YSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRY
MVTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGC
RLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK (SEQ ID
NO: 80)

[0347]

[0348] 페트로미존 마리누스 CDA1 (pmCDA1)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 81)

[0349]

[0350] 인간 APOBEC3G D316R_D317R

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPK
VTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQR
ELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERM
HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFT
SWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSE
FKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 82)

[0351]

[0352] 인간 APOBEC3G 쇄 A

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQ (SEQ ID NO: 83)

[0353]

[0354] 인간 APOBEC3G 쇄 A D120R_D121R

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQ (SEQ ID NO: 84)

[0355]

[0356] 일부 실시양태에서, 본원에 제공된 바와 같은 융합 단백질은 핵산 편집 효소의 전장 아미노산, 예를 들어, 상기 제공된 서열 중 하나를 포함한다. 그러나, 다른 실시양태에서, 본원에 제공된 바와 같은 융합 단백질은 핵산 편집 효소의 전장 서열을 포함하지 않고, 단지 그의 단편만을 포함한다. 예를 들어, 일부 실시양태에서, 본원에 제공된 융합 단백질은 Cas9 도메인 및 핵산 편집 효소의 단편을 포함하고, 예를 들어, 단편은 핵산 편집 도메인을 포함한다. 핵산 편집 도메인의 예시적인 아미노산 서열은 상기 서열에서 이탤릭체 문자로 나타나 있으며, 이러한 도메인의 추가의 적합한 서열은 관련 기술분야의 통상의 기술자에게 명백할 것이다.

[0357] 예를 들어, 뉴클레아제-불활성 Cas9 도메인에 융합될 수 있는 본 발명의 측면에 따라 사용될 수 있는 추가의 적합한 핵산 편집 효소 서열, 예를 들어, 데아미나제 효소 및 도메인 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다. 일부 실시양태에서, 이러한 추가의 효소 서열은 본원에 제공된 서열과 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 또는 적어도 99% 유사한 데아미나제 효소 또는 데아미나제 도메인 서열을 포함한다. 추가의 적합한 Cas9 도메인, 변이체, 및 서열은 또한 관련 기술분야의 통상의 기술자에게 명백할 것이다. 이러한 추가의 적합한 Cas9 도메인의 예는 D10A, D10A/D839A/H840A, 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하나 이에 제한되지는 않는다 (예를 들어, 문헌 [Prashant et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology. 2013; 31(9): 833-838] 참조; 그의 전체 내용은 본원에 참조로 포함됨). 일부 실시양태에서, Cas9는 서열식별번호: 6에 제공된 아미노산 서열의 위치 840에 히스티딘 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 촉매 잔기 H840의 존재는 Cas9의 활성을 복원하여 표적화된 C에 대향하는 G를 함유하는 비-편집된 가닥을 절단한다. H840의 복원은 C를 함유하는 표적 가닥의 절단을 발생시키지 않는다.

[0358] Cas9 도메인 및 데아미나제 도메인을 포함하는 융합 단백질을 생성하기 위한 추가의 적합한 전략은 관련 기술분야의 일반적 지식과 조합으로 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다. 링커를 사용한 또는 링커를 사용하지 않는 본 개시내용의 측면에 따른 융합 단백질을 생성하기 위한 적합한 전략은 또한 본 개시내용 및 관련 기술분야의 지식의 관점에서 관련 기술분야의 통상의 기술자에게 명백할 것이다. 예를 들어, 문헌 [Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013; 154(2):442-51]은 링커로서 2개의 NLS (SPKKKRKVEAS, 서열식별번호: 617)를 사용한 VP64와의 Cas9의 C-말단 융합이 전사 활성화에 사용될 수 있다는 것을 보여주었다. 문헌 [Mali et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol. 2013; 31(9):833-8]은 링커 없이 VP64와의 C-말단 융합물이 전사 활성화에 사용될 수 있다는 것을 보고하였다. 그리고, 문헌 [Maeder et al., CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013; 10: 977-979]은 Gly_4Ser (서열식별번호: 613) 링커를 사용한 VP64와의 C-말단 융합물이 전사 활성인자로서 사용될 수 있다는 것을 보고하였다. 최근, dCas9-FokI 뉴클레아제 융합물은 성공적으로 생성되었으며, 모 Cas9 효소와 비교하여 개선된 효소적 특이성을 나타낸다 (문헌 [Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82, 및 Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for

highly specific genome editing. Nat Biotechnol. 2014; 32(6):569-76]. PMID: 24770325 SGSETPGTSESATPES (서열식별번호: 604) 또는 GGGGS (서열식별번호: 607) 링커는 각각 FokI-dCas9 융합 단백질에 사용됨).

[0359] 본 개시내용의 일부 측면은 (i) Cas9 효소 또는 도메인 (예를 들어, 제1 단백질); 및 (ii) 핵산-편집 효소 또는 도메인 (예를 들어, 제2 단백질)을 포함하는 융합 단백질을 제공한다. 일부 측면에서, 본원에 제공된 융합 단백질은 (iii) 프로그램가능한 DNA-결합 단백질, 예를 들어, 아연-핑거 도메인, TALE, 또는 제2 Cas9 단백질 (예를 들어, 제3 단백질)을 추가로 포함한다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 프로그램가능한 DNA-결합 단백질 (예를 들어, 제2 Cas9 단백질)을 (i) Cas9 효소 또는 도메인 (예를 들어, 제1 단백질); 및 (ii) 핵산-편집 효소 또는 도메인 (예를 들어, 제2 단백질)을 포함하는 융합 단백질에 융합시키는 것은 표적 핵산 서열에 대한 융합 단백질의 특이성을 개선시키는 데, 또는 정규 PAM (NGG) 서열을 함유하지 않는 표적 핵산 서열에 결합하는 융합 단백질의 특이성 또는 결합 친화도를 개선시키는 데 유용할 수 있다. 일부 실시양태에서, 제3 단백질은 Cas9 단백질 (예를 들어, 제2 Cas9 단백질)이다. 일부 실시양태에서, 제3 단백질은 본원에 제공된 Cas9 단백질 중 임의의 것이다. 일부 실시양태에서, 제3 단백질은 Cas9 단백질 (예를 들어, 제1 단백질)에 대한 융합 단백질 N-말단에 융합된다. 일부 실시양태에서, 제3 단백질은 Cas9 단백질 (예를 들어, 제1 단백질)에 대한 융합 단백질 C-말단에 융합된다. 일부 실시양태에서, Cas9 도메인 (예를 들어, 제1 단백질) 및 제3 단백질 (예를 들어, 제2 Cas9 단백질)은 링커 (예를 들어, 제2 링커)를 통해 융합된다. 일부 실시양태에서, 링커는 $(GGGGS)_n$ (서열식별번호: 607), $(G)_n$ (서열식별번호: 608), $(EAAAK)_n$ (서열식별번호: 609), $(GGS)_n$ (서열식별번호: 610), $(SGGS)_n$ (서열식별번호: 606), SGSETPGTSESATPES (서열식별번호: 604), $SGGS(GGS)_n$ (서열식별번호: 612), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605), 또는 $(XP)_n$ (서열식별번호: 611) 모티프, 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이다. 일부 실시양태에서, 본원에 제공된 예시적인 napDNAbp 융합 단백질의 일반적 아키텍처는 구조:

[0360] [NH2]-[핵산-편집 효소 또는 도메인]-[napDNAbp]-[제3 단백질]-[COOH];

[0361] [NH2]-[제3 단백질]-[napDNAbp]-[핵산-편집 효소 또는 도메인]-[COOH];

[0362] [NH2]-[napDNAbp]-[핵산-편집 효소 또는 도메인]-[제3 단백질]-[COOH];

[0363] [NH2]-[제3 단백질]-[핵산-편집 효소 또는 도메인]-[napDNAbp]-[COOH];

[0364] [NH2]-[UGI]-[핵산-편집 효소 또는 도메인]-[napDNAbp]-[제3 단백질]-[COOH];

[0365] [NH2]-[UGI]-[제3 단백질]-[napDNAbp]-[핵산-편집 효소 또는 도메인]-[COOH];

[0366] [NH2]-[UGI]-[napDNAbp]-[핵산-편집 효소 또는 도메인]-[제3 단백질]-[COOH];

[0367] [NH2]-[UGI]-[제3 단백질]-[핵산-편집 효소 또는 도메인]-[napDNAbp]-[COOH];

[0368] [NH2]-[핵산-편집 효소 또는 도메인]-[napDNAbp]-[제3 단백질]-[UGI]-[COOH];

[0369] [NH2]-[제3 단백질]-[napDNAbp]-[핵산-편집 효소 또는 도메인]-[UGI]-[COOH];

[0370] [NH2]-[NapDNAbp]-[핵산-편집 효소 또는 도메인]-[제3 단백질]-[UGI]-[COOH];

[0371] [NH2]-[제3 단백질]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[UGI]-[COOH]; 또는

[0372] [NH2]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[제1 UGI 도메인]-[제2 UGI 도메인]-[COOH]

[0373] 를 포함하며; 여기서 NH2는 융합 단백질의 N-말단이고, COOH는 융합 단백질의 C-말단이다. 일부 실시양태에서, 상기 일반적 아키텍처에 사용된 "]-["는 임의적인 링커 서열의 존재를 나타낸다. 다른 예에서, 본원에 제공된 예시적인 NapDNAbp 융합 단백질의 일반적 아키텍처는 구조:

[0374] [NH2]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[제2 NapDNAbp 단백질]-[COOH];

[0375] [NH2]-[제2 NapDNAbp 단백질]-[NapDNAbp]-[핵산-편집 효소 또는 도메인]-[COOH];

[0376] [NH2]-[NapDNAbp]-[핵산-편집 효소 또는 도메인]-[제2 NapDNAbp 단백질]-[COOH];

[0377] [NH2]-[제2 NapDNAbp 단백질]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[COOH];

[0378] [NH2]-[UGI]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[제2 NapDNAbp 단백질]-[COOH],

[0379] [NH2]-[UGI]-[제2 NapDNAbp 단백질]-[NapDNAbp]-[핵산-편집 효소 또는 도메인]-[COOH];

[0380] [NH2]-[UGI]-[NapDNAbp]-[핵산-편집 효소 또는 도메인]-[제2 NapDNAbp 단백질]-[COOH];

[0381] [NH2]-[UGI]-[제2 NapDNAbp 단백질]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[COOH];

[0382] [NH2]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[제2 NapDNAbp 단백질]-[UGI]-[COOH];

[0383] [NH2]-[제2 NapDNAbp 단백질]-[NapDNAbp]-[핵산-편집 효소 또는 도메인]-[UGI]-[COOH];

[0384] [NH2]-[NapDNAbp]-[핵산-편집 효소 또는 도메인]-[제2 NapDNAbp 단백질]-[UGI]-[COOH]; 또는

[0385] [NH2]-[제2 NapDNAbp 단백질]-[핵산-편집 효소 또는 도메인]-[NapDNAbp]-[UGI]-[COOH]

[0386] 를 포함하며; 여기서 NH_2는 융합 단백질의 N-말단이고, COOH는 융합 단백질의 C-말단이다. 일부 실시양태에서, 상기 일반적 아키텍처에 사용된 "]-["는 임의적인 링커 서열의 존재를 나타낸다. 일부 실시양태에서, 제2 NapDNAbp는 dCas9 단백질이다. 일부 예에서, 본원에 제공된 예시적인 Cas9 융합 단백질의 일반적 아키텍처는 도 3에 제시된 바와 같은 구조를 포함한다. 예시적인 Cas9 융합 단백질의 일반적 아키텍처 중 임의의 것에 제공된 단백질 중 임의의 것은 본원에 제공된 링커 중 1종 이상에 의해 연결될 수 있다는 것이 인지되어야 한다. 일부 실시양태에서, 링커는 동일하다. 일부 실시양태에서, 링커는 상이하다. 일부 실시양태에서, 예시적인 Cas9 융합 단백질의 일반적 아키텍처 중 임의의 것에 제공된 단백질 중 1종 이상은 링커를 통해 융합되지 않는다. 일부 실시양태에서, 융합 단백질은 핵 표적화 서열, 예를 들어 핵 국재화 서열을 추가로 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질은 핵 국재화 서열 (NLS)을 추가로 포함한다. 일부 실시양태에서, NLS는 융합 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 융합 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 제3 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 제3 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 Cas9 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 Cas9 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 핵산-편집 효소 또는 도메인의 N-말단에 융합된다. 일부 실시양태에서, NLS는 핵산-편집 효소 또는 도메인의 C-말단에 융합된다. 일부 실시양태에서, NLS는 UGI 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 UGI 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 1종 이상의 링커를 통해 융합 단백질에 융합된다. 일부 실시양태에서, NLS는 링커 없이 융합 단백질에 융합된다.

[0387] 우라실 글리코실라제 억제제 융합 단백질

[0388] 본 개시내용의 일부 측면은 우라실 글리코실라제 억제제 (UGI) 도메인을 포함하는 융합 단백질에 관한 것이다. 일부 실시양태에서, Cas9 도메인 (예를 들어, 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 dCas9 도메인, 또는 Cas9 니카제)을 포함하는 본원에 제공된 융합 단백질 중 임의의 것은 직접적으로 또는 링커를 통해 UGI 도메인에 추가로 융합될 수 있다. 본 개시내용의 일부 측면은 증가된 핵염기 편집 효율을 갖는 데아미나제-dCas9 융합 단백질, 데아미나제-뉴클레아제 활성 Cas9 융합 단백질 및 데아미나제-Cas9 니카제 융합 단백질을 제공한다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, U:G 헤테로듀플렉스 DNA의 존재에 대한 세포의 DNA-복구 반응은 세포에서의 핵염기 편집 효율의 감소의 원인이 될 수 있다. 예를 들어, 우라실 DNA 글리코실라제 (UDG)는 세포에서의 DNA로부터의 U의 제거를 촉매하며, 이는 염기 절제 복구를 개시할 수 있고, 가장 흔한 결과물로서 U:G 쌍의 C:G 쌍으로의 복귀를 갖는다. 하기 실시예에서 입증된 바와 같이, 우라실 DNA 글리코실라제 억제제 (UGI)는 인간 UDG 활성을 억제할 수 있다. 따라서, 본 개시내용은 UGI 도메인에 추가로 융합된 dCas9-핵산 편집 도메인을 포함하는 융합 단백질을 고려한다. 본 개시내용은 또한 UGI 도메인에 추가로 융합된 Cas9 니카제-핵산 편집 도메인을 포함하는 융합 단백질을 고려한다. UGI 도메인의 사용은 C에서 U로의 변화를 촉매할 수 있는 핵산 편집 도메인의 편집 효율을 증가시킬 수 있다는 것이 이해되어야 한다. 예를 들어, UGI 도메인을 포함하는 융합 단백질은 C 잔기를 탈아미노화하는 데 보다 효율적일 수 있다. 일부 실시양태에서, 융합 단백질은 구조:

[0389] [데아미나제]-[임의적인 링커 서열]-[dCas9]-[임의적인 링커 서열]-[UGI];

[0390] [데아미나제]-[임의적인 링커 서열]-[UGI]-[임의적인 링커 서열]-[dCas9];

[0391] [UGI]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[dCas9];

[0392] [UGI]-[임의적인 링커 서열]-[dCas9]-[임의적인 링커 서열]-[데아미나제];

[0393] [dCas9]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[UGI];

[0394] [dCas9]-[임의적인 링커 서열]-[UGI]-[임의적인 링커 서열]-[데아미나제];

[0395] [데아미나제]-[임의적인 링커 서열]-[dCas9]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI];

[0396] [데아미나제]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]-[임의적인 링커 서열]-[dCas9];

[0397] [제1 UGI]-[임의적인 링커 서열]-[제2 UGI]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[dCas9];

[0398] [제1 UGI]-[임의적인 링커 서열]-[제2 UGi]-[임의적인 링커 서열]-[dCas9]-[임의적인 링커 서열]-[데아미나제];

[0399] [dCas9]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]; 또는

[0400] [dCas9]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]-[임의적인 링커 서열]-[데아미나제]

[0401] 를 포함한다.

[0402] 다른 실시양태에서, 융합 단백질은 구조:

[0403] [데아미나제]-[임의적인 링커 서열]-[Cas9 니카제]-[임의적인 링커 서열]-[UGI];

[0404] [데아미나제]-[임의적인 링커 서열]-[UGI]-[임의적인 링커 서열]-[Cas9 니카제];

[0405] [UGI]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[Cas9 니카제];

[0406] [UGI]-[임의적인 링커 서열]-[Cas9 니카제]-[임의적인 링커 서열]-[데아미나제];

[0407] [Cas9 니카제]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[UGI];

[0408] [Cas9 니카제]-[임의적인 링커 서열]-[UGI]-[임의적인 링커 서열]-[데아미나제];

[0409] [데아미나제]-[임의적인 링커 서열]-[Cas9 니카제]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI];

[0410] [데아미나제]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]-[임의적인 링커 서열]-[Cas9 니카제];

[0411] [제1 UGI]-[임의적인 링커 서열]-[제2 UGI]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[Cas9 니카제];

[0412] [제1 UGI]-[임의적인 링커 서열]-[제2 UGi]-[임의적인 링커 서열]-[Cas9 니카제]-[임의적인 링커 서열]-[데아미나제];

[0413] [Cas9 니카제]-[임의적인 링커 서열]-[데아미나제]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]; 또는

[0414] [Cas9 니카제]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]-[임의적인 링커 서열]-[데아미나제]

[0415] 를 포함한다.

[0416] 상기 기재된 융합 단백질 중 임의의 것은 (i) 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp); (ii) 시티딘 데아미나제 도메인; 및 (iii) 2개 이상의 UGI 도메인으로 구성될 수 있으며, 여기서 2개 이상의 UGI 도메인은 구축물에서 서로 인접할 수 있거나 (예를 들어, [제1 UGI]-[제2 UGI], 여기서 "-"는 임의적인 링커임), 또는 2개 이상의 UGI 도메인은 (i)의 napDNAbp 및/또는 (ii)의 시티딘 데아미나제 도메인에 의해 분리될 수 있다는 것 (예를 들어, [제1 UGI]-[데아미나제]-[제2 UGI], [제1 UGI]-[napDNAbp]-[제2 UGI], [제1 UGI]-[데아미나제]-[napDNAbp]-[제2 UGI] 등, 여기서 "-"는 임의적인 링커임)이 인지되어야 한다.

[0417] 또 다른 측면에서, 융합 단백질은 (i) Cas9 효소 또는 도메인; (ii) 핵산-편집 효소 또는 도메인 (예를 들어, 제2 단백질) (예를 들어, 시티딘 데아미나제 도메인); (iii) 제1 우라실 글리코실라제 억제제 도메인 (UGI) (예를 들어, 제3 단백질); 및 (iv) 제2 우라실 글리코실라제 억제제 도메인 (UGI) (예를 들어, 제4 단백질)을 포함한다. 제1 및 제2 우라실 글리코실라제 억제제 도메인 (UGI)은 동일하거나 상이할 수 있다. 일부 실시양태에서, Cas9 도메인 (예를 들어, 제1 단백질) 및 데아미나제 (예를 들어, 제2 단백질)는 링커를 통해 융합된다.

일부 실시양태에서, Cas9 도메인은 데아미나제의 C-말단에 융합된다. 일부 실시양태에서, Cas9 단백질 (예를 들어, 제1 단백질) 및 제1 UGI 도메인 (예를 들어, 제3 단백질)은 링커 (예를 들어, 제2 링커)를 통해 융합된다. 일부 실시양태에서, 제1 UGI 도메인은 Cas9 단백질의 C-말단에 융합된다. 일부 실시양태에서, 제1 UGI 도메인 (예를 들어, 제3 단백질) 및 제2 UGI 도메인 (예를 들어, 제4 단백질)은 링커 (예를 들어, 제3 링커)를 통해 융합된다. 일부 실시양태에서, 제2 UGI 도메인은 제1 UGI 도메인의 C-말단에 융합된다. 일부 실시양태에서, 링커는 $(GGGGS)_n$ (서열식별번호: 607), $(G)_n$ (서열식별번호: 608), $(EAAAK)_n$ (서열식별번호: 609), $(GGS)_n$ (서열식별번호: 610), $(SGGS)_n$ (서열식별번호: 606), SGSETPGTSESATPES (서열식별번호: 604), $SGGS(GGS)_n$ (서열식별번호: 612), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605) 또는 $(XP)_n$ (서열식별번호: 611) 모티프 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이다. 일부 실시양태에서, 제1 링커는 1-50개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 1-40개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 1-35개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 1-30개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 1-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 10-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 30-40개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 14, 16 또는 18개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 16개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 30, 32 또는 34개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 32개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제1 링커는 SGSETPGTSESATPES (서열식별번호: 604) 모티프를 포함한다. 일부 실시양태에서, 제1 링커는 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 605) 모티프를 포함한다. 일부 실시양태에서, 제2 링커는 1-50개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 1-40개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 1-35개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 1-30개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 1-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 2-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 2-10개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 10-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 2, 4 또는 6개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 7, 9 또는 11개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 14, 16 또는 18개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 4개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 9개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 16개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제2 링커는 $(SGGS)_n$ (서열식별번호: 606) 모티프를 포함하며, 여기서 n은 1 내지 30 (양끝값 포함)의 정수이다. 일부 실시양태에서, 제2 링커는 $(SGGS)_n$ (서열식별번호: 606) 모티프를 포함하며, 여기서 n은 1이다. 일부 실시양태에서, 제2 링커는 $SGGS(GGS)_n$ (서열식별번호: 612) 모티프를 포함하며, 여기서 n은 1 내지 30 (양끝값 포함)의 정수이다. 일부 실시양태에서, 제2 링커는 $SGGS(GGS)_n$ (서열식별번호: 612) 모티프를 포함하며, 여기서 n은 2이다. 일부 실시양태에서, 제3 링커는 포함하고 1-50개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 1-40개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 1-35개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 1-30개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 1-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 2-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 2-10개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 10-20개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 2, 4 또는 6개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 7, 9 또는 11개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 14, 16 또는 18개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 4개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 9개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 16개 아미노산의 아미노산 서열을 포함한다. 일부 실시양태에서, 제3 링커는 $(SGGS)_n$ (서열식별번호: 606) 모티프를 포함하며, 여기서 n은 1 내지 30 (양끝값 포함)의 정수이다. 일부 실시양태에서, 제3 링커는 $(SGGS)_n$ (서열식별번호: 606) 모티프를 포함하며, 여기서 n은 1이다. 일부 실시양태에서, 제3 링커는 $SGGS(GGS)_n$ (서열식별번호: 612) 모티프를 포함하며, 여기서 n은 1 내지 30 (양끝값 포함)의 정수이다. 일부 실시양태에서, 제3 링커는 $SGGS(GGS)_n$ (서열식별번호:

612) 모티프를 포함하며, 여기서 n은 2이다.

[0418] 일부 실시양태에서, 융합 단백질은 하기 구조:

[0419] [데아미나제]-[임의적인 링커 서열]-[dCas9]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI];

[0420] [데아미나제]-[임의적인 링커 서열]-[Cas9 니카제]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]; 또는

[0421] [데아미나제]-[임의적인 링커 서열]-[Cas9]-[임의적인 링커 서열]-[제1 UGI]-[임의적인 링커 서열]-[제2 UGI]

[0422] 를 포함한다.

[0423] 또 다른 측면에서, 융합 단백질은 (i) Cas9 효소 또는 도메인; (ii) 핵산-편집 효소 또는 도메인 (예를 들어, 제2 단백질) (예를 들어, 시티딘 데아미나제 도메인); (iii) 2개 초과의 우라실 글리코실라제 억제제 (UGI) 도메인을 포함한다.

[0424] 일부 실시양태에서, 본원에 제공된 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시양태에서, 임의적인 링커 서열 중 하나 또는 둘 다가 존재한다. 일부 실시양태에서, 임의적인 링커 서열 중 1, 2, 또는 3개가 존재한다.

[0425] 일부 실시양태에서, 상기 일반적 아키텍처에 사용된 "-"는 임의적인 링커 서열의 존재를 나타낸다. 일부 실시양태에서, UGI를 포함하는 융합 단백질은 핵 표적화 서열, 예를 들어 핵 국재화 서열을 추가로 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질은 핵 국재화 서열 (NLS)을 추가로 포함한다. 일부 실시양태에서, NLS는 융합 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 융합 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 UGI 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 UGI 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 Cas9 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 Cas9 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 데아미나제의 N-말단에 융합된다. 일부 실시양태에서, NLS는 데아미나제의 C-말단에 융합된다. 일부 실시양태에서, NLS는 제2 Cas9의 N-말단에 융합된다. 일부 실시양태에서, NLS는 제2 Cas9의 C-말단에 융합된다. 일부 실시양태에서, NLS는 1종 이상의 링커를 통해 융합 단백질에 융합된다. 일부 실시양태에서, NLS는 링커 없이 융합 단백질에 융합된다. 일부 실시양태에서, NLS는 본원에 제공되거나 언급된 NLS 서열 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, NLS는 서열식별번호: 614 또는 서열식별번호: 615에 제시된 바와 같은 아미노산 서열을 포함한다.

[0426] 일부 실시양태에서, UGI 도메인은 야생형 UGI 또는 서열식별번호: 134에 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: 134에 제시된 아미노산 서열의 단편을 포함한다. 일부 실시양태에서, UGI 단편은 서열식별번호: 134에 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: 134에 제시된 아미노산 서열에 상동인 아미노산 서열 또는 서열식별번호: 134에 제시된 아미노산 서열의 단편에 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 단편과 상동성을 공유한다. 예를 들어 UGI 변이체는 야생형 UGI 또는 서열식별번호: 134에 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, 적어도 99.9% 동일하다. 일부 실시양태에서, UGI 변이체는 단편이 야생형 UGI 또는 서열식별번호: 134에 제시된 바와 같은 UGI의 상응하는 단편과 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, 적어도 99.9% 동일하도록 UGI의 단편을 포함한다. 일부 실시양태에서, UGI는 하기 아미노산 서열을 포함한다:

[0427] >sp|P14739|UNGI_BPPB2 우라실-DNA 글리코실라제 억제제

[0428] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKML (SEQ ID NO: 134)

[0429] 적합한 UGI 단백질 및 뉴클레오티드 서열은 본원에 제공되며, 추가의 적합한 UGI 서열은 관련 기술분야의 통상의 기술자에게 공지되어 있고, 예를 들어, 문헌 [Wang et al., Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J. Biol. Chem. 264:1163-1171(1989); Lundquist et al., Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. J. Biol. Chem. 272:21408-21419(1997); Ravishankar et al., X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. Nucleic Acids Res. 26:4880-4887(1998); 및 Putnam et al., Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. J. Mol. Biol. 287:331-346(1999)]에 공개된 것들을 포함하며, 각각의 전체 내용은 본원에 참조로 포함된다.

[0430] 추가의 단백질은 우라실 글리코실라제 억제제일 수 있다는 것이 인지되어야 한다. 예를 들어, 우라실-DNA 글리코실라제 염기-절제 복구 효소를 억제할 (예를 들어, 입체적으로 차단할) 수 있는 다른 단백질은 본 개시내용의 범주 내에 있다. 추가적으로, 염기-절제 복구를 차단하거나 억제하는 임의의 단백질은 또한 본 개시내용의 범주 내에 있다. 일부 실시양태에서, 본원에 기재된 융합 단백질은 1개의 UGI 도메인을 포함한다. 일부 실시양태에서, 본원에 기재된 융합 단백질은 1개 초과의 UGI 도메인을 포함한다. 일부 실시양태에서, 본원에 기재된 융합 단백질은 2개의 UGI 도메인을 포함한다. 일부 실시양태에서, 본원에 기재된 융합 단백질은 2개 초과의 UGI 도메인을 포함한다. 일부 실시양태에서, DNA에 결합하는 단백질이 사용된다. 또 다른 실시양태에서, UGI에 대한 치환이 사용된다. 일부 실시양태에서, 우라실 글리코실라제 억제제는 단일-가닥 DNA에 결합하는 단백질이다. 예를 들어, 우라실 글리코실라제 억제제는 에르위니아 타스마니엔시스(Erwinia tasmaniensis) 단일-가닥 결합 단백질일 수 있다. 일부 실시양태에서, 단일-가닥 결합 단백질은 아미노산 서열 (서열식별번호: 135)을 포함한다. 일부 실시양태에서, 우라실 글리코실라제 억제제는 우라실에 결합하는 단백질이다. 일부 실시양태에서, 우라실 글리코실라제 억제제는 DNA에서 우라실에 결합하는 단백질이다. 일부 실시양태에서, 우라실 글리코실라제 억제제는 촉매적으로 불활성인 우라실 DNA-글리코실라제 단백질이다. 일부 실시양태에서, 우라실 글리코실라제 억제제는 DNA로부터 우라실을 절제하지 않는 촉매적으로 불활성인 우라실 DNA-글리코실라제 단백질이다. 예를 들어, 우라실 글리코실라제 억제제는 UdgX이다. 일부 실시양태에서, UdgX는 아미노산 서열 (서열식별번호: 136)을 포함한다. 또 다른 예로서, 우라실 글리코실라제 억제제는 촉매적으로 불활성인 UDG이다. 일부 실시양태에서, 촉매적으로 불활성인 UDG는 아미노산 서열 (서열식별번호: 137)을 포함한다. 다른 우라실 글리코실라제 억제제는 통상의 기술자에게 명백할 것이며, 본 개시내용의 범위 내에 있다는 것이 인지되어야 한다. 일부 실시양태에서, 우라실 글리코실라제 억제제는 서열식별번호: 135-137 또는 143-148 중 어느 하나와 상동인 단백질이다. 일부 실시양태에서, 우라실 글리코실라제 억제제는 서열식별번호: 135-137 또는 143-148 중 어느 하나와 적어도 50% 동일하거나, 적어도 55% 동일하거나, 적어도 60% 동일하거나, 적어도 65% 동일하거나, 적어도 70% 동일하거나, 적어도 75% 동일하거나, 적어도 80% 동일하거나, 적어도 85% 동일하거나, 적어도 90% 동일하거나, 적어도 95% 동일하거나, 적어도 96% 동일하거나, 적어도 98% 동일하거나, 적어도 99% 동일하거나, 적어도 99.5% 동일한 단백질이다.

[0431] 에르위니아 타스마니엔시스 SSB (열안정성 단일-가닥 DNA 결합 단백질)

MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETKEKTEWH
RVVLFGKLAEVAGEYLRKGSQVYIEGALQTRKWTDQAGVEKYTTEVVVNVGGTMQ
MLGGRSQGGGASAGGQNGGSNNGWGQPQQPQGGNQFSGGAQQQARPQQQPQQNN
APANNEPPIDFDDDIP (SEQ ID NO: 135)

[0432]

[0433] UdgX (DNA에서 우라실에 결합하지만 절제하지 않음)

MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPG
DKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTP
SRTEVVACRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDVP
GDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVAADVRP (SEQ ID NO: 136)

[0434]

[0435] UDG (촉매적으로 불활성인 인간 UDG, DNA에서 우라실에 결합하지만, 절제하지 않음)

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEP
GTPPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMG
FVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQEPYHGPNQAHGLCFSVQRPV
PPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERG
WEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRG
FFGCRHFSKTNELLQKSGKKPIDWKEL (SEQ ID NO: 137)

[0436]

[0437] UGI로서 사용될 수 있는 추가의 단일-가닥 DNA 결합 단백질이 하기 제시된다. 다른 단일-가닥 결합 단백질, 예를 들어 전체 내용이 본원에 참조로 포함된 문헌 [Dickey TH, Altschuler SE, Wuttke DS. Single-stranded DNA-binding proteins:multiple domains for multiple functions. Structure. 2013 Jul 2;21(7):1074-84. doi: 10.1016/j.str.2013.05.013. Review.; Marceau AH. Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. Methods Mol Biol. 2012;922:1-21. doi: 10.1007/978-1-62703-032-8_1.; Mijakovic, Ivan, et al.; Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. Nucleic Acids Res 2006; 34 (5): 1588-1596. doi: 10.1093/nar/gkj514; Mumtsidu E, Makhov AM, Konarev PV, Svergun DI, Griffith JD, Tucker PA. Structural features of the single-stranded DNA-binding protein of Epstein-Barrvirus. J Struct Biol. 2008 Feb;161(2):172-87. Epub 2007 Nov 1; Nowak M, Olszewski M, Spibida M, Kur J. Characterization of single-strandedDNA-binding proteins from the psychrophilic bacteria Desulfotalea psychrophila, Flavobacterium psychrophilum, Psychrobacter arcticus, Psychrobactercryohalolentis, Psychromonas ingrahamii, Psychroflexus torquis, and Photobacterium profundum. BMC Microbiol. 2014 Apr 14;14:91. doi: 10.1186/1471-2180-14-91; Tone T, Takeuchi A, Makino O. Single-stranded DNA binding protein Gp5 of Bacillus subtilis phage Φ29 is required for viral DNA replication in growth-temperature dependent fashion. Biosci Biotechnol Biochem. 2012;76(12):2351-3. Epub 2012 Dec 7; Wold. REPLICATION PROTEIN A:A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. Annual Review of Biochem. 1997; 66:61-92. doi: 10.1146/annurev.biochem.66.1.61; Wu Y, Lu J, Kang T. Human single-stranded DNA binding proteins: guardians of genome stability. Acta Biochim Biophys Sin (Shanghai). 2016 Jul;48(7):671-7. doi: 10.1093/abbs/gmw044. Epub 2016 May 23. Review]에 기재된 것들이 UGI로서 사용될 수 있다는 것이 인지되어야 한다.

[0438] mtSSB - SSBP1 단일 가닥 DNA 결합 단백질 1 [호모 사피엔스 (인간)] (유니프롯KB: Q04837; NP_001243439.1)

MFRRPVLQVLRQFVRHESETTTSLVLERSLNRVHLLGRVGQDPVLRQVEGKNPVTIFS
LATNEMWRSGDSEVYQLGDVSQKTTWHRISVFRPGLRDVAYQYVKKGSRIYLEGKI
DYGEYMDKNNVRRQATTIIADNIIFLSDQTKEKE (SEQ ID NO: 138)

[0439]

[0440] 단일-가닥 DNA 결합 단백질 3 이소형 A [무스 무스쿨루스] (유니프롯KB - Q9D032-1; NCBI Ref: NP_076161.2)

MFAKGKGSAVPSDGQAREKLALYVYEYLLHVGAQKSAQTFLSEIRWEKNITLGEPPG
FLHSWWCVFWDLYCAAPERRDTCEHSSEAKAFHDYSAAAAPSPVLGNIPPNDGMPG
GPIPPGFFQGPPGSQPSPHAQPPPHNPSSMMGPHSQPFMSPRYAGGPRPPIRMGNQPPG
GVPGTQPLLPNSMDPTRQQGHPNMGGSMQRMNPPRGMGPMGPGPQNYGSGMRPPP
NSLGPAMPGINMGPGAGRPWPNPNSANSIPYSSSSPGTYVGPPGGGGPPGTPIMPSPAD
STNSSDNIYTMINPVPPGGSRSNFPMGPGSDGPMGGMGGMEPHHMNGSLGSGDIDGL
PKNSPNNISGISNPPGTPRDDGELGGNFLHSFQNDNYSPSMTMSV (SEQ ID NO: 139)

[0441]

[0442] RPA 1 - 복제 단백질 A 70 kDa DNA 결합 서브유닛 (유니프롯KB: P27694; NCBI Ref: NM_002945.3)

MVGQLSEGAIAAIMQKGDTNIKPILQVINIRPITTGNSPPRYRLLMSDGLNTLSSFMLAT
QLNPLVEEEQLSSNCVCQIHRFIVNTLKDGRRVVILMELEVLKSAEAVGVKIGNPVPY
NE
GLGQPQVAPPAPAASPAASSRPQPQNGSSGMGSTVSKAYGASKTFGKAAGPSLSHTS
GGTQSKVVPIASLTPYQSKWTICARVTNKSQIRTWSNSRGEGKLFSLELVDESGEIRAT
AFNEQVDKFFPLIEVNKVYYFSKGTLKIANKQFTAVKNDYEMTFNNETSVMPCEDDH
HLPTVQFDFTGIDDLENKSKDSLVDIIGICKSYEDATKITVRSNNREVAKRNIYLMDTS
GKVVTATLWGEDADKFDGSRQPVLAIKGARVSDFGGRSLSVLSSSTIIANPDIPEAYKL
RGWFDAEGQALDGVSISDLKSGGVGGSNTNWKTLYEVKSENLGQGDKPDYFSSVAT
VVYLRKENCMYQACPTQDCNKKVIDQQNGLYRCEKCDTEFPNFKYRMILSVNIADFQ
ENQWVTCFQESAEAILGQNAAYLGELKDKNEQAFEEVFQNANFRSFIFRVRVKVETY
NDESRIKATVMDVKPVDYREYGRRLVMSIRRSALM (SEQ ID NO: 140)

[0443]

[0444] RPA 2 - 복제 단백질 A 32 kDa 서브유닛 (유니프롯KB: P15927; NCBI Ref: NM_002946)

MWNSGFESYGSSSYGGAGGYTQSPGGFGSPAPSQAEKKSRARAQHIVPCTISQLLSAT
LVDEVFRIGNVEISQVTIVGIIRHAEKAPTNIVYKIDDMTAAPMDVRQWVDTDDTSSE
NTVVPPETYVKVAGHLRSFQNKKSLVAFKIMPLEDMNEFTTHILEVINAHMVLSKAN
SQPSAGRAPISNPGMSEAGNFGGNSFMPANGLTVAQNQVLNLIKACPRPEGLNFQDL
KNQLKHMSVSSIKQAVDFLSNEGHIYSTVDDDHFKSTDAE (SEQ ID NO: 141)

[0445]

[0446] RPA 3 - 복제 단백질 A 14 kDa 서브유닛 (유니프롯KB: P35244; NCBI Ref: NM_002947.4)

MVDMMDLPRSRINAGMLAQFIDKPVCFVGRLEKIHPTGKMFILSDGEGKNGTIELMEP
LDEEISGIVEVVGRVTAKATILCTSYVQFKEDSHPFDLGLYNEAVKIIHDFPQFYPLGIV
QH
D (SEQ ID NO: 142)

[0447]

[0448] 박테리아 단일-가닥 DNA 결합 단백질:

[0449] ssbA - 단일-가닥 DNA-결합 단백질 [바실루스 서브틸리스 아종 서브틸리스 균주 168] (유니프롯KB: P37455; NCBI Ref:)

MLNRVVLVGR LTKDPELRYT PNGAAVATFT LAVNRTFTNQ SGEREADFIN
CVTWRRQAEN VANFLKKGSL AGVDGRLQTR NYENQQGQRV FVTEVQAESV
QFLEPKNGGG SGSGGYNEGN SGGGQYFGGG QNDNPFGGNQ NNQRRNQGNS
FNDDPFANDG KPIDISDDDLPF (SEQ ID NO: 143)

[0450]

[0451] 단일-가닥 DNA 결합 단백질 2 [스트렙토미세스 코엘리콜로르 A3(2)] (유니프롯KB: Q9X8U3; NCBI Ref: NP_628093.1)

MAGETVITVVGNLVDDPELRFTPSGAAVAKFRVASTPRTFDRQTNEWKDGESLFLTC
SVWRQAAENVAESLQRGMRVIVQGRLKQRSYEDREGVKRTVYELDVDEVGASLRSA
TAKVTKTSGQGRGGQGGYGGGGGGQGGGGWGGGPGGGQQGGGAPADDPWATGG
APAGGQQGGGGQGGGGWGGGSGGGGGYSDEPPF (SEQ ID NO: 144)

[0452]

[0453] 단일-가닥 DNA 결합 단백질 [스트렙토코쿠스 뉴모니아에 R6] (유니프롯KB: P66855; NCBI Ref: NP_358988.1)

MINNVVLVGRMTRDAELRYTPSNVAVATFTLAVNRTFKSQNGEREADFINVVMWRQ
QAENLANWAKKGSLIGVTGRIQTRSYDNQQGQRVYVTEVVAENFQMLESRSVREGH
TGGAYSAPTANYSAPTNSVPDFSRNENPFGATNPLDISDDDLPF (SEQ ID NO: 145)

[0454]

[0455] 바이러스 단일-가닥 DNA 결합 단백질:

[0456] 단일-가닥 DNA 결합 단백질 [인간 알파헤르페스바이러스 1] (유니프롯KB: P04296; NCBI Ref: YP_009137104.1)

METKPKTATTIKVPPGPLGYVYARACPSEGIELLALLSARSGDSDVAVAPLVVGLTVE
SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRL
CERARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINN
TFLHLGGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFN
RPLNRLLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEAS
QGKTPRGGRDGGGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPL
FEGQDTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSKPS
FYRFFLVPGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQEFAGEHLAMLCGFSPA
LLAKMLFYLERCDGGVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLM
RLRARHPKFASAARGAIGVFGTMNSMYSDCDVLGNYAAFSALKRADGSETARTIMQ
ETYRAATERVMAELETLQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREV
EQLMRNLVEGRNFKFRDGLGEANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDL
ALSQCHGVFAGQSVEGRNFRNQFQPVLRRRVMDMFNNGFLSAKTLTVALSEGAAIC
APSLTAGQTAPAESSFEGDVARVTLGFPKELRVKSRVLFAGASANASEAAKARVASL
QSAYQKPDKRVDILLGPLGFLLKQFHAAIFPNGKPPGSNQPNPQWFWTALQRNQLPA
RLLSREDIETIAFIKKFSLDYGAINFINLAPNNVSELAMYYMANQILRYCDHSTYFINTL
TAIIAGSRRPPSVQAAAAWSAQGGAGLEAGARALMDAVDAHPGAWTSMFASCNLLR
PVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLMGGKNACPLLIFDRTRKFVL
ACPRAGFVCAASSLGGGAHESSLCEQLRGIISEGGAAVASSVFVATVKSLGPRTQQLQI
EDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVAHEAEALVSQLGNAGE
VFNFGDFGCEDDNATPFGGPGAPGPAFAGRKRAFHGDDPFGEGPPDKKGDLTLDML
(SEQ ID NO: 146)

[0457]

[0458] 바실루스 바이러스 phi29로부터의 단일-가닥 DNA 결합 단백질 (유니프롯KB: Q38504.1; NCBI Ref: YP_002004532.1)

MENTNIVKATFDTETLEGQIKIFNAQTGGGQSFKNLPDGTIIEANAIAQYKQVSDTYGD
AKEETVTTIFAADGSLYSAISKTVAEAASDLIDLVTRHKLETFKVKVVQGTSSKGNVF
FSLQLSL (SEQ ID NO: 147)

[0459]

[0460] 단일 가닥 DNA 결합 단백질 [부르크홀데리아 바이러스 DC1] (유니프롯KB: I6NRL7; NCBI Ref: YP_006589943.1)

MASVNKVILVGNLGADPETRYLPSGDAISNIRLATTDRYKDKASGEMKESTEWHRVS
FFGRLAEIVDEYLRKGAPVYIEGRIRTRKWQDNAGQDRYTTEIVAEKMQMLGDRRDG
GERQQRAPQQQQQRTQRNGYADATGRAQPSQRPAAGGGFDEMDDDIPF (SEQ ID
NO: 148)

[0461]

[0462] 일부 실시양태에서, 핵산 편집 도메인은 데아미나제 도메인이다. 일부 실시양태에서, 데아미나제는 시토신 데아미나제 또는 시티딘 데아미나제이다. 일부 실시양태에서, 데아미나제는 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC1 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC2 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3A 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3B 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3C 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3D 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3E 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3F 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3G 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC3H 데아미나제이다. 일부 실시양태에서, 데아미나제는 APOBEC4 데아미나제이다. 일부 실시양태에서, 데아미나제는 활성화-유도된 데아미나제 (AID)이다. 일부 실시양태에서, 데아미나제는 래트 APOBEC1 (서열식별번호: 74)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC1 (서열식별번호: 76)이다. 일부 실시양태에서, 데아미나제는 페트로미존 마리누스 시티딘 데아미나제 1 (pmCDA1)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G (서열식별번호: 60)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G의 단편 (서열식별번호: 83)이다. 일부 실시양태에서, 데아미나제는 D316R_D317R 돌연변이를 포함하는 인간 APOBEC3G 변이체 (서열식별번호: 82)이다. 일부 실시양태에서, 데아미나제는 인간 APOBEC3G 및 서열식별번호: 60에서의 D316R_D317R 돌연변이에 상응하는 돌연변이를 포함하는 단편 (서열식별번호: 84)이다.

[0463] 일부 실시양태에서, 링커는 $(GGGS)_n$ (서열식별번호: 613), $(GGGGS)_n$ (서열식별번호: 607), $(G)_n$ (서열식별번호: 608), $(EAAAK)_n$ (서열식별번호: 609), $(GGS)_n$ (서열식별번호: 610), SGSETPGTSESATPES (서열식별번호: 604), 또는 $(XP)_n$ (서열식별번호: 611) 모티프, 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1

내지 30의 정수이다.

[0464] 적합한 UGI 단백질 및 뉴클레오티드 서열은 본원에 제공되며, 추가의 적합한 UGI 서열은 관련 기술분야의 통상의 기술자에게 공지되어 있고, 예를 들어, 문헌 [Wang et al., Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J. Biol. Chem. 264:1163-1171(1989)]; [Lundquist et al., Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. J. Biol. Chem. 272:21408-21419(1997)]; [Ravishankar et al., X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. Nucleic Acids Res. 26:4880-4887(1998)]; 및 [Putnam et al., Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. J. Mol. Biol. 287:331-346(1999)]에 공개된 것들을 포함하며, 그의 전체 내용은 본원에 참조로 포함된다. 일부 실시양태에서, 임의적 링커는 $(GGS)_n$ (서열식별번호: 610) 모티프를 포함하며, 여기서 n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 19, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20이다. 일부 실시양태에서, 임의적 링커는 $(GGS)_n$ (서열식별번호: 610) 모티프를 포함하며, 여기서 n은 1, 3, 또는 7이다. 일부 실시양태에서, 임의적 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 604)를 포함하며, 이는 또한 실시예에서 XTEN 링커로 지칭된다.

[0465] 일부 실시양태에서, Cas9 니카제는 그의 게놈이 생체내에서 편집되는 유기체에서 비-편집된 가닥 상의 염기의 제거를 추가로 용이하게 할 수 있다. 본원에 기재된 바와 같이 Cas9 니카제는 서열식별번호: 6에서의 D10A 돌연변이, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함할 수 있다. 일부 실시양태에서, 본 개시내용의 Cas9 니카제는 서열식별번호: 6의 돌연변이 840, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 잔기에 히스티딘을 포함할 수 있다. Cas9 니카제를 포함하는 이러한 융합 단백질은 표적 DNA 서열의 단일 가닥, 예를 들어, 편집되지 않는 가닥을 절단할 수 있다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 이 절단은 데아미나제에 의해 생성된 C에서 U로의 편집을 역전시키는 미스-매치 복구 매커니즘을 억제할 수 있다.

[0466] 가이드 RNA와의 Cas9 복합체

[0467] 본 개시내용의 일부 측면은 본원에 제공된 융합 단백질 중 임의의 것, 및 융합 단백질의 Cas9 도메인 (예를 들어, dCas9, 뉴클레아제 활성 Cas9, 또는 Cas9 니카제)에 결합된 가이드 RNA를 포함하는 복합체를 제공한다.

[0468] 일부 실시양태에서, 가이드 RNA는 15 내지 100개 뉴클레오티드 길이이고, 표적 서열에 상보적인 적어도 10개의 인접한 뉴클레오티드의 서열을 포함한다. 일부 실시양태에서, 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 길이이다. 일부 실시양태에서, 가이드 RNA는 표적 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 인접한 뉴클레오티드의 서열을 포함한다. 일부 실시양태에서, 표적 서열은 DNA 서열이다. 일부 실시양태에서, 표적 서열은 포유동물의 게놈에서의 서열이다. 일부 실시양태에서, 표적 서열은 인간의 게놈에서의 서열이다. 일부 실시양태에서, 표적 서열의 3' 말단은 정규 PAM 서열 (NGG)에 바로 인접한다. 일부 실시양태에서, 가이드 RNA는 질환 또는 장애와 연관된 서열에 상보적이다.

[0469] Cas9 융합 단백질을 사용하는 방법

[0470] 본 개시내용의 일부 측면은 본원에 제공된 Cas9 단백질, 융합 단백질, 또는 복합체를 사용하는 방법을 제공한다. 예를 들어, 본 개시내용의 일부 측면은 DNA 분자를 (a) 본원에 제공된 Cas9 단백질 또는 융합 단백질과, 및 적어도 1종의 가이드 RNA와 (여기서, 가이드 RNA는 약 15 내지 100개 뉴클레오티드 길이이고, 표적 서열에 상보적인 적어도 10개의 인접한 뉴클레오티드의 서열을 포함함); 또는 (b) 본원에 제공된 바와 같은 Cas9 단백질, Cas9 융합 단백질, 또는 적어도 1종의 gRNA를 갖는 Cas9 단백질 또는 융합 단백질 복합체와 접촉시키는 단계를 포함하는 방법을 제공한다. 일부 실시양태에서, 표적 서열의 3' 말단은 정규 PAM 서열 (NGG)에 바로 인접하지 않는다. 일부 실시양태에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG, 또는 CAA 서열에 바로 인접한다.

[0471] 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 서열을 포함한다. 일부 실시양태에서, 표적 DNA

서열은 질환 또는 장애와 연관된 점 돌연변이를 포함한다. 일부 실시양태에서, Cas9 단백질, Cas9 융합 단백질, 또는 복합체의 활성은 점 돌연변이의 교정을 발생시킨다. 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 T→C 점 돌연변이를 포함하고, 여기서, 돌연변이체 C 염기의 탈아미노화는 질환 또는 장애와 연관되지 않은 서열을 발생시킨다. 일부 실시양태에서, 표적 DNA 서열은 단백질을 코딩하고, 여기서, 점 돌연변이는 코돈에 있으며, 야생형 코돈에 비해 돌연변이체 코돈에 의해 코딩되는 아미노산의 변화를 발생시킨다. 일부 실시양태에서, 돌연변이체 C의 탈아미노화는 돌연변이체 코돈에 의해 코딩되는 아미노산의 변화를 발생시킨다. 일부 실시양태에서, 돌연변이체 C의 탈아미노화는 야생형 아미노산을 코딩하는 코돈을 발생시킨다. 일부 실시양태에서, 접촉은 대상체의 생체내에서 이루어진다. 일부 실시양태에서, 대상체는 질환 또는 장애를 갖거나, 이로 진단되었다. 일부 실시양태에서, 질환 또는 장애는 낭성 섬유증, 페닐케톤뇨, 표피박리성 과다각화증 (EHK), 샤르코-마리-투스병 유형 4J, 신경모세포종 (NB), 폰 빌레브란트병 (vWD), 선천성 근긴장증, 유전성 신장 아밀로이드증, 확장성 심근병증 (DCM), 유전성 림프부종, 가족성 알츠하이머병, HIV, 프리온 질환, 만성 영아 신경학적 피부 관절 증후군 (CINCA), 데스민-관련 근병증 (DRM), 돌연변이체 PI3KCA 단백질, 돌연변이체 CTNNB1 단백질, 돌연변이체 HRAS 단백질, 또는 돌연변이체 p53 단백질과 연관된 신생물성 질환이다.

[0472] 일부 실시양태는 본원에 제공된 Cas9 DNA 편집 융합 단백질을 사용하는 방법을 제공한다. 일부 실시양태에서, 융합 단백질은 표적 핵염기, 예를 들어, C 잔기를 탈아미노화함으로써 점 돌연변이를 핵산 내로 도입하는 데 사용된다. 일부 실시양태에서, 표적 핵염기의 탈아미노화는 유전적 결함의 교정, 예를 들어, 유전자 생성물에서의 기능의 상실을 발생시키는 점 돌연변이의 교정을 발생시킨다. 일부 실시양태에서, 유전적 결함은 질환 또는 장애, 예를 들어, 리소솜 축적 장애 또는 대사 질환, 예컨대, 예를 들어, 유형 I 당뇨병과 연관된다. 일부 실시양태에서, 본원에 제공된 방법은 탈활성화 점 돌연변이를 질환 또는 장애와 연관된 유전자 생성물을 코딩하는 유전자 또는 대립유전자 내로 도입하는 데 사용된다. 예를 들어, 일부 실시양태에서, (예를 들어, 증식성 질환의 치료에서) Cas9 DNA 편집 융합 단백질을 사용하여 탈활성화 점 돌연변이를 종양유전자 내로 도입하는 방법이 본원에 제공된다. 탈활성화 돌연변이는 일부 실시양태에서, 코딩 서열에서의 조기 정지 코돈을 생성할 수 있으며, 이는 말단절단된 유전자 산물, 예를 들어, 전장 단백질의 기능이 결여된 말단절단된 단백질의 발현을 발생시킨다.

[0473] 일부 실시양태에서, 본원에 제공된 방법의 목적은 게놈 편집을 통해 기능이상 유전자의 기능을 복원하는 것이다. 본원에 제공된 Cas9 데아미나제 융합 단백질은 예를 들어, 인간 세포 배양물에서 질환-연관된 돌연변이를 교정함으로써, 시험관내에서 유전자 편집-기반 인간 치료제에 대해 확인될 수 있다. 본원에 제공된 융합 단백질, 예를 들어, Cas9 도메인 및 핵산 데아미나제 도메인을 포함하는 융합 단백질은 임의의 단일 점 T -> C 또는 A -> G 돌연변이를 교정하는 데 사용될 수 있다는 것이 통상의 기술자에 의해 이해될 것이다. 첫번째 경우, 돌연변이체 C에서 다시 U로의 탈아미노화는 돌연변이를 교정하고, 후자의 경우, 돌연변이체 G와 염기-쌍 형성된 C의 탈아미노화, 그 후 반복 라운드는 돌연변이를 교정한다.

[0474] 제공된 융합 단백질에 의해 시험관내에서 또는 생체내에서 교정될 수 있는 예시적인 질환-관련 돌연변이는 PI3KCA 단백질에서의 H1047R (A3140G) 다형성이다. 포스포이노시티드-3-키나제, 즉, 촉매 알파 서브유닛 (PI3KCA) 단백질은 포스파티딜이노시톨의 이노시톨 고리의 3-OH 기를 인산화하는 작용을 한다. PI3KCA 유전자는 많은 상이한 암종에서 돌연변이되는 것으로 밝혀졌으며, 따라서 이는 강력한 종양유전자인 것으로 간주된다.[37] 사실, A3140G 돌연변이는 몇몇 NCI-60 암 세포주, 예컨대, 예를 들어, HCT116, SKOV3, 및 T47D 세포주에 존재하며, 이들은 아메리칸 타입 컬쳐 콜렉션(American Type Culture Collection) (ATCC)으로부터 용이하게 입수가능하다.[38]

[0475] 일부 실시양태에서, 교정되는 돌연변이를 운반하는 세포, 예를 들어, PI3KCA 유전자의 엑손 20에서 점 돌연변이, 예를 들어, A3140G 점 돌연변이를 운반하는 세포 (PI3KCA 단백질에서의 H1047R 치환을 발생시킴)는 Cas9 데아미나제 융합 단백질 및 융합 단백질을 표적화하는 적절하게 설계된 sgRNA를 코딩 PI3KCA 유전자에서의 각각의 돌연변이 부위에 코딩하는 발현 구축물과 접촉된다. sgRNA가 융합 효소를 PI3KCA 유전자 내에 있는 비-C 잔기에 표적화하도록 설계된 대조 실험이 수행될 수 있다. 처리된 세포의 게놈 DNA를 추출하고, PI3KCA 유전자의 관련 서열을 PCR 증폭시키고, 서열분석하여 인간 세포 배양물에서의 융합 단백질의 활성을 평가할 수 있다.

[0476] PI3KCA에서의 점 돌연변이를 교정하는 실시예는 예시 목적으로 제공되며, 본 개시내용을 제한하는 것을 의미하지 않는 것으로 이해될 것이다. 통상의 기술자는 본원에 개시된 DNA-편집 융합 단백질이 다른 점 돌연변이 및 다른 암과, 및 다른 증식성 질환을 포함한 암 이외의 질환과 연관된 돌연변이를 교정하는 데 사용될 수 있다는

것을 이해할 것이다.

[0477] 질환-연관된 유전자 및 대립유전자에서의 점 돌연변이의 성공적인 교정은 치료제 및 기본적 연구에서의 적용과 함께 유전자 교정을 위한 새로운 전략을 열고 있다. Cas9 및 데아미나제 효소 또는 도메인의 개시된 융합물과 같은 부위-특이적 단일-염기 변형 시스템은 또한 "역" 유전자 요법에서의 적용을 가지며, 여기서, 특정 유전자 기능은 고의로 억제되거나 폐지된다. 이들 경우에, Trp (TGG), Gln (CAA 및 CAG), 또는 Arg (CGA) 잔기를 조기 정지 코돈 (TAA, TAG, TGA)으로 부위-특이적으로 돌연변이시키는 것은 시험관내에서, 생체외에서, 또는 생체내에서 단백질 기능을 폐지하는 데 사용될 수 있다.

[0478] 본 개시내용은 본원에 제공된 Cas9 DNA 편집 융합 단백질에 의해 교정될 수 있는 점 돌연변이와 연관되거나 그에 의해 유발되는 질환으로 진단된 대상체의 치료 방법을 제공한다. 예를 들어, 일부 실시양태에서, 이러한 질환, 예를 들어, 상기 기재된 바와 같은 PI3KCA 점 돌연변이와 연관된 암을 갖는 대상체에게, 점 돌연변이를 교정하거나, 탈활성화 돌연변이를 질환-연관된 유전자 내로 도입하는 Cas9 데아미나제 융합 단백질의 유효량을 투여하는 단계를 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, 탈활성화 돌연변이를 질환-연관된 유전자 내로 도입함으로써 치료될 수 있는 다른 질환은 관련 기술분야의 통상의 기술자에게 공지되어 있을 것이며, 본 개시내용은 이와 관련하여 제한되지 않는다.

[0479] 본 개시내용은 추가의 질환 또는 장애, 예를 들어, 데아미나제-매개된 유전자 편집에 의해 교정될 수 있는 점 돌연변이와 연관되거나 그에 의해 유발되는 질환 또는 장애의 치료 방법을 제공한다. 일부 이러한 질환은 본원에 기재되어 있으며, 본원에 제공된 전략 및 융합 단백질로 치료될 수 있는 추가의 적합한 질환은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다. 예시적인 적합한 질환 및 장애는 하기에 열거된다. 각각의 서열에서의 특정 위치 또는 잔기의 넘버링은 특정한 단백질 및 사용되는 넘버링 스킴에 따라 달라지는 것으로 이해될 것이다. 넘버링은 예를 들어, 성숙 단백질의 전구체 및 성숙 단백질 그 자체에서 상이할 수 있으며, 종에서 종까지의 서열의 차이는 넘버링에 영향을 미칠 수 있다. 관련 기술분야이 통상의 기술자는 관련 기술분야에 널리 공지된 방법에 의해, 예를 들어 서열 정렬 및 상동 잔기의 측정에 의해 임의의 상동 단백질에서 및 각각의 코딩 핵산에서 각각의 잔기를 확인할 수 있을 것이다. 예시적인 적합한 질환 및 장애로는 비제한적으로 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 결함다는 것을 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 막 당단백질 IX에서의 위치 55 또는 상동 잔기에서 페닐알라닌에서 세린으로의 돌연변이, 또는 잔기 24 또는 상동 잔기에서 시스테인에서 아르기닌으로의 (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 (개시제 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-항티로신의 프로세싱된 형태에서의 위치 54 또는 55 (개시제 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 류신의 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; 82: 66-72] 참조, 또한 유니프롯 데이터베이스에서 수탁 번호 P04275 참조; 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.

[0480] 본원에 개시된 바와 같은 Cas9:핵산 편집 효소/도메인 융합 단백질을 표적 부위, 예를 들어, 편집되는 점 돌연변이를 포함하는 부위에 표적화하기 위해, 전형적으로 Cas9:핵산 편집 효소/도메인 융합 단백질을 가이드 RNA, 예를 들어, sgRNA와 함께 공동-발현시킬 필요가 있다는 것이 관련 기술분야의 통상의 기술자에게 명백할 것이다. 본원의 다른 곳에서 보다 상세히 설명된 바와 같이, 가이드 RNA는 전형적으로 Cas9 결합을 허용하는 tracrRNA 프레임워크, 및 가이드 서열을 포함하며, 이는 Cas9:핵산 편집 효소/도메인 융합 단백질에 서열 특이성을 부여한다. 일부 실시양태에서, 가이드 RNA는 구조 5'-[가이드 서열]-guuuuagagcuagaaauagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3' (서열식별번호: 618)를 포함하고, 여기서, 가이드 서열은 표적 서열에 상보적인 서열을 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오티드 길이이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특이적 게놈 표적 부위에 표적화하는 데 적합한 가이드 RNA의 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 명백할 것이다. 이러한 적합한 가이드 RNA 서열은 전형적으로 편집되는 표적 뉴클레오티드의 50개 뉴클레오티드 상류 또는 하류 내의 핵산 서열에 상보적인 가이드 서열을 포함한다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특이적 표적 서열에 표적화하는 데 적합한 일부 예시적인 가이드 RNA 서열은 하기에 제공된다.

[0481] 염기 편집제 효율

[0482] 본 개시내용의 일부 측면은 본원에 제공된 염기 편집제 중 임의의 것이 유의한 비율의 indel을 생성하지 않고 특이적 뉴클레오티드 염기를 변형시킬 수 있다는 인식에 기초한다. 본원에 사용된 "indel"은 핵산 내의 뉴클레오티드 염기의 삽입 또는 결실을 지칭한다. 이러한 삽입 또는 결실은 유전자의 코딩 영역 내의 격자 이동 돌연변이를 발생시킬 수 있다. 일부 실시양태에서, 핵산에서 다수의 삽입 또는 결실 (즉, indel)을 생성하지 않고, 핵산 내의 특이적 뉴클레오티드를 효율적으로 변형시키는 (예를 들어 돌연변이시키거나 탈아미노화하는) 염기 편집제를 생성하는 것이 바람직하다. 특정 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 보다 큰 비율의 의도된 변형 (예를 들어, 점 돌연변이 또는 탈아미노화) 대 indel을 생성할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제는 1:1 초과인 의도된 점 돌연변이 대 indel의 비를 생성할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제는 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 200:1, 적어도 300:1, 적어도 400:1, 적어도 500:1, 적어도 600:1, 적어도 700:1, 적어도 800:1, 적어도 900:1, 또는 적어도 1000:1, 또는 그보다 큰 의도된 점 돌연변이 대 indel의 비를 생성할 수 있다. 의도된 돌연변이 및 indel의 수는 임의의 적합한 방법, 예를 들어 하기 실시예에서 사용된 방법을 사용하여 측정될 수 있다.

[0483] 일부 실시양태에서, 본원에 제공된 염기 편집제는 핵산의 영역에서의 indel의 형성을 제한할 수 있다. 일부 실

시양태에서, 영역은 염기 편집제에 의해 표적화된 뉴클레오티드, 또는 염기 편집제에 의해 표적화된 뉴클레오티드의 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 뉴클레오티드 내의 영역에 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 핵산의 영역에서의 indel의 형성을 1% 미만, 1.5% 미만, 2% 미만, 2.5% 미만, 3% 미만, 3.5% 미만, 4% 미만, 4.5% 미만, 5% 미만, 6% 미만, 7% 미만, 8% 미만, 9% 미만, 10% 미만, 12% 미만, 15% 미만, 또는 20% 미만으로 제한할 수 있다. 핵산 영역에서 형성된 indel의 수는 핵산 (예를 들어, 세포의 게놈 내의 핵산)이 염기 편집제에 노출된 시간의 양에 따라 달라질 수 있다. 일부 실시양태에서, indel의 수 또는 비율은 핵산 (예를 들어, 세포의 게놈 내의 핵산)을 염기 편집제에 노출시킨 후 적어도 1시간, 적어도 2시간, 적어도 6시간, 적어도 12시간, 적어도 24시간, 적어도 36시간, 적어도 48시간, 적어도 3일, 적어도 4일, 적어도 5일, 적어도 7일, 적어도 10일, 또는 적어도 14일에 측정된다.

[0484] 본 개시내용의 일부 측면은 본원에 제공된 염기 편집제 중 임의의 것이 유의한 수의 비의도된 돌연변이, 예컨대 비의도된 점 돌연변이를 생성하지 않고 핵산 (예를 들어 대상체의 게놈 내의 핵산)에서 의도된 돌연변이, 예컨대 점 돌연변이를 효율적으로 생성할 수 있다는 인식에 기초한다. 일부 실시양태에서, 의도된 돌연변이는 의도된 돌연변이를 생성하도록 특이적으로 설계된 gRNA에 결합된 특이적 염기 편집제에 의해 생성된 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 질환 또는 장애와 연관된 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 질환 또는 장애와 연관된 시토신 (C)에서 티민 (T)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 질환 또는 장애와 연관된 구아닌 (G)에서 아데닌 (A)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 유전자의 코딩 영역 내의 시토신 (C)에서 티민 (T)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 유전자의 코딩 영역 내의 구아닌 (G)에서 아데닌 (A)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 정지 코돈, 예를 들어, 유전자의 코딩 영역 내의 조기 정지 코돈을 생성하는 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 정지 코돈을 제거하는 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 유전자의 스플라이싱을 변경시키는 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 유전자의 조절 서열 (예를 들어, 유전자 프로모터 또는 유전자 리프레서)을 변경시키는 돌연변이이다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 1:1 초과인 의도된 돌연변이 대 비의도된 돌연변이의 비 (예를 들어, 의도된 점 돌연변이:비의도된 점 돌연변이)를 생성할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 150:1, 적어도 200:1, 적어도 250:1, 적어도 500:1, 또는 적어도 1000:1, 또는 그보다 큰 의도된 돌연변이 대 비의도된 돌연변이의 비 (예를 들어, 의도된 점 돌연변이:비의도된 점 돌연변이)를 생성할 수 있다. 본원에서 "염기 편집제 효율" 섹션에 기재된 염기 편집제의 특징은 본원에 제공된 융합 단백질, 또는 융합 단백질을 사용하는 방법 중 임의의 것에 적용될 수 있다는 것이 인지되어야 한다.

[0485] 핵산을 편집하는 방법

[0486] 본 개시내용의 일부 측면은 핵산을 편집하는 방법을 제공한다. 일부 실시양태에서, 방법은 핵산의 핵염기 (예를 들어, 이중-가닥 DNA 서열의 염기 쌍)를 편집하는 방법이다. 일부 실시양태에서, 방법은 a) 핵산의 표적 영역 (예를 들어, 이중-가닥 DNA 서열)을 염기 편집제 (예를 들어, 시티딘 데아미나제 도메인에 융합된 Cas9 도메인) 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계이며, 여기서 표적 영역은 표적화된 핵염기 쌍을 포함하는 것인 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 표적 영역의 단일 가닥에서의 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, 및 d) 상기 표적 영역의 1개 이하의 가닥을 절단하는 단계이며, 여기서 제1 핵염기 염기에 상보적인 제3 핵염기는 제2 핵염기에 상보적인 제4 핵염기에 의해 대체된 것인 단계를 포함하고; 방법은 핵산에서 20% 미만의 indel 형성을 발생시킨다. 일부 실시양태에서, 단계 b는 생략됨이 인지되어야 한다. 일부 실시양태에서, 제1 핵염기는 시토신이다. 일부 실시양태에서, 제2 핵염기는 탈아미노화된 시토신, 또는 우라실이다. 일부 실시양태에서, 제3 핵염기는 구아닌이다. 일부 실시양태에서, 제4 핵염기는 아데닌이다. 일부 실시양태에서, 제1 핵염기는 시토신이고, 제2 핵염기는 탈아미노화된 시토신, 또는 우라실이고, 제3 핵염기는 구아닌이고, 제4 핵염기는 아데닌이다. 일부 실시양태에서, 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만, 또는 0.1% 미만의 indel 형성을 발생시킨다. 일부 실시양태에서, 방법은 제2 핵염기를 제4 핵염기에 상보적인 제5 핵염기로 대체함으로써, 의도된 편집된 염기 쌍 (예를 들어, C:G -> T:A)을 생성하는 것을 추가로 포함한다. 일부 실시양태에서, 제5 핵염기는 티민이다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 5%가 편집된다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 또는 50%가 편집된다.

[0487] 일부 실시양태에서, 표적 뉴클레오티드에서의 의도된 생성물 대 비의도된 생성물의 비는 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 또는 200:1, 또는 그보다 크다. 일부 실시양태에서, 의도된 점 돌연변이 대 indel 형성의 비는 1:1, 10:1, 50:1, 100:1, 500:1, 또는 1000:1 초과, 또는 그보다 크다. 일부 실시양태에서, 절단된 단일 가닥 (니킹된 가닥)은 가이드 핵산에 혼성화된다. 일부 실시양태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥에 대향한다. 일부 실시양태에서, 염기 편집제는 Cas9 도메인을 포함한다. 일부 실시양태에서, 제1 염기는 시토신이고, 제2 염기는 G, C, A, 또는 T가 아니다. 일부 실시양태에서, 제2 염기는 우라실이다. 일부 실시양태에서, 제1 염기는 시토신이다. 일부 실시양태에서, 제2 염기는 G, C, A, 또는 T가 아니다. 일부 실시양태에서, 제2 염기는 우라실이다. 일부 실시양태에서, 염기 편집제는 편집된 가닥의 염기 절제 복구를 억제한다. 일부 실시양태에서, 염기 편집제는 비-편집된 가닥을 보호하거나 그에 결합한다. 일부 실시양태에서, 염기 편집제는 UGI 활성을 포함한다. 일부 실시양태에서, 염기 편집제는 니카제 활성을 포함한다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 상류이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오티드 상류이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 하류이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오티드 하류 스트림이다. 일부 실시양태에서, 방법은 정규 (예를 들어, NGG) PAM 부위를 요구하지 않는다. 일부 실시양태에서, 핵염기 편집제는 링커를 포함한다. 일부 실시양태에서, 링커는 1 내지 25개 아미노산 길이이다. 일부 실시양태에서, 링커는 5 내지 20개 아미노산 길이이다. 일부 실시양태에서, 링커는 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 아미노산 길이이다. 일부 실시양태에서, 표적 영역은 표적 윈도우를 포함하고, 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 실시양태에서, 표적 윈도우는 1 내지 10개의 뉴클레오티드를 포함한다. 일부 실시양태에서, 표적 윈도우는 1 내지 9, 1 내지 8, 1 내지 7, 1 내지 6, 1 내지 5, 1 내지 4, 1 내지 3, 1 내지 2, 또는 1개 뉴클레오티드 길이이다. 일부 실시양태에서, 표적 윈도우는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오티드 길이이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 표적 윈도우 내에 있다. 일부 실시양태에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부 실시양태에서, 방법은 본원에 제공된 염기 편집제 중 임의의 것을 사용하여 수행된다. 일부 실시양태에서, 표적 윈도우는 탈아미노화 윈도우이다.

[0488] 일부 실시양태에서, 본 개시내용은 뉴클레오티드를 편집하는 방법을 제공한다. 일부 실시양태에서, 본 개시내용은 이중-가닥 DNA 서열의 핵염기 쌍을 편집하는 방법을 제공한다. 일부 실시양태에서, 방법은 a) 이중-가닥 DNA 서열의 표적 영역을 염기 편집제 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계이며, 여기서 표적 영역은 표적 핵염기 쌍을 포함하는 것인 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 표적 영역의 단일 가닥에서의 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, d) 상기 표적 영역의 1개 이하의 가닥을 절단하는 단계를 포함하고, 제1 핵염기 염기에 상보적인 제3 핵염기는 제2 핵염기에 상보적인 제4 핵염기에 의해 대체되고, 제2 핵염기는 제4 핵염기에 상보적인 제5 핵염기로 대체됨으로써, 의도된 편집된 염기 쌍을 생성하고, 의도된 편집된 염기 쌍을 생성하는 효율은 적어도 5%이다. 일부 실시양태에서, 단계 b는 생략됨이 인지되어야 한다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 5%가 편집된다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 또는 50%가 편집된다. 일부 실시양태에서, 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만, 또는 0.1% 미만의 indel 형성을 유발한다. 일부 실시양태에서, 표적 뉴클레오티드에서의 의도된 생성물 대 비의도된 생성물의 비는 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 또는 200:1, 또는 그보다 크다. 일부 실시양태에서, 의도된 점 돌연변이 대 indel 형성의 비는 1:1, 10:1, 50:1, 100:1, 500:1, 또는 1000:1 초과, 또는 그보다 크다. 일부 실시양태에서, 절단된 단일 가닥은 가이드 핵산에 혼성화된다. 일부 실시양태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥에 대향한다. 일부 실시양태에서, 제1 염기는 시토신이다. 일부 실시양태에서, 제2 핵염기는 G, C, A, 또는 T가 아니다. 일부 실시양태에서, 제2 염기는 우라실이다. 일부 실시양태에서, 염기 편집제는 편집된 가닥의 염기 절제 복구를 억제한다. 일부 실시양태에서, 염기 편집제는 비-편집된 가닥을 보호하거나 그에 결합한다. 일부 실시양태에서, 핵염기 편집제는 UGI 활성을 포함한다. 일부 실시양태에서, 핵염기 편집제는 니카제 활성을 포함한다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 상류이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오티드 상류이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 하류이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오티드 하류 스트림이다. 일부 실시양태에서, 방법은 정규 (예를 들어, NGG) PAM 부위를 요구하지 않는다. 일부 실시양태

에서, 핵염기 편집제는 링커를 포함한다. 일부 실시양태에서, 링커는 1 내지 25개 아미노산 길이이다. 일부 실시양태에서, 링커는 5 내지 20개 아미노산 길이이다. 일부 실시양태에서, 링커는 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 아미노산 길이이다. 일부 실시양태에서, 표적 영역은 표적 윈도우를 포함하고, 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 실시양태에서, 표적 윈도우는 1 내지 10개의 뉴클레오티드를 포함한다. 일부 실시양태에서, 표적 윈도우는 1 내지 9, 1 내지 8, 1 내지 7, 1 내지 6, 1 내지 5, 1 내지 4, 1 내지 3, 1 내지 2, 또는 1개 뉴클레오티드 길이이다. 일부 실시양태에서, 표적 윈도우는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오티드 길이이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 표적 윈도우 내에서 발생한다. 일부 실시양태에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부 실시양태에서, 핵염기 편집제는 본원에 제공된 염기 편집제 중 어느 하나이다.

[0489] 제약 조성물

[0490] 일부 실시양태에서, 본원에 기재된 융합 단백질, gRNA, 및/또는 복합체 중 임의의 것은 제약 조성물의 일부로서 제공된다. 일부 실시양태에서, 제약 조성물은 본원에 제공된 융합 단백질 중 임의의 것을 포함한다. 일부 실시양태에서, 제약 조성물은 본원에 제공된 복합체 중 임의의 것을 포함한다. 일부 실시양태에서, 제약 조성물은 gRNA 및 양이온성 지질과의 복합체를 형성하는 RNA-가이드된 뉴클레아제 (예를 들어, Cas9)를 포함하는 리보핵단백질을 포함한다. 일부 실시양태에서 제약 조성물은 gRNA, 핵산 프로그램가능한 DNA 결합 단백질, 양이온성 지질, 및 제약상 허용되는 부형제를 포함한다. 제약 조성물은 1종 이상의 추가의 치료 활성 물질을 임의로 포함할 수 있다.

[0491] 일부 실시양태에서, 본원에 제공된 조성물은 대상체 내에서 표적화된 게놈 변형을 실시하기 위해 대상체, 예를 들어 인간 대상체에게 투여된다. 일부 실시양태에서, 세포는 대상체로부터 수득되고 본원에 제공된 제약 조성물 중 임의의 것과 접촉된다. 일부 실시양태에서, 대상체로부터 제거되고, 생체외에서 제약 조성물과 접촉된 세포는 임의로 목적하는 게놈 변형이 세포에서 실시되거나 검출된 후, 대상체 내로 재-도입된다. 뉴클레아제를 포함하는 제약 조성물을 전달하는 방법은 공지되어 있고, 예를 들어, 미국 특허 번호 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; 및 7,163,824에 기재되며, 이들 모두의 개시내용은 그 전문이 본원에 참조로 포함된다. 본원에 제공된 제약 조성물의 설명은 원칙적으로 인간에의 투여에 적합한 제약 조성물에 관한 것이지만, 이러한 조성물은 일반적으로 모든 종류의 동물 또는 유기체에의 투여에 적합하다는 것이 통상의 기술자에 의해 이해될 것이다. 인간에 대한 투여에 적합한 제약 조성물을 다양한 동물에게 투여하기에 적합하게 만들기 위해 조성물을 변형시키는 것은 널리 이해되며, 통상의 숙련된 수의학 약리학자는, 존재하는 경우, 단지 통상의 실험에 의해 이러한 변형을 설계하고/거나 수행할 수 있다. 제약 조성물이 투여되는 것으로 고려되는 대상체는 인간 및/또는 다른 영장류; 포유동물, 가축, 애완용 동물, 및 상업적으로 관련된 포유동물, 예컨대 소, 돼지, 말, 양, 고양이, 개, 마우스, 및/또는 래트; 및/또는 조류 (상업적으로 관련된 조류, 예컨대 치킨, 오리, 거위 및/또는 칠면조 포함)를 포함하나 이에 제한되지는 않는다.

[0492] 본원에 기재된 제약 조성물의 제제는 약리학 기술분야에 공지되어 있거나 또는 이후에 개발된 임의의 방법에 의해 제조될 수 있다. 일반적으로, 이러한 제조 방법은 활성 성분(들)을 부형제 및/또는 1종 이상의 보조 성분과 회합시킨 후, 필요할 경우 및/또는 바람직할 경우, 생성물을 목적하는 단일- 또는 다중-투여 단위로 성형 및/또는 패키징하는 단계를 포함한다.

[0493] 제약 제제는 추가로 제약상 허용되는 부형제를 추가로 포함할 수 있고, 이는 본원에 사용된 바와 같은, 임의의 모든 용매, 분산 매질, 희석제, 또는 다른 액체 비히클, 분산액 또는 현탁 보조제, 표면 활성제, 등장화제, 증점 또는 유화제, 보존제, 고체 결합제, 윤활제 등을 목적하는 특정한 투여에 적합한 것으로서 포함한다. 그 전문이 본원에 참조로 포함된 문헌 [Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006)]에는 제약 조성물을 제형화하는 데 사용된 각종 부형제 및 그의 제조를 위한 공지된 기술이 개시되어 있다. 또한, 뉴클레아제를 포함하는 제약 조성물을 제조하기 위한 추가의 적합한 방법, 시약, 부형제 및 용매에 대해 그 전문이 본원에 참조로 포함된 PCT 출원 PCT/US2010/055131 (공개 번호 WO2011053982 A8, 2010년 11월 2일에 출원됨)을 참조한다. 임의의 통상적인 부형제 매질이, 예컨대 임의의 목적하지 않은 생물학적 효과를 생성하거나 또는 달리 제약 조성물의 임의의 다른 성분(들)과 유해한 방식으로 상호작용함으로써 물질 또는 그의 유도체와 비상용성인 경우를 제외하고, 그의 사용은 본 개시내용의 범주 내에 있는 것으로 고려된다.

[0494] 일부 실시양태에서, 본 발명에 따르는 조성물은 하기 중 1종 이상을 포함하나 이에 제한되지는 않는 다양한 질

환, 장애 및/또는 상태 중 임의의 것을 치료하는 데 사용될 수 있다: 자가면역 장애 (예를 들어, 당뇨병, 루푸스, 다발성 경화증, 건선, 류마티스성 관절염); 염증성 장애 (예를 들어, 관절염, 골반 염증성 질환); 감염성 질환 (예를 들어, 바이러스 감염 (예를 들어, HIV, HCV, RSV), 박테리아 감염, 진균 감염, 패혈증); 신경계 장애 (예를 들어, 알츠하이머병, 헌팅턴병; 자폐증; 뒤시엔느 근육 이영양증); 심혈관 장애 (예를 들어, 아테롬성 동맥경화증, 고콜레스테롤혈증, 혈전증, 응고 장애, 혈관신생 장애, 예컨대 황반 변성); 증식성 장애 (예를 들어, 암, 양성 신생물); 호흡기 장애 (예를 들어, 만성 폐쇄성 폐 질환); 소화 장애 (예를 들어, 염증성 장 질환, 궤양); 근골격 장애 (예를 들어, 섬유근육통, 관절염); 내분비, 대사 및 영양 장애 (예를 들어, 당뇨병, 골다공증); 비뇨기 장애 (예를 들어, 신질환); 심리 장애 (예를 들어, 우울증, 정신분열증); 피부 장애 (예를 들어, 상처, 습진); 혈액 및 림프성 장애 (예를 들어, 빈혈, 혈우병) 등.

[0495] 키트, 벡터, 세포

[0496] 본 개시내용의 일부 측면은 (a) 본원에 제공된 바와 같은 Cas9 단백질 또는 Cas9 융합 단백질을 코딩하는 뉴클레오티드 서열; 및 (b) (a)의 서열의 발현을 유도하는 이종 프로모터를 포함하는 핵산 구축물을 포함하는 키트를 제공한다. 일부 실시양태에서, 키트는 가이드 RNA 백본을 코딩하는 발현 구축물을 추가로 포함하고, 구축물은 표적 서열과 동일하거나 그에 상보적인 핵산 서열의 가이드 RNA 백본 내로의 클로닝을 허용하도록 위치된 클로닝 부위를 포함한다.

[0497] 본 개시내용의 일부 측면은 본원에 제공된 바와 같은 융합 단백질의 Cas9 단백질을 코딩하는 폴리뉴클레오티드를 제공한다. 본 개시내용의 일부 측면은 이러한 폴리뉴클레오티드를 포함하는 벡터를 제공한다. 일부 실시양태에서, 벡터는 폴리뉴클레오티드의 발현을 유도하는 이종 프로모터를 포함한다.

[0498] 본 개시내용의 일부 측면은 본원에 제공된 바와 같은 Cas9 단백질, 융합 단백질, 융합 단백질을 코딩하는 핵산 분자, Cas9 단백질 및 gRNA를 포함하는 복합체, 및/또는 벡터를 포함하는 세포를 제공한다.

[0499] 상기 리포터 시스템의 예시적인 실시양태의 설명은 단지 예시 목적으로 제공되며, 제한하는 것으로 의도되지 않는다. 추가의 리포터 시스템, 예를 들어, 상기 상세히 기재된 예시적인 시스템의 변형이 또한 본 개시내용에 의해 포함된다.

[0500] 실시예

[0501] 실시예 1: Cas9 데아미나제 융합 단백질

[0502] 다수의 Cas9:데아미나제 융합 단백질을 생성하고, 생성된 융합물의 데아미나제 활성을 특징화하였다. 하기 데아미나제를 시험하였다:

[0503] 인간 AID (hAID):

[0504] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPYLSLRIFTAR LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS VRLSRQLRRILLPLYEVDDLRDAFRTLGLLD (SEQ ID NO: 49)

[0505] 인간 AID-DC (hAID-DC, 7-배 증가된 활성을 갖는 hAID의 말단절단된 버전):

[0506] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENS VRLSRQLRRILL (SEQ ID NO: 50)

[0507] 래트 APOBEC1 (rAPOBEC1):

[0508] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 76)

[0509] 인간 APOBEC1 (hAPOBEC1)

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNT
TNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVA
RLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPP
LWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIH
[0510] PSVAWR (SEQ ID NO: 74)

[0511] 페트로미존 마리누스 (칠성장어) CDA1 (pmCDA1):

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
[0512] NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 81)

[0513] 인간 APOBEC3G (hAPOBEC3G):

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVT
LTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELF
EPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHN
DTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSW
SPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFK
[0514] HCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 60)

[0515] ssDNA에 대한 데아미나제 활성. 탈아미노화를 위한 USER (우라실-특이적 절제 시약) 효소-기반 검정을 사용하여 단일-가닥 DNA (ssDNA) 기질에 대한 다양한 데아미나제의 활성을 시험하였다. USER 효소는 뉴 잉글랜드 바이오랩스(New England Biolabs)로부터 입수하였다. 상이한 위치에 표적 시토신 잔기를 갖는 ssDNA 기질을 제공받았다. ssDNA 시토신 표적 잔기의 탈아미노화는 표적 시토신의 우라실로의 전환을 발생시킨다. USER 효소는 우라실 염기를 절제하고, 그 위치에서 ssDNA 백본을 절단하며, 이는 ssDNA 기질을 DNA의 2개의 보다 짧은 단편으로 절단한다. 일부 검정에서, ssDNA 기질은 한 말단 상에 염료로, 예를 들어, 5' Cy3 표지 (하기 식에서 *)로 표지된다. 탈아미노화, 절제, 및 가닥의 절단시, 기질은 전기영동으로 처리될 수 있으며, 기질 및 그로부터 방출된 임의의 단편은 표지를 검출함으로써 가시화될 수 있다. Cy5가 영상인 경우, 단지 표지를 갖는 단편만이 영상화를 통해 가시적일 것이다.

[0516] 한 USER 효소 검정에서, 시험된 다양한 데아미나제의 표적 서열을 매칭시킨 ssDNA 기질을 사용하였다. 시험되는 데아미나제를 코딩하는 발현 카세트를 실험실에서 이전에 사용된 CMV 백본 플라스미드 (애드진(Addgene) 플라스미드 52970) 내로 삽입하였다. 데아미나제 단백질을 TNT 퀵 커플드(Quick Coupled) 전사/번역 시스템 (프로메가(Promega))을 제조업체의 권장사항에 따라 사용하여 발현시켰다. 90분의 인큐베이션 후, 5 mL의 용해물을 5' Cy3-표지된 ssDNA 기질 및 1 단위의 USER 효소 (NEB)와 함께 3시간 동안 인큐베이션하였다. DNA를 10% TBE PAGE 겔 상에서 분해하고, DNA를 Cy-염료 영상화를 사용하여 영상화하였다. USER 효소 검정의 개략적 표현이 도 41에 나타나 있다.

[0517] 도 1은 ssDNA 기질, 예컨대 도엔치(Doench) 1, 도엔치 2, G7' 및 VEGF 표적 2에 대한 시험된 데아미나제의 데아미나제 활성을 나타낸다. rAPOBEC1 효소는 정규 NGG PAM으로는 단일-가닥 DNA 기질에 대해 실질적인 양의 탈아미노화를 나타내었지만, 음성 대조군 비-정규 NNN PAM으로는 그렇지 않았다. APOBEC 패밀리 데아미나제를 갖는 Cas9 융합 단백질을 생성하였다. 하기 융합 아키텍처를 구축하고, ssDNA에 대해 시험하였다:

rAPOBEC1-GGS-*dCas9* 1차 서열

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
[0518] YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR

LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGS*DKKY*
SIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT
FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA
VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 149)

rAPOBEC1-$(GGS)_3$-***dCas9*** 1차 서열

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGGSGGSG
GS*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG*
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKE
IFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM

[0519]

NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 150)

dCas9-GGS-rAPOBEC1

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGDGGSMSSETGPVAVDPTLRRRIEPHEFEVF FDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSIT WFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMT EQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQL TFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 151)

dCas9-GGS_3-rAPOBEC1

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE

[0520]

DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHD
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE
EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE
LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ
KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSMSSETGPVAVDPTLRRRIEP
HEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCP
NTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSG
VTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNIL
RRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 152)

rAPOBEC1-XTEN-***dCas9*** 1차 서열

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL
YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVR
LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPG
TSESATPES***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL***
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF
LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIA
QLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS
RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 153)

[0521]

[0522] 도 2는 N-말단 데아미나제 융합물이 단일 가닥 DNA 기질에 대해 유의한 활성을 나타내었음을 나타낸다. 이러한 이유로, 단지 N-말단 아키텍처만을 추가의 실험을 위해 선택하였다.

[0523] 도 3은 데아미나제-dCas9:sgRNA 복합체에 의한 이중 가닥 DNA 기질 결합을 예시한다. 다수의 이중 가닥 데아미나제 기질 서열을 생성하였다. 서열은 하기에 제공된다. 도 3에 따른 구조는 이들 서열에서 확인된다 (36bp: 밑줄표시됨, sgRNA 표적 서열: 볼드체; PAM: 박스표시됨; 21bp: 이탤릭체). 모든 기질은 5'-Cy3 표지로 표지하였다:

2:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTC**CCGCGGATTTATTTATTTA**
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 85)
3:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCT**TCCGCGGATTTATTTATTT**
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 86)
4:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TTCCGCGGATTTATTTATT**
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 87)
5:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTCCGCGGATTTATTTAT**
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 88)
6:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TATTCCGCGGATTTATTTA**
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 89)
7:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TTATTCCGCGGATTTATTT**
ATGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 90)
8:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTCCGCGGATTTATT**
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 91)
9:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TATTATTCCGCGGATTTAT**
TTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 92)
10:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATATTCCGCGGATTT**
ATTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 93)
11:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TATTATATTCCGCGGATT**
TATGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 94)
12:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TTATTATATTCCGCGGAT**
TTTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 95)
13:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTATATTCCGCGGA**
TTTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 96)
14:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**TATTATTATATTCCGCGG**
ATTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 97)
15:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTATTATTACCGCG**
GATGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 98)
18:GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTATTATTATTACCG**
CTGG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 99)
“-“:
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGTA**ATATTAATTTATTTATTTAA**T
GG*ATGACCTCTGGATCCATGGAC*-3’ (SEQ ID NO: 100)
8U:GTAGGTAGTTAGGATGAATGGAAGGTTGGTGTAG**ATTATTATCUGCGGATTTAT**
TGG*ATGACCTCTGGATCCATGGACAT*-3’ (SEQ ID NO: 101)

[0524]

[0525] *"8U"를 제외한 모든 기질에서, 도 3에서의 상부 가닥은 여기에 특정된 서열의 보체이다. "8U"의 경우, U 반대편에는 "G"가 있다.

[0526] 도 4는 이중 가닥 DNA 탈아미노화 검정의 결과를 나타낸다. 융합물을 발현시키고, Ni-NTA 및 세파로스 크로마토그래피 둘 다를 통해 N-말단 His6 태그로 정제하였다. dsDNA 기질에 대한 탈아미노화를 평가하기 위해, 이전의 슬라이드 상에 나타난 다양한 dsDNA 기질을 1:8 dsDNA:융합 단백질 비로 인큐베이션하고, 37℃에서 2시간 동안 인큐베이션하였다. 융합물의 dCas9 부분이 DNA에 결합하면, 이는 DNA에의 USER 효소의 접근을 차단한다. 따라서, 융합 단백질을 인큐베이션 후에 변성시키고, dsDNA를 스핀 칼럼 상에서 정제한 후, USER 효소와 함께 45분 동안 인큐베이션하고, 생성된 DNA 기질 및 기질 단편을 10% TBE-우레아 겔 상에서 분해하였다.

[0527] 도 5는 Cas9 융합물이 이중-가닥 DNA 표적 서열의 위치 3 내지 11 (도 3에서의 개략도에 따라 넘버링됨)을 표적화할 수 있다는 것을 입증한다. 상부 겔: 1 μM rAPOBEC1-GGS-dCas9, 125 nM dsDNA, 1 당량 sgRNA. 중간 겔: 1 μM rAPOBEC1-$(GGS)_3$-dCas9, 125 nM dsDNA, 1 당량 sgRNA. 하부 겔: 1.85 μM rAPOBEC1-XTEN-dCas9, 125 nM dsDNA, 1 당량 sgRNA. 이들 겔로부터의 데이터에 기초하여, 위치 3 내지 11 (도 3에서의 넘버링에 따름)은 시험된 융합 단백질에 의해 표적화되는 데아미나제의 활성에 충분히 노출된다. 다른 위치에의 데아미나제의 접근은 가장 가능성 있게는 dCas9 단백질에 의해 차단된다.

[0528] 데이터는 추가로 단지 3개 아미노산의 링커 (GGS)가 데아미나제가 DNA의 단일 가닥 부분에 접근하는 것을 허용하는 데 최적이지 않다는 것을 나타낸다. 9개 아미노산 링커 [$(GGS)_3$] (서열식별번호: 610) 및 보다 구조화된 16개 아미노산 링커 (XTEN)는 보다 효율적인 탈아미노화를 허용한다.

[0529] 도 6은 교정 가이드 RNA, 예를 들어, 교정 sgRNA가 데아미나제 활성에 요구된다는 것을 입증한다. 겔은 데아미나제를 dCas9, 데아미나제 효소에 융합시키는 것이 서열 특이적이 되고 (예를 들어, eGFP sgRNA와의 융합물을 사용하는 것이 탈아미노화를 발생시키지 않음), 또한 dsDNA를 탈아미노화하는 능력을 데아미나제에 부여한다는

것을 나타낸다. 데아미나제 효소의 천연 기질은 ssDNA이고, sgRNA가 첨가되지 않았을 경우 탈아미노화는 발생하지 않았다. 이는 APOBEC 데아미나제 단독은 dsDNA를 탈아미노화하지 않는다는 보고된 지식과 일치한다. 데이터는 Cas9가 짧은 윈도우 내의 이중-가닥 DNA 나선을 개방하고, 이는 그 후, 시티딘 탈아미노화를 위해 APOBEC 데아미나제에 접근가능한 단일-가닥 DNA를 노출시킨다는 것을 나타낸다. 사용된 sgRNA 서열은 하기에 제공된다. 서열 (36bp: 밑줄표시됨, sgRNA 표적 서열: 볼드체; PAM: 박스표시됨; 21bp: 이탤릭체)

[0530] DNA 서열 8:

[0531] 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATAGCC**ATTATTCCGCGGATTTATT**
TTGG*ATGACCTCTGGATCCATGGAC*-3' (SEQ ID NO: 102)

[0532] 교정 sgRNA 서열 (부분적 3' 서열):

[0533] 5'-**AUUAUUCCGCGGAUUUAUUU**GUUUUAGAGCUAG...-3' (SEQ ID NO: 103)

[0534] eGFP sgRNA 서열 (부분적 3'-서열):

[0535] 5'-**CGUAGGCCAGGGUGGUCACG**GUUUUAGAGCUAG...-3' (SEQ ID NO: 104)

[0536] 실시예 2: DNA 표적 서열의 탈아미노화

[0537] 예시적인 탈아미노화 표적. 본원에 기재된 dCas9:데아미나제 융합 단백질은 시험관내에서 또는 생체외에서 세포에, 또는 생체내에서 대상체에 전달될 수 있으며, 표적 뉴클레오티드가 PAM에 관하여 위치 3 내지 11에 있는 경우, C에서 T로의 또는 G에서 A로의 전위를 실시하는 데 사용될 수 있다. 예시적인 탈아미노화 표적은 비제한적으로 하기를 포함한다: CCR5 말단절단: CCR5의 Q93, Q102, Q186, R225, W86, 또는 Q261을 코딩하는 코돈 중 임의의 것은 탈아미노화되어 정지 코돈을 생성할 수 있으며, 이는 HIV 치료에서 적용되는 CCR5의 비관능적 말단절단을 발생시킨다. APOE4 돌연변이: C11R 및 C57R 돌연변이체 APOE4 단백질을 코딩하는 돌연변이체 코돈은 탈아미노화되어 알츠하이머 치료에서 적용되는 야생형 아미노산으로 복귀될 수 있다. eGFP 말단절단: Q158, Q184, Q185를 코딩하는 코돈 중 임의의 것은 탈아미노화되어 정지 코돈을 생성할 수 있거나, M1을 코딩하는 코돈은 탈아미노화되어 I을 코딩할 수 있으며, 이들 모두는 eGFP 형광의 상실을 발생시키고, 리포터 시스템에서 적용된다. eGFP 복원: 실질적 형광을 나타내지 않는 T65A 또는 Y66C 돌연변이체 GFP를 코딩하는 돌연변이체 코돈은 탈아미노화되어 야생형 아미노산을 복원하고, 형광을 부여할 수 있다. PIK3CA 돌연변이: K111E 돌연변이체 PIK3CA를 코딩하는 돌연변이체 코돈은 탈아미노화되어 암에서 적용되는 야생형 아미노산 잔기를 복원할 수 있다. CTNNB1 돌연변이: T41A 돌연변이체 CTNNB1을 코딩하는 돌연변이체 코돈은 탈아미노화되어 암에서 적용되는 야생형 아미노산 잔기를 복원할 수 있다. HRAS 돌연변이: Q61R 돌연변이체 HRAS를 코딩하는 돌연변이체 코돈은 탈아미노화되어 암에서 적용되는 야생형 아미노산 잔기를 복원할 수 있다. P53 돌연변이: Y163C, Y236C, 또는 N239D 돌연변이체 p53을 코딩하는 돌연변이체 코돈 중 임의의 것은 탈아미노화되어 암에서 적용되는 야생형 아미노산 서열을 코딩할 수 있다. 이중-가닥 DNA에서의 이들 표적 서열을 탈아미노화하는 것의 실시가능성은 도 7 및 8에서 입증된다. 도 7은 데아미나제-dCas9:sgRNA 복합체에 의한 생체내 표적 서열의 표적 DNA 결합의 메카니즘을 예시한다.

[0538] 도 8은 예시적인 질환-연관된 표적 서열의 성공적인 탈아미노화를 나타낸다. 상부 겔: CCR5 Q93: 위치 10에서의 코딩 가닥 표적 (위치 2, 5, 6, 8, 9에서의 잠재적 오프-타겟); CCR5 Q102: 위치 9에서의 코딩 가닥 표적 (위치 1, 12, 14에서의 잠재적 오프-타겟); CCR5 Q186: 위치 9에서의 코딩 가닥 표적 (위치 1, 5, 15에서의 잠재적 오프-타겟); CCR5 R225: 위치 6에서의 코딩 가닥 표적 (잠재적 오프-타겟 없음); eGFP Q158: 위치 5에서의 코딩 가닥 표적 (위치 1, 13, 16에서의 잠재적 오프-타겟); eGFP Q184 /185: 위치 4 및 7에서의 코딩 가닥 표적 (위치 3, 12, 14, 15, 16, 17, 18에서의 잠재적 오프-타겟); eGFP M1: 위치 12에서의 주형 가닥 표적 (위치 2, 3, 7, 9, 11에서의 잠재적 오프-타겟) (표적 위치 7 및 9를 작은 정도로 표적화함); eGFP T65A: 위치 7에서의 주형 가닥 표적 (위치 1, 8, 17에서의 잠재적 오프-타겟); PIK3CA K111E: 위치 2에서의 주형 가닥 표적 (위치 5, 8, 10, 16, 17에서의 잠재적 오프-타겟); PIK3CA K111E: 위치 13에서의 주형 가닥 표적 (위치 11, 16, 19에서의 잠재적 오프-타겟) X. 하부 겔: CCR5 W86: 위치 2 및 3에서의 주형 가닥 표적 (위치 1, 13에서의 잠재적 오프-타겟) X; APOE4 C11R: 위치 11에서의 코딩 가닥 표적 (위치 7, 13, 16, 17에서의 잠재적 오프-타겟); APOE4 C57R: 위치 5에서의 코딩 가닥 표적) (위치 7, 8, 12에서의 잠재적 오프-타겟); eGFP Y66C: 위치 11에서의 주형 가닥 표적 (위치 1, 4 ,6, 8, 9, 16에서의 잠재적 오프-타겟); eGFP Y66C: 위치 3에서의 주형 가닥 표적 (위치 1, 8, 17에서의 잠재적 오프-타겟); CCR5 Q261: 위치 10에서의 코딩 가닥 표적 (위치 3, 5, 6, 9, 18

에서의 잠재적 오프-타겟); CTNNB1 T41A: 위치 7에서의 주형 가닥 표적 (위치 1, 13, 15, 16에서의 잠재적 오프-타겟) X; HRAS Q61R: 위치 6에서의 주형 가닥 표적 (위치 1, 2, 4, 5, 9, 10, 13에서의 잠재적 오프-타겟); p53 Y163C: 위치 6에서의 주형 가닥 표적 (위치 2, 13, 14에서의 잠재적 오프-타겟); p53 Y236C: 위치 8에서의 주형 가닥 표적 (위치 2, 4에서의 잠재적 오프-타겟); p53 N239D: 위치 4에서의 주형 가닥 표적 (위치 6, 8에서의 잠재적 오프-타겟). 질환 표적의 예시적인 DNA 서열은 하기에 제공된다 (PAM (5'-NGG-3') 및 표적 위치는 박스표시됨):

[0539]

CCR5 **Q93**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACTAT**GCTGCCGCC**
CAGTGGGACTTTGG*AAATACAATGTGTCAACTCTT*-3' (SEQ ID NO: 105)
CCR5 **Q102**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAATA**CAATGTGT**
CAACTCTTGACAGGG*CTCTATTTTATAGGCTTCTTC*-3' (SEQ ID NO: 106)
CCR5 **Q186**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTTTC**CATACAGT**
CAGTATCAATTCTGG*AAGAATTTCCAGACATTAAAG*-3' (SEQ ID NO: 107)
CCR5 **R225**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCTTC**GGTGTC**GA
AATGAGAAGAAGAGG*CACAGGGCTGTGAGGCTTATC*-3' (SEQ ID NO: 108)

CCR5 **W86**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTGAG**C**CC**AGAAGG**
GGACAGTAAGAAGG*AAAAACAGGTCAGAGATGGCC*-3' (SEQ ID NO: 109)
CCR5 **Q261**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATCCTG**AACACCTT**
CC**AGGAATTCTT**TGG*CCTGAATAATTGCAGTAGCTC*-3' (SEQ ID NO: 110)
APOE4 **C11R**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGACAT**GGAGGAC**
GTGC**GCGGCCGCC**TGG*TGCAGTACCGCGGCGAGGTGC*-3' (SEQ ID NO: 111)
APOE4 **C57R**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGCA**GAAG**C**GC**
CTGGCAGTGTACCAGG*CCGGGGCCCGCGAGGGCGCCG*-3' (SEQ ID NO: 112)
eGFP **Q158**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGCCGA**CAAG**C**AGA**
AGAACGGCATCAAGG*TGAACTTCAAGATCCGCCACA*-3' (SEQ ID NO: 113)
eGFP **Q184/185**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAACCAC**TAC**C
AGC**AGAACACCCCCAT**CGG*CGACGGCCCCGTGCTGCTGCC*-3' (SEQ ID NO: 114)
eGFP **M1**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCTCG**CCCTTGCTCA**
CC**ATCTCGAGT**CGG*CCGCCAGTGTGATGGATATCT*-3' (SEQ ID NO: 115)
eGFP **T65A**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACGC**GTAGG**C**CA**
GGGTGGTCACGAGG*GTGGGCCAGGGCACGGGCAGC*-3' (SEQ ID NO: 116)
eGFP **Y66C**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAAGCA**CTGCACTC**
CGC**AGGTCAGGG**TGG*TCACGAGGGTTGGCCAGGGCA*-3' (SEQ ID NO: 117)
eGFP **Y66C**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACACTC**CG**C**AGGTC**
AGGGTGGTCACGAGG*GTTGGCCAGGGCACGGGCAGG*-3' (SEQ ID NO: 118)
PIK3CA **K111E**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGATC**T**C**TTC**
TTCACGGTTGCCTACTGG*TTCAATTACTTTTAAAAATGG*-3' (SEQ ID NO: 119)
PIK3CA **K111E**: 5'-Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTATTCTC**GATTG**
AGGATCTC**TTCTTCA**CGG*TTGCCTACTGGTTCAATTACT*-3' (SEQ ID NO: 120)
CTNNB1 **T41A**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAGGAG**CTGTGG**
CAGTGGCACCAGAA**TGG*ATTCCAGAGTCCAGGTAAGAC*-3' (SEQ ID NO: 121)
HRAS **Q61R**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGTACT**CCTCC**C**GG**
CCGGCGGTATCCAGG*ATGTCCAACAGGCACGTCTCC*-3' (SEQ ID NO: 122)
p53 **Y163C**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGACT**GCTTG**C**AG**
ATGGCCATGGCGCGG*ACGCGGGTGCCGGGCGGGGGT*-3' (SEQ ID NO: 123)
p53 **Y236C**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTGTT**ACACATG**C
AGTTGTAGTGGATGG*TGGTACAGTCAGAGCCAACCT*-3' (SEQ ID NO: 124)
p53 **N239D**: 5'-Cy3-
GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGAAC**TGT**C**ACAC**
ATGTAGTTGTAGTGG*ATGGTGGTACAGTCAGAGCCA*-3' (SEQ ID NO: 125)

[0540]

[0541] 실시예 3: 우라실 글리코실라제 억제제 융합물은 탈아미노화 효율을 개선시킨다

[0542] dCas9:데아미나제 융합 단백질에 의한 포유동물 세포에서의 직접적 프로그램가능한 핵염기 편집 효율은 우라실 글리코실라제 억제제 (UGI)를 dCas9:데아미나제 융합 단백질에 융합시킴으로써 유의하게 개선될 수 있다.

[0543] 도 9는 rAPOBEC1-XTEN-dCas9를 사용한 인간 HEK293 세포에서의 시험관내 C→T 편집 효율을 나타낸다:

rAPOBEC1-**XTEN**-dCas9-*NLS* 1차 서열

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTS**
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS*PKKKRKV* (SEQ ID NO: 126)

[0544]

[0545] 프로토스페이서 서열은 하기와 같았다:

EMX1: 5'- GAGT**C**$_5$**C**$_6$GAG**C**$_{10}$AGAAGAAGAAGGG -3' (SEQ ID NO: 127)

FANCF: 5'- GGAAT**C**$_6$**C**$_7$**C**$_8$TT**C**$_{11}$TGCAGCACCTGG -3' (SEQ ID NO: 128)

HEK293 부위 2: 5'- GAA**C**$_4$A**C**$_6$AAAG**C**$_{11}$ATAGACTGCGGG -3' (SEQ ID NO: 129)

HEK293 부위 3: 5'- GG**C**$_3$**C**$_4$**C**$_5$AGA**C**$_9$TGAGCACGTGATGG -3' (SEQ ID NO: 130)

HEK293 부위 4: 5'- GG**C**$_3$A**C**$_5$TG**C**$_8$GG**C**$_{11}$TGGAGGTGGGGG -3' (SEQ ID NO: 735)

RNF2: 5'- GT**C**$_3$AT**C**$_6$TTAGTCATTACCTGAGG -3' (SEQ ID NO: 132)

[0546]

[0547] *PAM은 박스표시되어 있고, 표적 윈도우 (위치 3 내지 11) 내의 C 잔기는 넘버링되고, 볼드체이다.

[0548] 도 10은 HEK293T 세포에서의 동일한 프로토스페이서 서열에 대한 C→T 편집 효율이 UGI 도메인이 rAPOBEC1:dCas9 융합 단백질에 융합되는 경우 크게 증진된다는 것을 입증한다.

rAPOBEC1-**XTEN**-dCas9-**UGI**-*NLS* 1차 서열

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTS**
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE
GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESI**
LMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGE
NKIKMLSGGS*PKKKRKV* (SEQ ID NO: 133)

[0549]

[0550] 도 9 및 10에서의 백분율은 표적 서열의 둘 다의 가닥의 서열분석으로부터 제시된다. 가닥 중 단지 1개만이 탈아미노화를 위한 기질이기 때문에, 이 검정에서 최대 가능한 탈아미노화 값은 50%이다. 따라서, 탈아미노화 효율은 표에 제시된 백분율의 2배이다. 예를 들어, 50%의 값은 이중-가닥 표적 서열의 100%의 탈아미노화와 관련된다. 우라실 글리코실라제 억제제 (UGI)가 dCas9:데아미나제 융합 단백질 (예를 들어, rAPOBEC1-XTEN-dCas9-[UGI]-NLS)에 융합된 경우, 세포에서의 편집 효율의 유의한 증가가 관찰되었다. 이 결과는 포유동물 세포에서, U:G 염기 쌍에서 우라실 염기를 절단해 내는 DNA 복구 기구가 DNA 편집에서의 속도-제한적 과정이라는 것을 나타낸다. UGI를 dVas9:데아미나제 융합 단백질에 테더링하는 것은 편집 수율을 크게 증가시킨다.

[0551] UGI 없이, 인간 세포에서의 전형적인 편집 효율은 ~2-14% 수율 범위였다 (도 9 및 도 10, "XTEN" 엔트리). UGI (도 10, "UGI" 엔트리)로는, 편집은 ~6-40% 범위에서 관찰되었다. 따라서, UGI 융합을 사용하는 것은 HDR을 통해 점 돌연변이를 교정하는 현재의 대안적인 방법보다 더 효율적이며, 이는 또한 점 돌연변이를 교정하는 것 이외에 과량의 indel을 생성한다. cas9:데아미나제:UGI 융합물로의 처리로부터 생성되는 indel은 관찰되지 않았다.

[0552] 실시예 4: 이중-가닥 DNA 절단 없는 게놈 DNA에서의 표적 뉴클레오티드의 직접적, 프로그램가능한 전환

[0553] 현재의 게놈-편집 기술은 유전자 교정에 대한 제1 단계로서 이중-가닥 DNA 파괴를 관심 표적 유전자좌에서 도입한다.[39,40] 대부분의 유전 질환은 단일 핵염기의 상이한 핵염기로의 돌연변이로부터 발생하지만, 이러한 변화를 복귀시키기 위한 현재의 접근법은 매우 비효율적이며, 전형적으로 이중-가닥 DNA 파괴에 대한 세포의 반응의 결과로서 표적 유전자좌에서 풍부한 무작위 삽입 및 결실 (indel)을 유도한다.[39,40] 핵염기 편집, 즉, 이중-가닥 DNA 백본 절단을 요구하지 않고, 프로그램가능한 방식으로 하나의 표적 핵염기의 또 다른 것으로의 직접적 전환을 가능하게 하는 게놈 편집을 위한 새로운 전략의 개발이 본원에 보고된다. CRISPR/Cas9의 융합물을 조작하였으며, 가이드 RNA로 프로그램되는 능력을 보유하는 시티딘 데아미나제 효소 APOBEC1은 이중-가닥 DNA 파괴를 유도하지 않고, 시티딘의 우라실로의 직접적 전환을 매개함으로써, 주형 가닥이 표적화되는 경우, DNA 복제, DNA 복구, 또는 전사 후의 C→T (또는 G→A) 치환을 실시한다. 생성된 "핵염기 편집제"는 대략 5개 뉴클레오티드의 윈도우 내에서 시티딘을 전환시키며, 시험관내에서 인간 질환과 관련된 다양한 점 돌연변이를 효율적으로 교정

할 수 있다. 4가지 형질전환된 인간 및 뮤린 세포주에서, 각각 우라실 글리코실라제 억제제 (UGI)를 융합시키고, 비-편집된 가닥을 표적화하는 Cas9 니카제를 사용하는 제2- 및 제3-세대 핵염기 편집제는 핵염기 편집에 대한 세포의 DNA 복구 반응을 극복할 수 있으며, 이는 최소 (전형적으로 1% 이하) indel 형성으로 인간 세포에서 총 세포의 DNA의 37% 이하 또는 (약 15 내지 75%)의 영구적 교정을 발생시킨다. 대조적으로, 동일한 표적 상의 정규 Cas9-매개된 HDR은 4% indel 형성으로 평균 0.7% 교정을 수득하였다. 핵염기 편집제를 사용하여 인간 유방암 및 림프종 세포에서 2가지 종양형성 p53 돌연변이를 야생형 대립유전자로 복귀시켰으며, 마우스 성상세포에서 ApoE4에서의 알츠하이머병 연관된 Arg 코돈을 비-질환-연관된 Cys 코돈으로 전환시켰다. 염기 편집은 점 돌연변이의 게놈 편집의 범위 및 효율을 확장시킨다.

[0554] 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) 시스템은 다양한 유기체 및 세포주에서 게놈 조작을 매개하도록 적응된 원핵생물 적응 면역계이다.[41] CRISPR/Cas9 단백질-RNA 복합체는 가이드 RNA와의 염기 쌍형성을 통해 표적 DNA 서열에 국재화하며, 가이드 RNA에 의해 특정된 유전자좌에서 자연적으로 DNA 이중-가닥 파괴 (DSB)를 생성한다. DSB에 대한 반응에서, 내인성 DNA 복구 과정은 주로 비-상동 말단 연결 (NHEJ)을 통해 DNA 절단의 부위에서 무작위 삽입 또는 결실 (indel)을 발생시킨다. 상동 DNA 주형의 존재 하에서, 절단 부위 주위의 DNA는 상동성-지정 복구 (HDR)를 통해 대체될 수 있다. 질환-연관된 유전자의 간단한 교란이 충분한 경우 (예를 들어, 일부 기능 획득 질환을 치료하기 위해), 표적화된 DNA 절단, 그 후 indel 형성은 효과적일 수 있다. 그러나, 대부분의 공지된 유전 질환에 대해, 유전자의 확률적 교란이라기 보다 표적 유전자좌에서의 점 돌연변이의 교정은 질환의 기저를 이루는 원인을 다루거나 연구하는 데 필요하다.[68]

[0555] 이 필요에 의해 동기부여되어, 연구자들은 HDR의 효율을 증가시키고, NHEJ를 억제하기 위해 깊은 노력을 투자하였다. 예를 들어, NHEJ 경로에서 필수적 효소인 리가제 IV의 소분자 억제제는 HDR 효율을 증가시키는 것으로 나타났다.[42,43] 그러나, 이 전략은 유사분열후 세포에서 도전중이며, 이는 전형적으로 HDR을 하향-조절하고, 그의 치료적 타당성은 비-표적 세포에서의 리가제 IV를 억제하는 것의 잠재적 위험에 의해 제한된다. 증진된 HDR 효율은 또한 화학적으로 동기화된 세포 내로의 Cas9-가이드 RNA 복합체의 시의적절한 전달에 의해 달성될 수 있으며, 이는 HDR 효율이 매우 세포-주기 의존성이기 때문이다.[44] 그러나, 이러한 접근법은 세포를 동기화하는 것이 매우 방해적이기 때문에 세포 배양물에서의 연구 적용에 제한된다. 이들 개발에도 불구하고, 대부분의 맥락에서 HDR을 사용하여 점 돌연변이를 대체하는 현재의 전략은 특히 비변형된, 비-분열하는 세포에서 매우 비효율적이다 (전형적으로 약 0.1 내지 5%).[42,43,45,46,75] 또한, HDR은 이중-가닥 파괴의 해결 동안 NHEJ와 경쟁하며, indel은 일반적으로 유전자 대체보다 더 풍부한 결과물이다. 이들 관찰은 이중-가닥 DNA 파괴를 생성하는 것에 의존하지 않는 게놈 DNA에서의 특이적 변형을 설치하는 대안적인 접근법을 개발할 필요를 강조한다. NHEJ 경로에서 필수적인 효소인 리가제 IV의 소분자 억제제는 HDR 효율을 증가시키는 것으로 나타났다.[42,43] 그러나, 이 전략은 유사분열후 세포에서 도전중이며, 이는 전형적으로 HDR을 하향-조절하고, 그의 치료적 타당성은 비-표적 세포에서 리가제 IV를 억제하는 것의 잠재적 위험에 의해 제한된다. 증진된 HDR 효율은 또한 화학적으로 동기화된 세포 내로의 Cas9-가이드 RNA 복합체의 시의적절한 전달에 의해 달성될 수 있으며, 이는 HDR 효율이 매우 세포-주기 의존성이기 때문이다.[44] 그러나, 이러한 접근법은 세포를 동기화하는 것이 매우 방해적이기 때문에 세포 배양물에서의 연구 적용에 제한된다. 일부의 경우, 성공적인 HDR의 생성물이 PAM 서열에서의 돌연변이를 함유하고, 따라서 더 이상 후속의 Cas9 변형을 위한 기질이 아니도록 HDR 주형을 설계하는 것이 가능하며, 이는 이러한 접근법이 생성물 서열에 대한 제약을 부과하긴 하지만, HDR 생성물의 전반적 수율을 증가시킨다.[75] 최근, 이 전략은 비-표적 가닥에 상보적인 ssDNA 공여자의 사용 및 HDR의 효율을 실질적으로 증가시키는 고-효율 리보핵단백질 (RNP) 전달에 결합되었지만, 심지어 이들 경우에도, HDR 대 NHEJ 결과의 비는 상대적으로 낮다 (< 2).[83]

[0556] DNA 백본 절단을 요구하지 않는 프로그램가능한 표적 유전자좌에서 하나의 핵염기의 또 다른 것으로의 전환의 직접적 촉매작용은 관심 유전자좌에서 목적하지 않은 무작위 indel을 도입하지 않는 HDR에 비한 유전자 교정의 효율을 증가시킬 수 있는 것으로 고려되었다. 그의 뉴클레아제 활성을 불활성화시키는 Asp10Ala 및 His840Ala 돌연변이를 함유하는 촉매적 데드 Cas9 (dCas9)는 가이드 RNA-프로그램된 방식으로 DNA에 결합하는 그의 능력을 보유하지만, DNA 백본을 절단하지 않는다.[16,47] 원칙적으로, 하나의 핵염기의 또 다른 것으로의 직접적 전환을 매개하는 효소적 또는 화학적 촉매와의 dCas9의 접합은 RNA-프로그램된 핵염기 편집을 가능하게 할 수 있다. 시토신 (C)의 탈아미노화는 시티딘 데아미나제에 의해 촉매되고[29], 우라실 (U)을 발생시키며, 이는 티민 (T)의

염기 쌍형성 특성을 갖는다. 가이드 RNA-특정된 DNA 유전자좌에서 C를 U로 전환시키는 그의 능력을 시험하기 위해, dCas9를 시티딘 데아미나제 효소에 융합시켰다. 대부분의 공지된 시티딘 데아미나제는 RNA 상에서 작동하며, DNA를 수용하는 것으로 공지된 소수의 예는 단일-가닥 DNA를 요구한다.[48] dCas9-표적 DNA 복합체에 대한 최근의 연구는 대체된 DNA 가닥의 적어도 9개의 뉴클레오티드가 Cas9:가이드 RNA:DNA "R-루프" 복합체의 형성시 쌍형성되지 않는다는 것을 밝혀낸다.[12] 사실, Cas9 R-루프 복합체의 구조에서, 대체된 DNA 가닥 상의 프로토스페이서의 최초 11개의 뉴클레오티드는 무질서하며, 이는 그의 이동이 매우 제한되지 않음을 시사한다.[76] 또한, 비-주형 가닥에서의 시토신에서 Cas9 니카제-유도된 돌연변이는 세포의 시티딘 데아미나제 효소에 의한 그의 접근가능성으로부터 발생할 수 있다고 추측되었다.[77] dCas9-표적 DNA 복합체에 대한 최근의 연구는 비-주형 가닥 상의 적어도 26개의 염기가 Cas9가 그의 표적 DNA 서열에 결합하는 경우 쌍형성되지 않음을 밝혀내었다.[49] R-루프에서의 단일-가닥 DNA의 이 스트레치의 하위세트는 dCas9-테더링된 시티딘 데아미나제가 DNA에서 C의 U로의 직접적, 프로그램가능한 전환을 실시하기 위한 기질로서 기능할 수 있다고 추론되었다 (도 11a).

[0557] 4가지 상이한 시티딘 데아미나제 효소 (hAID, hAPOBEC3G, rAPOBEC1, 및 pmCDA1)를 포유동물 세포 용해물-유래된 시험관내 전사-번역 시스템에서 발현시키고, ssDNA 탈아미노화에 대해 평가하였다. 4가지 효소 중에서, rAPOBEC1은 시험된 조건 하에서 가장 높은 데아미나제 활성을 나타내었으며, dCas9 융합 실험을 위해 선택되었다 (도 36a). rAPOBEC1을 dCas9의 C-말단에 첨부하는 것은 데아미나제 활성을 폐지하지만, dCas9의 N-말단에의 융합은 ssDNA에 대한 데아미나제 활성을 비융합된 효소의 것과 대등한 수준에서 보존한다. 4가지 rAPOBEC1-dCas9 융합물을 발현시키고, 상이한 길이 및 조성의 링커로 정제하고 (도 36b), 각각의 융합물을 단일 가이드 RNA (sgRNA)-프로그램된 dsDNA 탈아미노화에 대해 시험관내에서 평가하였다 (도 11a 내지 11c 및 도 15a 내지 15d).

[0558] 효율적인, 서열-특이적, sgRNA-의존성 C에서 U로의 전환이 시험관내에서 관찰되었다 (도 11a 내지 11c). 전환 효율은 9개 아미노산 길이에 걸친 rAPOBEC1-dCas9 링커를 사용한 것이 가장 컸다. 탈아미노화에 감수성인 위치 (탈아미노화 "활성 윈도우")의 수는 링커 길이가 3에서 21개 아미노산으로 확장됨에 따라 증가한다 (도 36c 내지 36f, 15a 내지 15d). 16-잔기 XTEN 링커[50]는 이들 2가지 특징 사이의 유망한 균형을 제공하는 것으로 밝혀졌으며, 프로토스페이서-인접 모티프 (PAM)에 원위인 말단을 위치 1로서 카운팅하여 프로토스페이서 내의 위치 4 내지 8에서 대략 5개의 뉴클레오티드의 효율적인 탈아미노화 윈도우를 갖는다. rAPOBEC1-XTEN-dCas9 단백질은 제1-세대 핵염기 편집제 (NBE1)로서 기능하였다.

[0559] 이론적으로 C에서 T로의 핵염기 편집에 의해 교정될 수 있는 인간 질환과 관련된 7가지 돌연변이를 선택하고, 상응하는 서열의 이중-가닥 DNA 80량체를 합성하고, 시험관내에서 이들 돌연변이를 교정하는 NBE1의 능력을 평가하였다 (도 16a 내지 16b). NBE1은 시험관내에서 이들 7개의 표적 중 6개에서, 표적 C의, 또는 다중 C가 존재한 경우 활성 윈도우 내의 적어도 1개의 C의 효율적인 편집과 일치하는 생성물을 수득하였으며, 평균 겉보기 편집 효율은 44%였다 (도 16a 내지 16b). 다중 C가 탈아미노화 윈도우 내에 존재한 3가지 경우에, 이들 시토신의 일부 또는 전부의 탈아미노화의 증거가 관찰되었다. 시험된 7가지 경우 중 단지 1가지에서, 편집된 생성물의 실질적인 수율이 관찰되었다 (도 16a 내지 16b). APOBEC1 기질에 대한 바람직한 서열 컨텍스트는 CC 또는 TC인 것으로 보고되어 있지만,[51] 표적 유전자좌에 결합하는 dCas9에 의해 매개되는 데아미나제 및 그의 단일-가닥 DNA 기질의 증가된 유효 몰농도는 이 제한을 완화시킬 수 있는 것으로 예상되었다. NBE1의 서열 컨텍스트 일반성을 조명하기 위해, 프로토스페이서 내의 위치 7에 단일 고정된 C를 함유하는 60량체 이중-가닥 DNA 올리고뉴클레오티드, 뿐만 아니라 프로토스페이서 염기 1 내지 6 및 8 내지 13이 다른 3개의 염기의 각각으로 개별적으로 다양화된 모든 36개의 단일 돌연변이된 변이체를 편집하는 그의 능력을 검정하였다. 이들 37개의 서열의 각각을 1.9 μM NBE1, 1.9 μM의 상응하는 sgRNA, 및 125 nM DNA로 2시간 동안, 시험관내 Cas9 검정에 대한 표준 조건과 유사하게 처리하였다[52]. 고처리량 DNA 서열분석 (HTS)은 표적화된 가닥의 50 내지 80% C에서 U로의 전환을 밝혀내었다 (둘 다의 DNA 가닥 (이들 중 하나는 NBE1에 대한 기질이 아님)으로부터 발생한 총 서열 판독물의 25 내지 40%) (도 12a). 표적 C 주위의 뉴클레오티드는 표적 C의 바로 5'의 염기가 G가 아니라면 서열 컨텍스트에 독립적인 편집 효율에 거의 효과를 갖지 않았으며, 이 경우, 편집 효율은 실질적으로 더 낮았다 (도 12a 내지 12b). 시험관내에서의 NBE1 활성을 프로토스페이서 내의 위치 1 내지 8에서 모든 4개의 NC 모티프에 대해 평가하였다 (도 12a 내지 12b). 일반적으로, 기질에 대한 NBE1 활성은 순서 TC ≥ CC ≥ AC >GC를 따르는 것으로 관찰되었으며, 최대 편집 효율은 표적 C가 위치 7에 또는 부근에 있는 경우 달성되었다. 또한, 핵염기 편집제는 매우 혁신적인 것으로 관찰되었으며, 5-염기 활성 윈도우 내에서 동일한 DNA 가닥 상에서 모든 C의 대

부분을 U로 효율적으로 전환시킬 것이다 (도 17).

[0560] BE1은 시험 튜브에서 기질을 효율적으로 프로세싱하는 반면, 세포에서, 가능한 DNA 복구 결과물의 트리는 염기 편집의 초기 U:G 생성물의 운명을 결정한다 (도 29a). 인간 세포에서의 핵염기 편집의 유효성을 시험하기 위해, NBE1 코돈 용법을 포유동물 발현에 대해 최적화하고, C-말단 핵 국재화 서열 (NLS)을 첨부하고,[53] 인간 게놈 전반에 걸쳐 6개의 널리 연구된 표적 부위에서 14개의 C 상의 인간 세포에서 C를 T로 전환시키는 그의 능력을 검정하였다 (도 37a).[54] NBE1을 6개의 상이한 게놈 부위에 상응하는 합성 80량체와 함께 인큐베이션한 후, HTS함으로써 편집가능한 C를 시험관내에서 각각의 프로토스페이서 내에서 확인하였다 (도 13a 내지 13c, 도 29b 및 도 25). 다음으로, HEK293T 세포를 NBE1 및 6개의 표적 sgRNA 중 1개를 코딩하는 플라스미드로 형질감염시키고, 3일 동안 핵염기 편집이 일어나게 하고, 게놈 DNA를 세포로부터 추출하고, 유전자좌를 HTS에 의해 분석하였다. 표적 유전자좌에서 세포에서의 C에서 T로의 편집은 모든 6가지 경우에 대해 관찰되었지만, 핵염기 편집의 효율은 총 DNA 서열의 1.1% 내지 6.3% 또는 0.8% 내지 7.7% (표적화된 가닥의 2.2% 내지 12.6%에 상응함)이며, 시험관내 핵염기 편집의 그것에 비해 효율의 6.3-배 내지 37-배 또는 5-배 내지 36-배 감소였다 (도 13a 내지 13c, 도 29b 및 도 25). T가 기질 C에 선행하는 경우 위치 4 내지 8의 전형적인 윈도우의 외부의 일부 염기 편집이 관찰되었으며, 본 발명자들은 이를 TC 기질에 대한 APOBEC1의 비통상적으로 높은 활성에 기인한다고 본다.[48]

[0561] U:G 헤테로듀플렉스 DNA의 존재에 대한 세포의 DNA 복구 반응이 세포에서의 핵염기 편집 효율의 큰 감소의 원인인지 질문하였다 (도 29a). 우라실 DNA 글리코실라제 (UDG)는 세포에서 DNA로부터 U의 제거를 촉매하며, 염기 절제 복구 (BER)를 개시하고, U:G 쌍의 C:G 쌍으로의 복귀가 가장 통상적인 결과물이다 (도 29a).[55] 비. 서브틸리스 박테리오파지 PBS1로부터의 83-잔기 단백질인 우라실 DNA 글리코실라제 억제제 (UGI)는 인간 UDG 활성을 강력하게 차단한다 (IC_{50} = 12 pM).[56] UGI를 NBE1의 C-말단에 융합시켜 제2-세대 핵염기 편집제 NBE2를 생성하고, 모든 6개의 게놈 유전자좌에 대한 편집 검정을 반복하였다. 인간 세포에서의 편집 효율은 NBE1에서보다 NBE2에서 평균 3-배 더 높았으며, 이는 서열분석된 총 DNA의 22.8% 이하 (표적화된 가닥의 45.6% 이하)의 유전자 전환 효율을 발생시킨다 (도 13a 내지 13c 및 도 29b). 인간 세포에서의 염기 편집을 시험하기 위해, BE1 코돈 용법을 포유동물 발현에 대해 최적화하고, C-말단 핵 국재화 서열 (NLS)을 첨부하였다.[53]

[0562] UGI를 과다발현하는 별개의 플라스미드를 NBE1과 공동-형질감염시켰을 경우, 유사한 편집 효율이 관찰되었다 (도 18a 내지 18h). 그러나, NBE1에의 UGI의 직접적 융합은 모니터링된 비-표적화된 게놈 위치에서 C에서 T로의 돌연변이의 유의한 증가를 발생시키지 않은 반면, 비융합된 UGI의 과다발현은 게놈의 다른 곳에서 C에서 T로의 돌연변이의 빈도를 검출가능하게 증가시켰다 (도 18a 내지 18h). NBE2-매개된 핵염기 편집의 일반성을 U2OS 세포에서의 동일한 6개의 게놈 표적에 대한 편집 효율을 평가함으로써 확인하고, HEK293T 세포에서의 것들과 유사한 결과를 관찰하였다 (도 19). 중요하게는, NBE2는 전형적으로 임의의 검출가능한 indel을 발생시키지 않았으며 (도 13c 및 도 29c), 이는 이중-가닥 DNA 파괴에 대한 NHEJ의 공지된 기계론적 의존성과 일치한다.[57,78] 함께, 이들 결과는 UGI를 NBE1에 접합시키는 것이 인간 세포에서의 핵염기 편집의 효율을 크게 증가시킬 수 있다는 것을 나타낸다.

[0563] 인간 세포에서의 핵염기 편집의 영구성을, 시험된 게놈 유전자좌 중 2개에서 HEK293T 세포에서의 다중 세포 분열에 걸친 편집 효율을 모니터링함으로써 확인하였다. 게놈 DNA를 2개의 시점에서 수거하였다: NBE2 및 적절한 sgRNA를 발현하는 플라스미드로의 형질감염 후 3일, 및 세포를 계대하고, 그들을 추가의 4일 동안 성장시킨 (대략 5회의 후속적 세포 분열) 후. 비-계대된 세포 (편집은 3개의 상이한 표적 C에 대한 표적화된 가닥의 4.6% 내지 6.6%에서 관찰됨) 및 계대된 세포 (편집은 동일한 3개의 표적 C에 대한 표적화된 가닥의 4.6% 내지 6.4%에서 관찰됨) 사이에 편집 효율의 유의한 변화는 관찰되지 않았으며, 이는 핵염기 편집이 세포 분열 후 영구적이 된다는 것을 확인시켜 준다 (도 20). indel은 드문 경우에 세포의 복구 과정에 의한 U:G 병변의 프로세싱으로부터 발생할 것이며, 이는 indel을 발생시키는 것으로 공지된 단일-가닥 파괴 중간체를 포함한다.[84] 수백개의 내인성 U:G 병변이 자발적 시티딘 데아미나제로부터 인간 세포당 매일 생성된다는 것을 고려할 때,[85] U:G 병변 복구로부터의 총 indel 빈도는 단일 표적 유전자좌에서 BE1 또는 BE2 활성을 증가시킬 가능성이 적을 것으로 예상되었다.

[0564] 세포에서의 핵염기 편집의 효율을 추가로 증가시키기 위해, 비-편집된 가닥을 니킹시키는 것은, 진핵생물 미스

매치 복구 기구가 미스매칭된 듀플렉스의 임의의 파괴된 가닥에 대한 직접적 DNA 복구에 대한 가닥 불연속성을 사용하기 때문에, 세포에 의해 제거되는 보다 작은 분율의 편집된 U를 발생시킬 수 있을 것으로 예상되었다 (도 29a).[58, 79, 80] 촉매 His 잔기는 Cas9 HNH 도메인에서의 위치 840에서 복원되었으며,[47,59] 이는 표적화된 C에 대향하는 G를 함유하는 비-편집된 가닥을 니킹하지만, C를 함유하는 표적 가닥을 절단하지 않는 제3-세대 핵염기 편집제 NBE3을 발생시킨다. NBE3은 Cas9에서의 Asp10Ala 돌연변이를 여전히 함유하기 때문에, 이는 이중-가닥 DNA 절단을 유도하지 않는다. 비-편집된 가닥을 니킹하는 이 전략은 인간 세포에서의 핵염기 편집 효율을 NBE2에 비해 추가의 1.4- 내지 4.8-배 증강시켰으며, 이는 HEK293T 세포에서의 동일한 6개의 인간 게놈 표적 상에 표적화된 C에서 T로의 전환을 함유하는 총 DNA 서열의 36.3% 이하를 발생시킨다 (도 13a 내지 13c 및 도 29b). 중요하게는, 평균 0.8% (6개의 상이한 유전자좌에 대해 0.2% 내지 1.6%의 범위)의 단지 작은 빈도의 indel이 NBE3 처리로부터 관찰되었다 (도 13c, 도 29c, 및 도 34). 대조적으로, 세포를 야생형 Cas9, sgRNA, 및 단일-가닥 DNA 공여자 주형으로 처리하여 이들 유전자좌 중 3개에서 HDR을 매개한 경우, 평균 단지 0.7%의 C에서 T로의 전환 효율이 관찰되었으며, 평균 3.9%의 indel 형성에 비해 훨씬 더 높았다 (도 13a 내지 13c 및 도 29c). 대립유전자 전환 대 NHEJ 결과물의 비는 BE2에 대해 평균 >1,000, BE3에 대해 23, 및 야생형 Cas9에 대해 0.17 초과였다 (도 3c). 본 발명자들은 HEK293 부위 3 및 4 게놈 유전자좌에서 HEK293T 세포에서의 다중 세포 분열에 걸친 편집 효율을 모니터링함으로써 인간 세포에서의 염기 편집의 영구성을 확인하였다 (도 38). 이들 결과는 집합적으로 핵염기 편집이 Cas9-매개된 HDR보다 인간 세포에서 훨씬 더 효율적인 표적화된 단일-염기 편집을 실시할 수 있으며, 훨씬 적거나 (NBE3), 없는 (NBE2) indel 형성을 갖는다는 것을 확립한다.

[0565] 다음으로, 인간 세포에서의 NBE1, NBE2, 및 NBE3의 오프-타겟 활성을 평가하였다. Cas9, dCas9, 및 Cas9 니카제의 오프-타겟 활성은 광범위하게 연구되었다 (도 23 내지 24 및 31 내지 33).[54,60-62] rAPOBEC1의 서열 선호도는 도 12a 내지 12b에서 관찰된 서열 컨텍스트 독립성과 일치하는 표적 C로부터 1개 초과의 염기의 DNA 염기에 독립적인 것으로 나타났기 때문에,[63] 핵염기 편집제의 잠재적 오프-타겟 활성은 오프-타겟 Cas9 결합으로부터 발생하는 것으로 가정되었다. 단지 한 부분의 Cas9 오프-타겟 부위만이 핵염기 편집을 위한 활성 윈도우 내에 C를 가질 것이기 때문에, 오프-타겟 핵염기 편집 부위는 정규 Cas9 변이체의 오프-타겟 부위의 하위세트이어야 한다. 연구된 6개의 부위의 각각에 대해, GUIDE-seq 방법을 사용하여 이전에 측정된 인간 세포에서의 상부 10개의 공지된 Cas9 오프-타겟 유전자좌를 서열분석하였다 (도 23 내지 27 및 31 내지 33).[54, 61] 공지된 dCas9 오프-타겟 유전자좌의 단지 하위세트에서의 검출가능한 오프-타겟 핵염기 편집 (NBE1 및 NBE2에 대해 16/34, 47%, 및 NBE3에 대해 17/34, 50%)이 관찰되었다. 모든 경우, 오프-타겟 염기-편집 기질은 5-염기 표적 윈도우 내에 C를 함유하였다. 일반적으로, 오프-타겟 C에서 T로의 전환은 오프-타겟 Cas9 뉴클레아제-매개된 게놈 변형 빈도와 평행하였다 (도 23 내지 27). 또한, 시험된 6개의 온-타겟 및 34개의 오프-타겟 유전자좌 주위의 2,500개의 별개의 시토신에서 C에서 T로의 전환을 모니터링하였으며, 이는 대략 $1.8x10^6$개의 세포로부터 유래된 총 14,700,000개의 서열 판독물을 나타내고, 비처리된 세포의 것에 비해 NBE1, NBE2, 또는 NBE3 처리시 이들 다른 부위 중 임의의 것에서 C에서 T로의 전환의 검출가능한 증가는 관찰되지 않았다 (도 28). 종합하면, 이들 발견은 핵염기 편집제의 오프-타겟 기질이 Cas9 오프-타겟 기질의 하위세트를 포함하고, 인간 세포에서의 핵염기 편집제가 여기서 사용된 방법에 의해 검출될 수 있는 수준에서 게놈 전반에 걸쳐 비표적화된 C에서 T로의 전환을 유도하지 않음을 시사한다. 비-계대된 HEK293T 세포 (편집은 BE2로는 3개의 표적 C에 대해 서열분석된 가닥의 1.8% 내지 2.6%, 및 BE3으로는 6.2% 내지 14.3%에서 관찰됨) 및 염기 편집 후 대략 5회의 세포 분열을 겪은 세포 (편집은 BE2로는 동일한 표적 C에 대해 서열분석된 가닥의 1.9% 내지 2.3%, 및 BE3으로는 6.4% 내지 14.5%에서 관찰됨) 사이에 편집 효율의 실질적인 변화는 관찰되지 않았으며, 이는 이들 세포에서의 염기 편집이 내구성이 있다는 것을 확인시켜 준다 (확장된 데이터 도 6).

[0566] 마지막으로, 포유동물 세포에서의 3가지 질환-관련 돌연변이를 교정하는 핵염기 편집의 잠재력을 시험하였다. 아포지단백질 E 유전자 변이체 APOE4는 아미노산 위치 112 및 158에서 2개의 Arg 잔기를 코딩하며, 후기-발병 알츠하이머병에 대한 가장 크고 가장 통상적인 유전적 위험 인자이다.[64] APOE2 (Cys112/Cys158), APOE3 (Cys112/Arg158), 및 APOE3' (Arg112/Cys158)를 포함한 위치 112 또는 158에 Cys 잔기를 갖는 ApoE 변이체는 APOE4보다 실질적으로 더 낮은 알츠하이머병 위험을 부여하는 것으로 나타났거나,[65] 그렇게 추정되었다.[81] 시험관내에서 APOE4를 APOE3'로 전환시키는 NBE1의 능력에 의해 고무되어 (도 16a 내지 16b), 이 전환은 내인성 뮤린 APOE 유전자가 인간 APOE4에 의해 대체된 불멸화 마우스 성상세포 (타코닉(Taconic))에서 시도되었다. NBE3 및 적절한 sgRNA를 코딩하는 DNA를 뉴클레오펙션에 의해 이들 성상세포 내로 전달하고 (25%의 뉴클레오펙션 효

율), 게놈 DNA를 2일 후 모든 처리된 세포로부터 추출하고, 편집 효율을 HTS에 의해 측정하였다. Arg158의 Cys158로의 전환이 총 DNA 서열분석 판독물의 58 내지 75% (뉴클레오펙션된 성상세포의 44%)에서 관찰되었다 (도 14a 내지 14c 및 도 30a). 또한, 이들 C 중 모든 3개가 활성 핵염기 편집 윈도우 내에 있기 때문에 예상된 바와 같이, 코돈 158의 제3 위치에서 총 DNA의 36 내지 50% 편집 및 Leu159의 제1 위치에서 총 DNA의 38 내지 55% 편집이 관찰되었다. 그러나, 다른 2개의 C→T 전환 중 어느 것도 ApoE3' 단백질의 아미노산 서열의 변화를 발생시키지 않는데, 이는 TGC 및 TGT 둘 다가 Cys를 코딩하고, CTG 및 TTG 둘 다가 Leu를 코딩하기 때문이다. $1x10^6$개의 세포로부터 유래된 1,500,000개 초과의 서열분석 판독물로부터, NBE3 처리 후 표적화된 유전자좌에서 1.7% indel의 증거가 관찰되었다 (도 35). 대조적으로, 야생형 Cas9 및 공여자 ssDNA로의 성상세포의 동일한 처리는 표적화된 유전자좌에서 0.1 내지 0.3% APOE4 교정 및 26 내지 40% indel을 발생시켰으며, 이는 Cas9 및 HDR을 사용한 단일-염기 교정의 이전의 보고와 일치하는 효율이다[45,75] (도 30a 및 도 40a). 동일하게 처리되었지만 VEGFA 유전자좌를 표적화하는 sgRNA를 갖는 성상세포는 APOE4 염기 편집의 증거를 나타내지 않았다 (도 34 및 도 40a). 이들 결과는 어떻게 핵염기 편집제가, 심지어 그의 프로세싱성이 게놈 DNA에서 1개 초과의 뉴클레오티드 변화를 발생시키는 경우에도, 편집의 주요 생성물로서 단백질의 코딩 서열의 정확한, 단일-아미노산 변화를 실시할 수 있는 지를 입증한다. Cas9, dCas9, 및 Cas9 니카제의 오프-타겟 활성은 광범위하게 연구되었다.[54, 60-62] 일반적으로, BE1, BE2, 및 BE3에 의한 오프-타겟 C에서 T로의 전환은 오프-타겟 Cas9 뉴클레아제-매개된 게놈 변형 빈도와 평행하였다.

[0567] 우성-음성 p53 돌연변이 Tyr163Cys 및 Asn239Asp는 여러 유형의 암과 강하게 연관된다.[66-67] 이들 돌연변이 둘 다는 주형 가닥 상에서 C에서 T로의 전환에 의해 교정될 수 있다 (도 16a 내지 16b). p53 Tyr163Cys 돌연변이에 대해 동형접합성인 인간 유방암 세포주 (HCC1954 세포)를 Tyr163Cys를 교정하도록 프로그램된 NBE3 및 sgRNA을 코딩하는 DNA로 뉴클레오펙션시켰다. HCC1954 세포의 뉴클레오펙션 효율은 10% 미만이었기 때문에, IRFP를 발현하는 플라스미드를 이들 세포 내로 공동-뉴클레오펙션시켜, 처리 2일 후에 형광-활성화된 세포 분류에 의해 뉴클레오펙션된 세포의 단리를 가능하게 하였다. 게놈 DNA의 HTS는 뉴클레오펙션된 HCC1954 세포의 7.6%에서 Tyr163Cys 돌연변이의 교정을 밝혀내었다 (도 30b 및 도 40a 내지 40b). 또한, p53 Asn239Asp에 대해 이형접합성인 인간 림프종 세포주 (ST486 세포)를 Asn239Asp를 교정하도록 프로그램된 NBE2 및 sgRNA를 코딩하는 DNA로 92% 뉴클레오펙션 효율로 뉴클레오펙션시켰다. Asn239Asp 돌연변이의 교정은 처리된 ST486 세포의 11% (뉴클레오펙션된 ST486 세포의 12%)에서 관찰되었다. HEK 세포에서의 발견과 일치하게, ST486 세포의 NBE2로의 처리로부터 indel이 관찰되지 않았으며, HCC1954 세포의 NBE3으로의 처리로부터 0.6% indel 형성이 관찰되었다. 프로토스페이서의 양 측의 적어도 50개 염기 쌍 내에 다른 DNA 변화는 $2x10^5$개의 세포로부터 유래된 2,000,000개 초과의 서열분석 판독물 중에서 비처리된 대조군의 것 초과의 빈도에서 검출되지 않았다 (도 14a 내지 14c, 도 30b). 이들 결과는 집합적으로, 본 발명자들의 지식으로, 다른 방법을 사용하여 현재 달성가능하지 않은 효율 및 다른 게놈 변형 사건의 결여를 갖는, 게놈 DNA에서의 3가지 질환-연관된 대립유전자의 그의 야생형 형태로의 전환을 나타낸다.

[0568] 인간 유전 질환을 다루기 위한 핵염기 편집제의 잠재적 타당성을 조명하기 위해, NCBI 클린바 데이터베이스[68]를 원칙적으로 이 접근법에 의해 교정될 수 있는 공지된 유전 질환에 대해 조사하였다. 클린바를, 먼저 단지 단일 뉴클레오티드 다형성 (SNP)을 조사한 후, 임의의 비병원성 변이체를 제거함으로써 필터링하였다. 24,670개의 병원성 SNP 중에서, 3,956개는 T에서 C로의, 또는 A에서 G로의 치환 중 어느 하나에 의해 유발된다. 이 목록을, SNP를 탈아미노화 활성 윈도우 내에 위치시킬 인근의 NGG PAM을 갖는 변이체만을 포함하도록 필터링하였으며, 이는 원칙적으로 여기에 기재된 핵염기 편집제에 의해 교정될 수 있는 1,089개의 임상적으로 관련된 병원성 유전자 변이체를 발생시킨다 (도 21).

[0569] 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 질환 또는 장애를 치료하는 데 사용될 수 있다. 예를 들어, 본원에 제공된 임의의 염기 편집제는 본원에 제공된 질환 또는 장애 중 임의의 것과 연관된 1개 이상의 돌연변이를 교정하는 데 사용될 수 있다. 치료될 수 있는 예시적인 질환 또는 장애는 비제한적으로 3-메틸글루타콘산 산뇨 유형 2, 46,XY 생식선 이발생증, 4-알파-히드록시페닐피루베이트 히드록실라제 결핍, 6-피루보일-테트라하이드로프테린 신타제 결핍, 완전색맹, 산-불안정성 서브유닛 결핍, 선단이골증, 말단홍색각피증, ACTH 저항성, ACTH-독립성 대결절성 부신 증식증, 활성화 PI3K-델타 증후군, 급성 간헐성 포르피린증, 급성 골수성 백혈병, 아담스-올리버 증후군 1/5/6, 아데닐로숙시네이트 리아제 결핍, 부신백질이영양증, 성인 신경 세로이드 리포푸신증, 안구운동 실시증을 갖는 성인 발병 운동실조, 전진성 수면 위상 증후군, 연령-관련 황반 변

성, 알라질 증후군, 알렉산더병, 알란-헌던-더들리 증후군, X-연관 열성 알포트 증후군, 소아기 교대 편마비, 폐정맥의 오정렬을 갖는 폐포 모세관 이형성증, 불완전 법랑질형성, 아밀로이드생성 트랜스티레틴 아밀로이드증, 근위축성 측삭 경화증, 빈혈 (G6PD 결핍으로 인한 비구형 용혈성), 빈혈 (상염색체 열성 피리독신-불응성 철적아구성), 무조증, 항트롬빈 III 결핍, 대동맥류, 재생불량성 빈혈, 아포지단백질 C2 결핍, 겉보기 미네랄로코르티코이드 과잉, 아로마타제 결핍, 부정맥유발성 우심실 심근병증, 가족성 비대성 심근병증, 비대성 심근병증, 선천성 다발성 관절구축, 아스파르틸글라이코사민뇨, 질식형 가슴 이영양증, 비타민 E 결핍을 갖는 운동실조, 운동실조 (경직성), 심방 세동, 심방 중격 결손, 비정형 용혈성-요독성 증후군, 상염색체 우성 CD11C+/CD1C+ 수지상 세포 결핍, 미토콘드리아 DNA 결실을 갖는 상염색체 우성 진행성 외안근마비, 바라이트서-윈터 증후군, 바터 증후군, 기저 신경절 석회화, 베크위드-위드만 증후군, 양성 가족성 신생아 발작, 양성 견갑골 비골 근육 이영양증, 버나드 술리에 증후군, 베타 중간성 지중해빈혈, 베타-D-만노시드축적증, 비에티 결정질 각막망막 이영양증, 담즙산 흡수장애, 바이오티니다제 결핍, 보제슨-포스만-레만 증후군, 부셰 노이하우저 증후군, 보언-콘라디 증후군, 단지증, 브라운-비알레토-반 라에레 증후군, 브루가다 증후군, 심장 부정맥, 심장얼굴피부 증후군, 심근병증, 카르네발레 증후군, 카르니틴 팔미토일트랜스퍼라제 II 결핍, 카펜터 증후군, 백내장, 카테콜아민 다형성 심실성 빈맥, 중심핵병, 염색체 1,9 및 16의 중심립 불안정성 및 면역결핍, 뇌 상염색체 우성 동맥병증, 뇌-눈-얼굴-골격 증후군, 세로이드 리포푸신증, 샤르코-마리-투스병, 콜레스타놀 축적 질환, 연골석회증, 연골영양장애, 만성 진행성 다발성 경화증, 조효소 Q10 결핍, 코헨 증후군, 인자 V 및 인자 VIII의 복합 결핍, 복합 면역결핍, 복합 산화성 인산화 결핍, 복합 부분적 17-알파-히드록실라제/17,20-리아제 결핍, 보체 인자 d 결핍, 완전 복합 17-알파-히드록실라제/17,20-리아제 결핍, 추체-간체 이영양증, 선천성 구축성 거미가락증, 글리코실화의 선천성 장애, 선천성 지방종증성 과도성장, 난소의 신생물, PIK3CA 관련된 과도성장 스펙트럼, 선천성 긴 QT 증후군, 선천성 근육 이영양증, 선천성 근육 비대-뇌 증후군, 선천성 근무력 증후군, 섬유 유형 불균형을 갖는 선천성 근병증, 아이쉬펠트 유형 선천성 근육 이영양증, 선천성 비진행성 야맹증, 각막 이영양증, 코넬리아 드 랑쥐 증후군, 두개골골간단 이형성증, 크리글러 나자르 증후군, 크루존 증후군, 골이영양증을 갖는 이완성 피부증, 청색증, 낭성 섬유증, 시스틴축적증, 시토크롬-c 옥시다제 결핍, 미토콘드리아 복합체 I 결핍, D-2-히드록시글루타르산 산뇨, 다논병, 미로 무형성 작은귀증 및 작은치아증을 갖는 난청 (LAMM), 난청, 아세틸-코에이 아세틸트랜스퍼라제의 결핍, 페록시다제의 결핍, UDP글루코스-헥소스-1-포스페이트 유리딜일트랜스퍼라아제의 결핍, 데제린-소타스 질환, 데스부쿠오이스 증후군, DFNA, 당뇨병 유형 2, 당뇨병-난청 증후군, 다이아몬드-블랙판 빈혈, 이영양성 이형성증, 디하이드로프테리딘 리덕타제 결핍, 디하이드로피리미디나제 결핍, 확장성 심근병증, 파종성 비정형 미코박테리아 감염, 원위 관절구축, 원위 유전성 운동신경 뉴런병증, 돈나이 바로우 증후군, 뒤시엔느 근육 이영양증, 베커 근육 이영양증, 전신성 유전 색소이상, 선천성 이상각화증, 이상긴장증, 조기 영아 간질성 뇌병증, 엘러스-단로스 증후군, 아이쉬펠트 유형 선천성 근육 이영양증, 에머리-드레이퓨스 근육 이영양증, 에나멜-신장 증후군, 역성 위축성 수포성 표피박리증, 포진성 수포성 표피박리증, 간질, 간헐성 운동실조, 가변성 홍반각피증, 적혈구생성 프로토포르피린증, 운동 불내성, 삼출성 유리체망막병증, 파브리병, 인자 V 결핍, 인자 VII 결핍, 인자 xiii 결핍, 가족성 선종성 폴립증, 유방암, 난소암, 한랭두드러기, 만성 영아 신경학적 피부 관절 증후군, 편마비성 편두통, 고콜레스테롤혈증, 비대성 심근병증, 저알파지단백혈증, 저칼륨혈증-저마그네슘혈증, 소아 통풍, 고지단백혈증, 내장 아밀로이드증, 저인산혈증 비타민 D 불응성 구루병, FG 증후군, 외안근의 섬유증, 핀란드 선천성 신증후군, 초점성 간질, 초점성 분절성 사구체경화증, 이마코 이형성증, 전두측두엽 치매, 프룩토스-바이포스파타제 결핍, 감스토르프-볼파르트 증후군, 강글리오시드 시알리다제 결핍, GATA-1-관련 혈소판감소증, 고셔병, 거대 축삭 신경병증, 글라즈만 혈소판무력증, 사구체낭성 신장 질환, 사구체병증, 글루코코르티코이드 저항성, 글루코스-6-포스페이트 수송 결손, 글루타르산 산뇨, 글리코겐 축적 질환, 골린 증후군, 완전전뇌증, 그라실 증후군, 출혈성 모세혈관확장증, 혈색소증, 헤모글로빈 H 질환, 용혈성 빈혈, 혈구포식성 림프조직구증식증, 결장의 암종, 미레 증후군, 백질뇌병증, 유전성 인자 IX 결핍 질환, 유전성 인자 VIII 결핍 질환, 유전성 인자 XI 결핍 질환, 유전성 프룩토스뇨증, 유전성 비용종증 결장직장 신생물, 유전성 췌장염, 유전성 열변형 적혈구증, 타원적혈구증, 내장위치이상증, 이소증, 조직구성 수질성 망상세포증다증, 조직구증-림프절병증 플러스 증후군, 알돌라제 A 결핍으로 인한 HNSHA, 호로카르복실라제 신테타제 결핍, 호모시스테인증, 호웰-에반스 증후군, 포상 기태, 고칼슘뇨성 고칼슘혈증, 과면역글로불린 D, 메발론산 산뇨, 고인슐린성 저혈당증, 고칼륨혈증성 주기성 마비, 폰 오일렌부르크의 선천성 이상근긴장증, 고지단백혈증, 고망간혈증, 고메티오닌혈증, 고포스파타제혈증, 고혈압, 저마그네슘혈증, 저베타지단백혈증, 저칼슘혈증, 저생식선자극성 생식선기능저하증, 저생식선자극성 생식선기능저하증, 땀저하 외배엽 이형성증, 고-IgM 면역결핍, 땀저하 X-연관 외배엽 이형성증, 저마그네슘혈증, 부갑상선기능저하증, 특발성 섬유화 폐포염, 면역결핍, 이뮤노글로불린 A 결핍, 영아 저포스파타제증, 영아 파킨슨증-이상긴장증, 인슐린 의존성 당뇨병, 중간성 메이플 시럽뇨 질환, 좌골슬개 이형성증, 도세포 증식증, 분리성 성장 호르몬 결핍, 분리성 루트로핀

결핍, 이소발레르산 산혈증, 주버트 증후군, 소아 폴립증 증후군, 소아 망막층간분리, 칼만 증후군, 카르타게너 증후군, 쿠겔베르그-벨란더 질환, 레티스 각막 이영양증, 레베르 선천성 흑암시, 레베르 시신경 위축, 좌심실 비조밀, 라이병, 미토콘드리아 복합체 I 결핍, 요정증 증후군, 관절구축, 전각 세포 질환, 백혈구 부착 결핍, 백질이영양증, 백질뇌병증, 난소백질이영양증, L-페리틴 결핍, 리-프라우메니 증후군, 지대형 근육 이영양증-디스트로글리칸병증, 로에이스-디에쯔 증후군, 긴 QT 증후군, 대두증/자폐 증후군, 황반 각막 이영양증, 황반 이영양증, 악성 고체온 민감성, 전립선의 악성 종양, 메이플 시럽뇨 질환, 마르덴 워커 유사 증후군, 마르팡 증후군, 마리 운나 유전성 털감소증, 비만 세포 질환, 태변성 장폐쇄증, 중간-쇄 아실-코엔자임 A 데히드로게나제 결핍, 멜닉-프레이저 증후군, 정신 지체, 메로신 결함성 선천성 근육 이영양증, 중피종, 이염성 백질이영양증, 골간단 연골영양장애, 메트헤모글로빈혈증, 메틸말론산 산뇨, 호모시스틴뇨증, 소두증, 맥락망막병증, 림프부종, 소안구증, 경도 비-PKU 고페닐알라닌혈증, 미첼-릴리 증후군, 미토콘드리아 3-히드록시-3-메틸글루타릴-코에이 신타제 결핍, 미토콘드리아 복합체 I 결핍, 미토콘드리아 복합체 III 결핍, 미토콘드리아 근병증, 점액지질증 III, 뮤코폴리사카라이드증, 다중 술파타제 결핍, 근무력 증후군, 미코박테리움 투베르쿨로시스, 미엘로퍼옥시다제 결핍, 미레 증후군, 근간대성 간질, 근원섬유성 근병증, 미오글로빈뇨증, 근병증, 근시, 선천성 근긴장증, 나바조 신경간병증, 네말린 근병증, 위의 신생물, 신원성 요붕증, 신결석증, 신증후군, 신경섬유종증, 중성 지질 축적 질환, 니만-피크 질환, 비-케톤성 고글라이신혈증, 누난 증후군, 누난 증후군-유사 장애, 노룸병, 황반 변성, N-말단 아세틸트랜스퍼라제 결핍, 눈피부 백색증, 눈코치아골격 이형성증, 오도 증후군, 시신경 무형성증, 오르니틴 카르바모일트랜스퍼라제 결핍, 입얼굴손발가락 증후군, 골형성 부전증, 골화석증, 난소 이상발육, 손발톱비대증, 비표피박리성 수장족저 각피증, 파피용-르페브르 증후군, 하임-멍크 증후군, 치주염, 피부 박피 증후군, 펜드레드 증후군, 퍼옥시솜 지방 아실-코에이 리덕타제 1 장애, 퍼옥시솜 생물발생 장애, 파이퍼 증후군, 페닐케톤뇨, 페닐케톤뇨, 비-PKU 고페닐알라닌혈증, 뇌하수체 호르몬 결핍, 모공성 홍색 비강진, 결절성 다발동맥염, 다낭성 신장 질환, 다낭성 지방막형성 골발육부전, 다왜소뇌회증, 교뇌소뇌 저형성증, 한공각화증, 후측 칼럼 운동실조, 원발성 지단홍통증, 과옥살산뇨증, 진행성 가족성 간내 담즙정체, 진행성 가성 류마티스 이형성증, 프로피온산 산혈증, 가성반음양증, 가성저알도스테론증, 가성황색종 탄성섬유-유사 장애, 퓨린-뉴클레오시드 포스포릴라제 결핍, 피리독살 5-포스페이트-의존성 간질, 신장 이형성증, 망막 색소형 이영양증, 소뇌 운동실조, 골격 이형성증, 망상 이발생증, 색소성 망막염, 어셔 증후군, 망막모세포종, 망막병증, RRM2B-관련 미토콘드리아 질환, 루빈스타인-타이비 증후군, 신더 결정질 각막 이영양증, 피지종양, 중증 선천성 호중구감소증, 중증 영아기 근간대성 간질, 중증 X-연관 근세관성 근병증, 손발톱이형성증, 얼굴 이형증, 털감소증, 단락-늑골 가슴 이형성증, 시알산 축적 질환, 시알리다제결핍증, 철적혈모구성 빈혈, 소섬유 신경병증, 스미스-마제니스 증후군, 소르스비 안저 이영양증, 경직성 운동실조, 경직성 하반신마비, 정자발생 부전, 구상적혈구증, 스핑고미엘린/콜레스테롤 지질증, 척수소뇌성 운동실조, 갈림-손/발 기형, 척추골단골간단 이형성증, 편평추 치사성 골격 이형성증, 두경부의 편평 세포 암종, 스타르가르트병, 수크라제-이소말타제 결핍, 영아 돌연사 증후군, 판상부 대동맥 협착, 계면활성제 대사 기능장애, 탄지에르병, 타톤-브라운-라아만 증후군, 가슴 대동맥류 및 대동맥 박리, 혈전성향증, 갑상선 호르몬 저항성, TNF 수용체-연관 주기성 발열 증후군 (TRAPS), 치아 무발생증, 토르사드 드 포인트, 대동맥의 전위, 트리처 콜린스 증후군, 결절성 경화증 증후군, 티로시나제-음성 눈피부 백색증, 티로시나제-양성 눈피부 백색증, 티로신혈증, UDP글루코스-4-에피머라제 결핍, 울리히 선천성 근육 이영양증, 베스렘 근병증 어셔 증후군, UV-민감성 증후군, 반 데르 우드 증후군, 슬와 익상편 증후군, 초장쇄 아실-코에이 데히드로게나제 결핍, 방광요관 역류, 유리망막맥락막증, 폰 히펠-린다우 증후군, 폰 빌레브란트병, 바르덴부르크 증후군, 바르샤바 손상 증후군, WFS1-관련 장애, 윌슨병, 색소성 건피증, X-연관 무감마글로불린혈증, X-연관 유전성 운동 및 감각 신경병증, X-연관 중증 복합 면역결핍, 및 젤베거 증후군을 포함한다.

[0570] 핵염기 편집의 개발은 게놈 편집의 범위 및 유효성 둘 다를 진전시킨다. 여기에 기재된 핵염기 편집제는 연구자들에게 사실상 indel 형성을 갖지 않는 편집 (NBE2), 또는 낮은 빈도 (여기서, 전형적으로 1% 이하)의 indel 형성을 갖는 보다 효율적인 편집 (NBE3)의 선택을 제공한다. 염기 편집의 생성물이 정의상 더 이상 기질이 아닌 것은 전통적인 Cas9 적용을 방해할 수 있는 후속의 생성물 형질전환을 방지함으로써 편집 효율에 기여할 가능성이 있다. 세포 상태 및 세포 유형에 의해 크게 다양한 이중-가닥 DNA 절단 및 확률적 DNA 복구 과정에 대한 의존성을 제거함으로써, 핵염기 편집은 깨끗하게 설치될 수 있는 게놈 변형의 유형, 및 이들 변형의 효율, 편집을 잘 받아들이는 세포의 유형을 확장하는 잠재력을 갖는다. 개선된 DNA 특이성을 갖는 최근의 조작된 Cas9 변이체[69,70, 82] 또는 전달 방법[71], 뿐만 아니라 변경된 PAM 특이성을 갖는 Cas9 변이체[72]는 추가의 핵염기 편집제에게 개선된 DNA 특이성을 제공하는 이 전략 내로 통합될 수 있거나, 훨씬 더 넓은 범위의 질환-연관된 돌연변이를 표적화할 수 있는 가능성이 있다. 이들 발견은 또한 추가의 핵염기 형질전환을 촉매하는 효소로

dCas9의 추가의 융합물을 조작하는 것이 핵염기 편집을 통해 생성될 수 있는 가능한 DNA 염기 변화의 분율을 증가시킬 것임을 시사한다. 이들 결과는 또한 추가의 유형의 프로그램가능한 게놈 및 에피게놈 염기 편집을 가능하게 할 수 있는 메틸라제 및 데메틸라제를 포함한 다른 DNA-변형 효소의 융합을 위한 아키텍처를 시사한다.

[0571] 물질 및 방법

[0572] 클로닝. 이 명세서에 사용된 모든 구축물의 DNA 서열 및 프라이머는 보충적 서열에 열거된다. NBE1, NBE2, 및 NBE3을 코딩하는 유전자를 함유하는 플라스미드는 애드진으로부터 입수가능할 것이다. PCR을 베라섹 울트라 (VeraSeq ULtra) DNA 폴리머라제 (엔지매틱스), 또는 Q5 핫 스타트(Hot Start) 고-충실도 DNA 폴리머라제 (뉴 잉글랜드 바이오랩스)를 사용하여 수행하였다. NBE 플라스미드를 USER 클로닝 (뉴 잉글랜드 바이오랩스)을 사용하여 구축하였다. 데아미나제 유전자를 g블록스(gBlocks) 유전자 단편 (인테그레이티드 DNA 테크놀로지스 (Integrated DNA Technologies))로서 합성하고, Cas9 유전자를 이전에 보고된 플라스미드로부터 수득하였다.[18] 데아미나제 및 융합 유전자를 pCMV (포유동물 코돈-최적화된) 또는 pET28b (이. 콜라이 코돈-최적화된) 백본 내로 클로닝하였다. sgRNA 발현 플라스미드를 부위-지정 돌연변이유발을 사용하여 구축하였다. 간략하게, 보충적 서열에 열거된 프라이머를 T4 폴리뉴클레오티드 키나제 (뉴 잉글랜드 바이오랩스)를 제조업체의 지침에 따라 사용하여 5' 인산화하였다. 다음으로, PCR을, 인산화된 프라이머 및 주형으로서 플라스미드 pFYF1320 (EGFP sgRNA 발현 플라스미드)을 갖는 Q5 핫 스타트 고-충실도 폴리머라제 (뉴 잉글랜드 바이오랩스)를 제조업체의 지침에 따라 사용하여 수행하였다. PCR 생성물을 DpnI (20 U, 뉴 잉글랜드 바이오랩스)과 함께 37℃에서 1시간 동안 인큐베이션하고, 퀴아프렙(QIAprep) 스핀 칼럼 (퀴아젠 (Qiagen)) 상에서 정제하고, 퀵리가제 (QuickLigase) (뉴 잉글랜드 바이오랩스)를 제조업체의 지침에 따라 사용하여 라이게이션하였다. DNA 벡터 증폭을 Mach1 적격 세포 (써모피셔 사이언티픽(ThermoFisher Scientific))를 사용하여 수행하였다.

[0573] ssDNA에 대한 시험관내 데아미나제 검정. 모든 ssDNA 기질의 서열은 보충적 서열에 열거된다. 모든 Cy3-표지된 기질은 인테그레이티드 DNA 테크놀로지스 (IDT)로부터 얻었다. 데아미나제를 TNT T7 퀵 커플드 전사/번역 키트 (프로메가)를 제조업체의 지침에 따라 사용하여 1 μg의 플라스미드를 사용하여 시험관내에서 발현시켰다. 단백질 발현 후, 5 μL의 용해물을 컷스마트(CutSmart) 완충제 (뉴 잉글랜드 바이오랩스) (50 mM 아세트산칼륨, 29 mM 트리스아세테이트, 10 mM 아세트산마그네슘, 100 ug/mL BSA, pH 7.9) 중 35 μL의 ssDNA (1.8 μM) 및 USER 효소 (1 유닛)와 합하고, 37℃에서 2시간 동안 인큐베이션하였다. 절단된 U-함유 기질을 10% TBE-우레아 겔 (바이오-라드(Bio-Rad)) 상의 전장 비변형된 기질로부터 분할하였다.

[0574] His_6-rAPOBEC1-링커-dCas9 융합물의 발현 및 정제. 이. 콜라이 BL21 STAR (DE3)-적격 세포 (써모피셔 사이언티픽)를 GGS, $(GGS)_3$, (서열식별번호: 610) XTEN, 또는 $(GGS)_7$ (서열식별번호: 610) 링커를 갖는 pET28b-His_6-rAPOBEC-링커-dCas9를 코딩하는 플라스미드로 형질전환시켰다. 생성된 발현 균주를 밤새 100 μg/mL의 카나마이신을 함유하는 루리아-베르타니 (LB) 브로쓰에서 37℃에서 성장시켰다. 세포를 동일한 성장 배지 내로 1:100 희석하고, 37℃에서 OD_{600} = 약 0.6으로 성장시켰다. 배양물을 2시간의 기간에 걸쳐 4℃로 냉각시키고, 이소프로필-β-D-1-티오갈락토피라노시드 (IPTG)를 0.5 mM에서 첨가하여 단백질 발현을 유도하였다. 약 16시간 후, 세포를 4,000 g에서의 원심분리에 의해 수집하고, 용해 완충제 (50 mM 트리스(히드록시메틸)-아미노메탄 (트리스 (Tris))-HCl, pH 7.0, 1 M NaCl, 20% 글리세롤, 10 mM 트리스(2-카르복시에틸)포스핀 (TCEP, 솔텍 벤처스 (Soltec Ventures)))에 재현탁시켰다. 세포를 초음파처리 (20초 펄스-온, 20초 펄스-오프, 6 W 출력에서 총 8분 동안)에 의해 용해시키고, 용해물 상청액을 25,000 g에서 15분 동안의 원심분리 후 단리하였다. 용해물을 His-Pur 니켈-니트릴로아세트산 (니켈-NTA) 수지 (써모피셔 사이언티픽)와 함께 4℃에서 1시간 동안 인큐베이션하여 His-태그부착된 융합 단백질을 포획하였다. 수지를 칼럼에 옮기고, 40 mL의 용해 완충제로 세척하였다. His-태그부착된 융합 단백질을 285 mM 이미다졸로 보충된 용해 완충제에서 용리시키고, 한외여과 (아미콘-밀리포어(Amicon-Millipore), 100-kDa 분자량 컷-오프)에 의해 1 mL 총 부피로 농축시켰다. 단백질을 50 mM 트리스(히드록시메틸)-아미노메탄 (트리스)-HCl, pH 7.0, 0.1 M NaCl, 20% 글리세롤, 10 mM TCEP를 함유하는 저염 정제 완충제에 20 mL로 희석하고, SP 세파로스 고속 유동 수지 (지이 라이프 사이언시스(GE Life Sciences)) 상에 로딩하였다. 수지를 40 mL의 이 저염 완충제로 세척하고, 단백질을 50 mM 트리스(히드록시메틸)-아미노메탄 (트리스)-HCl, pH 7.0, 0.5 M NaCl, 20% 글리세롤, 10 mM TCEP를 함유하는 활성 완충제 5 mL로 용리시켰다. 용리된 단백질을 SDSPAGE 겔 상에서 정량화하였다.

[0575] sgRNA의 시험관내 전사. 20-bp sgRNA 표적 서열에 이어진 T7 프로모터를 함유하는 선형 DNA 단편을 보충적 서열에 열거된 프라이머를 사용하여 트랜스크립트에이드(TranscriptAid) T7 고 수율 전사 키트 (써모피셔 사이언티픽)로 제조업체의 지침에 따라 시험관내에서 전사하였다. sgRNA 생성물을 메가클리어(MEGAclear) 키트 (써모

피셔 사이언티픽)를 제조업체의 지침에 따라 사용하여 정제하고, UV 흡광도에 의해 정량화하였다.

[0576] Cy3-접합된 dsDNA 기질의 제조. 80-뉴클레오티드 비표지된 가닥의 서열은 보충적 서열에 열거되며, IDT로부터 PAGE-정제된 올리고뉴클레오티드로서 주문하였다. 보충적 서열에 열거된 25-nt Cy3-표지된 프라이머는 각각의 80-nt 기질의 3' 말단에 상보적이다. 이 프라이머는 IDT로부터 HPLC-정제된 올리고뉴클레오티드로서 주문하였다. Cy3-표지된 dsDNA 기질을 생성하기 위해, 80-nt 가닥 (5 μL의 100 μM 용액)을 dNTP (0.75 μL의 100 mM 용액)를 갖는 NE완충제 2 (38.25 μL의 50 mM NaCl, 10 mM 트리스-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9 용액, 뉴 잉글랜드 바이오랩스) 중 Cy3-표지된 프라이머 (5 μL의 100 μM 용액)와 합하고, 95℃로 5분 동안 가열한 후, 0.1℃/초의 속도로 45℃로 점차 냉각시켰다. 이 어닐링 기간 후, 클레노브(Klenow) 엑소$^-$ (5 U, 뉴 잉글랜드 바이오랩스)를 첨가하고, 반응물을 37℃에서 1시간 동안 인큐베이션하였다. 용액을 완충제 PB (250 μL, 퀴아젠) 및 이소프로판올 (50 μL)로 희석하고, 50 μL의 트리스 완충제로 용리시키는 퀴아프렙 스핀 칼럼 (퀴아젠) 상에서 정제하였다.

[0577] dsDNA에 대한 데아미나제 검정. 정제된 융합 단백질 (활성 완충제 중 1.9 μM 20 μL)을 1 당량의 적절한 sgRNA와 합하고, 주위 온도에서 5분 동안 인큐베이션하였다. Cy3-표지된 dsDNA 기질을 125 nM의 최종 농도로 첨가하고, 생성된 용액을 37℃에서 2시간 동안 인큐베이션하였다. dsDNA를 완충제 PB (100 μL, 퀴아젠) 및 이소프로판올 (25 μL)의 첨가에 의해 융합물로부터 분리하고, 20 μL의 컷스마트 완충제 (뉴 잉글랜드 바이오랩스)로 용리시키는 이코노스핀(EconoSpin) 마이크로 스핀 칼럼 (에포치 라이프 사이언스(Epoch Life Science)) 상에서 정제하였다. USER 효소 (1 U, 뉴 잉글랜드 바이오랩스)를 정제된, 편집된 dsDNA에 첨가하고, 37℃에서 1시간 동안 인큐베이션하였다. Cy3-표지된 가닥을, 5 μL의 반응 용액을 15 μL의 DMSO-기반 로딩 완충제 (5 mM 트리스, 0.5 mM EDTA, 12.5% 글리세롤, 0.02% 브로모페놀 블루, 0.02% 크실렌 시안, 80% DMSO)와 합함으로써 그의 보체로부터 완전히 변성시켰다. 전장 C-함유 기질을 10% TBE-우레아 겔 (바이오-라드) 상의 임의의 절단된, U-함유 편집된 기질로부터 분리하고, 지이 아머샴 타이푼(GE Amersham Typhoon) 영상화기 상에서 영상화하였다.

[0578] 고처리량 서열분석 (HTS)을 위한 시험관내-편집된 dsDNA의 제조. 보충적 서열에 열거된 올리고뉴클레오티드는 IDT로부터 수득하였다. 상보적 서열을 트리스 완충제 중에서 합하고 (5 μL의 100 μM 용액), 95℃로 5분 동안 가열함으로써 어닐링한 후, 0.1℃/초의 속도로 45℃로 점차 냉각시켜 60-bp dsDNA 기질을 생성하였다. 정제된 융합 단백질 (활성 완충제 중 1.9 μM 20 μL)을 1 당량의 적절한 sgRNA와 합하고, 주위 온도에서 5분 동안 인큐베이션하였다. 60량체 dsDNA 기질을 125 nM의 최종 농도로 첨가하고, 생성된 용액을 37℃에서 2시간 동안 인큐베이션하였다. dsDNA를 완충제 PB (100 μL, 퀴아젠) 및 이소프로판올 (25 μL)의 첨가에 의해 융합물로부터 분리하고, 20 μL의 트리스 완충제로 용리시키는 이코노스핀 마이크로 스핀 칼럼 (에포치 라이프 사이언스) 상에서 정제하였다. 생성된 편집된 DNA (1 μL는 주형으로서 사용됨)를 보충적 서열로 특정된 HTS 프라이머 쌍 및 베라섹 울트라 (엔지매틱스)를 제조업체의 지침에 따라 사용하여 13 주기의 증폭으로 PCR에 의해 증폭시켰다. PCR 반응 생성물을 래피드팁스(RapidTips) (디피니티 게노믹스(Diffinity Genomics))를 사용하여 정제하고, 정제된 DNA를 서열분석 어댑터를 함유하는 프라이머로 PCR에 의해 증폭시키고, 정제하고, 이전에 기재된 바와 같이 MiSeq 고처리량 DNA 시퀀서 (일루미나(Illumina)) 상에서 서열분석하였다.[73]

[0579] 세포 배양. HEK293T (ATCC CRL-3216), U2OS (ATCC-HTB-96) 및 ST486 세포 (ATCC)를 10% (v/v) 소 태아 혈청 (FBS) 및 페니실린/스트렙토마이신 (1x, 암레스코(Amresco))으로 보충된 둘베코 변형 이글 배지(Dulbecco's Modified Eagle's Medium) 플러스 글루타맥스(GlutaMax) (써모피셔)에서 37℃에서 5% CO_2로 유지하였다. HCC1954 세포 (ATCC CRL-2338)를 상기 기재된 바와 같이 보충된 RPMI-1640 배지 (써모피셔 사이언티픽)에서 유지하였다. APOE 유전자의 ApoE4 이소형을 함유하는 불멸화 래트 성상세포 (타코닉 바이오사이언시스(Taconic Biosciences))를 10% (v/v) 소 태아 혈청 (FBS) 및 200 μg/mL 제네티신 (써모피셔 사이언티픽)으로 보충된 둘베코 변형 이글 배지 플러스 글루타맥스 (써모피셔 사이언티픽)에서 배양하였다.

[0580] 형질감염. HEK293T 세포를 48-웰 콜라겐-코팅된 바이오코트(BioCoat) 플레이트 (코닝(Corning)) 상에 시딩하고, 대략 85% 융합도에서 형질감염시켰다. 간략하게, 750 ng의 NBE 및 250 ng의 sgRNA 발현 플라스미드를 웰당 1.5 μL의 리포펙타민 2000 (써모피셔 사이언티픽)을 제조업체의 프로토콜에 따라 사용하여 형질감염시켰다. 성상세포, U2OS, HCC1954, HEK293T 및 ST486 세포를 적절한 아막사 뉴클레오펙터(AMAXA NUCLEOFECTOR)™ II 프로그램을 제조업체의 지침에 따라 사용하여 형질감염시켰다. 40 ng의 적외선 RFP (애드진 플라스미드 45457)[74]를 뉴클레오펙션 용액에 첨가하여 이들 세포주에서의 뉴클레오펙션 효율을 평가하였다. 성상세포,

U2OS, 및 ST486 세포에 대해, 뉴클레오펙션 효율은 각각 25%, 74%, 및 92%였다. HCC1954 세포에 대해, 뉴클레오펙션 효율은 10% 미만이었다. 따라서, 트립신화 후, HCC1954 세포를 40 마이크로미터 여과기 (피셔 사이언티픽(Fisher Scientific))를 통해 여과하고, 뉴클레오펙션된 HCC1954 세포를 iRFP 신호 (흡수 643 nm, 방출 670 nm)를 사용하여 베크만 쿨터(Beckman Coulter) 모플로 XDP 세포 분류기 상에서 수집하였다. 다른 세포는 뉴클레오펙션된 세포의 풍부화 없이 사용하였다.

[0581] 게놈 DNA 샘플의 고처리량 DNA 서열분석. 형질감염된 세포를 3일 후에 수거하고, 게놈 DNA를 아젠코트(Agencourt) DNAdvance 게놈 DNA 단리 키트 (베크만 쿨터)를 제조업체의 지침에 따라 사용하여 단리하였다. 관심 온-타겟 및 오프-타겟 게놈 영역을 보충적 서열에 열거된 플랭킹 HTS 프라이머 쌍으로 PCR에 의해 증폭시켰다. PCR 증폭은 푸션(Phusion) 고-충실도 DNA 폴리머라제 (써모피셔)로 제조업체의 지침에 따라 5 ng의 게놈 DNA를 주형으로서 사용하여 수행하였다. 주기 수는 각각의 프라이머 쌍에 대해 별개로 측정하여, 반응이 선형 범위의 증폭에서 중단되도록 하였다 (EMX1, FANCF, HEK293 부위 2, HEK293 부위 3, HEK293 부위 4, 및 RNF2 프라이머에 대해 각각 30, 28, 28, 28, 32, 및 32 주기). PCR 생성물을 래피드팁스 (디피니티 게노믹스)를 사용하여 정제하였다. 정제된 DNA를 서열분석 어댑터를 함유하는 프라이머로 PCR에 의해 증폭시켰다. 생성물을 겔-정제하고, 퀀트-잇(QUANT-IT)™ 피코그린 dsDNA 검정 키트 (써모피셔) 및 카파 라이브러리 정량화 키트-일루미나 (카파 바이오시스템즈(KAPA Biosystems))를 사용하여 정량화하였다. 샘플을 이전에 기재된 바와 같이 일루미나 MiSeq 상에서 서열분석하였다.[73]

[0582] 데이터 분석. 서열분석 판독물을 MiSeq 리포터 (일루미나)를 사용하여 자동으로 역다중화하고, 개별 FASTQ 파일을 보충 노트에 제공된 주문제작 매트랩 스크립트로 분석하였다. 각각의 판독물을 스미스-워터만(Smith-Waterman) 알고리즘을 사용하여 적절한 참조 서열에 대해 쌍별로 정렬하였다. 31 미만의 Q-점수를 갖는 염기 호출을 N으로 대체하고, 따라서 뉴클레오티드 빈도를 계산하는 데 있어서 제외하였다. 이 처리는 1,000 중 대략 1의 예상된 MiSeq 염기-호출 오차율을 생성한다. 판독물 및 참조 서열이 갭을 함유하지 않은 정렬된 서열을 염기 빈도가 각각의 유전자좌에 대해 표로 제시될 수 있는 정렬 표에 저장하였다.

[0583] indel 빈도를 이전에 기재된 기준[71]을 사용하여 보충 노트에 제시된 주문제작 매트랩 스크립트로 정량화하였다. 서열분석 판독물을 indel이 일어날 수 있는 윈도우의 양 측에 플랭킹된 2개의 10-bp 서열에의 정확한 매치를 위해 스캐닝하였다. 정확한 매치가 위치하지 않은 경우, 판독물을 분석으로부터 제외하였다. 이 indel 윈도우의 길이가 참조 서열을 정확하게 매칭시킨 경우, 판독물을 indel을 함유하지 않는 것으로서 분류하였다. indel 윈도우가 참조 서열보다 더 길거나 더 짧은 2개 이상의 염기인 경우, 서열분석 판독물을 각각 삽입 또는 결실로서 분류하였다.

[0584] 본원에, 예를 들어, 배경기술, 요약, 상세한 설명, 실시예, 및/또는 참고문헌 섹션에 언급된 모든 공개, 특허, 특허 출원, 공개, 및 데이터베이스 엔트리 (예를 들어, 서열 데이터베이스 엔트리)는 각각의 개별 공개, 특허, 특허 출원, 공개, 및 데이터베이스 엔트리가 구체적으로 및 개별적으로 본원에 참조로 포함되는 것처럼 그 전문이 본원에 참조로 포함된다. 상충하는 경우, 본원의 임의의 정의를 포함하여 본 출원이 우선할 것이다.

[0585] 보충 서열

[0586] sgRNA 형질감염 플라스미드를 생성하는 데 사용된 프라이머. rev_sgRNA_플라스미드를 모든 경우에 사용하였다. pFYF1320 플라스미드를 물질 및 방법 섹션에 언급된 바와 같이 주형으로서 사용하였다. 서열식별번호: 187-196은 각각 하기에 위에서 아래로 나타난다.

rev_sgRNA_플라스미드	GGTGTTTCGTCCTTTCCACAAG
fwd _p53 _Y163C	GCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_p53_N239D	TGTCACACATGTAGTTGTAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_APOE4 _C158R	GAAGCGCCTGGCAGTGTACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_EMX1	GAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_FANCF	GGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_2	GAACACAAAGCATAGACTGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_3	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_HEK293_4	GGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
fwd_RNF2	GTCATCTTAGTCATTACCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

[0587]

[0588] 시험관내 데아미나제 검정에 사용된 모든 ssDNA 기질의 서열. 서열식별번호: 197-199는 각각 하기에 위에서 아래로 나타난다.

rAPOBEC1 기질	Cy3-ATTATTATTATTCCGCGGATTTATTTATTTATTTATTTATTT
hAID/pmCDA1 기질	Cy3-ATTATTATTATTAGCTATTTATTTATTTATTTATTTATTT
hAPOBEC3G 기질	Cy3-ATTATTATTATTCCCGGATTTATTTATTTATTTATTTATTT

[0589]

[0590] 겔-기반 데아미나제 검정을 위한 sgRNA의 T7 전사를 위한 기질로서 기능하는 PCR 생성물을 생성하는 데 사용된 프라이머. rev_gRNA_T7을 모든 경우에 사용하였다. pFYF1320 플라스미드를 물질 및 방법 섹션에 언급된 바와 같이 주형으로서 사용하였다. 서열식별번호: 200-223은 각각 하기에 위에서 아래로 나타난다.

rev_sgRNA_T7	AAAAAAAGCACCGACTCGGTG
fwd_sgRNA_T7_dsDNA_2	TAATACGACTCACTATAGGCCGCGGATTTATTTATTTAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_3	TAATACGACTCACTATAGGTCCGCGGATTTATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_4	TAATACGACTCACTATAGGTTCCGCGGATTTATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_5	TAATACGACTCACTATAGGATTCCGCGGATTTATTTATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_6	TAATACGACTCACTATAGGTATTCCGCGGATTTATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_7	TAATACGACTCACTATAGGTTATTCCGCGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_8	TAATACGACTCACTATAGGATTATTCCGCGGATTTATTTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_9	TAATACGACTCACTATAGGTATTATTCCGCGGATTTATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_10	TAATACGACTCACTATAGGATTATTATCCGCGGATTTATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_11	TAATACGACTCACTATAGGTATTATATTCCGCGGATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_12	TAATACGACTCACTATAGGTTATTATATTCCGCGGATTTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_13	TAATACGACTCACTATAGGATTATTATATTCCGCGGATTGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_14	TAATACGACTCACTATAGGTATTATTATATTCCGCGGATGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_15	TAATACGACTCACTATAGGATTATTATTATTACCGCGGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_18	TAATACGACTCACTATAGGATTATTATTATTATTACCGCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_noC	TAATACGACTCACTATAGGATATTAATTTATTTATTTAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_APOE4_C112R	TAATACGACTCACTATAGGGGAGGACGTGCGCGGCCGCCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_APOE4_C158R	TAATACGACTCACTATAGGGAAGCGCCTGGCAGTGTACCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_CTNNB1_T41A	TAATACGACTCACTATAGGCTGTGGCAGTGGCACCAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_HRAS_Q61R	TAATACGACTCACTATAGGCCTCCCGGCCGGCGGTATCCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_Y163C	TAATACGACTCACTATAGGGCTTGCAGATGGCCATGGCGGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_Y236C	TAATACGACTCACTATAGGACACATGCAGTTGTAGTGGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_dsDNA_53_N239D	TAATACGACTCACTATAGGTGTCACACATGTAGTTGTAGGTTTTAGAGCTAGAAATAGCA

[0591]

[0592] 겔-기반 dsDNA 데아미나제 검정에 사용된 80-뉴클레오티드 비표지된 가닥 및 Cy3-표지된 유니버셜 프라이머의 서열. 서열식별번호: 224-248은 각각 하기에 위에서 아래로 나타난다.

Cy3-프라이머	Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTA
dsDNA_2	GTCCATGGATCCAGAGGTCATCCATTAAATAAATAAATCCGCGGGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_3	GTCCATGGATCCAGAGGTCATCCATAAATAAATAAATCCGCGGAAGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_4	GTCCATGGATCCAGAGGTCATCCATAATAAATAAATCCGCGGAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_5	GTCCATGGATCCAGAGGTCATCCAAATAAATAAATCCGCGGAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_6	GTCCATGGATCCAGAGGTCATCCAATAAATAAATCCGCGGAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_7	GTCCATGGATCCAGAGGTCATCCATAAATAAATCCGCGGAATAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_8	GTCCATGGATCCAGAGGTCATCCAAAATAAATCCGCGGAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_9	GTCCATGGATCCAGAGGTCATCCAAATAAATCCGCGGAAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_10	GTCCATGGATCCAGAGGTCATCCAATAAATCCGCGGATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_11	GTCCATGGATCCAGAGGTCATCCATAAATCCGCGGAATATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_12	GTCCATGGATCCAGAGGTCATCCAAAATCCGCGGAATATAATAAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_13	GTCCATGGATCCAGAGGTCATCCAAATCCGCGGAATATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_14	GTCCATGGATCCAGAGGTCATCCAATCCGCGGAATATAATAATAGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_15	GTCCATGGATCCAGAGGTCATCCATCCGCGGTAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_16	GTCCATGGATCCAGAGGTCATCCAGCGGTAATAATAATAATAATGGCTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_noC	GTCCATGGATCCAGAGGTCATCCATTAAATAAATAAATTAATATTACTATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_5U	5Cy3-GTAGGTAGTTAGGATGAATGGAAGGTTGGTGTAGATTATTATCUGCGGATTTATTGGATGACCTCTGGATCCATGGACAT
dsDNA_APOE_C112R	GCACCTCGCCGCGGTACTGCACCAGGCGGCCGCGCACGTCCTCCATGTCTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_APOE_C158R	CGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCGCTTCTGCAGTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_CTNNB1_T41A	GTCTTACCTGGACTCTGGAATCCATTCTGGTGCCACTGCCACAGCTCCTTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_HRAS_Q61R	GGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGCCGGGAGGAGTACTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_Y163C	ACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCATACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_Y236C	AGGTTGGCTCTGACTGTACCACCATCCACTACAACTACATGTGTAACAGTACCAACCTTCCATTCATCCTAACTACCTAC
dsDNA_p53_N239D	TGGCTCTGACTGTACCACCATCCACTACAACTACATGTGTGACAGTTCCTACCAACCTTCCATTCATCCTAACTACCTAC

[0593]

[0594] 고처리량 서열분석을 위한 sgRNA의 T7 전사를 위한 기질로서 기능하는 PCR 생성물을 생성하는 데 사용된 프라이머. rev_gRNA_T7 (상기)을 모든 경우에 사용하였다. pFYF1320 플라스미드를 물질 및 방법 섹션에 언급된 바와 같이 주형으로서 사용하였다. 서열식별번호: 249-300은 각각 하기에 위에서 아래로 나타낸다.

[0595]

fwd_sgRNA_T7_HTS_ 염기	TAATACGACTCACTATAGGTTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 1A	TAATACGACTCACTATAGGATATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 1C	TAATACGACTCACTATAGGCTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 1G	TAATACGACTCACTATAGGGTATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 2A	TAATACGACTCACTATAGGTAATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 2C	TAATACGACTCACTATAGGTCATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 2G	TAATACGACTCACTATAGGTGATTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 3T	TAATACGACTCACTATAGGTTTTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 3C	TAATACGACTCACTATAGGTTCTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 3G	TAATACGACTCACTATAGGTTGTTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 4A	TAATACGACTCACTATAGGTTAATTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 4C	TAATACGACTCACTATAGGTTACTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 4G	TAATACGACTCACTATAGGTTAGTTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 5A	TAATACGACTCACTATAGGTTATATCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 5C	TAATACGACTCACTATAGGTTATCTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 5G	TAATACGACTCACTATAGGTTATGTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 6A	TAATACGACTCACTATAGGTTATTACGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 6C	TAATACGACTCACTATAGGTTATTCCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 6G	TAATACGACTCACTATAGGTTATTGCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 8A	TAATACGACTCACTATAGGTTATTTCATGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 8T	TAATACGACTCACTATAGGTTATTTCTTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 8C	TAATACGACTCACTATAGGTTATTTCCTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 9A	TAATACGACTCACTATAGGTTATTTCGAGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 9C	TAATACGACTCACTATAGGTTATTTCGCGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 9G	TAATACGACTCACTATAGGTTATTTCGGGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 10A	TAATACGACTCACTATAGGTTATTTCGTAGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 10T	TAATACGACTCACTATAGGTTATTTCGTTGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 10C	TAATACGACTCACTATAGGTTATTTCGTCGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 11A	TAATACGACTCACTATAGGTTATTTCGTGAATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 11T	TAATACGACTCACTATAGGTTATTTCGTGTATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 11C	TAATACGACTCACTATAGGTTATTTCGTGCATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 12T	TAATACGACTCACTATAGGTTATTTCGTGGTTTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ 12C	TAATACGACTCACTATAGGTTATTTCGTGGCTTTATTTAGTTTTAGAGCTAGAAATAGCA

fwd_sgRNA_T7_HTS_12G	TAATACGACTCACTATAGGTTATTTCGTGGGTTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_13A	TAATACGACTCACTATAGGTTATTTCGTGGAATTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_13C	TAATACGACTCACTATAGGTTATTTCGTGGACTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_13G	TAATACGACTCACTATAGGTTATTTCGTGGAGTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ multiC	TAATACGACTCACTATAGGTTCCCCCCCCGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ TCGCACCC_홀수	TAATACGACTCACTATAGGCGCACCCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ CCTCGCAC_홀수	TAATACGACTCACTATAGGCTCGCACGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ ACCCTCGC_홀수	TAATACGACTCACTATAGGCCCTCGCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ GCACCCTC_홀수	TAATACGACTCACTATAGGCACCCTCGTGGATTTATTTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ TCGCACCC_짝수	TAATACGACTCACTATAGGTCGCACCCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ CCTCGCAC_짝수	TAATACGACTCACTATAGGCCTCGCACGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ ACCCTCGC_짝수	TAATACGACTCACTATAGGACCCTCGCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ GCACCCTC_짝수	TAATACGACTCACTATAGGGCACCCTCGTGGATTTATTAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ EMX1	TAATACGACTCACTATAGGGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ FANCF	TAATACGACTCACTATAGGGGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_부위2	TAATACGACTCACTATAGGGAACACAAAGCATAGACTGCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_부위3	TAATACGACTCACTATAGGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ HEK293_부위4	TAATACGACTCACTATAGGGGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_HTS_ RNF2	TAATACGACTCACTATAGGGTCATCTTAGTCATTACCTGGTTTTAGAGCTAGAAATAGCA

[0596]

[0597] 고처리량 서열분석 (HTS)을 위한 시험관내-편집된 dsDNA의 서열. 편집된 가닥의 서열이 제시된다. 제시된 모든 서열의 역방향 보체를 또한 얻었다. dsDNA 기질을 물질 및 방법에 기재된 바와 같이 상보적 가닥을 어닐링함으로써 얻었다. EMX1, FANCF, HEK293 부위 2, HEK293 부위 3, HEK293 부위 4, 및 RNF2 유전자좌를 나타내는 올리고뉴클레오티드는 원래 겔-기반 데아미나제 검정에 사용하기 위해 설계되었으며, 따라서 그의 5'-말단 상에 동일한 25-nt 서열을 갖는다 (Cy3-프라이머의 것과 매칭시킴). 서열식별번호: 301-352는 각각 하기에 위에서 아래로 나타난다.

[0598]

염기 서열	
	ACGTAAACGGCCACAAGTTCTTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1A	ACGTAAACGGCCACAAGTTCATATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1C	ACGTAAACGGCCACAAGTTCCTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
1G	ACGTAAACGGCCACAAGTTCGTATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2A	ACGTAAACGGCCACAAGTTCTAATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2C	ACGTAAACGGCCACAAGTTCTCATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
2G	ACGTAAACGGCCACAAGTTCTGATTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3T	ACGTAAACGGCCACAAGTTCTTTTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3C	ACGTAAACGGCCACAAGTTCTTCTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
3G	ACGTAAACGGCCACAAGTTCTTGTTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4A	ACGTAAACGGCCACAAGTTCTTAATTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4C	ACGTAAACGGCCACAAGTTCTTACTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
4G	ACGTAAACGGCCACAAGTTCTTAGTTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5A	ACGTAAACGGCCACAAGTTCTTATATCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5C	ACGTAAACGGCCACAAGTTCTTATCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
5G	ACGTAAACGGCCACAAGTTCTTATGTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6A	ACGTAAACGGCCACAAGTTCTTATTACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6C	ACGTAAACGGCCACAAGTTCTTATTCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
6G	ACGTAAACGGCCACAAGTTCTTATTGCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
8A	ACGTAAACGGCCACAAGTTCTTATTTCATGGATTTATTTATGGCATCTTCTTCAAGGACG
8T	ACGTAAACGGCCACAAGTTCTTATTTCTTGGATTTATTTATGGCATCTTCTTCAAGGACG
8C	ACGTAAACGGCCACAAGTTCTTATTTCCTGGATTTATTTATGGCATCTTCTTCAAGGACG
9A	ACGTAAACGGCCACAAGTTCTTATTTCGAGGATTTATTTATGGCATCTTCTTCAAGGACG
9C	ACGTAAACGGCCACAAGTTCTTATTTCGCGGATTTATTTATGGCATCTTCTTCAAGGACG
9G	ACGTAAACGGCCACAAGTTCTTATTTCGGGGATTTATTTATGGCATCTTCTTCAAGGACG
10A	ACGTAAACGGCCACAAGTTCTTATTTCGTAGATTTATTTATGGCATCTTCTTCAAGGACG
10T	ACGTAAACGGCCACAAGTTCTTATTTCGTTGATTTATTTATGGCATCTTCTTCAAGGACG
10C	ACGTAAACGGCCACAAGTTCTTATTTCGTCGATTTATTTATGGCATCTTCTTCAAGGACG
11A	ACGTAAACGGCCACAAGTTCTTATTTCGTGAATTTATTTATGGCATCTTCTTCAAGGACG
11T	ACGTAAACGGCCACAAGTTCTTATTTCGTGTATTTATTTATGGCATCTTCTTCAAGGACG
11C	ACGTAAACGGCCACAAGTTCTTATTTCGTGCATTTATTTATGGCATCTTCTTCAAGGACG

12T	ACGTAAACGGCCACAAGTTCTTATTTCGTGGTTTTATTTATGGCATCTTCTTCAAGGACG
12C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGCTTTATTTATGGCATCTTCTTCAAGGACG
12G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGGTTTATTTATGGCATCTTCTTCAAGGACG
13A	ACGTAAACGGCCACAAGTTCTTATTTCGTGGAATTATTTATGGCATCTTCTTCAAGGACG
13C	ACGTAAACGGCCACAAGTTCTTATTTCGTGGACTTATTTATGGCATCTTCTTCAAGGACG
13G	ACGTAAACGGCCACAAGTTCTTATTTCGTGGAGTTATTTATGGCATCTTCTTCAAGGACG
multiC	ACGTAAACGGCCACAAGTTCTTCCCCCCCCGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC_홀수	ACGTAAACGGCCACAAGTTTCGCACCCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
CCTCGCAC_홀수	ACGTAAACGGCCACAAGTTCCTCGCACGTGGATTTATTTATGGCATCTTCTTCAAGGACG
ACCCTCGC_홀수	ACGTAAACGGCCACAAGTTACCCTCGCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
GCACCCTC_홀수	ACGTAAACGGCCACAAGTTGCACCCTCGTGGATTTATTTATGGCATCTTCTTCAAGGACG
TCGCACCC_짝수	ACGTAAACGGCCACAAGTATTCGCACCCGTGGATTTATTATGGCATCTTCTTCAAGGACG
CCTCGCAC_짝수	ACGTAAACGGCCACAAGTATCCTCGCACGTGGATTTATTATGGCATCTTCTTCAAGGACG
ACCCTCGC_짝수	ACGTAAACGGCCACAAGTATACCCTCGCGTGGATTTATTATGGCATCTTCTTCAAGGACG
GCACCCTC_짝수	ACGTAAACGGCCACAAGTATGCACCCTCGTGGATTTATTATGGCATCTTCTTCAAGGACG
EMX1_시험관내	GTAGGTAGTTAGGATGAATGGAAGGTTGGTAGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTG
FANCF_시험관내	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGCTTCTGG
HEK293_부위2_시험관내	GTAGGTAGTTAGGATGAATGGAAGGTTGGTAAACTGGAACACAAAGCATAGACTGCGGGGCGGGCCAGCCTGAATAGCTG
HEK293_부위3_시험관내	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAGCCCTGCT
HEK293_부위4_시험관내	GTAGGTAGTTAGGATGAATGGAAGGTTGGTACCGGTGGCACTGCGGCTGGAGGTGGGGGTTAAAGCGGAGACTCTGGTGC
RNF2_시험관내	GTAGGTAGTTAGGATGAATGGAAGGTTGGTATGGCAGTCATCTTAGTCATTACCTGAGGTGTTCGTTGTAACTCATATAA

[0599]

[0600] 시험관내 편집된 dsDNA의 HTS를 위한 프라이머. 서열식별번호: 353-361은 각각 하기에 위에서 아래로 나타난다.

fwd_시험관내_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACGTAAACGGCCACAA
rev_시험관내_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGTCCTTGAAGAAGATGC
fwd_시험관내_HEK_표적	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTAGGTAGTTAGGATGAATGGAA
rev_EMX1_시험관내	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCGGTTGATGTGATGG
rev_FANCF_시험관내	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGAAGCTCGGAAAAGC
rev_HEK293_부위2_시험관내	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCTATTCAGGCTGGC
rev_HEK293_부위3_시험관내	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGCAGGGCTTCCTTTC
rev_HEK293_부위4_시험관내	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCACCAGAGTCTCCG
rev_RNF2_시험관내	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTATATGAGTTACAACGAACACC

[0601]

[0602] 모든 포유동물 세포 배양 실험으로부터의 온-타겟 및 오프-타겟 부위의 HTS를 위한 프라이머. 서열식별번호: 362-469는 각각 하기에 위에서 아래로 나타난다.

[0603]

fwd_EMX1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_EMX1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_FANCF_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
rev_FANCF_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_HEK293_부위2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGCCCCATCTGTCAAACT
rev_HEK293_부위2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAATGGATTCCTTGGAAACAATGA
fwd_HEK293_부위3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
rev_HEK293_부위3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
fwd_HEK293_부위4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAACCCAGGTAGCCAGAGAC
rev_HEK293_부위4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
fwd_RNF2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTCTTCTTTATTTCCAGCAATGT
rev_RNF2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTTTTCATGTTCTAAAAATGTATCCCA
fwd_p53_Y163C_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTACAGTACTCCCCTGCCCTC
rev_p53_Y163C_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCTCACCATCGCTATCT
fwd_p53_N239D_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTCATCTTGGGCCTGTGTT
rev_p53_N239D_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAAATCGGTAAGAGGTGGGCC
fwd_APOE4_C158R_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCGGACATGGAGGACGTG
rev_APOE4_C158R_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTTCCACCAGGGGCCC
fwd_EMX1_오프1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGCCCAATCATTGATGCTTTT
rev_EMX1_오프1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGAAACATTTACCATAGACTATCACCT
fwd_EMX1_오프2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAGCCTCTTTCTCAATGTGC
rev_EMX1_오프2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAAGGATGCAGTCT
fwd_EMX1_오프3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTAGACTCCGAGGGGA
rev_EMX1_오프3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCTGCTCTCACTT
fwd_EMX1_오프4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAGGCTGAAGAGGAAGACCA
rev_EMX1_오프4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCAGCTGTGCATTCTAT

fwd_EMX1_오프6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAAGAGGGCCAAGTCCTG
rev_EMX1_오프6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCGAGGAGTGACAGCC
fwd_EMX1_오프7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACTCCACCTGATCTCGGGG
rev_EMX1_오프7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGAGGAGGGAGGGAGCAG
fwd_EMX1_오프8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCACAAATGCCCAAGAGAC
rev_EMX1_오프8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGACACAGTCAAGGGCCGG
fwd_EMX1_오프9_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCACCTTTGAGGAGGCAAA
rev_EMX1_오프9_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCATCTGAGAAGAGAGTGGT
fwd_EMX1_오프10_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCATACCTTGGCCCTTCCT
rev_EMX1_오프10_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCTAGGCCCACACCAG
fwd_FANCF_오프1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCCACTGAAGAAGCAGGG
rev_FANCF_오프1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAATCCGGCTCCAT
fwd_FANCF_오프2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCAGTGTTTCCATCCCGAA
rev_FANCF_오프2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCTCCACAACTCT
fwd_FANCF_오프3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGGTACAGTTCTGCGTGT
rev_FANCF_오프3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCACTCTGAGCATCGCCAAG
fwd_FANCF_오프4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTTTAGAGCCAGTGAACTAGAG
rev_FANCF_오프4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAAGACAAAATCCTCTTTATACTTTG
fwd_FANCF_오프5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGAGGGGACGGCCTTAC
rev_FANCF_오프5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCTGGCGAACATGGC
fwd_FANCF_오프6_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTGGTTAAGAGCATGGGC
rev_FANCF_오프6_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATTGAGTCCCCACAGCACA
fwd_FANCF_오프7_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGTGTTTCCCATCCCCAA
rev_FANCF_오프7_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACCTCCACAACTGGAAAAT
fwd_FANCF_오프8_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTTCCAGACCCACCTGAAG
rev_FANCF_오프8_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCGAGGAAAATTGCTTGTCG
fwd_HEK293_부위2_오프1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGTGGAGAGTGAGTAAGCCA
rev_HEK293_부위2_오프1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGGTAGGATGATTTCAGGCA
fwd_HEK293_부위2_오프2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAAAGCAGTGTAGCTCAGG
rev_HEK293_부위2_오프2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTTGGTACTCGAGTGTTATTCAG
fwd_HEK293_부위3_오프1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_부위3_오프1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK293_부위3_오프2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
rev_HEK293_부위3_오프2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_부위3_오프3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_부위3_오프3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT
fwd_HEK293_부위3_오프4_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTAGCACTTTGGAAGGTCG
rev_HEK293_부위3_오프4_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCTGCTCAGCC
fwd_HEK293_부위3_오프5_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAAAGGAGCAGCTCTTCCTGG
rev_HEK293_부위3_오프5_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGCACCATCTCCCACAA

[0604]

fwd_HEK293_부위4_오프1_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCATGGCTTCTGAGACTCA
rev_HEK293_부위4_오프1_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCACTCCCTGTCTTT
fwd_HEK293_부위4_오프2_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGGCAATGGAGGCATTGG
rev_HEK293_부위4_오프2_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGAGGCTGCCCATGAGAG
fwd_HEK293_부위4_오프3_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGTCTGAGGCTCGAATCCTG
rev_HEK293_부위4_오프3_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCATATCCCTG
fwd_HEK293_부위4_오프4_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTCCACCAGAACTCAGCCC
rev_HEK293_부위4_오프4_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCCTCCACAACAC
fwd_HEK293_부위4_오프5_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACGGGAAGGACAGGAGAAC
rev_HEK293_부위4_오프5_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGGGGAGGGATAAAGCAG
fwd_HEK293_부위4_오프6_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCACGGGAGATGGCTTATGT
rev_HEK293_부위4_오프6_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTCACATCCTCACTGTGCCACT
fwd_HEK293_부위4_오프7_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGTCTCGGCCCCTCA
rev_HEK293_부위4_오프7_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCACTGTAAAGCTCTTGGG
fwd_HEK293_부위4_오프8_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGGGTAGAGGGACAGAGCTG
rev_HEK293_부위4_오프8_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTGGACCCCACATAGTCAGTGC
fwd_HEK293_부위4_오프9_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTGTCAGCCCTATCTCCATC
rev_HEK293_부위4_오프9_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGGCAATTAGGACAGGGAC
fwd_HEK293_부위4_오프10_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCAGCGGAGGAGGTAGATTG
rev_HEK293_부위4_오프10_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAGTACCTGGAGTCCCGA
fwd_HEK2_ChIP_오프1_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGACAGGCTCAGGAAAGCTGT
rev_HEK2_ChIP_오프1_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTACACAAGCCTTTCTCCAGGG
fwd_HEK2_ChIP_오프2_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAATAGGGGGTGAGACTGGGG
rev_HEK2_ChIP_오프2_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCAGACGAGACTTGAGG
fwd_HEK2_ChIP_오프3_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCCAGCAGGAAAGGAATCT
rev_HEK2_ChIP_오프3_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACTGCACCTGTAGCCATG
fwd_HEK2_ChIP_오프4_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCAAGGAAATCACCCTGCCC
rev_HEK2_ChIP_오프4_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTAACTTCCTTGGTGTGCAGCT
fwd_HEK2_ChIP_오프5_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGGGCTCAGCTACGTCATG
rev_HEK2_ChIP_오프5_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTAATAGCAGTGTGGTGGGCAA
fwd_HEK3_ChIP_오프1_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCACATCCCTTGTCTCTCT
rev_HEK3_ChIP_오프1_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTCTACTGGAGCACACCCCAAG
fwd_HEK3_ChIP_오프2_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGGGTCACGTAGCTTTGGTC
rev_HEK3_ChIP_오프2_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGGCCATGTGCAACTAA
fwd_HEK3_ChIP_오프3_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTACTACGTGCCAGCTCAGG
rev_HEK3_ChIP_오프3_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTACCTCCCCTCCTCACTAACC
fwd_HEK3_ChIP_오프4_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCTCAGCTCCATTTCCTGT
rev_HEK3_ChIP_오프4_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTAACCTTTATGGCACCAGGGG
fwd_HEK3_ChIP_오프5_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTCAGCATTAGCAGGCT
rev_HEK3_ChIP_오프5_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCCTGGCTTTCCGATTCCC

[0605]

fwd_HEK4_ChIP_오프1_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGCAATTGGAGGAGGAGCT
rev_HEK4_ChIP_오프1_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCAGCTACAGGCAGAACA
fwd_HEK4_ChIP_오프3_HTS ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTACCCCCAACACAGATGG
rev_HEK4_ChIP_오프3_HTS TGGAGTTCAGACGTGTGCTCTTCCGATCTCCACACAACTCAGGTCCTCC

[0606]

[0607] HDR 연구에 사용된 단일-가닥 올리고뉴클레오티드 공여자 주형 (ssODN)의 서열.

[0608] EMX1 센스 (서열식별번호: 470)

TCATCTGTGCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCAGAACCGGAGGACAAAGTACA
AACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTTTGAGCAGAAGAAGAAGGGCTCCCATCACATC
AACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACTAG

[0609] GGT

[0610] EMX1 안티센스 (서열식별번호: 471)

[0611] ACCCTAGTCATTGGAGGTGACATCGATGTCCTCCCCATTGGCCTGCTTCGTGGCAATGCGCCACCG GTTGATGTGATGGGAGCCCTTCTTCTTCTGCTCAAACTCAGGCCCTTCCTCCTCCAGCTTCTGCCGT TTGTACTTTGTCCTCCGGTTCTGGAACCACACCTTCACCTGGGCCAGGGAGGGAGGGGCACAGATG A

[0612] HEK293 부위 3 센스 (서열식별번호: 472)

[0613] CATGCAATTAGTCTATTTCTGCTGCAAGTAAGCATGCATTTGTAGGCTTGATGCTTTTTTTCTGCTTCT CCAGCCCTGGCCTGGGTCAATCCTTGGGGCTTAGACTGAGCACGTGATGGCAGAGGAAAGGAAGC CCTGCTTCCTCCAGAGGGCGTCGCAGGACAGCTTTTCCTAGACAGGGGCTAGTATGTGCAGCTCCT

[0614] HEK293 부위 3 안티센스 (서열식별번호: 473)

[0615] AGGAGCTGCACATACTAGCCCCTGTCTAGGAAAAGCTGTCCTGCGACGCCCTCTGGAGGAAGCAGG GCTTCCTTTCCTCTGCCATCACGTGCTCAGTCTAAGCCCCAAGGATTGACCCAGGCCAGGGCTGGA GAAGCAGAAAAAAAGCATCAAGCCTACAAATGCATGCTTACTTGCAGCAGAAATAGACTAATTGCATG

[0616] HEK 부위 4 센스 (서열식별번호: 474)

[0617] GGCTGACAAAGGCCGGGCTGGGTGGAAGGAAGGGAGGAAGGGCGAGGCAGAGGGTCCAAAGCAG GATGACAGGCAGGGGCACCGCGGCGCCCCGGTGGCATTGCGGCTGGAGGTGGGGGTTAAAGCGG AGACTCTGGTGCTGTGTGACTACAGTGGGGGCCCTGCCCTCTCTGAGCCCCCGCCTCCAGGCCTGT GTGTGT

[0618] HEK 부위 4 안티센스 (서열식별번호: 475)

[0619] ACACACACAGGCCTGGAGGCGGGGGCTCAGAGAGGGCAGGGCCCCCACTGTAGTCACACAGCACC AGAGTCTCCGCTTTAACCCCCACCTCCAGCCGCAATGCCACCGGGGCGCCGCGGTGCCCCTGCCT GTCATCCTGCTTTGGACCCTCTGCCTCGCCCTTCCTCCCTTCCTTCCACCCAGCCCGGCCTTTGTCA GCC

[0620] APOE4 센스 (서열식별번호: 476)

[0621] AGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCG CGATGCCGATGACCTGCAGAAGTGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAG CGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCTGGTGGAACAGGGCCGCGTGCGGGCCG CCACTGT

[0622] APOE4 안티센스 (서열식별번호: 477)

[0623] ACAGTGGCGGCCCGCACGCGGCCCTGTTCCACCAGGGGCCCCAGGCGCTCGCGGATGGCGCTGA GGCCGCGCTCGGCGCCCTCGCGGGCCCCGGCCTGGTACACTGCCAGGCACTTCTGCAGGTCATCG GCATCGCGGAGGAGCCGCTTACGCAGCTTGCGCAGGTGGGAGGCGAGGCGCACCCGCAGCTCCT CGGTGCT

[0624] p53 Y163C 센스 (서열식별번호: 478)

[0625] ACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGA TTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACATGAC GGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTGAGCAGCTGGGGC TG

[0626] p53 Y163C 안티센스 (서열식별번호: 479)

[0627] CAGCCCCAGCTGCTCACCATCGCTATCTGAGCAGCGCTCATGGTGGGGGCAGCGCCTCACAACCTC CGTCATGTGCTGTGACTGCTTGTAGATGGCCATGGCGCGGACGCGGGGTGCCGGGCGGGGGTGTGG AATCAACCCACAGCTGCACAGGGCAGGTCTTGGCCAGTTGGCAAAACATCTTGTTGAGGGCAGGGG AGT

[0628] 데아미나제 유전자 g블록스 유전자 단편

[0629] hAID (서열식별번호: 169)

[0630] rAPOBEC1 (포유동물)(서열식별번호: 170)

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGAGCTCAGAGACTGGCCCAGTGGCTGTGGACC
CCACATTGAGACGGCGGATCGAGCCCCATGAGTTTGAGGTATTCTTCGATCCGAGAGAGCTCCGCA
AGGAGACCTGCCTGCTTTACGAAATTAATTGGGGGGGCCGGCACTCCATTTGGCGACATACATCACA
GAACACTAACAAGCACGTCGAAGTCAACTTCATCGAGAAGTTCACGACAGAAAGATATTTCTGTCCG
AACACAAGGTGCAGCATTACCTGGTTTCTCAGCTGGAGCCCATGCGGCGAATGTAGTAGGGCCATC
ACTGAATTCCTGTCAAGGTATCCCCACGTCACTCTGTTTATTTACATCGCAAGGCTGTACCACCACGC
TGACCCCCGCAATCGACAAGGCCTGCGGGATTTGATCTCTTCAGGTGTGACTATCCAAATTATGACT
GAGCAGGAGTCAGGATACTGCTGGAGAAACTTTGTGAATTATAGCCCGAGTAATGAAGCCCACTGG
CCTAGGTATCCCCATCTGTGGGTACGACTGTACGTTCTTGAACTGTACTGCATCATACTGGGCCTGC
CTCCTTGTCTCAACATTCTGAGAAGGAAGCAGCCACAGCTGACATTCTTTACCATCGCTCTTCAGTCT
TGTCATTACCAGCGACTGCCCCCACACATTCTCTGGGCCACCGGGTTGAAATGAGCGGCCGCTCGA
TTGGTTTGGTGTGGCTCTAA

[0631]

[0632] pmCDA1 (서열식별번호: 171)

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGACAGACGCTGAATATGTTAGGATCCATGAAAA
ACTGGATATCTATACATTTAAGAAGCAGTTCTTCAATAACAAAAAGTCAGTATCTCACAGATGCTATGT
CCTGTTCGAACTCAAGAGAAGAGGAGAAAGGCGGGCCTGTTTCTGGGGGTACGCGGTTAATAAACC
CCAGTCCGGGACCGAGAGGGGGATTCACGCCGAGATCTTTTCAATTAGGAAGGTTGAAGAGTATCT
TCGCGACAATCCCGGTCAGTTCACAATTAACTGGTACAGCTCCTGGAGCCCTTGCGCTGATTGCGCC
GAGAAAATACTCGAATGGTACAACCAGGAGTTGAGAGGCAATGGCCACACTCTCAAGATTTGGGCTT
GCAAGCTTTACTACGAGAAGAACGCGAGAAATCAGATTGGCTTGTGGAACCTCAGGGACAACGGGG
TCGGGTTGAATGTTATGGTGTCCGAACATTACCAGTGCTGTAGAAAGATCTTCATTCAGTCCAGTCAC
AATCAGCTGAACGAGAACAGATGGCTGGAGAAAACACTGAAACGGGCAGAGAAAAGGCGCTCAGAG
CTGAGTATCATGATCCAGGTCAAAATCCTGCATACAACCAAAAGCCCGGCTGTATAAGCGGCCGCTC
GATTGGTTTGGTGTGGCTCTAA

[0633]

[0634] haPOBEC3G (서열식별번호: 172)

CATCCTTGGTACCGAGCTCGGATCCAGCCACCATGGAGCTGAAGTATCACCCTGAGATGCGGTTTTT
CCACTGGTTTAGTAAGTGGCGCAAACTTCATCGGGATCAGGAGTATGAAGTGACCTGGTATATCTCT
TGGTCTCCCTGCACAAAATGTACACGCGACATGGCCACATTTCTGGCCGAGGATCCAAAGGTGACG
CTCACAATCTTTGTGGCCCGCCTGTATTATTTCTGGGACCCGGATTATCAGGAGGCACTTAGGTCAT
TGTGCCAAAAGCGCGACGGACCACGGGCGACTATGAAAATCATGAATTATGACGAATTCCAGCATTG
CTGGAGTAAGTTTGTGTACAGCCAGCGGGAGCTGTTCGAGCCCTGGAACAATCTTCCCAAGTACTAC
ATACTGCTTCACATTATGTTGGGGGAGATCCTTCGGCACTCTATGGATCCTCCTACCTTTACGTTTAA
CTTTAATAATGAGCCTTGGGTTCGCGGGCGCCATGAAACCTATTTGTGCTACGAGGTCGAGCGGATG
CATAATGATACGTGGGTCCTGCTGAATCAGAGGAGGGGGTTTCTGTGTAACCAGGCTCCACATAAAC
ATGGATTTCTCGAGGGGCGGCACGCCGAACTGTGTTTCCTTGATGTGATACCTTTCTGGAAGCTCGA
CCTTGATCAAGATTACAGGGTGACGTGTTTCACCTCCTGGTCACCCTGCTTCAGTTGCGCCCAAGAG
ATGGCTAAATTTATCAGTAAGAACAAGCATGTGTCCCTCTGTATTTTTACAGCCAGAATTTATGATGAC
CAGGGCCGGTGCCAGGAGGGGCTGCGGACACTCGCTGAGGCGGGCGCGAAGATCAGCATAATGA
CATACTCCGAATTCAAACACTGTTGGGACACTTTTGTGGACCACCAGGGCTGCCCATTTCAGCCGTG
GGATGGGCTCGACGAACATAGTCAGGATCTCTCAGGCCGGCTGCGAGCCATATTGCAGAACCAGGA
GAATTAGGCGGCCGCTCGATTGGTTTGGTGTGGCTCTAA

[0635]

[0636] rAPOBEC1(이. 콜라이) (서열식별번호: 173)

GGCCGGGGATTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGATGTCTTCTGAAA
CCGGTCCGGTTGCGGTTGACCCGACCCTGCGTCGTCGTATCGAACCGCACGAATTCGAAGTTTTCT
TCGACCCGCGTGAACTGCGTAAAGAAACCTGCCTGCTGTACGAAATCAACTGGGGTGGTCGTCACT
CTATCTGGCGTCACACCTCTCAGAACACCAACAAACACGTTGAAGTTAACTTCATCGAAAAATTCACC
ACCGAACGTTACTTCTGCCCGAACACCCGTTGCTCTATCACCTGGTTCCTGTCTTGGTCTCCGTGCG
GTGAATGCTCTCGTGCGATCACCGAATTCCTGTCTCGTTACCCGCACGTTACCCTGTTCATCTACATC
GCGCGTCTGTACCACCACGCGGACCCGCGTAACCGTCAGGGTCTGCGTGACCTGATCTCTTCTGGT
GTTACCATCCAGATCATGACCGAACAGGAATCTGGTTACTGCTGGCGTAACTTCGTTAACTACTCTCC
GTCTAACGAAGCGCACTGGCCGCGTTACCCGCACCTGTGGGTTCGTCTGTACGTTCTGGAACTGTA
CTGCATCATCCTGGGTCTGCCGCCGTGCCTGAACATCCTGCGTCGTAAACAGCCGCAGCTGACCTT

CTTCACCATCGCGCTGCAGTCTTGCCACTACCAGCGTCTGCCGCCGCACATCCTGTGGGCGACCGG
TCTGAAAGGTGGTAGTGGAGGGAGCGGCGGTTCAATGGATAAGAAATAC

[0637]

[0638] NBE1, NBE2 및 NBE3의 아미노산 서열.

[0639] 이. 콜라이 발현을 위한 NBE1 (His_6-rAPOBEC1-XTEN-dCas9) (서열식별번호: 154)

MGSSHHHHHHMSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQ
LTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

[0640]

[0641] 포유동물 발현을 위한 NBE1 (rAPOBEC1-XTEN-dCas9-NLS) (서열식별번호: 155)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV

[0642]

[0643] 인간 APOBEC1로의 포유동물 발현을 위한 대안적인 NBE1 (hAPOBEC1-XTEN-dCas9-NLS) (서열식별번호: 158)

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNT
TNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVA
RLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPP
LWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPS
VAWRGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH
RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL
AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP
KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
[0644]

KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE
HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKK
RKV
[0645]

[0646] NBE2 (rAPOBEC1-XTEN-dCas9-UGI-NLS) (서열식별번호: 156)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0647]

[0648] NBE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS) (서열식별번호: 157)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0649]

[0650] pmCDA1-XTEN-dCas9-UGI (박테리아) (서열식별번호: 159)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSI
GLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNL
IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR
ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES
ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDSGGSMTNLSDIIEKETGKQLVIQESILMLPEEVEE
VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

[0651]

[0652] pmCDA1-XTEN-nCas9-UGI-NLS (포유동물 구축물) (서열식별번호: 160)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKP
QSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGN
GHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL
NENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSI
GLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNL
IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK
QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR
ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM
YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES
ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVI
GNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPK
KKRKV

[0653]

[0654] huAPOBEC3G-XTEN-dCas9-UGI (박테리아) (서열식별번호: 161)

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[0655]

QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSMTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDE
NVMLLTSDAPEYKPWALVIQDSNGENKIKML

[0656]

[0657] huAPOBEC3G-XTEN-nCas9-UGI-NLS (포유동물 구축물) (서열식별번호: 162)

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0658]

[0659] huAPOBEC3G (D316R_D317R)-XTEN-nCas9-UGI-NLS (포유동물 구축물) (서열식별번호: 163)

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGF
LEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIF
TARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEH
SQDLSGRLRAILQSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKK
FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP
AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGI
KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW
DKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0660]

[0661] 염기 호출 매트랩 스크립트

WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCG
GCCAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGC
CGATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATC
CGCGAGCG CCTGGGGCCCCTGGTGGAACAG'(SEQ ID NO: 164);

```
%cycle through fastq files for different samples   files=dir('*.fastq');
for d=1:20
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength=length(seqs);                                    % number of sequences  seqsFile=
strrep(filename,'.fastq','');                               % trims off .fastq
%create a directory with the same name as fastq file  ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before   moving on.');
end
mkdir(seqsFile);                                            % make directory
wtLength=length(WTnuc);                                         % length of wildtype sequence
%% aligning back to the wildtype nucleotide sequence
%
```

[0662]

```
% AlN is a matrix of the nucleotide alignment  window=1:wtLength;
sBLength = length(seqs);                                    % number of sequences
% counts number of skips  nSkips = 0;
ALN=repmat('',[sBLength wtLength]);
% iterate through each sequencing read  for i = 1:sBLength
%If you only have forward read fastq files leave as is
%If you have R1 foward and R2 is reverse fastq files uncomment the
%next four lines of code and the subsequent end statement
%                   if mod(d,2)==0;
%                           reverse=seqrcomplement(seqs{i});
%                           [score,alignment,start]=
swalign(reverse,WTnuc,'Alphabet','NT');
%                   else

[score,alignment,start]=swalign(seqs{i},WTnuc,'Alphabet','NT');
%                   end

% length of the sequencing read  len=
length(alignment(3,:));
% if there is a gap in the alignment , skip = 1 and we will
% throw away the entire read  skip = 0;
for j = 1:len
if (alignment(3,j) == '-' || alignment(1,j) == '-')  skip = 1;
                                break;
end
%in addition if the qscore for any given base in the read is
%below 31 the nucleotide is turned into an N (fastq qscores that  are not letters)
ifisletter(qscore{i}(start(1)+j-1)) else
alignment(1,j) = 'N';
end

end
if skip == 0 && len>10
ALN(i,    start(2):(start(2)+length(alignment)-1))=alignment(1,:);

            end
end
```

[0663]

```
% with the alignment matrices we can simply tally up the occurrences of
% each nucleotide at each column in the alignment these
% tallies ignore bases annotated as N
% due to low qscores
TallyNTD=zeros(5,wtLength); fori=1:wtLength

TallyNTD(:,i)=[sum(ALN(:,i)=='A'),sum(ALN(:,i)=='C'),sum(ALN(:,i)=='G'),sum(A
LN(:,i)=='T'),sum(ALN(:,i)=='N')];
end
% we then save these tally matrices in the respective folder for
% further processing

save(strcat(seqsFile,'/TallyNTD'),'TallyNTD');  dlmwrite(strcat(seqsFile,'/TallyNTD.txt'),TallyNTD,'precision',
'%.3f', 'newline', 'pc');  end
```

[0664]

[0665] INDEL 검출 매트랩 스크립트

```
WTnuc='GCGGACATGGAGGACGTGCGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGC
CAGA
GCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGATGCCG
ATGAC
CTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCTCAGCGCCATCCGC
GAGCG CCTGGGGCCCCTGGTGGAACAG'(SEQ ID NO: 164);
%cycle through fastq files for different samples  files=dir('*.fastq');
%specify start and width of indel window as well as length of each flank  indelstart=154;
width=30; flank=10;

for d=1:3
filename=files(d).name;
%read fastq file
[header,seqs,qscore]=fastqread(filename);
seqsLength=length(seqs);                                    % number of sequences  seqsFile
=strcat(strrep(filename,'.fastq',''),'_INDELS');
%create a directory with the same name as fastq file+_INDELS  ifexist(seqsFile,'dir');
error('Directory already exists. Please rename or move it before  moving on.');
end
mkdir(seqsFile);                                            % make directory
wtLength=length(WTnuc);                                     % length of wildtype sequence  sBLength=
length(seqs);                                               % number of sequences

% initialize counters and cell arrays
nSkips = 0; notINDEL=0;
ins={};
dels={}; NumIns=0;
NumDels=0;
% iterate through each sequencing read  for i = 1:sBLength
%search for 10BP sequences that should flank both sides of the "INDEL  WINDOW"
windowstart=strfind(seqs{i},WTnuc(indelstart-flank:indelstart));

                    windowend=strfind(seqs{i},WTnuc(indelstart+width:indelstart+width+flank
));
%if the flanks are found proceed
```

[0666]

```
iflength(windowstart)==1&&length(windowend)==1
%if the sequence length matches the INDEL window length save as
%not INDEL
if  windowend-windowstart==width+flank notINDEL=notINDEL+1;
%if the sequence is two or more bases longer than the INDEL
%window length save as an Insertion
elseif  windowend-windowstart>=width+flank+2 NumIns=NumIns+1;
ins{NumIns}=seqs{i};
%if the sequence is two or more bases shorter than the INDEL
%window length save as a Deletion
elseif  windowend-windowstart<=width+flank-2  NumDels=NumDels+1;
dels{NumDels}=seqs{i};
%keep track of skipped sequences that are either one base
%shorter or longer than the INDEL window width  else
nSkips=nSkips+1;
end
%keep track of skipped sequences that do not possess matching flank
%sequences else
nSkips=nSkips+1;

            end
end
```

[0667]

```
fid=fopen(strcat(seqsFile,'/summary.txt'),'wt');
fprintf(fid,'Skipped reads %i\n not INDEL %i\n Insertions %i\n Deletions
%i\n', [nSkips, notINDEL, NumIns, NumDels]);  fclose(fid);
save(strcat(seqsFile,'/nSkips'),'nSkips');  save(strcat(seqsFile,'/notINDEL'),'notINDEL');
save(strcat(seqsFile,'/NumIns'),'NumIns');  save(strcat(seqsFile,'/NumDels'),'NumDels');
save(strcat(seqsFile,'/dels'),'dels');
C=dels;
fid=fopen(strcat(seqsFile,'/dels.txt'),'wt');  fprintf(fid,'"%s"\n',C{:});
fclose(fid);
save(strcat(seqsFile,'/ins'),'ins');  C=ins;
fid=fopen(strcat(seqsFile,'/ins.txt'),'wt');  fprintf(fid,'"%s"\n',C{:});
fclose(fid);

end
```

[0668]

[0669] 실시예 5: Cas9 변이체 서열

[0670] 본 개시내용은 1개 이상의 돌연변이 (예를 들어, dCas9 또는 Cas9 니카제를 생성하기 위한)를 포함할 수 있는 1종 이상의 유기체로부터의 Cas9 변이체, 예를 들어 Cas9 단백질을 제공한다. 일부 실시양태에서, Cas9 단백질의 별표에 의해 하기 확인되는 아미노산 잔기 중 1개 이상은 돌연변이될 수 있다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 및/또는 H840 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이는 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이는 D를 제외한 임의의 아미노산 잔기로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이는 A로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 H840 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 잔기는 H이다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 H840 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이는 H를 제외한 임의의 아미노산 잔기로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 H840 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이는 A로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 잔기, 또는 임의의 Cas9 단백질, 예컨대 본원에 제공된 바와 같은 Cas9 아미노산 서열 중 어느 하나에서의 상응하는 잔기는 D이다.

[0671] 다양한 종으로부터의 다수의 Cas9 서열을 정렬하여 서열식별번호: 6 또는 서열식별번호: 567의 D10 및 H840의 상응하는 상동 아미노산 잔기가 다른 Cas9 단백질에서 확인되어, 상동 아미노산 잔기의 상응하는 돌연변이를 갖는 Cas9 변이체의 생성을 허용할 수 있는지 측정하였다. 정렬은 NCBI 제약-기반 다중 정렬 도구(Constraint-based Multiple Alignment Tool) (COBALT (st-va.ncbi.nlm.nih.gov/tools/cobalt에서 접근가능함))를 사용하여 하기 파라미터로 수행하였다. 정렬 파라미터: 갭 패널티 -11,-1; 말단-갭 패널티 -5,-1. CDD 파라미터: RPS BLAST 온을 사용; 블라스트 E-값 0.003; 보존된 칼럼 발견 및 리컴퓨트 온. 쿼리 클러스터링 파라미터: 쿼리 클러스터 온 사용; 워드 크기 4; 최대 클러스터 거리 0.8; 알파벳 규칙적.

[0672] 4가지 Cas9 서열의 예시적인 정렬이 하기에 제공된다. 정렬에서 Cas9 서열은 서열 1 (S1): 서열식별번호: 567 | WP_010922251| gi 499224711 | 유형 II CRISPR RNA-가이드된 엔도뉴클레아제 Cas9 [스트렙토코쿠스 피오게네스]; 서열 2 (S2): 서열식별번호: 568 | WP_039695303 | gi 746743737 | 유형 II CRISPR RNA-가이드된 엔도뉴클레아제 Cas9 [스트렙토코쿠스 갈롤리티쿠스 (Streptococcus gallolyticus)]; 서열 3 (S3): 서열식별번호: 569 | WP_045635197 | gi 782887988 | 유형 II CRISPR RNA-가이드된 엔도뉴클레아제 Cas9 [스트렙토코쿠스 미티스 (Streptococcus mitis)]; 서열 4 (S4): 서열식별번호: 570 | 5AXW_A | gi 924443546 | 스타필로코쿠스 아우레우스 Cas9이다. HNH 도메인 (볼드체 및 밑줄표시됨) 및 RuvC 도메인 (박스표시됨)은 4가지 서열의 각각에 대해 확인된다. S1에서의 아미노산 잔기 10 및 840 및 정렬된 서열에서의 상동 아미노산은 하기 각각의 아미노산 잔기에서 별표로 확인된다.

```
S1 1    --MDKK-YSIGLD*IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI--GALLFDSG--ETAEATRLKRTARRRYT  73
S2 1    --MTKKNYSIGLD*IGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLL--GALLFDSG--ETAEATRLKRTARRRYT  74
S3 1    --M-KKGYSIGLD*IGTNSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLI--GALLFDEG--TTAEARRLKRTARRRYT  73
S4 1    GSHMKRNYILGLD*IGITSVGYGII--DYET----------------RDVIDAGVRLFKEANVENNEGRRSKRGARRLKR  61

S1 74   RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL 153
S2 75   RRKNRLRYLQEIFANEIAKVDESFFQRLDESFLTDDDKTFDSHPIFGNKAEEDAYHQKFPTIYHLRKHLADSSEKADLRL 154
S3 74   RRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLIPEDKRESKYPIFATLTEEKEYHKQFPTIYHLRKQLADSKEKTDLRL 153
S4 62   RRRHRIQRVKKLL--------------FDYNLLTD--------------------HSELSGINPYEARVKGLSQKLSEEE 107
```

[0673]

```
S1 154  IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK  233
S2 155  VYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFVGVYNRTFDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTEK  234
S3 154  IYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEK  233
S4 108  FSAALLHLAKRRG----------------------VHNVNEVEEDT-----------------------------------  131

S1 234  KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT 313
S2 235  KNTLFGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDAILLSGILTVDDNST 314
S3 234  STGLFSEFLKLIVGNQADFKKHFDLEDKAPLQFSKDTYDEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPST 313
S4 132  -----GNELS------------------TKEQISRN-------------------------------------------- 144

S1 314  KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM--DGTEELLV  391
S2 315  KAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKNKNGYAGYIENGVKQDEFYKYLKNILSKIKIDGSDYFLD  394
S3 314  KAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKDGYAGYIDGKTTQETFYKYIKNLLSKF--EGTDYFLD  391
S4 145  ----SKALEEKYVAELQ-----------------------------------------------LERLKKDG------  165

S1 392  KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE  471
S2 395  KIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKEKQDRIEKILTFRIPYYVGPLVRKDSRFAWAEYRSDE  474
S3 392  KIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEYYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDE  471
S4 166  --EVRGSINRFKTSD--------YVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGP--GEGSPFGW------K  227

S1 472  TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL 551
S2 475  KITPWNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETYAVYNELTKIKYVNEQGKE-SFFDSNMKQEIFDH 553
S3 472  AIRPWNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQ 551
S4 228  DIKEW----------------YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEK---LEYYEKFQIIEN 289

S1 552  LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED  628
S2 554  VFKENRKVTKEKLLNYLNKEFPEYRIKDLIGLDKENKSFNASLGTYHDLKKIL-DKAFLDDKVNEEVIEDIIKTLTLFED  632
S3 552  LFKENRKVTEKDIIHYLHN-VDGYDGIELKGIEKQ---FNASLSTYHDLLKIIKDKEFMDDAKNEAILENIVHTLTIFED  627
S4 290  VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF---TNLKVYHDIKDITARKEII---ENAELLDQIAKILTIYQS  363

S1 629  REMIEERLKTYAHLFDDKVMKQLKR-RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED 707
S2 633  KDMIHERLQKYSDIFTANQLKKLER-RHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDTLPFKQI 711
S3 628  REMIKQRLAQYDSLFDEKVIKALTR-RHYTGWGKLSAKLINGICDKQTGNTILDYLIDDGKINRNFMQLINDDGLSFKEI 706
S4 364  SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE------LWHTNDNQIAIFNRLKLVP---------- 428

S1 708  IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT------QKGQKNSRERM 781
S2 712  IQKSQVVGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMG-GNPDNIVIEMARENQTT------NRGRSQSQQRL 784
S3 707  IQKAQVIGKTDDVKQVVQELSGSPAIKKGILQSIKIVDELVKVMG-HAPESIVIEMARENQTT------ARGKKNSQQRY 779
S4 429  -KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG--LPNDIIIELAREKNSKDAQKMINEMQKRNRQTN 505

S1 782  KRIEEGIKELGSQIL-------KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD----YDVDH*IVPQSFLKDD  850
S2 785  KKLQNSLKELGSNILNEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMYTGDELDIDHLSD----YDIDH*IIPQAFIKDD  860
S3 780  KRIEDSLKILASGL---DSNILKENPTDNNQLQNDRLFLYYLQNGKDMYTGEALDINQLSS----YDIDH*IIPQAFIKDD  852
S4 506  ERIEEIIRTTGK----------------ENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDH*IIPRSVSFDN  570

S1 851  SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN-LTKAERGGL-SELD------KAGFIKRQLV  922
```

[0674]

```
S2 861  SIDNRVLTSSAKNRGKSDDVPSLDIVRARKAEWVRLYKSGLISKRKFDN-LTKAERGGL-TEAD------KAGFIKRQLV 932
S3 853  SLDNRVLTSSKDNRGKSDNVPSIEVVQKRKAFWQQLLDSKLISERKFNN-LTKAERGGL-DERD------KVGFIKRQLV 924
S4 571  SFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV 650

S1 923  ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP 1002
S2 933  ETRQITKHVAQILDARFNTEHDENDKVIRDVKVITLKSNLVSQFRKDFEFYKVREINDYHHAHDAYLNAVVGTALLKKYP 1012
S3 925  ETRQITKHVAQILDARYNTEVNEKDKKNRTVKIITLKSNLVSNFRKEFRLYKVREINDYHHAHDAYLNAVVAKAILKKYP 1004
S4 651  DTRYATRGLMNLLRSYFRVN-------NLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIA---------- 712

S1 1003 KLESEFVYGDYKVYDVRKMIAKSEQ--EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG--- 1077
S2 1013 KLASEFVYGEYKKYDIRKFITNSSD------KATAKYFFYSNLMNFFKTKVKYADGTVFERPIIETNAD-GEIAWNKQ--- 1083
S3 1005 KLEPEFVYGEYQKYDLKRYISRSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYSKDTGEIAWNKE--- 1081
S4 713  --NADFIFKEWKKLDKAKKVMENQM-------------------------FEEKQAESMPEIETEQEYKEIFITPHQIK 764

S1 1078 -----RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD---WDPKKYGGFDSPTVAYSVLVVAKV 1149
S2 1084 -----IDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKVYWDTKKYGGFDSPTVAYSVFVVADV 1158
S3 1082 -----KDFAIIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILLDTTKYGGFDSPVIAYSILLIADI 1156
S4 765  HIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL----KKLIN-KSP----EKLLMYHH 835

S1 1150 EKGKSKKLKSVKELLGITIMERSSFEKNPI-DFLEAKG-----YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG 1223
S2 1159 EKGKAKKLKTVKELVGISIMERSFFEENPV-EFLENKG-----YHNIREDKLIKLPKYSLFEFEGGRRRLLASASELQKG 1232
S3 1157 EKGKAKKLKTVKTLVGITIMEKAAFEENPI-TFLENKG-----YHNVRKENILCLPKYSLFELENGRRRLLASAKELQKG 1230
S4 836  DPQTYQKLK--------LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKV 907

S1 1224 NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH------ 1297
S2 1233 NEMVLPGYLVELLYHAHRADNF-----NSTEYLNYVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSM------ 1301
S3 1231 NEIVLPVYLTTLLYHSKNVHKL-----DEPGHLEYIQKHRNEFKDLLNLVSEFSQKYVLADANLEKIKSLYADN------ 1299
S4 908  VKLSLKPYRFD-VYLDNGVYKFV-----TVKNLDVIK--KENYYEVNSKAYEEAKKLKKISNQAEFIASFYNNDLIKING 979

S1 1298 RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT--------GLYETRI----DLSQL 1365
S2 1302 DNFSIEEISNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSIT--------GLYETRI----DLSKL 1369
S3 1300 EQADIEILANSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEILNATLIHQSIT--------GLYETWI----DLSKL 1367
S4 980  ELYRVIGVNNDLLNRIEVNMIDITYR-EYLENMNDKRPPRIIKTIASKT---QSIKKYSTDILGNLYEVKSKKHPQIIKK 1055

S1 1366 GGD  1368
S2 1370 GEE  1372
S3 1368 GED  1370
S4 1056 G--  1056
```

[0675]

[0676] 정렬은 참조 Cas9 아미노산 서열 또는 아미노산 잔기에 상동인 아미노산 서열 및 아미노산 잔기가, 정렬되는 아미노산 서열 또는 잔기를 관련 기술분야에 공지된 정렬 프로그램 및 알고리즘을 사용하여 참조 서열 또는 참조 잔기로 확인함으로써, 상이한 종으로부터의 Cas9 서열을 포함하나 이에 제한되지 않는 Cas9 서열 변이체에 걸쳐 확인될 수 있다는 것을 입증한다. 본 개시내용은 서열식별번호: 567-570에서 별표에 의해 확인되는 아미노산 잔기 중 1개 이상 (예를 들어, 각각 S1, S2, S3, 및 S4)이 본원에 기재된 바와 같이 돌연변이된 Cas9 변이체를 제공한다. 서열식별번호: 567-570에서 별표에 의해 확인되는 잔기에 상응하는 서열식별번호: 6의 Cas9에서의 잔기 D10 및 H840은 본원에서 "상동" 또는 "상응하는" 잔기로 지칭된다. 이러한 상동 잔기는 예를 들어, 상기 기재된 바와 같은 서열 정렬에 의해, 및 정렬되는 서열 또는 잔기를 참조 서열 또는 잔기로 확인함으로써 확인될 수 있다. 유사하게, 본원에서 서열식별번호: 6에서 확인되는 돌연변이, 예를 들어, 서열식별번호: 6에서의 잔기 10, 및 840의 돌연변이에 상응하는 Cas9 서열에서의 돌연변이는 본원에서 "상동" 또는 "상응하는" 돌연변이로 지칭된다. 예를 들어, 상기 4개의 정렬된 서열에 대해 서열식별번호: 6 또는 S1 (서열식별번호: 567)에서의 D10A 돌연변이에 상응하는 돌연변이는 S2에 대해 D11A, S3에 대해 D10A, 및 S4에 대해 D13A이고; 서열식별번호: 6 또는 S1 (서열식별번호: 567)에서의 H840A에 대한 상응하는 돌연변이는 S2에 대해 H850A, S3에 대해 H842A, 및 S4에 대해 H560A이다.

[0677] 실시예 6: 차세대 C에서 T로의 편집제

[0678] 염기 편집제 3 (BE3) 구축물에 대한 대안으로서 시티딘 데아미나제의 다른 패밀리를 조사하였다. 좁거나 상이한 편집 윈도우, 표적화가능한 기질을 확장하는 대안의 서열 특이성을 갖고, 보다 높은 활성을 갖는 상이한 C에서 T로의 편집제를 개발하였다.

[0679] 실시예 4에 기재된 방법을 사용하여, HeK-3 부위에서의 pmCDA1 (페트로미존 마리누스로부터의 시티딘 데아미나

제 1) 활성을 평가한다 (도 42). pmCDA1-nCas9-UGI-NLS (nCas9는 본원에 기재된 Cas9 니카제를 지시함) 구축물은 rAPOBEC1 (BE3)로 접근가능하지 않은 일부 부위 (예를 들어, 위치 9, 5, 4, 및 3에서 상보적 가닥 상의 C 염기) 상에서 활성이다.

[0680] HeK-2 부위에서의 pmCDA1 활성은 도 43에 주어진다. pmCDA1-XTEN-nCas9-UGI-NLS 구축물은 "G"에 인접한 부위 상에서 활성인 반면, rAPOBEC1 유사체 (BE3 구축물)는 "G"에 인접한 "C", 예를 들어, 상보적 가닥 상의 위치 11에서의 C 염기 상에서 낮은 활성을 갖는다.

[0681] T로 전환된 표적 C (도 44), A로 전환된 C (도 45), 및 G로 전환된 C (도 46)를 갖는 총 서열분석 판독물의 퍼센트가 CDA 및 APOBEC1 (BE3 구축물)에 대해 나타나 있다.

[0682] HeK-2 부위에서의 huAPOBEC3G 활성은 도 47에 나타나 있다. 2가지 구축물을 사용하였다: huAPOBEC3G-XTEN-nCas9-UGI-NLS 및 huAPOBEC3G*(D316R_D317R)-XTEN-nCas9-UGI-NLS. huAPOBEC3G-XTEN-nCas9-UGI-NLS 구축물은 도 47에 나타난 바와 같이 rAPOBEC1 (BE3)과 상이한 서열 특이성을 가지며, 편집 윈도우는 APOBEC1에 비해 위치 4에서 APOBEC3G의 감소된 활성에 의해 지시된 바와 같이 좁게 나타난다. huAPOBEC3G에서 생성된 돌연변이 (D316R 및 D317R)는 ssDNA 결합을 증가시켰으며, 편집된 부위를 확장하는 데 대해 관찰가능한 효과를 발생시켰다 (도 47에서 APOBEC3G를 APOBEC3G_RR과 비교함). 돌연변이는 APOBEC3G 결정 구조에 기초하여 선택하였고, 문헌 [Holden et al., Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implication. Nature. (2008); 121-4]을 참조하며, 그의 전체 내용은 본원에 참조로 포함된다.

[0683] 실시예 7: 이. 콜라이에서의 pmCDA1/huAPOBEC3G/rAPOBEC1 연구

[0684] A에서 I로의 전환을 위한 LacZ 선택 최적화를 F 플라스미드 상에서 코딩된 lacZ를 갖는 박테리아 균주를 사용하여 수행하였다. 중요한 글루탐산 잔기를, 시티딘 데아미나제에 의한 G에서 A로가 lacZ 활성을 복원하도록 돌연변이시켰다 (예를 들어, GAG에서 GGG로의, Glu에서 Gly로의 돌연변이) (도 48). 균주 CC102를 선택 검정을 위해 선택하였다. APOBEC1 및 CDA 구축물을 G에서 A로의 전환을 최적화하기 위해 선택 검정에 사용하였다.

[0685] lacZ 복귀 빈도에 대한 데아미나제 구축물을 코딩하는 플라스미드의 카피 수의 효과를 평가하기 위해, CDA 및 APOBEC1 데아미나제를 상이한 복제 원점 (따라서 상이한 카피 수)을 갖는 4가지의 플라스미드, 즉 SC101, CloDF3, RSF1030, 및 PUC (카피 수: PUC>RSF1030>CloDF3>SC101) 내로 클로닝하고, 유도성 프로모터 아래에 정치하였다. 플라스미드를 돌연변이된 LacZ 유전자를 함유하는 F 플라스미드를 갖는 이. 콜라이 세포 내로 개별적으로 형질전환시켰다. 데아미나제의 발현을 유도하고, LacZ 활성을 각각의 구축물에 대해 검출하였다 (도 49). 도 49에 나타난 바와 같이, CDA는 데아미나제가 클로닝된 플라스미드 카피 수와 무관하게, 모든 예에서 APOBEC1보다 유의하게 더 높은 활성을 나타내었다. 또한, 카피 수의 관점에서, 데아미나제 활성은 클로닝되는 플라스미드의 카피 수와 양의 상관관계가 있었다 (즉, PUC> CloDF3>SC101).

[0686] LacZ 복귀를 lacZ 유전자좌에서 게놈 DNA의 서열분석에 의해 확인하였다. 교정된 LacZ 유전자를 함유하는 게놈 DNA를 얻기 위해, 세포를 X-gal을 함유하는 배지에서 성장시켰으며, 여기서, LacZ 활성을 갖는 세포는 청색 콜로니를 형성한다. 청색 콜로니를 선택하고, 락토스를 함유하는 최소 배지에서 성장시켰다. 세포를 스핀 다운시키고, 세척하고, 최소 배지 플레이트 (락토스) 상에 재-플레이팅하였다. 그 후, 최고 희석에서의 청색 콜로니를 선택하고, 그의 게놈 DNA를 lacZ 유전자좌에서 서열분석하였다 (도 50).

[0687] 상이한 시티딘 데아미나제 (예를 들어, CDA, 및 APOBEC1)의 활성을 시험하기 위한 클로람페니콜 복귀 검정을 설계하였다. 박테라아에 대한 클로람페니콜 저항성을 부여하는 돌연변이체 CAT1 유전자를 갖는 플라스미드를 복제 원점으로서 RSF1030으로 구축하였다. 돌연변이체 CAT1 유전자는 단백질 불활성을 제공하는 H195R (CAC에서 CGC로의) 돌연변이를 갖는 CAT1 단백질을 코딩한다 (도 51). CGC 코돈에서의 G 염기에 염기-쌍형성된 C의 탈아미노화는 코돈을 다시 CAC 코돈으로 전환시킬 것이며, 이는 단백질의 활성을 복원한다. 도 52에 나타난 바와 같이, CDA는 이. 콜라이에서 클로람페니콜 저항성 유전자의 활성을 복원하는 데 있어서 rAPOBEC를 능가한다. 선택 플라스미드 (pNMG_ch_5)를 갖는 S1030에서의 chlor의 최소 억제 농도 (MIC)는 대략 1 μg/mL였다. rAPOBEC-XTEN-dCas9-UGI 및 CDA-XTEN-dCas9-UGI 둘 다는 선택 플라스미드 상에서 DNA 교정을 유도하였다 (도 53).

[0688] 다음으로, huAPOBEC3G-XTEN-dCas9-UGI 단백질을 동일한 검정에서 시험하였다. 흥미롭게도, huAPOBEC3G-XTEN-dCas9-UGI는 rAPOBEC1-XTEN-dCas9-UGI 융합 단백질과 상이한 서열 특이성을 나타내었다. rAPOBEC11-XTEN-dCas9-UGI 융합물 (여기서, 위치 3, 6, 및 8이 편집되었음)에 비해, 단지 위치 8만을 APOBEC3G-XTEN-dCas9-UGI

융합물로 편집하였다 (도 54).

[0689] 실시예 8: 더 적은 오프 타겟 편집을 갖는 C에서 T로의 염기 편집제

[0690] 현재의 염기 편집 기술은 게놈 DNA에서 C:G 염기 쌍의 T:A 염기 쌍으로의 서열-특이적 전환을 허용한다. 이는 시티딘 데아미나제 효소에 의한 시토신의 우라실로의 직접적 촉매 전환을 통해 수행되며, 따라서, 전통적인 게놈 편집 기술과는 달리 제1 단계로서 DNA 내로 이중-가닥 DNA 파괴 (DSB)를 도입하지 않는다. 문헌 [Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016), "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." Nature 533, 420-424]을 참조하며; 그의 전체 내용은 본원에 참조로 포함된다. 대신, 촉매적으로 사멸된 SpCas9 (dCas9) 또는 SpCas9 니카제 (dCas9(A840H))를 시티딘 데아미나제 효소, 예컨대 rAPOBEC1, pmCDA1, 또는 hAPOBEC3G에 테더링한다. 관심 게놈 유전자좌는 sgRNA에 의해 코딩되며, DNA 결합 및 국부 변성은 융합물의 dCas9 부분에 의해 용이해진다. 그러나, 야생형 dCas9 및 야생형 Cas9는 오프-타겟 DNA 결합 및 절단을 나타내는 것처럼, 현재의 염기 편집제는 또한 Cas9 오프-타겟 유전자좌에서 C에서 T로의 편집을 나타내며, 이는 그의 치료적 유용성을 제한한다.

[0691] 그의 표적 DNA의 단백질 및 당-포스페이트 백본 사이의 비특이적 정전기적 상호작용을 중화시키는 단지 3 내지 4개의 돌연변이의 SpCas9 내로의 도입은 SpCas9의 DNA 결합 특이성을 증가시키는 것으로 보고되었다. 문헌 [Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016) "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature 529, 490-495]; 및 [Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2015) "Rationally engineered Cas9 nucleases with improved specificity. Science 351, 84-88]을 참조하며; 각각의 전체 내용은 본원에 참조로 포함된다. 따라서, 4가지 보고된 중화 돌연변이를 초기에 보고된 염기 편집제 BE3 (서열식별번호: 48) 내로 혼입하였고, 이 효소의 오프-타겟 C에서 T로의 편집은 또한 극적으로 감소되는 것으로 밝혀졌으며 (도 55), 온-타겟 편집의 감소는 없었다 (도 56).

[0692] 도 55에 나타난 바와 같이, HEK293T 세포를 BE3 또는 HF-BE3 및 EMX1 서열을 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, GUIDE-seq 방법을 사용하여 종 및 동료들에 의해 이전에 측정된 바와 같이 EMX1 sgRNA에 대한 온-타겟 유전자좌, 플러스 상부 10개의 공지된 Cas9 오프-타겟 유전자좌에서 고처리량 DNA 서열분석에 의해 분석하였다. 문헌 [Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015) "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." Nat Biotech 33, 187-197]을 참조하며, 그의 전체 내용은 본원에 참조로 포함된다. EMX1 오프-타겟 5 유전자좌는 증폭하지 않았으며, 나타내지 않는다. 온-타겟 및 오프-타겟 프로토스페이서 및 프로토스페이서 인접 모티프 (PAM)의 서열이 제시된다 (도 55). 프로토스페이서 내의 원래 C의 각각의 위치에 T를 갖는 총 DNA 서열분석 판독물의 백분율로서 정의되는 세포의 C에서 T로의 전환 백분율이 BE3 및 HF-BE3에 대해 나타나 있다.

[0693] 도 56에서, HEK293T 세포를 BE3 또는 HF-BE3 및 지시된 게놈 유전자좌를 매칭시키는 sgRNA를 발현하는 플라스미드로 리포펙타민 2000을 사용하여 형질감염시켰다. 형질감염 3일 후에, 게놈 DNA를 추출하고, PCR에 의해 증폭시키고, 온-타겟 유전자좌에서 고처리량 DNA 서열분석에 의해 분석하였다. 각각의 프로토스페이서 내의 표적 C에 모든 4개의 염기를 갖는 총 DNA 서열분석 판독물의 백분율이 BE3 또는 HF-BE3으로의 처리에 대해 나타나 있다 (도 56). indel 형성의 빈도가 또한 나타나 있다.

[0694] HF-BE3의 1차 단백질 서열 (서열식별번호: 48):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTE
QESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRL
PPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT
YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL
QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFM
ALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY
DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

[0695]

[0696] 실시예 9: Cas9 변이체를 사용하는 염기 편집제의 개발 및 염기 편집 기술의 표적 범위 및 정확성을 증가시키기 위한 염기 편집제 프로세싱성의 조정

[0697] 전통적인 게놈 편집 플랫폼과는 달리, 염기 편집 기술은 이중-가닥 파괴 (DSB)를 유도하지 않고 DNA에서의 정확한 단일 뉴클레오티드 변화를 허용한다. 문헌 [Komor, A. C. et al. Nature 533, 420-424 (2016)]을 참조한다. 현세대의 염기 편집제는 NGG PAM을 배타적으로 사용한다. 이는 게놈 내에서 목적하는 염기를 편집하는 그의 능력을 제한하는데, 이는 염기 편집제가 표적 염기가 PAM의 대략 15 염기 상류의 4-염기 영역 ('탈아미노화 윈도우') 내에 놓이는 정확한 위치에 놓일 필요가 있기 때문이다. 문헌 [Komor, A. C. et al. Nature 533, 420-424 (2016)]을 참조한다. 더욱이, 시티딘 데아미나제의 높은 프로세싱성으로 인해, 염기 편집제는 그의 탈아미노화 윈도우 내에서 모든 시티딘을 티미딘으로 전환시킬 수 있으며, 이는 연구자에 의해 목적하는 것 외의 아미노산 변화를 유도할 수 있다. 문헌 [Komor, A. C. et al. Nature 533, 420-424 (2016)]을 참조한다.

[0698] Cas9 변이체를 갖는 염기 편집제의 개발을 통한 염기 편집의 범위의 확장

[0699] 상이한 PAM 특이성을 갖는 Cas9 상동체 및 다른 RNA-가이드된 DNA 결합제를 염기 편집제 아키텍처 내로 혼입하였다. 문헌 [Kleinstiver, B. P. et al. Nature 523, 481-485 (2015)]; [Kleinstiver, B. P. et al. Nature Biotechnology 33, 1293-1298 (2015)]; 및 [Zetsche, B. et al. Cell 163, 759-771 (2015)]을 참조하며; 각각의 전체 내용은 본원에 참조로 포함된다. 더욱이, 다양한 Cas9 단백질의 PAM 특이성을 넓힌 혁신을 또한 도입하여 염기 편집제의 표적 도달을 훨씬 더 확장시켰다. 문헌 [Kleinstiver, B. P. et al. Nature 523, 481-485 (2015)]; 및 [Kleinstiver, B. P. et al. Nature Biotechnology 33, 1293-1298 (2015)]을 참조한다. 염기 편집제의 현재의 팔레트는 표 4에 요약된다.

[0700] 표 4. Cas9 변이체로부터 제조된 새로운 염기 편집제

종	PAM	염기 편집제 명칭	Cas9 변이체에 대한 참조
에스. 피오게네스	...NGG	BE3	야생형
	...NGA	VQR BE3 또는 EQR BE3	Kleinstiver, B. P. *et al.*
	...NGCG	VRER BE3	Kleinstiver, B. P. *et al.*
에스. 아우레우스	...NNGRRT	SaBE3	야생형
	...NNNRRT	SaKKH BE3	Kleinstiver, B. P. *et al.*
엘. 박테리움	TTTN...	dCpf1 BE2	Zetsche, B. *et al.*

[0701]

[0702] rAPOBEC1의 부위-지정 돌연변이유발을 통한 염기 편집제의 프로세싱성의 조정

[0703] 염기 편집제의 프로세싱성은 데아미나제 효소에서 점 돌연변이를 생성함으로써 조정될 수 있다. 염기 편집제가 여전히 평균 1 라운드의 시티딘 탈아미노화를 촉매할 수 있었지만, 관련 기간 내에 또 다른 탈아미노화를 접근하고 촉매할 가능성은 적었던 데아미나제의 촉매 활성을 약간 감소시키는 돌연변이의 혼입이 추구되었다. 사실

상, 생성된 염기 편집제는 보다 좁은 탈아미노화 윈도우를 가질 것이다.

[0704] 이 연구에서 탐색된 rAPOBEC1 돌연변이는 표 5에 열거된다. 돌연변이의 일부는 rAPOBEC1 촉매작용의 약간 명백한 손상을 발생시켰으며, 이는 다중 시티딘이 탈아미노화 윈도우 내에서 발견되는 경우, 한 시티딘의 또 다른 것에 비한 우선적 편집으로서 분명해졌다. 이들 돌연변이의 일부를 조합하는 것은 상가 효과를 가졌으며, 이는 염기 편집제가 보다 높은 엄격성으로 기질 시티딘을 식별하는 것을 허용한다. 이중 돌연변이체 및 삼중 돌연변이체의 일부는 서로에 대해 바로 옆에 있는 다중 시티딘 중에서 1개의 시티딘의 선택적 편집을 허용하였다 (도 57).

[0705] 표 5. 조사된 rAPOBEC1 점 돌연변이

이 작업에서 연구된 rAPOBEC1 돌연변이	APOBEC3G에서의 상응하는 돌연변이	참조
H121R / H122R	D315R / D316R	Holden, L. G. *et al.*
R126A	R320A	Chen, K-M. *et al.*
R126E	R320E	Chen, K-M. *et al.*
R118A	R313A	Chen, K-M. *et al.*
W90A	W285A	Chen, K-M. *et al.*
W90Y	W285Y	
R132E	R326E	

[0706]

[0707] 염기 편집제 PAM 확장 및 프로세싱성 조정

[0708] 다른 RNA-가이드된 DNA 결합제를 사용함으로써 게놈에서 편집가능한 시티딘을 확장하기 위해 차세대의 염기 편집제를 설계하였다 (도 58). NGG PAM을 사용하는 것은 단지 "윈도우" 내에서 단일 표적을 허용하는 반면, 다중 상이한 PAM의 사용은 Cas9가 선택적 탈아미노화를 실시하기 위해 어디에도 위치되는 것을 허용한다. 다양한 새로운 염기 편집제를 Cas9 변이체로부터 생성하였다 (도 59 및 표 4). 상이한 PAM 부위 (NGA, 도 60; NGCG, 도 61; NNGRRT, 도 62; 및 NNHRRT, 도 63)를 탐구하였다. 선택적 탈아미노화는 시티딘 데아미나제 점 돌연변이유발의 동역학 조정을 통해 성공적으로 달성되었다 (도 65 및 표 5).

[0709] 그 후, 탈아미노화 윈도우에 대한 다양한 돌연변이의 효과를 세포 배양물에서 다중 시티딘을 갖는 스페이서를 사용하여 조사하였다 (도 66 및 67).

[0710] 또한, 제한된 수의 시티딘을 갖는 상이한 게놈 부위에 대한 다양한 돌연변이의 효과를 조사하였다 (도 68 내지 71). 대략 1개의 시티딘이 스페이서에서의 탈아미노화 윈도우 내에서 편집될 것인 반면, 나머지 시티딘은 무손상으로 남아 있을 것임이 밝혀졌다. 전체적으로, 편집에 대한 선호도는 하기와 같다: $C_6 > C_5 >> C_7 \approx C_4$.

[0711] Cpf1을 사용한 염기 편집

[0712] Cas9 상동체인 Cpf1은 AsCpf1, LbCpf1로서, 또는 임의의 다른 종으로부터 얻을 수 있다. BE2 및 BE3 등가물을 포함한 융합 구축물의 개략도는 도 73에 나타나 있다. BE2 등가물은 Cas9 대신 촉매적으로 불활성인 Cpf2 효소 (dCpf1)를 사용하는 반면, BE3 등가물은 표적 가닥을 니킹하는 Cpf1 돌연변이체를 포함한다. 하부 개략도는 그것 위에 예시된 2가지 혁신을 조합하는 상이한 융합 아키텍처를 나타낸다 (도 73). Cpf1 BE2를 사용한 HEK293T 세포 TTTN PAM 부위의 염기 편집 결과를 상이한 스페이서로 조사하였다 (도 64a 내지 64c). 일부 실시양태에서, Cpf1은 본원에 제공된 염기 편집제 중 임의의 것에서의 Cas9 도메인 대신 사용될 수 있다. 일부 실시양태에서, Cpf1은 서열식별번호: 9와 적어도 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, 또는 99.5% 동일한 단백질이다.

[0713] Cpf1의 전체 단백질 서열 (서열식별번호: 9):

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISED
LLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQS
KDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYE
SLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNG
ENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKT
VEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKE
QELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQG
KKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKI
RNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGE
GYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYK
QSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSK
GRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIK
DKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNI
IGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFK
RGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTS
KICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFR
NSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEL
DYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQ
NRNN

[0714]

[0715] 실시예 10: 염기 편집의 증가된 충실도

[0716] BE3 및 HF-BE3의 플라스미드 전달 사이의 차이를 조사하여, 2개가 대등한 효율로 온-타겟 유전자좌를 편집한다는 것을 밝혀내었다 (도 74 및 75). 그러나, HF-BE3은 오프-타겟 유전자좌를 BE3보다 훨씬 더 적게 편집하였으며, 이는 HF-BE3이 BE3보다 훨씬 더 높은 DNA 특이성을 갖는다는 것을 의미한다 (도 76). HEK 세포에의 데아미나제 단백질 리포펙션은 BE3의 단백질 전달이 BE3의 플라스미드 DNA 전달과 대등한 온-타겟 활성, 그러나 그보다 훨씬 더 양호한 특이성을 발생시킨다는 것을 입증하였다. 개선된 형질감염 절차 및 보다 양호한 플라스미드를 사용하여 (n=2), 실험은 하기 조건을 사용하였다: 단백질 전달은 125 nM Cas9:sgRNA 복합체였고, 플라스미드 전달은 750 ng BE3/HF-BE3 플라스미드 + 250 ng sgRNA 플라스미드였고, 리포펙션은 웰당 1.5μL의 리포펙타민 2000으로였다. EMX-1 오프 타겟 부위 2 및 FANCF 오프-타겟 부위 1은 검정된 모든 오프-타겟에 비해, BE3으로 가장 큰 오프-타겟 편집을 나타낸 반면 (도 77 및 78), HEK-3은 전달 방법 중 임의의 것에 대해 오프-타겟에서 유의한 편집을 나타내지 않았다 (도 79). HEK-4는 온-타겟 부위 상에서 일부 C에서 G로의 편집을 나타내는 반면, 그의 오프-타겟 부위 1, 3, 및 4는 모든 검정된 부위의 가장 큰 오프-타겟 편집을 나타내었다 (도 80).

[0717] 제브라피시에의 미세-주사를 통한 BE3 단백질의 전달

[0718] TYR 가이드 RNA를 sgRNA 활성에 대해 시험관내 검정에서 시험하였다 (도 81 및 82). % HTS 판독물은 얼마나 많은 C 잔기가 정제된 BE3 단백질과의 2시간 인큐베이션 및 생성된 생성물의 PCR 동안 T 잔기로 전환되었는지를 나타낸다. 실험은 60bp의 그의 게놈 컨텍스트에서 표적 탈아미노화 부위를 갖는 80량체 합성 DNA 기질을 사용하였다. 이는 단지 1개의 가닥이 니킹되고, 따라서 생성물이 PCR에 의해 증폭되지 않기 때문에 % 편집된 DNA 가닥과 동일하지 않다. 편집된 HTS 판독물의 비율은 x/(2-x)와 동일하며, 여기서, x는 편집된 THS 판독물의 실제 비율이다. 60% 편집에 대해, 편집된 염기의 실제 비율은 75%이다. "오프 타겟"은 오프-타겟 sgRNA에 결합되는 동안 동일한 DNA 기질과 함께 인큐베이션된 BE3을 나타낸다. sgRNA sgRH_13, sgHR_17, 및 가능하게는 sgHR_16은 생체내 주사 실험을 위한 유망한 표적으로 나타난 것으로 밝혀졌다.

[0719] BE3 단백질의 전달을 제브라피쉬에서 생체내에서 시험하였다. 제브라피시 배아 (n=16 내지 24)를 스크램블링된 sgRNA, sgHR_13, sgHR_16, 또는 sgHR_17 중 어느 하나로 주사하고, BE3을 정제하였다. 각각의 조건으로부터의 3개의 배아를 독립적으로 분석하고 (단일 배아), 각각의 조건에 대해, 모든 주사된 배아를 풀링하고, 풀로서 서열분석하였다. 결과는 도 83 내지 85에 나타나 있다.

[0720] 실시예 11: 질환을 치료하기 위한 염기 편집제의 용도

[0721] 본원에 제공된 염기 편집제 또는 복합체 (예를 들어, BE3)는 핵산을 변형시키는 데 사용될 수 있다. 예를 들어, 염기 편집제는 핵산 (예를 들어, DNA)에서 시토신을 티민으로 변화시키는 데 사용될 수 있다. 이러한 변화는 그 중에서도 단백질의 아미노산 서열을 변경시키거나, 출발 코돈을 파괴하거나 생성하거나, 정지 코돈을 생성하거나, 스플라이싱 공여자를 교란시키거나, 스플라이싱 수용자를 교란시키거나, 조절 서열을 편집하기 위해 생성될 수 있다. 가능한 뉴클레오티드 변화의 예는 도 86에 나타나 있다.

[0722] 본원에 제공된 염기 편집제 또는 복합체 (예를 들어, BE3)는 대상체에서 아포지단백질 E의 이소형을 편집하는 데 사용될 수 있다. 예를 들어, 아포지단백질 E 이소형은 알츠하이머병을 발달시킬 보다 낮은 위험과 연관된

이소형을 생성하도록 편집될 수 있다. 아포지단백질 E는 아미노산 112 및 158에서 상이한 4개의 이소형을 갖는다. APOE4는 후기-발병 알츠하이머병에 대한 가장 크고 가장 통상적인 유전적 위험 인자이다. 핵산 서열 CGC에 의해 코딩되는 APOE4의 아르기닌 잔기 158은 염기 편집제 (예를 들어, BE3)를 사용함으로써 시스테인으로 변화되어 CGC 핵산 서열을 TGC로 변화시킬 수 있으며, 이는 잔기 158에서 시스테인을 코딩한다. 이 변화는 보다 낮은 알츠하이머병 위험과 연관된 APOE3r 이소형을 생성한다. 도 87을 참조한다.

[0723] 염기 편집제 BE3이 마우스 성상세포에서 APOE4를 APOE3r로 편집하는 데 사용될 수 있는지 시험하였다 (도 88). APOE4 마우스 성상세포를 APOE4의 아르기닌 158을 코딩하는 핵산을 표적화하는 Cas9 + 주형 또는 BE3으로 뉴클레오펙션시켰다. Cas9 + 주형은 26% indel로 단지 0.3% 편집을 수득한 반면, BE3은 5% indel로 75% 편집을 수득하였다. 2개의 추가의 염기-편집된 시토신은 침묵되고, 아미노산 서열에 대한 변화를 생성하지 않는다 (도 88).

[0724] 본원에 제공된 염기 편집제 또는 복합체는 예를 들어, 돌연변이를 PRNP 유전자 내로 도입함으로써 프리온 단백질 질환, 예컨대 크로이츠펠트-야콥병 및 태아 가족성 불면증을 치료하는 데 사용될 수 있다. PRNP 돌연변이를 복귀시키는 것은 치료적 결과를 생성하지 않을 수 있으며, PRNP에서의 indel은 병원성일 수 있다. 따라서, PRNP를 염기 편집제 (예를 들어, BE3)를 사용하여 돌연변이시켜 PRNP 유전자에서 조기 정지 코돈을 도입할 수 있는지 시험하였다. 그의 가이드 RNA와 회합된 BE3을 HEK 세포 또는 교모세포종 세포 내로 도입하고, PRNP 유전자를 편집하여 잔기 37에서 코딩된 아르기닌을 정지 코돈으로 변화시킬 수 있었다. BE3은 41% 편집을 수득하였다 (도 89).

[0725] 편집될 수 있는 추가의 유전자는 하기를 포함한다: 증가된 알츠하이머 위험을 치료하기 위한 Arg 112 및 Arg 158의 APOE 편집; 알츠하이머 위험을 감소시키기 위한 Ala 673의 APP 편집; 태아 가족성 불면증 및 다른 프리온 단백질 질환을 치료하기 위한 Arg 37의 PRNP 편집; 뒤시엔느 근육 이영양증을 치료하기 위한 엑손 23 및 51 스플라이스 부위의 DMD 편집; 비만 위험을 치료하기 위한 인트론 1의 FTO 편집; 펜드레드 증후군 (유전적 난청)을 치료하기 위한 엑손 8의 PDS 편집; 선천성 청력 상실을 치료하기 위한 엑손 8의 TMC1 편집; 만성 육아종 질환을 치료하기 위한 다양한 환자-관련 돌연변이의 CYBB 편집. 본원에 제공된 염기 편집제를 사용하여 치료될 수 있는 추가의 질환은 하기 표 6에 나타나 있다.

[0726] UGI는 또한 핵심적 역할을 한다. UDG (UGI가 억제하는)를 녹아웃시키는 것은 C에서 T로의 염기 편집의 청결도 및 효율을 극적으로 개선시키는 것으로 나타났다 (도 90). 게다가, 니카제를 갖고 UGI가 없는 염기 편집제는 매우 높은 indel 비율로 결과물의 혼합물을 생성하는 것으로 나타났다 (도 91).

[0727] 실시예 12: 염기 편집의 표적화 범위의 확장

[0728] 염기 편집은 이중-가닥 DNA 파괴를 생성하지 않고, 공여자 DNA 주형을 요구하지 않고, 과량의 확률적 삽입 및 결실을 유도하지 않고, DNA에서 프로그램가능한, 단일-뉴클레오티드 C→T (또는 G→A) 변화를 유도하는 촉매적으로 결함성 스트렙토코쿠스 피오게네스 Cas9, 시티딘 데아미나제, 및 염기 절제 복구의 억제제를 함유하는 융합 단백질을 사용하는 게놈 편집에 대한 새로운 접근법이다[1]. 염기 편집에 의해 표적화될 수 있는 부위의 수를 2.5-배 확장시키기 위해 상이한 프로토스페이서-인접 모티프 (PAM) 특이성을 갖는 천연 및 조작된 Cas9 변이체를 사용하는 5가지 새로운 C→T (또는 G→A) 염기 편집제의 개발이 본원에 기재되어 있다. 추가적으로, 대략 5 뉴클레오티드 내지 1 또는 2개 뉴클레오티드의 겉보기 편집 윈도우의 폭을 좁히는 돌연변이된 시티딘 데아미나제 도메인을 함유하는 새로운 염기 편집제를 조작하여, 이전에 대등한 효율로 편집되었을 이웃하는 C 뉴클레오티드의 식별을 가능하게 하였다. 함께, 이들 개발은 염기 편집의 표적화 범위를 실질적으로 증가시킨다.

[0729] CRISPR-Cas9 뉴클레아제는 표적화된 게놈 편집을 매개하는 데 광범위하게 사용되었다[2]. 대부분의 게놈 편집 적용에서, Cas9는 단일 가이드 RNA (sgRNA)와 복합체를 형성하며, sgRNA 서열에 의해 특정된 표적 부위에서 이중-가닥 DNA 파괴 (DSB)를 유도한다. 세포는 주로 비-상동 말단-연결 (NHEJ) 복구 경로를 통해 이 DSB에 반응하며, 이는 유전자를 교란시키는 격자이동 돌연변이를 유발할 수 있는 확률적 삽입 또는 결실 (indel)을 발생시킨다. DSB에 플랭킹된 서열과 높은 정도의 상동성을 갖는 공여자 DNA 주형의 존재 하에서, 유전자 교정은 상동성 지정 복구 (HDR)로 공지된 대안적인 경로를 통해 달성될 수 있다.[3,4] 불행하게도, 대부분의 비-섭동 조건 하에서, HDR은 비효율적이고, 세포 상태 및 세포 유형에 의존하며, 보다 큰 빈도의 indel에 의해 지배된다.[3,4] 인간 질환과 연관된 공지된 유전적 변이의 대부분은 점 돌연변이이기 때문에[5], 정확한 점 돌연변이를 보다 효율적이고 명확하게 생성할 수 있는 방법이 필요하다.

[0730] DSB를 유도하지 않고 프로그램가능한 방식으로 C:G 염기 쌍의 T:A 염기 쌍으로의 표적화된 대체를 가능하게 하는 염기 편집[1]이 최근 기재되었다. 염기 편집은 스트렙토코쿠스 피오게네스 Cas9 (SpCas9)의 촉매적으로 불활성화된 (dCas9) 또는 니카제 형태, 시티딘 데아미나제, 예컨대 APOBEC1, 및 염기 절제 복구의 억제제, 예컨대 우라실 글리코실라제 억제제 (UGI) 사이의 융합 단백질을 사용하여 sgRNA에 의해 특정된 5-뉴클레오티드 윈도우 내에서 시티딘을 우리딘으로 전환시킨다.[1] 제3-세대 염기 편집제인 BE3은 다른 게놈 편집 방법을 사용하여 달성될 수 있는 것보다 더 높은 효율 및 더 낮은 indel 빈도로 다양한 세포주에서 질환-관련 점 돌연변이를 포함하여 C:G 염기 쌍을 T:A 염기 쌍으로 전환시킨다[1]. 후속 연구는 다양한 설정에서의 데아미나제-dCas9 융합 접근법을 확인하였다[6,7].

[0731] BE3에 의한 효율적인 편집은 표적 C를 프로토스페이서의 PAM-원위 말단 (위치 4 내지 8, PAM을 위치 21 내지 23으로서 카운팅함) 부근의 5-뉴클레오티드 윈도우 내에 정치시키는 NGG PAM의 존재를 요구한다[1]. 이 PAM 요구는 BE3에 의해 효율적으로 표적화될 수 있는 인간 게놈에서의 부위의 수를 실질적으로 제한하며, 이는 관심 많은 부위가 표적 C의 13- 내지 17- 뉴클레오티드 하류에 NGG가 결여되기 때문이다. 더욱이, BE3의 높은 활성 및 프로세싱성은 편집 윈도우 내의 모든 C의 T로의 전환을 발생시키며, 이는 표적 유전자좌에 대한 목적하지 않은 변화를 잠재적으로 도입할 수 있다. 본원에는, 이들 제한 둘 다를 다루는 새로운 C:G에서 T:A로의 염기 편집제가 기재되어 있다.

[0732] DNA에 결합하고, 단일-가닥 DNA 버블을 함유하는 "R-루프" 복합체[8]를 형성하는 임의의 Cas9 상동체는 원칙적으로 염기 편집제로 전환될 수 있는 것으로 생각되었다. 이들 새로운 염기 편집제는 비-NGG PAM 부위가 편집되는 것을 허용함으로써 표적화가능한 유전자좌의 수를 확장시킬 것이다. 스타필로코쿠스 아우레우스로부터의 Cas9 상동체 (SaCas9)는 SpCas9보다 상당히 더 작고 (1053 대 1368 잔기), 포유동물 세포에서 효율적인 게놈 편집을 매개할 수 있으며, NNGRRT PAM을 요구한다[9]. SpCas9를 BE3에서 SaCas9로 대체하여 SaBE3을 생성하고, HEK293T 세포를 SaBE3 및 6개의 인간 게놈 유전자좌를 표적화하는 sgRNA를 코딩하는 플라스미드로 형질감염시켰다 (도 92a 및 92b). 3일 후, 게놈 유전자좌를 고처리량 DNA 서열분석 (HTS)으로 처리하여 염기 편집 효율을 정량화하였다. SaBE3은 인간 세포에서의 다양한 게놈 부위에서 표적 C의 C에서 T로의 염기 편집을, 위치 6 내지 11에서의 C의 표적화로부터 발생한 매우 높은 전환 효율 (형질감염된 세포에 대한 풍부화 없이, C에서 T로 전환된 총 DNA 서열의 대략 50 내지 75%)로 가능하게 하였다. NNGRRT-함유 표적 부위에 대한 SaBE3의 효율은 일반적으로 NGG-함유 표적 부위에 대한 BE3의 것을 초과하였다[1]. 아마도 그의 보다 높은 평균 효율로 인해, SaBE3은 또한 정규 BE3 활성 윈도우의 외부의 위치에서의 표적 C에서 검출가능한 염기 편집을 발생시킬 수 있다 (도 92c). 이에 비해, BE3은 공지된 SpCas9 PAM 선호도에 따라, 동일한 조건 하에서 유의하게 감소된 편집을 나타내었다 (0 내지 11%) (도 106a)[10]. 이들 데이터는 SaBE3이 BE3에 접근가능하지 않은 부위에서 매우 효율적인 염기 편집을 용이하게 할 수 있다는 것을 나타낸다.

[0733] 염기 편집제의 표적화 범위는 PAM 특이성을 확장하거나 변경시키는 최근에 조작된 Cas9 변이체를 적용함으로써 추가로 확장되었다. 종 및 동료들은 최근 NGA (VQR-Cas9), NGAG (EQR-Cas9), 또는 NGCG(VRER-Cas9) PAM 서열을 수용하는 3개의 SpCas9 돌연변이체를 보고하였다[11]. 또한, 종 및 동료들은 NNNRRT에 대한 그의 PAM 요건을 완화시키는 3개의 돌연변이를 함유하는 SaCas9 변이체 (SaKKH-Cas9)를 조작하였다[12]. BE3의 SpCas9 부분을 이들 4가지 Cas9 변이체로 대체하여 VQR-BE3, EQR-BE3, VRER-BE3, 및 SaKKH-BE3을 생성하였으며, 이는 각각 NNNRRT, NGA, NGAG, 및 NGCG PAM을 표적화한다. HEK293T 세포를 이들 구축물 및 각각의 새로운 염기 편집제에 대한 6개의 게놈 유전자좌를 표적화하는 sgRNA를 코딩하는 플라스미드로 형질감염시키고, HTS를 사용하여 C에서 T로의 염기 전환을 측정하였다.

[0734] SaKKH-BE3은 처리된, 비-풍부화된 세포의 62% 이하의 효율로 NNNRRT PAM을 갖는 부위를 편집하였다 (도 92d). 예상된 바와 같이, SaBE3은 NNNHRRT (여기서, H = A, C, 또는 T)인 PAM을 함유하는 표적을 효율적으로 편집할 수 없었다 (도 92d). VQR-BE3, EQR-BE3, 및 VRER-BE3은 BE3의 것과 유사한 편집 윈도우를 갖는 예상된 PAM 요건을 갖는 게놈 유전자좌에서 처리된, 비-풍부화된 세포의 50% 이하의 보다 보통의, 그러나 여전히 실질적인 염기 편집 효율을 나타내었다 (도 92e 및 92f). VQR-BE3, EQR-BE3, 및 VRER-BE3의 염기 편집 효율은 일반적으로 상응하는 Cas9 뉴클레아제의 보고된 PAM 요구와 밀접하게 평행하며; 예를 들어, EQR-BE3은 NGAH PAM 서열을 함유하는 표적을 효율적으로 편집할 수 없었다 (도 92f). 대조적으로, BE3은 아마도 그의 PAM 제한으로 인해,

NGA 또는 NGCG PAM을 갖는 부위를 효율적으로 편집할 수 없었다 (0 내지 3%) (도 106b).

[0735] 집합적으로, SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, 및 VRER-BE3의 특성은 Cas9 상동체 및 조작된 변이체를 이용하는 그의 능력을 용이하게 하는 조정성을 나타내는 염기 편집제를 확립한다.

[0736] 다음으로, 변경된 활성 윈도우 폭을 갖는 염기 편집제를 개발하였다. BE3의 활성 윈도우 내의 모든 C는 T로 효율적으로 전환될 수 있다[1]. 이 윈도우의 폭을 조정하는 능력은 단지 BE3 활성 윈도우에 존재하는 C의 하위세트를 편집하는 것이 중요한 경우에 유용할 것이다.

[0737] APOBEC1 및 dCas9 사이의 링커의 길이는 시험관내에서 APOBEC1에 의해 접근가능한 염기의 수를 조정하는 것으로 이전에 관찰되었다[1]. 그러나, HEK293T 세포에서, 링커 길이를 다양화하는 것은 편집 윈도우의 폭을 유의하게 조정하지 않았으며, 이는 복잡한 세포 환경에서, dCas9 및 시티딘 데아미나제의 상대적 배향 및 유연성이 링커 길이에 의해 강하게 결정되지 않음을 시사한다 (도 96). 다음으로, sgRNA의 5' 말단을 말단절단하는 것은 RNA-DNA 헤테로듀플렉스의 형성시 데아미나제에 접근가능한 단일-가닥 DNA의 길이를 감소시킴으로써 염기 편집 윈도우를 좁힐 수 있는 것으로 생각되었다. HEK293T 세포를 BE3 및 편집 윈도우에서의 다중 C를 갖는 유전자좌를 표적화하는 상이한 스페이서 길이의 sgRNA를 코딩하는 플라스미드로 공동-형질감염시켰다. 17- 내지 19-염기 스페이서를 갖는 말단절단된 sgRNA를 사용하는 경우 염기 편집의 폭의 일치하는 변화는 관찰되지 않았다 (도 95a 내지 95c). sgRNA 스페이서를 17 염기 미만으로 말단절단하는 것은 활성의 큰 상실을 발생시켰다 (도 95a).

[0738] 대안적인 접근법으로서, 데아미나제 도메인에 대한 돌연변이는 다중 가능한 메카니즘을 통해 편집 윈도우의 폭을 좁힐 수 있는 것으로 생각되었다. 첫째로, 일부 돌연변이는 기질 결합, 결합된 DNA의 형태, 또는 활성 부위에 대한 기질 접근성을 데아미나제 활성 부위에 대한 C의 비-최적 제시에 대한 내약성을 감소시키는 방식으로 변경시킬 수 있다. 둘째로, APOBEC1의 높은 활성은 DNA 결합 사건당 다중 C의 탈아미노화에 기여할 가능성이 있기 때문에,[1,13,14] 염기 편집제의 데아미나제 도메인의 촉매 효율을 감소시키는 돌연변이는 그것을 DNA로부터 해리되기 전에 연속적 라운드의 탈아미노화를 촉매하는 것으로부터 방지할 수 있다. 일단 임의의 C:G에서 T:A로의 편집 사건이 발생하면, sgRNA는 표적 DNA 서열을 더 이상 완벽하게 매칭시킬 수 없으며, 표적 유전자좌에의 염기 편집제의 재-결합은 덜 선호되어야 한다. 원래 편집 윈도우 내의 다중 시티딘 중에서 구별하는 새로운 염기 편집제를 개발하기 위한 노력으로 둘 다의 전략을 시험하였다.

[0739] 이용가능한 APOBEC1 구조의 부재를 고려하여, APOBEC3G의 촉매 활성을 조정하는 몇몇 돌연변이가 이전에 보고되었으며, APOBEC1의 그것과 그의 활성 부위-함유 도메인의 42% 서열 유사성을 공유하는 동일한 패밀리로부터의 시티딘 데아미나제가 확인되었다[15]. 상응하는 APOBEC1 돌연변이를 BE3 내로 혼입하고, HEK293T 세포에서 위치 3, 4, 5, 6, 8, 9, 10, 12, 13, 및 14 (부위 A)에 C를 함유하거나; 위치 5 ,6, 7, 8, 9, 10, 11, 및 13 (부위 B)에 C를 함유하는 2개의 C-풍부 게놈 부위에서 염기 편집 효율 및 편집 윈도우 폭에 대한 그의 효과를 평가하였다.

[0740] APOBEC1 돌연변이 R118A 및 W90A 각각은 염기 편집 효율의 극적인 상실을 발생시켰다 (도 97c). R132E는 편집 효율의 일반적 감소를 발생시켰지만, 편집 윈도우의 형상을 실질적으로 좁히도록 변화하지 않았다 (도 97c). 대조적으로, 실질적인 편집 효율을 유지하면서 편집 윈도우의 폭을 좁힌 몇몇 돌연변이가 밝혀졌다 (도 93a 및 97c). "편집 윈도우 폭"은 그 내에서 편집 효율이 그 표적에 대한 반수-최대 값을 초과하는 인위적으로 계산된 윈도우 폭을 나타나는 것으로 정의되었다. 시험된 2개의 C-풍부 게놈 부위에 대한 BE3의 편집 윈도우 폭은 5.0개 (부위 A) 및 6.1개 (부위 B) 뉴클레오티드였다.

[0741] APOBEC1에서의 R126은 ssDNA의 포스페이트 백본과 상호작용할 것으로 예측된다[13]. 이전의 연구는 APOBEC3G 내로 상응하는 돌연변이를 도입하는 것이 촉매작용을 적어도 5-배 감소시켰음을 보여주었다[14]. 흥미롭게도, BE3에서의 APOBEC1 내로 도입되는 경우, R126A 및 R126E는 다른 위치에서의 편집 활성을 감소시키면서, 가장 강하게 편집된 위치 (C5, C6, 및 C7)에서 BE3에 비해 활성을 증가시키거나 유지하였다 (도 93a 및 97c). 따라서, 이들 2가지 돌연변이의 각각은 부위 A 및 부위 B에서의 편집 윈도우의 폭을 각각 4.4 및 3.4개 뉴클레오티드 (R126A)로, 또는 4.2 및 3.1개 뉴클레오티드 (R126E)로 좁혔다 (도 93a 및 97c).

[0742] APOBEC1에서의 W90 (APOBEC3G에서의 W285에 상응함)은 APOBEC3G 활성 부위에서 소수성 포켓을 형성하고, 기질 결합에 보조할 것으로 예측된다[13]. 이 잔기를 Ala로 돌연변이시키는 것은 APOBEC3G의 촉매 활성을 폐지하였다

[13]. BE3에서, W90A는 염기 편집 효율을 거의 완전히 폐지하였다 (도 97c). 대조적으로, W90Y는 부위 A 및 부위 B에서의 편집 윈도우 폭을 각각 3.8 및 4.9개 뉴클레오티드로 좁히면서 염기 편집 활성을 단지 보통으로 감소시킨 것으로 밝혀졌다 (도 93a). 이들 결과는 시티딘 데아미나제 도메인에 대한 돌연변이가 상응하는 염기 편집제의 활성 윈도우 폭을 좁힐 수 있다는 것을 입증한다.

[0743] 염기 편집 활성을 대폭 감소시키지 않고 편집 윈도우를 좁힌 3가지 돌연변이인 W90Y, R126E, 및 R132E를 이중으로 및 삼중으로 돌연변이된 염기 편집제 내로 조합하였다. 이중 돌연변이체 W90Y+R126E는 BE3-유사 최대 편집 효율을 갖는 염기 편집제 (YE1-BE3)를 발생시켰지만, 편집 윈도우 폭을 실질적으로 좁혔다 (부위 A 및 부위 B에서의 폭 = 각각 2.9 및 3.0개 뉴클레오티드) (도 93a). W90Y+R132E 염기 편집제 (YE2-BE3)는 보통으로 더 낮은 편집 효율을 나타내었으며 (BE3에 비해 시험된 5개의 부위에 걸쳐 평균 1.4-배 더 낮은 최대 편집 수율), 또한 편집 윈도우 폭을 실질적으로 좁혔다 (부위 A 및 부위 B에서의 폭 = 각각 2.7 및 2.8개 뉴클레오티드) (도 97c). R126E+R132E 이중 돌연변이체 (EE-BE3)는 YE2-BE3과 유사한 최대 편집 효율 및 편집 윈도우 폭을 나타내었다 (도 97c). 삼중 돌연변이체 W90Y+R126E+R132E (YEE-BE3)는 2.0-배 더 낮은 평균 최대 편집 수율, 그러나 C6 위치 넘어 매우 적은 편집 및 부위 A 및 부위 B에 대해 각각 2.1 및 1.4개 뉴클레오티드의 편집 윈도우 폭을 나타내었다 (도 97c). 종합하면 이들 데이터는 시티딘 데아미나제 도메인에서의 돌연변이가, 일부의 경우 편집 효율에 대한 최소 또는 단지 보통의 효과로 편집 윈도우 폭에 강하게 영향을 미칠 수 있다는 것을 나타낸다.

[0744] BE3, YE1-BE3, YE2-BE3, EE-BE3, 및 YEE-BE3의 염기 편집 결과물을 BE3 활성 윈도우 내에 다중 C를 함유하는 4개의 널리-연구된 인간 게놈 부위를 표적화하는 HEK293T 세포에서 추가로 비교하였다[1]. 이들 표적 유전자좌는 위치 4 및 5 (HEK 부위 3), 위치 4 및 6 (HEK 부위 2), 위치 5 및 6 (EMX1), 또는 위치 6, 7, 8, 및 11 (FANCF)에 표적 C를 함유하였다. BE3은 위치 4 내지 8 활성 윈도우 내의 임의의 C를 편집하는 데 대해 거의 (1.2-배 미만) 선호도를 나타내지 않았다. 대조적으로, YE1-BE3은 C4에 비해 C5를 편집하는 데 1.3-배 선호도 (HEK 부위 3), C4에 비해 C6에 대해 2.6-배 선호도 (HEK 부위 2), C6에 비해 C5에 대해 2.0-배 선호도 (EMX1), 및 C7에 비해 C6에 대해 1.5-배 선호도 (FANCF)를 나타내었다 (도 93b). YE2-BE3 및 EE-BE3은 YE1-BE3보다 다소 더 큰 위치적 특이성 (더 좁은 활성 윈도우)을 나타내었으며, C4에 비해 C5를 편집하는 데 평균 2.4-배 선호도 (HEK 부위 3), C4에 비해 C6에 대해 9.5-배 선호도 (HEK 부위 2), C6에 비해 C5에 대해 2.9-배 선호도 (EMX1), 및 C6에 비해 C7에 대해 2.6-배 선호도 (FANCF)였다 (도 93b). YEE-BE3은 가장 큰 위치적 선택성을 나타내었으며, C4에 비해 C5를 편집하는 데 2.9-배 선호도 (HEK 부위 3), C4에 비해 C6에 대해 29.7-배 선호도 (HEK 부위 2), C6에 비해 C5에 대해 7.9-배 선호도 (EMX1), 및 C6에 비해 C7에 대해 7.9-배 선호도 (FANCF)를 가졌다 (도 93b). 상기 발견은 돌연변이체 염기 편집제가 둘 다의 뉴클레오티드가 BE3 편집 윈도우 내에 있는 경우에도 인접한 C 사이를 식별할 수 있다는 것을 확립한다.

[0745] 이들 4가지 돌연변이체 및 BE3의 생성물 분포를 HTS에 의해 추가로 분석하여 그의 겉보기 프로세싱성을 평가하였다. BE3은 처리된 HEK293T 세포에서 현저하게 T4-T5 (HEK 부위 3), T4-T6 (HEK 부위 2), 및 T5-T6 (EMX1) 생성물을 생성하였으며, 이는 평균적으로 단일 T를 함유하는 생성물보다 7.4-배 더 많은 2개의 T를 함유하는 생성물을 발생시킨다. 대조적으로, YE1-BE3, YE2-BE3, EE-BE3, 및 YEE-BE3은 단독을 편집된 C4-T5, C4-T6, 및 T5-C6 생성물에 대해 실질적으로 더 높은 선호도를 나타내었다 (도 93c). YE1-BE3은 1.4의 평균 단일-T 대 이중-T 생성물 비를 갖는 생성물을 생성하였다. YE2-BE3 및 EE-BE3은 각각 4.3 및 5.1의 평균 단일-T 대 이중-T 생성물 비를 갖는 생성물을 생성하였다 (도 93c). 상기 결과와 일치하게, YEE-BE3 삼중 돌연변이체는 3개의 게놈 유전자좌에 걸쳐 단일-T 생성물을 평균 14.3-배 선호하였다 (도 93c). 단지 1개의 C가 표적 윈도우 내에 있는 표적 부위 (HEK 부위 4, 위치 C5에서)에 대해, 모든 4가지 돌연변이체는 BE3과 대등한 편집 효율을 나타내었다 (도 98). 이들 발견은 이들 BE3 돌연변이체가 감소된 겉보기 프로세싱성을 가지며, BE3 편집 윈도우 내에 다중 C를 함유하는 표적 부위에서의 단지 단일 C의 전환을 선호할 수 있다는 것을 나타낸다. 이들 데이터는 또한 이들 돌연변이체 염기 편집제에 대해 C5 > C6 > C7 ≈ C4의 위치적 선호도를 시사하지만, 이 선호도는 표적 서열에 따라 상이할 수 있다.

[0746] APOBEC1에서의 윈도우-조정 돌연변이를 VQR-BE3에 적용하였으며, 이는 NGA PAM에 의해 표적화된 부위에서 기질의 선택적 염기 편집을 허용하였다 (도 107a). 그러나, 이들 돌연변이를 SaKKH-BE3에 적용한 경우, 기질 선택성의 개선 없이 염기 편집 효율의 선형 감소가 관찰되었으며, 이는 BE3 및 그의 변이체의 것들과는 이 염기 편집제의 상이한 동역학적 평형 및 기질 접근성을 시사한다 (도 107b).

[0747] 이 연구에 기재된 변경된 PAM 특이성을 갖는 5가지 염기 편집제는 함께, 원칙적으로 염기 편집에 의해 2.5-배

교정될 수 있는 클린바 데이터베이스에서의 질환-연관된 돌연변이의 수를 증가시킨다 (도 94a 및 94b). 유사하게, 좁혀진 편집 윈도우를 갖는 염기 편집제의 개발은 비-표적 C의 대등한 변형 없이 염기 편집에 의해 교정될 수 있는 적절하게 위치된 NGG PAM을 갖는 클린바 엔트리의 분율을 대략 배가시킨다 (BE3에 대해 31% 내지 YEE-BE3에 대해 59%) (도 94a 및 94b).

[0748] 요약하면, 염기 편집의 표적화 범위는 상이한 PAM 특이성을 갖는 Cas9 변이체를 사용하는 염기 편집제를 개발함으로써, 및 다양한 편집 윈도우 폭을 갖는 데아미나제 돌연변이체의 집합을 개발함으로써 실질적으로 확장되었다. 이론적으로, 염기 편집은 단일-가닥-특이적 뉴클레오티드 데아미나제 효소에 대한 기질로서 기능할 수 있는 단일-가닥 DNA의 버블을 생성하는 다른 프로그램가능한 DNA-결합 단백질 (예컨대 Cpf1[16])을 사용하여 가능해야 할 것이다.

[0749] 물질 및 방법

[0750] 클로닝. PCR을 Q5 핫 스타트 고-충실도 DNA 폴리머라제 (뉴 잉글랜드 바이오랩스)를 사용하여 수행하였다. BE 및 sgRNA에 대한 플라스미드를 이전에 보고된 플라스미드[1]로부터 수득된 USER 클로닝 (뉴 잉글랜드 바이오랩스)을 사용하여 구축하였다. DNA 벡터 증폭을 NEB 10베타 적격 세포 (뉴 잉글랜드 바이오랩스)를 사용하여 수행하였다.

[0751] 세포 배양. HEK293T (ATCC CRL-3216)를 10% (v/v) 소 태아 혈청 (FBS)으로 보충된 둘베코 변형 이글 배지 플러스 글루타맥스 (써모피셔)에서 37℃에서 5% CO_2로 배양하였다. APOE 유전자의 ApoE4 이소형을 함유하는 불멸화 래트 성상세포 (타코닉 바이오사이언시즈)를 10% (v/v) 소 태아 혈청 (FBS) 및 200 μg/mL 제네티신 (써모피셔 사이언티픽)으로 보충된 둘베코 변형 이글 배지 플러스 글루타맥스 (써모피셔 사이언티픽)에서 유지하였다.

[0752] 형질감염. HEK293T 세포를 48-웰 콜라겐-코팅된 바이오코트 플레이트 (코닝) 상에 시딩하고, 대략 85% 융합도로 형질감염시켰다. 750 ng의 BE 및 250 ng의 sgRNA 발현 플라스미드를 웰당 1.5 μl의 리포펙타민 2000 (써모피셔 사이언티픽)을 제조자의 프로토콜에 따라 사용하여 형질감염시켰다.

[0753] 게놈 DNA 샘플의 고처리량 DNA 서열분석. 형질감염된 세포를 3일 후에 수거하고, 게놈 DNA를 아젠코트 DN어드밴스 게놈 DNA 단리 키트 (베크만 쿨터)를 제조업체의 지침에 따라 사용하여 단리하였다. 관심 게놈 영역을 보충적 서열에 열거된 플랭킹 HTS 프라이머 쌍으로 PCR에 의해 증폭시켰다. PCR 증폭을 푸션 핫-스타트 II DNA 폴리머라제 (써모피셔)로 제조업체의 지침에 따라 수행하였다. PCR 생성물을 래피드팁스 (디피니티 게노믹스)를 사용하여 정제하였다. 2차 PCR을 수행하여 서열분석 어댑터를 부착시켰다. 생성물을 겔-정제하고, 카파 라이브러리 정량화 키트-일루미나 (카파 바이오시스템즈)를 사용하여 정량화하였다. 샘플을 이전에 기재된 바와 같이 일루미나 MiSeq 상에서 서열분석하였다[1].

[0754] 데이터 분석. 뉴클레오티드 빈도를 이전에 기재된 매트랩 스크립트를 사용하여 평가하였다[1]. 간략하게, 판독물을 스미스-워터만 알고리즘을 통해 참조 서열에 대해 정렬하였다. 30 미만의 Q-점수를 갖는 염기 호출을 플레이스홀더 뉴클레오티드 (N)로 대체하였다. 이 품질 역치는 1000 중 1의 예상된 이론적 오차 비율을 갖는 뉴클레오티드 빈도를 발생시킨다.

[0755] 염기 편집 프로세싱성의 분석을 주문제작 피톤 스크립트를 사용하여 수행하였다. 이 프로그램은 표적 부위에 플랭킹되어야 하는 7개 뉴클레오티드 서열에 대한 완벽한 매치에 의해 측정되는 바와 같이 서열분석 판독물을 20개 뉴클레오티드 프로토스페이서 서열에 대해 트리밍한다. 그 후, 이들 표적을 통합하고, 풍부도에 의해 분류하여 염기 편집 생성물의 빈도를 평가하였다.

[0756] 인간 질환-연관된 돌연변이의 클린바 데이터베이스의 생물정보학적 분석을 작은 수정을 제외하고 이전에 기재된 것과 유사한 방식으로 수행하였다[1]. 이들 수정은 주문제작가능한 길이 및 서열의 PAM을 갖는 표적의 확인을 가능하게 한다. 또한, 이 개선된 스크립트는 표적 C 위치의 우선순위 랭킹 (C5 > C6 > C7 >C8 ≈ C4)을 포함하며, 따라서 온-타겟 C가 윈도우 내에서 단지 시토신이거나, 편집 윈도우 내에서 임의의 오프-타겟 C보다 더 높은 예측된 편집 효율을 갖는 위치에 놓이는 표적 부위의 확인을 가능하게 한다.

[0757] 실시예 12에 대한 참고문헌

1 Komor, A. C. *et al.* Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
2 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355 (2014).
3 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
4 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281-2308 (2013).
5 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2015).
6 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729-1-8 (2016).
7 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* doi:10.1038/nmeth.4027 (2016).
8 Jiang, F. *et al.* Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-71 (2016).

[0758]

9 Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191 (2015).
10 Zhang, Y. *et al.* Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* **4**, (2014).
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.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293-1298 (2015).
13 Holden, L. G. et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **452**, 121-124 (2008).
14 Chen, K.-M. et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116-119 (2008).
15 Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell* **10**, 1247-1253 (2002).
16 Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

[0759]

[0760] 실시예 13

[0761] 개선된 형질감염 절차 및 보다 양호한 플라스미드를 사용하여, 생물학적 반복 (n=3)을 사용하여 4가지 HF 돌연변이를 BE3의 Cas9 부분 내로 설치하였다. 돌연변이는 플라스미드 전달로 온-타겟팅 편집을 유의하게 실시하지 않는다 (도 99). 시험된 농도에서, BE3 단백질 전달은 작용하지만; 그러나, 온-타겟 편집은 플라스미드 전달에 대해서보다 더 낮다 (도 100). 설치된 HF 돌연변이를 갖는 BE3의 단백질 전달은 온-타겟팅 편집 효율을 감소시키지만, 여전히 일부 편집된 세포를 생성한다 (도 101).

[0762] 리포펙션 및 HF 돌연변이의 설치 둘 다는 오프-타겟 탈아미노화 사건을 감소시키는 것으로 나타났다. 도 102에 제시된 4개의 부위에 대해, 가장 높은 GUIDE-Seq 판독물 및 탈아미노화 사건을 갖는 오프-타겟 부위 (OT)를 검정하였다 (Komor et al., Nature, 2016). 가장 가까운 상응하는 C에서의 오프-타겟 편집을 온-타겟 편집으로 나눔으로써 특이성 비를 계산하였다. 오프-타겟 편집이 검출가능하지 않은 경우, 비는 100으로 설정되었다. 따라서, 보다 높은 특이성 비는 보다 많은 특이적 구축물을 나타낸다. BE3 플라스미드 전달은 BE3의 단백질 전달, HF-BE3의 플라스미드 전달, 또는 HF-BE3의 단백질 전달보다 훨씬 더 높은 오프-타겟/온-타겟 편집을 나타내었다 (도 102 및 105).

[0763] 정제된 단백질 HF-BE3 및 BE3을 가장 허용적인 모티프를 갖는 스페이서에서의 상이한 위치에서 C를 T 잔기로 전환시키는 그의 능력에 대해 시험관내에서 분석하였다. BE3 및 HF-BE3 단백질 둘 다는 염기 편집에 대해 동일한 "윈도우"을 갖는 것으로 밝혀졌다 (도 103 및 104).

[0764] 질환 표적의 목록이 표 8에 주어진다. 표 8에서 편집되는 염기는 볼드체 및 밑줄표시로 지시된다.

[0765] 표 8. 염기 편집제 질환 표적

유전자	질환	스페이서	SEQ ID NO	PAM	편집제	결함	세포
RB1	RETINOBLASTOMA	AATCTAGTAAATAAATTGATGT	571	AAAAGT	SAKKH-BE3	SPLICING IMPAIRMENT	J82
PTEN	CANCER	GACCAACGGCTAAGTGAAGA	572	TGA	VQR-BE3	W111R	MC116
PIK3CA	CANCER	TCCTTTCTTCACGGTTGCCT	573	ACTGGT	SAKKH-BE3	K111R	CRL-5853
PIK3CA	CANCER	CTCCTGCTCAGTGATTTCAG	574	AGA	VQR-BE3	Q546R	CRL-2505
TP53	CANCER	TGTCACACATGTAGTTGTAG	575	TGG	YEE-BE3	N239D	SNU475
HRAS	CANCER	CCTCCCGGCCGGCGGTATCC	576	AGG	YEE-BE3	Q61R	MC/CAR

[0766]

[0767] 표 6. 염기 편집제를 사용하여 치료될 수 있는 예시적인 질환. 프로토스페이서 및 PAM 서열 (서열식별번호: 577-589)은 sgRNA (PAM) 열에 나타나 있다. PAM 서열은 괄호 안에 및 밑줄표시됨으로써 지시되는 편집되는 염기와 함께 나타나 있다.

질환 표적	유전자 기호	변화된 염기	sgRNA (PAM)	염기 편집제
프리온 질환	PRNP	R37*	GGCAGCCGATACCCGGGGCA(GGG)	BE3
			GGGCAGCCGATACCCGGGGC(AGG)	
펜드레드 증후군	Slc26a4	c.919-2A>G	TTATTGTCCGAAATAAAAGA(AGA)	BE3
			ATTGTCCGAAATAAAAGAAG(AGG)	(VQR
			TTGTCCGAAATAAAAGAAGA(GGA)	SaCas9)
			GTCCGAAATAAAAGAAGAGGAAAA(AAT)	
			GTCCGAAATAAAAGAAGAGGAAAAA(ATT)	
선천성 난청	Tmc1	c.545A>G	CAGGAAGCACGAGGCCACTG(AGG)	BE3
			AACAGGAAGCACGAGGCCAC(TGA)	YE-BE3
			AGGAAGCACGAGGCCACTGA(GGA)	YEE-BE3
후천성 난청	SNHL	S33F	TTGGATTCTGGAATCCATTC(TGG)	BE3
알츠하이머병	APP	A673T	TCTGCATCCATCTTCACTTC(AGA)	BE3 VQR
C형 니만-픽병	NPC1	I1061T	CTTACAGCCAGTAATGTCAC(CGA)	BE3 VQR

[0768]

[0769] 실시예 14: 염기 편집 구축물의 시험

[0770] BE3, BE4-pmCDA1, BE4-hAID, BE4-3G, BE4-N, BE4-SSB, BE4-$(GGS)_3$, BE4-XTEN, BE4-32aa, BE4-2xUGI, 및 BE4를 포함한 여러 염기 편집 구축물을 상이한 표적 서열 (즉, EMX1, FANCF, HEK2, HEK3, HEK4, 및 RNF2) 내의 시토신 (C) 잔기를 편집하는 그의 능력에 대해 시험하였다. 예를 들어, 이들 구축물이 C에서 T로의 돌연변이를 생성할 수 있는지 시험하였다. 염기 편집 구축물의 개략적 표현이 도 109에 제시된다. 시험된 표적 서열이 또한 도 109에 제시되고, 표적화된 시토신은 넘버링되고 적색으로 표시된다.

[0771] 하기 아미노산 서열을 본 실시예의 염기 편집 구축물에서 사용하였다:

UGI:

TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP EYKPWALVIQDSNGENKIKML (SEQ ID NO: 736)

rAPOBEC1:

SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKH VEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHA DPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLEL YCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 737)

pmCDA1:

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQ SGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHT LKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENR WLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 81)

hAID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHV ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFC EDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSR QLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 49)

[0772]

hAPOBEC3G:

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTL TIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEP WNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTW VLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTF VDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 60)

SSB (단일-가닥 DNA 결합 단백질):

ASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKATGEMKEQTEWHRV VLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNVGGTMQMLGG RQGGGAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPAAPSNEPPMDF DDDIPF (SEQ ID NO: 590)

링커 서열:

XTEN:

SGSETPGTSESATPES (SEQ ID NO: 604)

32aa:

SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605)

SGGS:

SGGS (SEQ ID NO: 606)

$(GGS)_3$:

GGSGGSGGS (SEQ ID NO: 610)

[0773]

[0774] 도 109에 제시된 구축물의 아미노산 서열이 하기에 제시된다:

BE3:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN

[0775]

SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 174)

BE4-pmCDA1:

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGY
AVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQEL
RGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQ
LNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSGSETPGTSESATPESDKKYSIG
LAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL

[0776]

VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQED
FYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK
LINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP
KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIE
KETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA
LVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 175)

BE4-hAID:
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHV
ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFC
EDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSR
QLRRILLPLYEVDDLRDAFRTLGLSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICY
LQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI

[0777]

NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIK
MLSGGSPKKKRKV (SEQ ID NO: 176)

BE4-3G:

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTL
TIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEP
WNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVVRGRHETYLCYEVERMHNDTW
VLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC
AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTF
VDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQENSGSETPGTSESATPESDKKYSIGLAI
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR

[0778]

RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ
TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ
EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY
PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKET
GKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVI
QDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 177)

BE4-N:

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKMLGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKE
TCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGE
CSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNF
VNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQ

[0779]

RLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKF
KVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS
ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL
NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV (SEQ ID NO: 178)

BE4-SSB:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET

[0780]

AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSGGSGGSASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKATGEM
KEQTEWHRVVLFGKLAEVASEYLRKGSQVYIEGQLRTRKWTDQSGQDRYTTEVVVNV
GGTMQMLGGRQGGGAPAGGNIGGGQPQGGWGQPQQPQGGNQFSGGAQSRPQQSAPA
APSNEPPMDFDDDIPFSGGSPKKKRKV (SEQ ID NO: 179)

BE4-$(GGS)_3$:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL

[0781]

ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 180)

BE4-XTEN:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL

[0782]

ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GSETPGTSESATPESTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE
STDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 181)

BE4-32aa:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL

[0783]

ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSET
PGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILV
HTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID
NO: 182)

BE4-2XUGI:

[0784] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH

ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESAT
PESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDS
GGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKMLSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKR
KV (SEQ ID NO: 183)

[0785] BE4:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSET
PGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL
SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP
ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTN
LSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEY
KPWALVIQDSNGENKIKMLSGGSPKKKRKV (SEQ ID NO: 184)

[0786]

[0787] EMX1, FANCF, HEK2, HEK3, HEK4, 및 RNF2의 표적 시토신 잔기를 돌연변이시키는 도 109의 염기 편집 구축물의 능력이 도 110-115에 제시된다. 편집된 표적 시토신의 백분율 (T, A 또는 G로 돌연변이된 C 잔기의 비율 포함), 뿐만 아니라 생성된 indel의 %가 도 110-115에 제시된다. 편집된 표적 시토신의 백분율을 하기 식을 사용하여 계산하였다: 100-[C를 갖는 서열분석 판독물의 %]. 도 110-115의 파이 차트는 표시된 다양한 염기를 갖는 판독물의 분포를 보여주며, 이는 모든 염기 편집된 판독물 (표시된 염기에서 C 이외의 뉴클레오티드를 갖는 판독물) 중 A, G, 또는 T를 갖는 것들의 백분율이 얼마인지 살펴보는 것을 의미한다. 하기 표 9 내지 표 14는 도 110-115의 파이 차트에 의해 표시되는 돌연변이 백분율의 값을 나타낸다.

[0788] 관찰된 C에서 비-T로의 편집은 UDG (우라실 DNA 글리코실라제)로 인한 것일 가능성이 있다. 예를 들어, C가 우라실 중간체로 전환되면, UDG는 이를 무염기성 부위로 전환시킬 수 있다. 이어서, 이러한 무염기성 부위는 다른 내인성 효소에 의해 프로세싱되고, 결국 indel 또는 다른 염기 (예컨대 G 또는 A)가 C를 대체하는 것을 야기한다. 본 발명자들은 UDG 녹-아웃 세포주에서 indel을 거의 내지 전혀 갖지 않는 증가된 C에서 T로의 편집을 보여준다는 것을 밝혀냈다.

표 9 EMX1

EMX1		C_5
BE3	A	1.9%
	C	51.9%
	G	6.0%
	T	40.3%
pmCDA1	A	0.2%
	C	88.9%
	G	0.8%
	T	10.0%
hAID	A	0.3%
	C	83.6%
	G	0.6%
	T	15.4%
hAPOBEC3G	A	0.0%
	C	98.9%

[0789]

	G	0.2%
	T	1.0%
BE4-N	A	1.4%
	C	85.6%
	G	3.5%
	T	9.5%
EMX1		
BE4-SSB	A	0.3%
	C	96.7%
	G	0.7%
	T	2.3%
BE4-$(GGS)_3$	A	1.7%
	C	36.8%
	G	4.1%
	T	57.4%
BE4-XTEN	A	2.1%
	C	45.7%
	G	5.1%
	T	47.1%
BE4-32aa	A	1.6%
	C	46.3%
	G	4.6%
	T	47.5%
BE4-2XUGI	A	0.6%
	C	60.2%
	G	1.9%
	T	37.3%

표 10 FANCF

FANCF		C_8
BE3	A	1.6%
	C	74.4%
	G	0.7%
	T	23.3%
pmCDA1	A	0.5%
	C	88.8%
	G	0.3%
	T	10.3%

[0790]

hAID	A	0.3%
	C	89.6%
	G	0.2%
	T	9.9%
hAPOBEC3G	A	2.1%
	C	75.4%
	G	10.4%
	T	12.1%
BE4-N	A	1.8%
	C	83.5%
	G	0.6%
	T	14.2%
BE4-SSB	A	0.3%
	C	97.1%
	G	0.1%
	T	2.6%
BE4-$(GGS)_3$	A	2.3%
	C	45.7%
	G	1.4%
	T	50.6%
BE4-XTEN	A	1.1%
	C	56.3%
	G	0.4%
	T	42.1%
BE4-32aa	A	1.2%
	C	70.0%
	G	0.6%
	T	28.2%
BE4-2XUGI	A	0.9%
	C	57.8%
	G	1.1%
	T	40.2%

표 11 HEK2

HEK2		C_6
BE3	A	0.9%

[0791]

	C	28.2%
	G	52.9%
	T	18.1%
pmCDA1	A	1.8%
	C	73.5%
	G	3.5%
	T	21.2%
hAID	A	2.1%
	C	56.9%
	G	7.7%
	T	33.3%
hAPOBEC3G	A	0.1%
	C	86.5%
	G	9.9%
	T	3.5%
BE4-N	A	1.0%
	C	37.6%
	G	57.0%
	T	4.4%
BE4-SSB	A	0.2%
	C	78.4%
	G	20.0%
	T	1.4%
BE4-$(GGS)_3$	A	0.6%
	C	11.1%
	G	40.6%
	T	47.7%
BE4-XTEN	A	1.2%
	C	24.8%
	G	44.6%
	T	29.4%
BE4-32aa	A	1.1%
	C	26.3%
	G	41.8%
	T	30.7%
BE4-2XUGI	A	0.8%

[0792]

	C	37.0%
	G	21.6%
	T	40.6%

표 12 HEK3

HEK3		C_5
BE3	A	2.23%
	C	38.06%
	G	12.77%
	T	46.95%
pmCDA1	A	0.21%
	C	76.57%
	G	0.12%
	T	23.09%
hAID	A	0.28%
	C	60.23%
	G	1.03%
	T	38.45%
hAPOBEC3G	A	3.11%
	C	33.89%
	G	28.59%
	T	34.41%
BE4-N	A	2.6%
	C	64.1%
	G	13.5%
	T	19.8%
BE4-SSB	A	0.4%
	C	92.9%
	G	2.8%
	T	3.9%
BE4-$(GGS)_3$	A	1.3%
	C	9.9%
	G	7.9%
	T	80.8%
BE4-XTEN	A	2.3%
	C	15.9%
	G	12.2%

[0793]

	T	69.6%
BE4-32aa	A	1.3%
	C	14.9%
	G	9.9%
	T	73.9%
BE4-2XUGI	A	0.6%
	C	23.4%
	G	3.8%
	T	72.2%

표 13 HEK4

HEK4		C_5
BE3	A	8.40%
	C	41.89%
	G	24.54%
	T	25.17%
pmCDA1	A	0.50%
	C	87.53%
	G	0.01%
	T	11.95%
hAID	A	0.93%
	C	71.32%
	G	0.69%
	T	27.06%
hAPOBEC3G	A	0.12%
	C	99.37%
	G	0.35%
	T	0.16%
BE4-N	A	7.3%
	C	56.6%
	G	25.7%
	T	10.3%
BE4-SSB	A	2.1%
	C	86.8%
	G	5.8%
	T	5.2%

[0794]

BE4-$(GGS)_3$	A	6.7%
	C	13.0%
	G	19.8%
	T	60.5%
BE4-XTEN	A	7.5%
	C	19.7%
	G	25.4%
	T	47.4%
BE4-32aa	A	7.9%
	C	21.8%
	G	25.1%
	T	45.3%
BE4-2XUGI	A	3.4%
	C	22.2%
	G	12.4%
	T	62.0%

표 14 RNF2

RNF2		C_6
BE3	A	2.46%
	C	46.65%
	G	19.87%
	T	31.03%
pmCDA1	A	0.60%
	C	83.52%
	G	1.33%
	T	14.55%
hAID	A	0.36%
	C	75.03%
	G	3.20%
	T	21.40%
hAPOBEC3G	A	0.10%
	C	86.60%
	G	3.70%
	T	9.59%
BE4-N	A	5.1%
	C	50.0%

[0795]

	G	28.8%
	T	16.2%
BE4-SSB	A	1.1%
	C	89.9%
	G	4.9%
	T	4.1%
BE4-$(GGS)_3$	A	2.0%
	C	23.0%
	G	14.0%
	T	61.0%
BE4-XTEN	A	2.6%
	C	32.4%
	G	16.0%
	T	49.0%
BE4-32aa	A	2.2%
	C	29.2%
	G	18.5%
	T	50.0%
BE4-2XUGI	A	0.7%
	C	45.0%
	G	6.5%
	T	47.8%

[0796]

[0797] 실시예 15: 표적화된 가닥을 절단하는 Cpf1 니카제를 포함하는 염기 편집제.

[0798] 상기 논의된 바와 같이, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 Cpf1 단백질일 수 있다. 일부 실시양태에서, Cpf1 단백질은 Cpf1 니카제 (nCpf1)이다. Cpf1 니카제, 예를 들어 비-표적 가닥을 절단하는 Cpf1 니카제 (AsCpf1 내의 R1225A; 및 LbCpf1 내의 R1138A)는 문헌 [Yamano et al., "Crystal structure of Cpf1 in complex with guide RNA and target DNA." Cell (165) 2016, p. 949-962]에 기재되어 있으며, 그의 전체 내용은 본원에 참조로 포함된다. 그러나, 표적 가닥을 절단하는 니카제 (예를 들어, 염기 편집제 단백질의 Cpf1 니카제)는 염기 편집 효율을 개선시킬 것으로 예상된다.

[0799] 형광 표지된 DNA를 사용하여 비-표적 가닥보다는 표적 가닥을 우선적으로 니킹하는 Cpf1 돌연변이체를 확인하였다 (도 116 참조). 도 116에서, DNA 구축물 1-3의 상부 가닥은 비-표적 가닥이고, 하부 가닥은 표적 가닥이다. 야생형 LbCpf1, R836A (LbCpf1), 및 R1138A (LbCpf1)를 사용하여 시험관내 검정을 수행한다. R836A (LbCpf1)는 "손상된" 니카제인 것으로 보이며, 이는 그것이 비-표적 가닥보다 더 효율적으로 표적 가닥을 절단한다는 것을 의미한다. 도 117에 제시된 바와 같이, 비-표적 가닥은 절단되지 않았고, 형광 350 조각이 관찰되지 않았다. 2시간 후, 가닥 둘 다를 절단하였다. 상이한 강도는 비-표적 가닥보다 더 많은 표적 가닥이 절단된다는 것을 시사한다.

[0800] AsCp1-BE3을 사용한 염기 편집 윈도우의 확립

[0801] napDNAbp로서 LbCpf1(R836A) 또는 AsCpf1(R912A)을 갖는 염기 편집 단백질 (예를 들어, BE3 (SpCas9-BE3))은 낮은 효율로 (0.1% 내지 0.4%) 염기를 편집하는 것으로 밝혀졌다. AsCpf1 (R912A) napDNAbp를 갖는 염기 편집제는 EMX1, FANCF, HEK3 및 HEK4 부위에서 표적 C를 더 효율적으로 돌연변이시켰다. 시험된 구축물의 편집 윈도우는 제7 염기에서 제11 염기인 것으로 보인다. 숫자는 가장 높은 수를 갖는 BE3 및 보다 낮은 수를 갖는 자기-무력화 BE (비-표적 가닥을 절단하는 APOBEC-AsCpf1(R1225A)-UGI)와 경향이 일치한다. 도 118 참조 EMX1 상에 Cas9-BE3를 갖는 양성 대조군: 5-6%. AsCpf1에 대한 indel 값: >20%. AsCpf1 내의 R912는 Cpf1 패밀리의 많은 구성원에 걸쳐 보존된다. LbCpf1 내의 상응하는 잔기는 R836이고, 이는 R이 A로 돌연변이되는 경우에 "손상된" 니카제인 것으로 여겨진다.

[0802] Cpf1-BE (링커)의 최적화

[0803] indel 데이터는 Cpf1이 DNA 표적 부위에 접근할 수 있다는 것을 시사한다. 따라서, Cpf1 염기 편집 단백질의 최적화는 특정 APOBEC 단백질, 링커, 및/또는 UGI 도메인에 초점을 맞추었다. LbCpf1 (R836A) 및 AsCpf1 (R912A) 둘 다를 사용하여 다양한 링커를 사용하여 도 119에 제시된 구축물을 시험하였다. 간략하게, APOBEC와

Cpf1 도메인 (예를 들어, AsCpf1 또는 LbCpf1) 사이의 상이한 링커 서열 (즉, XTEN, GGS, $(GGS)_3$ (서열식별번호: 610), $(GGS)_5$ (서열식별번호: 610), 및 $(GGS)_7$ (서열식별번호: 610))을 시험하였다. 도 120을 참조한다. 구축물을 $TGCTTCTC_8CAGCCCTGGCCTGG$ (서열식별번호: 592)인 HEK3 부위의 C_8 잔기를 돌연변이시키는 그의 능력에 대해 시험하였다. AsCpf1을 사용한 염기 편집 단백질에 대한 편집 수준은 1% 초과에 도달한 반면, LbCpf1을 사용한 염기 편집 단백질은 염기 편집 효율의 상대 감소를 보여주었다. 도 121에 제시된 바와 같이, 통합 생물정보학 VU 센터에 의해 유지된 데이터베이스로부터의 링커는 AsCpf1-BE3에 대해 GGS-유형 링커만큼 유의한 개선을 보여주지 않았다. 도 121에 제시된 링커는 하기에 제시된다:

PDB_코드	길이 (aa)	서열
1au7A_1	10	KRRTTISIAA (SEQ ID NO: 593)
1c1kA_1	19	ALVFYREYIGRLKQIKFKF (SEQ ID NO: 594)
1c20A_1	14	LPIMAKSVLDLYEL (SEQ ID NO: 595)
1ee8A_1	5	LLRLG (SEQ ID NO: 596)
1f1zA_1	15	TDKEINPVVKENIEW (SEQ ID NO: 597)
1ignA_1	8	PPSIKRKF (SEQ ID NO: 598)
1jmcA_1	9	LPTVQFDFT (SEQ ID NO: 599)
1sfe_1	14	LPLDIRGTAFQQQV (SEQ ID NO: 600)
2ezx_1	8	AYVVLGQF (SEQ ID NO: 601)
2reb_1	8	INFYGELV (SEQ ID NO: 602)

[0804]

[0805] Cpf1-BE (배향)의 최적화

[0806] Cas9는 C와 N 말단 사이의 아미노산의 스트레치를 갖는 반면 (적색 사각형, 도 123 참조), AsCpf1은 그렇지 않다 (도 122 참조). 또한, AsCpf1은 N과 C 말단 사이의 거리가 더 짧다. 이러한 관찰은 UGI가 비-표적 가닥에 대한 APOBEC 접근을 방해할 수 있는 (N 말단 상의) APOBEC와 (C 말단 상의) UGI 사이의 잠재적 간섭을 나타낸다. 하나의 해결책은 APOBEC 및 UGI를 N 또는 C의, 동일한 말단 상으로 이동시키는 것이다. 따라서, 아키텍처 NLS-UGI-APOBEC-XTEN-AsCpf1; UGI-APOBEC-XTEN-AsCpf1-NLS; 및 AsCpf1-XTEN-APOBEC-NLS를 갖는 구축물이 시험될 것이다.

[0807] Cpf1-BE (내부 말단절단)의 최적화

[0808] 비-표적 가닥이 풀려 있는 Cpf1의 공지된 결정 구조는 존재하지 않는다 (도 124, 시안색 참조). 편집 윈도우는 도 124에 제시된 바와 같은 적색 원 내에 있어야 한다고 여겨진다. APOBEC를 방해할 수 있는 나선 영역이 존재한다 (도 124의 정사각형 참조). 이러한 영역은 아미노산 서열 K(661)KTGDQK(667) (서열식별번호: 603)을 포함한다.

[0809] 2, 4 또는 6개의 잔기의 제거가 염기 편집 효율을 개선시키는지 시험하기 위해, 표적 부위로서 HEK3을 사용하여 AsCpf1(R912A) napDNAbp를 갖는 염기 편집제로 실험을 수행하였다. 편집 수준은 T663 및 D665가 결실된 경우에 대조군 수준으로부터 대략 2.6%로 증가한다 (~ 6배 증가) (하기 표 7 참조). 본 실험에서 사용된 구축물은 APOBEC-XTEN-AsCpf1(R912A)-SGGS-UGI였다.

[0810] 표 7

결실	C8에서의 편집	C9에서의 편집
T663, D665	2.59%	1.29%
K662, T663, D665, Q666	0.15%	0.15%
K661, K662, T663, D665, Q666, K667	0.22%	0.21%

[0811]

[0812] 참고문헌

1. Humbert O, Davis L, Maizels N. Targeted gene therapies: tools, applications, optimization. *Crit Rev Biochem Mol.***2012;** 47(3):264-81. PMID: 22530743.
2. Perez-Pinera P, Ousterout DG, Gersbach CA. Advances in targeted genome editing. *Curr Opin Chem Biol.* **2012;** 16(3-4):268-77. PMID: 22819644.
3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet.* **2010;** 11(9):636-46. PMID: 20717154.
4. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol.* **2013;** 14(1):49-55. PMID: 23169466.
5. Charpentier E, Doudna JA. Biotechnology: Rewriting a genome. *Nature.* **2013;** 495, (7439):50-1. PMID: 23467164.
6. Pan Y, Xia L, Li AS, Zhang X, Sirois P, Zhang J, Li K. Biological and biomedical applications of engineered nucleases. *Mol Biotechnol.* **2013;** 55(1):54-62. PMID: 23089945.
7. De Souza, N. Primer: genome editing with engineered nucleases. *Nat Methods.* **2012;** 9(1):27. PMID: 22312638.
8. Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A.* **2008;** 105(15):5809-14. PMID: 18359850.
9. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley

[0813]

GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet.* **1999;** 22(3):231-8. PMID: 10391209.
10. Jansen R, van Embden JD, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol.* **2002;** 43(6):1565-75. PMID: 11952905.
11. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods.* **2013;** 10(10):957-63. PMID: 24076990.
12. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Shou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol.* **2011;** 18(5):529-36. PMID: 21460843.
13. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* **2010;** 327(5962):167-70. PMID: 20056882.
14. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* **2012;** 482(7385):331-8. PMID: 22337052.
15. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.* **2013;** 21(11):562-7. PMID: 24095303.
16. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* **2013;** 152(5):1173-83. PMID: 23452860.
17. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods.* **2013;** 10(10):973-6. PMID: 23892895.
18. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* **2013;** 31(9):833-8. PMID: 23907171.
19. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* **2013;** 154(2):442-51. PMID: 23849981.
20. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc.* **2013;** 8(11):2180-96. PMID: 24136345.
21. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* **2013;** 339(6121):823-6. PMID: 23287722.

[0814]

22. Cole-Strauss A, Yoon K, Xiang Y, Byrne BC, Rice MC, Gryn J, Holloman WK, Kmiec EB. Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science.* **1996;** 273(5280):1386-9. PMID: 8703073.
23. Tagalakis AD, Owen JS, Simons JP. Lack of RNA-DNA oligonucleotide (chimeraplast) mutagenic activity in mouse embryos. *Mol Reprod Dev.* **2005;** 71(2):140-4. PMID: 15791601.
24. Ray A, Langer M. Homologous recombination: ends as the means. *Trends Plant Sci.* **2002;** 7(10):435-40. PMID 12399177.
25. Britt AB, May GD. Re-engineering plant gene targeting. *Trends Plant Sci.* **2003;** 8(2):90-5. PMID: 12597876.
26. Vagner V, Ehrlich SD. Efficiency of homologous DNA recombination varies along the Bacillus subtilis chromosome. *J Bacteriol.* **1988;** 170(9):3978-82. PMID: 3137211.
27. Saleh-Gohari N, Helleday T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* **2004;** 32(12):3683-8. PMID: 15252152.
28. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol.* **2007;** 25(11):1298-306. PMID: 17965707.
29. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **2008;** 9(6):229. PMID: 18598372.
30. Reynaud CA, Aoufouchi S, Faili A, Weill JC. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol.* **2003;** 4(7):631-8.
31. Bhagwat AS. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair (Amst).* **2004;** 3(1):85-9. PMID: 14697763.
32. Navaratnam N, Sarwar R. An overview of cytidine deaminases. *Int J Hematol.* **2006;** 83(3):195-200. PMID: 16720547.
33. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature.* **2008;** 456(7218):121-4. PMID: 18849968.
34. Chelico L, Pham P, Petruska J, Goodman MF. Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem.* **2009;** 284(41). 27761-5. PMID: 19684020.
35. Pham P, Bransteitter R, Goodman MF. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry.* **2005;** 44(8):2703-15. PMID 15723516.
36. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* **2013;** 65(10):1357-69. PMID: 23026637.
37. Lee JW, Soung YH, Kim SY, Lee HW, Park WS, Nam SW, Kim SH, Lee JY, Yoo NJ, Lee SH. PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. *Oncogene.* **2005;** 24(8):1477-80. PMID: 15608678.

[0815]

38. Ikediobi ON, Davies H, Bignell G, Edkins S, Stevens C, O'Meara S, Santarius T, Avis T, Barthorpe S, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Hunter C, Jenkinson A, Jones D, Kosmidou V, Lugg R, Menzies A, Mironenko T, Parker A, Perry J, Raine K, Richardson D, Shepherd R, Small A, Smith R, Solomon H, Stephens P, Teaque J, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Reinhold W, Weinstein JN, Stratton MR, Futreal PA, Wooster R. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther.* **2006;** 5(11):2606-12. PMID: 17088437.

39. Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* **21**, 121-131, doi:10.1038/nm.3793 (2015).

40. Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome research* **25**, 1442-1455, doi:10.1101/gr.190124.115 (2015).

41. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355, doi:10.1038/nbt.2842 (2014).

42. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33**, 538-542, doi:10.1038/nbt.3190 (2015).

43. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology* **33**, 543-548, doi:10.1038/nbt.3198 (2015).

44. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766, doi:10.7554/eLife.04766 (2014).

45. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

46. Rong, Z., Zhu, S., Xu, Y. & Fu, X. Homologous recombination in human embryonic stem cells using CRISPR/Cas9 nickase and a long DNA donor template. *Protein & cell* **5**, 258-260, doi:10.1007/s13238-014-0032-5 (2014).

47. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

48. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell* **10**, 1247-1253 (2002).

49. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997, doi:10.1126/science.1247997 (2014).

50. Schellenberger, V. *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nature biotechnology* **27**, 1186-1190, doi:10.1038/nbt.1588 (2009).

[0816]

51. Saraconi, G., Severi, F., Sala, C., Mattiuz, G. & Conticello, S. G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome biology* **15**, 417- (2014).
52. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).
53. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
54. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR- Cas nucleases. *Nature biotechnology* **33**, 187-197, doi:10.1038/nbt.3117 (2015).
55. Kunz, C., Saito, Y. & Schar, P. DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cellular and molecular life sciences : CMLS* **66**, 1021-1038, doi:10.1007/s00018-009- 8739-9 (2009).
56. D., M. C. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82**, 701-708 (1995).
57. Caldecott, K. W. Single-strand break repair and genetic disease. *Nature reviews. Genetics* **9**, 619-631, doi:10.1038/nrg2380 (2008).
58. Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *Journal of nucleic acids* **2010**, doi:10.4061/2010/260512 (2010).
59. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109**, E2579–E2586, doi:10.1073/pnas.1208507109 (2012).
60. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191, doi:10.1038/nature14299 (2015).
61. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology* **32**, 677-683, doi:10.1038/nbt.2916 (2014).
62. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* **32**, 670-676, doi:10.1038/nbt.2889 (2014).
63. Beale, R. C. L. *et al.* Comparison of the Differential Context-dependence of DNA Deamination by APOBEC Enzymes: Correlation with Mutation Spectra in Vivo. *Journal of Molecular Biology* **337**, 585-596, doi:10.1016/j.jmb.2004.01.046 (2004).
64. Kim, J., Basak, J. M. & Holtzman, D. M. The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63**, 287-303, doi:10.1016/j.neuron.2009.06.026 (2009).
65. Liu, C. C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature reviews. Neurology* **9**, 106-118, doi:10.1038/nrneurol.2012.263 (2013).
66. Sjöblom, T. *et al.* The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science* **314**, 268-274, doi:10.1126/science.1133427 (2006).

[0817]

67. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-404, doi:10.1038/nature11017 (2012).
68. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, doi:10.1093/nar/gkv1222 (2015).
69. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science*, doi:10.1126/science.aad5227 (2015).
70. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nature chemical biology* **11**, 316-318, doi:10.1038/nchembio.1793 (2015).
71. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33**, 73-80, doi:10.1038/nbt.3081 (2015).
72. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485, doi:10.1038/nature14592 (2015).
73. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA- programmed Cas9 nuclease specificity. *Nature Biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
74. Shcherbakova, D. M. & Verkhusha, V. V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nature Methods* **10**, 751-754, doi:10.1038/nmeth.2521 (2013).
75. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
76. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, doi:10.1126/science.aad8282 (2016).
77. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotech* **32**, 569-576, doi:10.1038/nbt.2908 (2014).
78. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* **4**, 712-720 (2003).
79. Heller, R. C. & Marians, K. J. Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol* **7**, 932-943 (2006).
80. Pluciennik, A. *et al.* PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16066-16071, doi:10.1073/pnas.1010662107 (2010).
81. Seripa, D. *et al.* The missing ApoE allele. *Annals of human genetics* **71**, 496-500, doi:10.1111/j.1469-1809.2006.00344.x (2007).
82. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).
83. 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 Biotech* **34**, 339-344, doi:10.1038/nbt.3481 (2016).

[0818]

84. Simonelli, V., Narciso, L., Dogliotti, E. & Fortini, P. Base excision repair intermediates are mutagenic in mammalian cells. *Nucleic acids research* **33**, 4404-4411, doi:10.1093/nar/gki749 (2005).
85. Barnes, D. E. & Lindahl, T. Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells. *Annual Review of Genetics* **38**, 445-476, doi:doi:10.1146/annurev.genet.38.072902.092448 (2004).

[0819]

[0820] 실시예 16: 단백질 조작 및 단백질 전달을 통한 염기 편집의 DNA 특이성 및 적용가능성 개선

[0821] 염기 편집, 이중 가닥 DNA 절단이 없는 한 염기 쌍의 또 다른 것으로의 프로그램가능한 전환을 가능하게 하는 게놈 편집 접근법, 과도한 확률적 삽입 및 결실, 또는 상동성-지정 복구에 대한 의존이 개발되었다. 염기 편집의 적용은 오프-타겟 활성 및 세포내 DNA 전달에 대한 의존에 의해 제한된다. 여기서, 이러한 한계를 해결하는 2가지 진전이 기재된다. 먼저, 오프-타겟 염기 편집은 고충실도 염기 편집제 (HF-BE3)를 생성하기 위해 돌연변이를 제3 세대 염기 편집제 (BE3) 내에 설치함으로써 감소되었다. 다음으로, BE3 및 HF-BE3을 정제하고, 리보핵단백질 (RNP) 복합체로서 포유동물 세포 내로 전달하고, DNA-무함유 염기 편집을 확립한다. BE3의 RNP 전달은 대등한 온-타겟 편집 수준을 유지하면서 HF-BE3의 플라스미드 형질감염보다 훨씬 더 높은 특이성을 부여한다. 마지막으로, 이들 진전을 적용하여 BE3 RNP를 제브라피쉬 배아 및 살아있는 마우스의 내이 둘 다로 전달하여, 생체내에서 특이적인 DNA-무함유 염기 편집을 달성한다.

[0822] 도입

[0823] 전통적인 게놈 편집제는 게놈 편집의 제1 단계로서 이중-가닥 DNA 파괴 (DSB)를 도입한다[1-4]. 세포는 주로 비-상동 말단 연결 (NHEJ)을 통해 DSB에 반응하여, 절단 부위에서 확률적 삽입 또는 결실 (indel)을 발생시킨다[1,5]. 게놈 DNA에서 보다 정확한 변화를 생성하기 위해, 상동성-지정 복구 (HDR)는 절단 부위를 둘러싼 게놈 DNA를 외인성으로 공급된 DNA 공여자 주형의 것으로 대체하는 데 사용될 수 있다[6-8]. 불행히도, HDR은 전형적으로 NHEJ와 경쟁하여 생성된 과량의 indel을 동반하고, 이는 주로 유사분열 세포로 제한된다. 또한, 대부분의 게놈 편집 방법은 외인성 플라스미드 또는 바이러스 DNA의 포유동물 세포 내로의 전달에 이어서 작용제의 세포내 발현에 의존한다[9-12]. 이들 전달 방법은 온-타겟 유전자좌가 편집된 후에도 연속적인 제어되지 않은 Cas9 및 sgRNA 발현을 발생시켜, 오프-타겟 유전자좌에서의 게놈 편집의 기회를 증가시킨다[1,13].

[0824] 염기 편집, C:G 염기 쌍의 T:A 염기 쌍으로의 직접적 프로그램가능한 표적화된 전환을 가능하게 하는 게놈 편집에 대한 상이한 접근법이 최근에 기재되었다[3,14]. 제3 세대 염기 편집제 BE3은 단일 단백질 내에 (i) 가이드 RNA-특정된 유전자좌에서 소형 단일 가닥 DNA 버블을 개방하는 촉매적으로 손상된 Cas9, (ii) 단일 가닥 DNA 버블 내의 대략 5개 뉴클레오티드의 윈도우 내에서 C를 U로 전환시키는 테더링된 단일 가닥-특이적 시티딘 데아미나제, (iii) 염기 절제 복구를 억제함으로써 염기 편집의 효율 및 생성물 선택성을 개선시키는 우라실 글리코실라제 억제제 (UGI), 및 (iv) G-함유 DNA 가닥을 대체하는 것으로 세포 미스매치 복구를 조작하기 위한 니카제 활성을 함유한다. 이들 성분의 조합은 최소의 indel을 형성하면서 포유동물 세포에서 효율적이고 영구적인 C에서 T로의 (또는 G에서 A로의) 전환을 가능하게 한다. 이전에 보고된 이후로[14], 다른 연구자들이 포유동물 세포[15-17] 및 식물[18]에서의 Cas9-지정 C에서 T로의 전환을 용이하게 하는 이러한 전략 및 관련 접근법의 능력을 확인하였다.

[0825] 여기서, 염기 편집의 DNA 특이성을 크게 개선시키고, 시험관내에서 및 생체내에서 외인성 DNA를 공급하지 않으면서, 숙주 게놈과의 재조합 및 세포독성의 위험과 연관된 염기 편집을 허용하는 2가지 진전이 기재된다[18,19]. 먼저, 온-타겟 편집 활성에서 오직 보통의 감소를 가지면서 오프-타겟 편집 사건을 감소시키는 Cas9[20]의 DNA 친화도를 감소시키는 것으로 공지된 돌연변이를 혼입시키는 BE3의 돌연변이체 형태가 조작된다. 이어서, 가이드 RNA와 복합체화된 염기 편집제 단백질의 지질-매개 전달이 플라스미드 DNA 전달과 비교하여 온-타겟 염기 편집에서 명백한 감소 없이 훨씬 더 큰 특이성 증진을 발생시킨다는 것이 밝혀졌다. RNP로서의 염기 편집제의 전달은 전형적으로, 배양된 인간 및 마우스 세포에서 고도로 반복되는 게놈 DNA 서열을 표적화하는 상당히 혼재성인 가이드 RNA에 대해서도, 오프-타겟 편집을 측정가능한 수준 미만으로 감소시킨다. 이러한 진전은 마우스 및 제브라피쉬에서 고도로 특이적인 DNA-무함유 생체내 염기 편집이 입증되도록 한다.

[0826] 결과

[0827] 고충실도 염기 편집제의 조작

[0828] Cas9 뉴클레아제 및 그의 연관된 융합 구축물은 오프-타겟 게놈 유전자좌에서 DNA에 결합하고 그를 절단하는 것으로 밝혀졌다[21-24]. 종 및 동료들은 4개의 점 돌연변이 (N497A, R661A, Q695A, Q926A)를 함유하는 고충실도 SpCas9 변이체인 HF-Cas9을 개발하였으며, 이는 그의 DNA 절단 특이성을 크게 증가시킨 TALEN에서의 비-특이적 DNA 상호작용의 이전 폐기와 일치하는[25], Cas9와 DNA 표적 가닥의 포스페이트 백본 사이의 비-특이적 상호작용을 제거하도록 설계되었다 (도 125a)[20]. 염기 편집제가 Cas9에 의해 생성되는 단일 가닥 DNA 버블 내에서 비-표적 가닥 상에 작용하기 때문에[14], "HF-BE3"을 생성하기 위해 HF-Cas9에서 BE3으로 이들 4개의 점 돌연변이를 도입하는 것이 그의 염기 전환 능력을 변경시키지 않으면서 오프-타겟 염기 편집을 감소시킬 수 있는 것으로 가설화될 수 있다 (도 125a 및 125c).

[0829] His_6-태그부착된 단백질로서의 BE3 및 HF-BE3을 코딩하는 플라스미드를 이. 콜라이에서 과다발현시키고, 먼저 니켈 친화도 크로마토그래피에 이어서 양이온 교환 크로마토그래피에 의해 정제하였다 (도 126a-126b). 발현 및 정제 조건의 광범위한 최적화 후에, BE3 및 HF-BE3 단백질은 상용적으로 배양 배지 리터당 ~2 mg의 수율로 생성될 수 있다 (도 126a-126c).

[0830] 정제된 염기 편집제 단백질을 사용하여 염기 편집 효율 및 HF-BE3 및 BE3의 편집 윈도우의 폭을 생화학적으로 비교하였다. TC 반복부로 구성된 프로토스페이서를 갖는 합성 dsDNA 79량체에서 시험관내 C에서 U로의 전환 효율을 측정하였다. 표적 dsDNA (250 nM)를 30분 동안 37℃에서 BE3:sgRNA 또는 HF-BE3:sgRNA (2 μM)와 함께 인큐베이션하였다. 인큐베이션 후, 편집된 DNA를 우라실-내성 폴리머라제를 사용하여 증폭시키고, 고처리량 DNA 서열분석 (HTS)에 의해 서열분석하였다. 시험관내에서 HF-BE3 및 BE3에 대한 비교 편집 효율 및 활성 윈도우 폭이 관찰되었다 (도 125b). 이러한 발견은 고충실도 돌연변이를 BE3에 도입하는 것이 내재된 온-타겟 염기 편집 효율을 손상시키지 않거나, 시험관내에서 생성된 HF-BE3 단백질의 편집 윈도우의 폭을 변화시킨다는 것을 나타낸다.

[0831] HF-BE3은 DNA 형질감염 후 편집 특이성을 증진시킨다

[0832] 다음으로, 플라스미드 DNA 형질감염 후에 포유동물 세포에서의 BE3 및 HF-BE3의 염기 편집 효율, 특이성 및 편집 윈도우 폭을 비교하였다. 4개의 널리 연구된 내인성 게놈 유전자좌 (HEK293 부위 3, FANCF, EMX1 및 VEGFA 부위 2)를 포유동물 세포에서 온- 및 오프-타겟 염기 편집을 조사하기 위해 선택하였다[14,24]. VEGFA 부위 2는 20개의 프로토스페이서 뉴클레오티드 중 14개 C를 함유하여 고도로 반복적이고, 상당히 높은 비율의 공지된 오프-타겟 게놈 편집과 연관된다[20,22,24,26]. 이 부위는 그것이 강력한 특이성 문제를 제기하기 때문에 포함되도록 선택되었다. 대부분의 뉴클레아제-기반 게놈 편집 적용과는 대조적으로, 염기 편집은 프로토스페이서의 정확한 위치에 의존하여 편집 윈도우 내에 표적 뉴클레오티드를 위치시키고, 통상적으로 가이드 RNA의 선택에서 이용가능한 가요성이 거의 또는 전혀 없다. 따라서, 고도로 반복적인 혼재성 sgRNA 표적에 대해서도 증진된 특이성을 갖는 염기 편집제의 개발이 중요하다[3,14].

[0833] 온-타겟 유전자좌 및 공지된 오프-타겟 유전자좌를 PCR에 의해 증폭시키고, 4개의 염기 편집제:sgRNA 쌍 각각으로의 플라스미드 형질감염[24] 후 HTS에 의해 분석하였다. 이들 4개의 내인성 게놈 유전자좌에 대한 HEK293T 세포에서의 온-타겟 편집은 HF 돌연변이의 도입에 의해 약간 감소되었고; 편집은 BE3에서 29±5%, 및 HF-BE3에서 21±3%로 평균내어졌다 (n = 3개의 생물학적 반복에 대해 평균 ± s.e.m.) (도 127a-127d, 128a).

[0834] 3개의 표준, 비-반복적 표적 부위 (HEK293 부위 3, FANCF, 및 EMX1) 각각에 대해, Cas9 및 동일한 가이드 RNA를 사용한 처리로부터 변형된 것으로 이전에 보고된 오프-타겟 유전자좌로부터의 편집 윈도우 내의 C를 함유하는 3개의 가장 빈번하게 변형된 오프-타겟 유전자좌를 시험하였다 (표 15)[24]. 세포를 BE3 플라스미드로 형질감염시킨 경우에, HEK293 부위 3, FANCF, 및 EMX1에 대한 9개의 가장 빈번하게 변형된 Cas9 오프-타겟 유전자좌에 걸친 C→T 전환은 1.1±0.3%로 평균내어졌다 (도 127a-c; n = 3개의 생물학적 반복에 대해 평균 ± s.d.). HF 돌연변이의 설치는 평균 오프-타겟 편집의 절대 수준을 0.03±0.005%로 37배만큼 감소시켰고, 단 1개의 경우의 측정가능한 오프-타겟 C→T 전환을 갖는다 (도 127a; 오프-타겟 1에서 EMX1 C_5).

부위	서열	SEQ ID NO	GUIDE-Seq 카운트
EMX1 온-타겟	GAGTC$_5$C$_6$GAGCAGAAGAAGAA**GGG**	480	4,521
EMX1 오프-타겟 1	GAGTC$_5$TAAGCAGAAGAAGAA**GAG**	481	1,445
EMX1 오프-타겟 2	GAGGC$_3$C$_6$GAGCAGAAGAAAGA**CGG**	482	700
EMX1 오프-타겟 3	GAGTC$_5$C$_6$TAGCAGGAGAAGAA**GAG**	483	390
HEK293 부위 3 온-타겟	GGCC$_4$C$_5$AGACTGAGCACGTGA**TGG**	484	2,074
HEK293 부위 3 오프-타겟 1	CACC$_4$C$_5$AGACTGAGCACGTGC**TGG**	485	327
HEK293 부위 3 오프-타겟 2	GACAC$_5$AGACTGGGCACGTGA**GGG**	486	306
HEK293 부위 3 오프-타겟 3	AGCTC$_5$AGACTGAGCAAGTGA**GGG**	487	136
VEGFA 부위 2 온-타겟	GAC$_3$C$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCCCGCCTCC**GG**	488	540
VEGFA 부위 2 오프-타겟 1	CTAC$_4$C$_5$C$_6$C$_7$TC$_9$C$_{10}$ACCCCGCCTC**CG** **G**	489	1,925
VEGFA 부위 2 오프-타겟 2	ATTC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCGCCTCA**G** **G**	490	1,549
VEGFA 부위 2 오프-타겟 3	ACAC$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCGCCTC**A** **GG**	491	1,178
VEGFA 부위 2 오프-타겟 4	TGC$_3$C$_4$C$_5$C$_6$C$_7$C$_8$C$_9$C$_{10}$ACCCCACCTC**T** **GG**	492	1,107
FANCF 온-타겟	GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACC**TGG**	493	4,816
FANCF 오프-타겟 1	GGAAC$_5$C$_6$C$_7$C$_8$GTC$_{11}$TGCAGCACCA**G** **G**	494	2,099
FANCF 오프-타겟 2	GGAGTC$_6$C$_7$C$_8$TC$_{10}$C$_{11}$TACAGCACCA**G** **G**	495	524
FANCF 오프-타겟 3	AGAGGC$_6$C$_7$C$_8$C$_9$TC$_{11}$TGCAGCACCA**G** **G**	496	150

[0835]

[0836] 표 15: 본 작업에서 연구된 온- 및 오프-타겟 인간 게놈 유전자좌에 대한 프로토스페이서 및 PAM 서열. 오프-타겟 부위를 그의 GUIDE-Seq 판독 카운트에 기초하여 선택하였다[45]. 특정한 sgRNA에 대한 편집 윈도우 내의 시토신을 넘버링하였다. PAM 서열은 볼드체로 나타나 있다. 그의 각각의 온-타겟 유전자좌와 상이한 오프-타겟 유전자좌 내의 프로토스페이서 염기가 밑줄표시되어 있다. 게놈 서열에 대해 뮤린 샘플에서 조사하였다 (도 132e 참조).

[0837] 극도로 어려운 부위에 대한 HF-BE3 특이성을 특징화하기 위해, 고도로 반복적인 VEGFA 부위 2 유전자좌를 표적화하는 경우에 BE3 및 HF-BE3 오프-타겟 활성을 비교하였다. BE3 처리는 이러한 sgRNA와 연관된 4개의 시험된 오프-타겟 부위의 활성 윈도우에 위치한 시토신의 평균 15±5% 편집을 야기한다 (이 단락에서 인용된 모든 평균 값은 n = 3개의 생물학적 반복에 대해 평균 ± s.d.를 나타냄). 대조적으로, HF-BE3은 동일한 오프-타겟 부위에서 절대 오프-타겟 편집 (5.0±2.3%)의 3배 감소를 생성한다 (도 127d). BE3의 형질감염과 비교한 경우에, HF-BE3은 유의하게 ($p < 0.05$, 양측 스튜던트 t 검정) 오프-타겟 유전자좌에 위치한 57개의 시토신 중 27개에서 오프-타겟 편집을 감소시킨 반면 (표 16), HF-BE3 처리는 16개의 조사된 온-타겟 시토신 잔기 중 오직 3개에서 온-타겟 편집의 유의한 감소 ($p < 0.05$ 양측 스튜던트 t 검정)를 야기하였다.

유전자좌 및 시토신 위치	열거된 처리 사이의 비교를 위한 P 값 (스튜던트 양측 t-검정)									
	플라스미드 BE3 vs 플라스미드 HF-BE3	플라스미드 BE3 vs 단백질 BE3	플라스미드 BE3 vs 단백질 HF-BE3	플라스미드 BE3 vs 대조군	플라스미드 HF-BE3 vs 단백질 BE3	플라스미드 HF-BE3 vs 단백질 HF-BE3	플라스미드 HF-BE3 vs 대조군	단백질 BE3 vs 단백질 HF-BE3	단백질 BE3 vs 대조군	단백질 HF-BE3 vs 대조군
EMX1, C5	0.053	0.000	0.001	0.000	0.416	0.056	0.003	0.023	0.000	0.004
EMX1, C6	0.065	0.000	0.000	0.000	0.445	0.023	0.004	0.001	0.000	0.003
FANCF, C6	0.152	0.017	0.003	0.000	0.137	0.003	0.000	0.002	0.000	0.004
FANCF, C7	0.591	0.554	0.007	0.000	0.914	0.007	0.000	0.008	0.000	0.006
FANCF, C8	0.011	0.026	0.004	0.000	0.958	0.018	0.000	0.023	0.000	0.007
FANCF, C11	0.524	0.948	0.019	0.001	0.363	0.010	0.000	0.010	0.000	0.021
HEK 부위 3, C3	0.061	0.001	0.071	0.000	0.002	0.002	0.005	0.00199	0.001	0.003
HEK 부위 3, C4	0.048	0.924	0.010	0.004	0.001	0.001	0.001	0.00004	0.000	0.001
HEK 부위 3, C5	0.291	0.592	0.016	0.006	0.243	0.243	0.002	0.00022	0.000	0.001
VEGFA 부위 2, C3	0.060	0.416	0.239	0.010	0.042	0.280	0.002	0.475	0.002	0.018
VEGFA 부위 2, C4	0.036	0.191	0.047	0.004	0.032	0.803	0.002	0.066	0.000	0.005
VEGFA 부위 2, C5	0.098	0.650	0.028	0.003	0.044	0.169	0.002	0.004	0.000	0.001
VEGFA 부위 2, C6	0.452	0.781	0.118	0.004	0.165	0.239	0.002	0.013	0.000	0.001
VEGFA 부위 2, C7	0.401	0.683	0.172	0.003	0.120	0.454	0.002	0.026	0.000	0.001
VEGFA 부위 2, C9	0.225	0.308	0.254	0.004	0.504	0.828	0.004	0.624	0.000	0.002
VEGFA 부위 2, C10	0.064	0.061	0.023	0.005	0.732	0.257	0.009	0.057	0.000	0.003
EMX1, C5 오프-타겟 1	0.003	0.003	0.002	0.002	0.119	0.036	0.035	0.269	0.255	0.643
EMX1, C5 오프-타겟 2	0.013	0.013	0.013	0.013	0.158	0.294	0.521	0.390	0.058	0.054
EMX1, C6 오프-타겟 2	0.024	0.024	0.024	0.024	0.285	0.560	0.954	0.420	0.103	0.306
EMX1, C5 오프-타겟 3	0.022	0.019	0.019	0.019	0.297	0.297	0.300	>0.99999	0.882	0.815
EMX1, C6 오프-타겟 3	0.017	0.015	0.015	0.015	0.296	0.296	0.328	>0.99999	0.051	0.025
FANCF, C5 오프-타겟 1	0.031	0.031	0.031	0.031	0.314	0.530	0.333	0.337	0.349	0.618
FANCF, C6 오프-타겟 1	0.016	0.016	0.016	0.016	0.347	0.786	0.930	0.338	0.344	0.678
FANCF, C7 오프-타겟 1	0.028	0.028	0.028	0.027	0.374	0.039	0.106	0.353	0.346	0.639
FANCF, C8 오프-타겟 1	0.014	0.014	0.014	0.014	0.341	0.932	0.685	0.343	0.318	0.605
FANCF, C11 오프-타겟 1	0.007	0.007	0.007	0.007	0.374	0.001	0.000	>0.99999	0.475	0.016
FANCF, C6 오프-타겟 2	0.099	0.099	0.032	0.036	0.599	0.475	0.912	0.914	0.393	0.060
FANCF, C7 오프-타겟 2	0.080	0.080	0.027	0.030	0.898	0.638	0.819	0.539	0.859	0.530
FANCF, C8 오프-타겟 2	0.123	0.123	0.045	0.050	0.789	0.538	0.960	0.047	0.539	0.252
FANCF, C10 오프-타겟 2	0.093	0.093	0.029	0.033	0.630	0.509	0.847	0.768	0.670	0.482
FANCF, C11 오프-타겟 2	0.264	0.264	0.127	0.107	0.599	0.658	0.326	>0.99999	0.047	0.345
FANCF, C6 오프-타겟 3	0.872	0.872	0.492	0.108	0.239	0.493	0.129	0.469	0.584	0.252
FANCF, C7 오프-타겟 3	>0.99999	>0.99999	0.859	0.016	0.537	0.866	0.116	0.572	0.272	0.595
FANCF, C8 오프-타겟 3	0.886	0.886	0.246	0.757	>0.99999	0.001	0.648	0.495	0.780	0.650
FANCF, C10 오프-타겟 3	0.566	0.566	0.284	0.202	0.053	0.387	0.260	0.453	0.913	0.541
FANCF, C11 오프-타겟 3	0.422	0.422	0.145	0.145	0.495	0.230	0.230	0.658	0.658	>0.99999
HEK293 부위 3, C3 오프-타겟 1	>0.99999	0.910	0.412	0.326	0.910	0.412	0.326	0.293	0.223	0.480
HEK293 부위 3, C4 오프-타겟 1	>0.99999	0.994	0.437	0.391	0.994	0.437	0.391	0.495	0.451	0.614
HEK293 부위 3, C5 오프-타겟 1	>0.99999	0.616	0.814	0.337	0.616	0.814	0.337	0.459	0.116	0.481
HEK293 부위 3, C3 오프-타겟 2	0.285	0.473	0.100	0.473	0.141	0.735	0.141	0.038	>0.99999	0.038
HEK293 부위 3, C5 오프-타겟 2	0.375	0.294	0.177	0.294	0.687	0.428	0.687	0.064	>0.99999	0.064
HEK293 부위 3, C9 오프-타겟 2	0.053	0.624	0.374	0.624	0.554	0.154	0.554	0.872	>0.99999	0.872
HEK293 부위 3, C3 오프-타겟 3	0.067	0.116	0.768	0.435	0.519	0.230	0.561	0.349	0.768	0.643
HEK293 부위 3, C5 오프-타겟 3	0.011	0.011	0.011	0.011	0.016	0.643	0.435	0.184	0.025	0.346
HEK293 부위 3, C9 오프-타겟 3	>0.99999	0.374	0.652	0.811	0.132	0.539	0.776	0.609	0.643	0.893
VEGFA 부위 2, C4 오프-타겟 1	0.101	0.015	0.014	0.012	0.041	0.032	0.025	0.117	0.012	0.001
VEGFA 부위 2, C5 오프-타겟 1	0.060	0.013	0.012	0.010	0.078	0.062	0.044	0.201	0.009	0.012
VEGFA 부위 2, C6 오프-타겟 1	0.019	0.005	0.005	0.004	0.080	0.062	0.045	0.087	0.002	0.012
VEGFA 부위 2, C7 오프-타겟 1	0.017	0.004	0.004	0.003	0.080	0.060	0.037	0.076	0.001	0.002
VEGFA 부위 2, C9 오프-타겟 1	0.230	0.088	0.037	0.011	0.667	0.256	0.051	0.134	0.004	0.007
VEGFA 부위 2, C10 오프-타겟 1	0.535	0.136	0.106	0.035	0.283	0.211	0.050	0.717	0.028	0.010
VEGFA 부위 2, C4 오프-타겟 2	0.038	0.004	0.003	0.003	0.087	0.051	0.048	0.063	0.048	0.134
VEGFA 부위 2, C5 오프-타겟 2	0.033	0.004	0.004	0.004	0.078	0.061	0.059	0.028	0.020	0.248
VEGFA 부위 2, C6 오프-타겟 2	0.026	0.005	0.005	0.004	0.051	0.038	0.038	0.043	0.038	0.783
VEGFA 부위 2, C7 오프-타겟 2	0.053	0.006	0.005	0.005	0.072	0.056	0.055	0.078	0.064	0.704
VEGFA 부위 2, C8 오프-타겟 2	0.071	0.006	0.006	0.006	0.079	0.065	0.065	0.118	0.107	0.703
VEGFA 부위 2, C9 오프-타겟 2	0.193	0.008	0.007	0.006	0.103	0.090	0.084	0.068	0.007	0.217
VEGFA 부위 2, C10 오프-타겟 2	0.063	0.003	0.003	0.002	0.116	0.107	0.090	0.545	0.016	0.346
VEGFA 부위 2, C4 오프-타겟 3	0.005	0.003	0.003	0.003	0.091	0.031	0.030	0.116	0.107	0.158
VEGFA 부위 2, C5 오프-타겟 3	0.011	0.007	0.005	0.005	0.211	0.048	0.042	0.220	0.177	0.001
VEGFA 부위 2, C6 오프-타겟 3	0.020	0.005	0.003	0.003	0.142	0.038	0.033	0.193	0.149	0.015
VEGFA 부위 2, C7 오프-타겟 3	0.045	0.006	0.003	0.003	0.101	0.035	0.030	0.093	0.060	0.083
VEGFA 부위 2, C8 오프-타겟 3	0.069	0.007	0.005	0.005	0.087	0.045	0.039	0.120	0.067	0.041
VEGFA 부위 2, C9 오프-타겟 3	0.093	0.006	0.005	0.005	0.041	0.032	0.028	0.396	0.195	0.005
VEGFA 부위 2, C10 오프-타겟 3	0.342	0.011	0.008	0.007	0.109	0.081	0.069	0.273	0.098	0.036
VEGFA 부위 2, C3 오프-타겟 4	0.001	0.001	0.001	0.001	0.374	0.374	0.230	0.271	0.358	0.633
VEGFA 부위 2, C4 오프-타겟 4	0.007	0.006	0.006	0.006	0.137	0.137	0.137	0.592	0.862	0.690
VEGFA 부위 2, C5 오프-타겟 4	0.007	0.007	0.007	0.007	0.026	0.017	0.018	0.461	0.655	0.279
VEGFA 부위 2, C6 오프-타겟 4	0.005	0.004	0.004	0.004	0.021	0.018	0.018	0.398	0.546	0.149
VEGFA 부위 2, C7 오프-타겟 4	0.007	0.006	0.006	0.006	0.051	0.048	0.050	0.373	0.720	0.029
VEGFA 부위 2, C8 오프-타겟 4	0.007	0.006	0.006	0.006	0.092	0.092	0.092	0.325	0.014	0.275
VEGFA 부위 2, C9 오프-타겟 4	0.016	0.007	0.007	0.007	0.150	0.150	0.150	0.502	1.000	0.615
VEGFA 부위 2, C10 오프-타겟 4	0.213	0.009	0.009	0.009	0.261	0.261	0.261	0.653	0.575	0.660

[0838]

[0839] 표 16: 본 연구에서 평가된 모든 유전자좌에서의 상이한 처리 조건 하의 염기 편집의 차이에 대한 P-값. 물질 및 방법에 기재된 바와 같은 스튜던트 양측 t-검정을 사용하여 p-값을 계산하였다. p-값이 유의한 차이를 나타내는 경우에 ($p < 0.05$), 상응하는 엔트리가 강조되었다.

[0840] 오프-타겟 유전자좌에서의 절대 편집 사이의 차이를 고려하는 것 이외에도, 관찰된 온-타겟 효율을 오프-타겟 효율로 나누어 온-타겟:오프-타겟 편집 특이성 비를 또한 계산하였다 (도 129a-129b). 이러한 메트릭은 HF-돌연변이의 설치와 연관된 온-타겟 편집의 임의의 감소를 고려하고, 염기 편집의 효율 및 특이성 둘 다에 대해 감수성인 적용에 유용하다. HF-BE3에 의한 오프-타겟 편집은 여러 오프-타겟 유전자좌에 대한 고처리량 서열분석의 검출 한계 미만이었다. 이들 경우에 대해, 검출 상한과 동일한 보존적 오프-타겟 편집 효율을 가정하였다 (0.025% C→T 전환; 방법 참조). 이러한 분석에 기초하여, 본원에서 조사된 모든 34개의 표적 시토신에 걸친 HF 돌연변이의 설치시 특이성 비율의 평균 개선은 2개의 구축물의 플라스미드 전달이 수행된 경우에 19배였다. 이러한 결과는 비-반복적 부위 (도 129a) 뿐만 아니라 고도로 반복적인 부위 (도 129b)에 대해, HF-BE3이 BE3과 비교하여 온-타겟 편집 효율의 보통의 감소만을 가지면서 실질적으로 증진된 염기 편집 특이성을 발생시킨다는 것을 집합적으로 확립한다.

[0841] BE3의 RNP 전달은 DNA-무함유 염기 편집을 가능하게 한다

[0842] 다음으로, 배양된 인간 세포 내로 직접 전달된 경우에 염기 편집을 매개하는 DNA-무함유, RNP 형태의 BE3의 능력을 연구하였다. 최근에 양이온성 지질 시약이 음으로 하전된 단백질 또는 단백질:핵산 복합체를 리보핵단백질 (RNP) 복합체를 포함하는 포유동물 세포 내로 강력하게 전달할 수 있고, RNP 전달은 오프-타겟 게놈 편집을 실질적으로 감소시킬 수 있는 것으로 확립되었다[27-29].

[0843] 상업적으로 입수가능한 양이온성 지질 리포펙타민 2000을 EMX1, HEK293 부위 3, FANCF, 또는 VEGFA 부위 2 유전자좌를 표적화하는 가이드 RNA와 사전-복합체화한 후 정제된 BE3 단백질 또는 HF-BE3 단백질과 조합하고, 생성된 지질:RNP 복합체를 HEK293T 세포와 함께 인큐베이션하였다. 72시간 후, 게놈 DNA를 수거하고, 온-타겟 및 오프-타겟 염기 편집을 고처리량 DNA 서열분석에 의해 분석하였다. 모든 Cas9-기반 기술에서와 같이, 실질적인 변이가 상이한 게놈 유전자좌에서의 편집 효율에서 관찰되었다 (도 127 및 130). 상이한 처리 사이의 온-타겟 편집 효율과 연관된 경향을 나타내기 위해, 평균 온-타겟 염기 편집 효율을 4개의 시험된 유전자좌에서 계산하였다 (도 128a). BE3 (200nM)의 단백질 전달은 플라스미드 형질감염에서 관찰된 것과 대등한 온-타겟 편집 효율을 생성한다 (각각 26±4% 대 29±5%; n = 3개의 생물학적 반복에 대해 평균 ± s.e.m.; 도 128a).

[0844] 대조적으로, HF-BE3의 단백질 전달은 연구된 4개의 게놈 유전자좌에서의 BE3의 단백질 전달과 비교하여 온-타겟 편집을 감소시켰다 (각각 13±3% 대 26±4%의 평균 편집 효율; n = 3의 생물학적 반복에 대해 평균 ± s.e.m.; 도 128a). HF-BE3 및 BE3은 시험 튜브에서 대등한 편집 효율을 갖고 (도 125b), 편집은 HF-BE3이 HEK293T 세포에서 플라스미드로부터 발현되는 경우에 단지 약간 감소되기 때문에 (도 127a-d), HF-BE3 단백질 전달로부터의 편집의 감소된 효율이 포유동물 세포에서의 감소된 HF-BE3 안정성의 결과일 수 있다는 것을 추측하는 것이 그럴듯하다. 보다 낮은 안정성은 플라스미드로부터의 연속적인 발현에 의해 오프셋될 수 있지만, 1회 단백질 전달 후에는 그렇지 않다. 이러한 관찰은 $CD34^+$ 조혈 줄기 및 전구 세포로 뉴클레오펙션된 경우에 정제된 Cas9와 비교하여 정제된 HF-Cas9를 사용한 감소된 온-타겟 indel 형성의 최근 보고와 일치하였다[30]. 본 연구를 검토하는 동안, 킴(Kim) 등은 전기천공법을 사용하여 BE3의 마우스 배아로의 RNP 전달을 입증하였다[31]. 본 발명자들의 지식이 미치는 한, 본 접근법은 전기천공 없이 포유동물 세포에서의 개별 뉴클레오티드에 대한 정확한 변화를 생성할 수 있는 제1 DNA-무함유 기술이고, 이는 제한된 생체내 치료적 관련성을 갖는다.

[0845] 염기 편집제의 RNP 전달은 DNA 특이성을 크게 증진시킨다

[0846] 중요하게는, BE3 및 HF-BE3의 RNP 전달이 실질적인 온-타겟 염기 편집을 야기하는 반면, 어떠한 경우의 측정가능한 염기 편집 (< 0.025%)도 EMX1, FANCF 및 HEK293 부위 3과 연관된 9개의 시험된 오프-타겟 유전자좌 중 임의의 것에서 관찰되지 않았다 (도 130a-130c). 대조적으로, BE3의 플라스미드 전달은 염기 편집 활성 윈도우 내의 모든 서열분석된 시토신에 걸쳐 평균 1.1±0.3% (n = 3개의 생물학적 반복에 대해 평균 ± s.d.) 오프-편집 및, 이들 9개의 오프-타겟 유전자좌에 위치한 16개 오프-타겟 시토신 중 11개에서 검출가능한 오프-타겟 편집을 야기한다 (도 127a-127d). 시험된 3개의 비-반복적 유전자좌의 오프-타겟 유전자좌에서, BE3 단백질 전달은 플라스미드 전달의 것보다 26배 더 높은 평균 특이성 비를 야기한다 (도 127a). 이러한 결과는 염기 편집제의 RNP 전달이 염기 편집의 DNA 특이성을 극적으로 증가시킨다는 것을 밝혀냈다.

[0847] BE3 또는 HF-BE3의 단백질 전달은 또한 BE3 또는 HF-BE3의 플라스미드 전달과 비교하여 고도로 혼재성인 VEGFA 부위 2 유전자좌에서 크게 개선된 염기 편집 특이성을 발생시켰다 (도 127 및 130 참조; 표 16 참조). 이러한 부위와 연관된 오프-타겟 유전자좌에서의 염기 편집의 절대 빈도는 BE3 (플라스미드 전달: 15±4% 오프-타겟 편집; 단백질 전달: 1.3±0.4% 오프-타겟 편집; 이 단락에서의 모든 값은 n = 3개의 생물학적 반복에 대해 평균 ± s.d.를 나타냄) 및 HF-BE3 (플라스미드 전달: 5±2% 오프-타겟 편집; 단백질 전달: 0.5±0.1% 오프-타겟 편집) 둘 다에 대해 단백질 전달시 적어도 10배 감소되었다. 모든 4개의 연구된 유전자좌에 걸쳐, 온-타겟:오프-타겟 편집에 대한 염기 편집 특이성 비는 BE3의 플라스미드 전달과 비교하여 BE3의 단백질 전달에 대해 평균 66배 증가하였다 (도 129). 집합적으로, 이들 결과는 반복적 및 비-반복적 표적 부위 둘 다에 대해, DNA 전달에 비해 RNP 전달이 고충실도 Cas9 돌연변이의 존재 또는 부재보다 염기 편집 특이성의 더 강한 결정인자라는 것을 밝혀냈다.

[0848] HF 돌연변이의 도입도 전달 방법도 염기 편집과 연관된 낮은 indel 비율을 실질적으로 변경시키지 않았다. 본 연구에서 모든 처리 조건에 걸친 모든 온-타겟 유전자좌에서의 indel 빈도는 낮게 유지되었고 (전형적으로 ≤ 5%; 도 131a), 편집:indel 비는 최적화된 HDR 프로토콜을 사용한 이전 연구에서보다 시험된 모든 경우에서 더

높게 유지되었다 (전형적으로 ≥ 10-배; 도 131b)[30,32,33]. 비-반복적 sgRNA의 경우, BE3의 플라스미드 전달이 VEGFA 부위 2와 연관된 오프-타겟 유전자좌에 대해 최대 5% indel을 생성하였다는 것에 주목하더라도 (도 131c), 매우 적은 indel이 오프-타겟 유전자좌에서 관찰되었다 (도 131c).

[0849] 종합하면, 이들 결과는 염기 편집제의 단백질 전달이 온-타겟 염기 편집 효율을 유지하고 플라스미드 DNA의 전달에 비해 편집 특이성을 크게 증진시킨다는 것을 확립한다.

[0850] RNP 전달은 온- 및 오프-타겟 편집을 탈커플링시킨다

[0851] BE3의 단백질 전달과 연관된 염기 편집 특이성의 현저한 증진을 고려할 때, 이러한 개선이 세포 내로 전달된 활성 게놈 편집제의 총량을 감소시키는지 조사하였다. EMX1을 표적화하는 sgRNA를 사용하여, 플라스미드 (도 128b) 및 단백질 전달 (도 128c)에 대한 용량 반응 연구를 수행하였다. 처리 조건 사이의 형질감염 효율을 최대화하기 위해, 리포펙타민 2000의 부피는 모든 시험에 대해 1.5 μL였고, 염기 편집제 단백질:sgRNA 몰비는 단백질 전달을 위해 1:1.1로 유지하였다. 플라스미드 전달을 위해, 1:3 (몰비 ~1:1)의 sgRNA 플라스미드:BE3 플라스미드의 질량 비 및 1.5 μL의 리포펙타민 2000을 사용하였다. 오프-타겟 염기 편집은 플라스미드 전달에 대해 시험된 모든 조건 하에 관찰되었지만 (도 128b), 시험된 모든 단백질 전달 조건 하에서는 실질적으로 오프-타겟 편집이 관찰되지 않았다 (도 128c).

[0852] 플라스미드 및 단백질 전달에 대한 온- 및 오프-타겟 편집 사이의 관계를 평가하기 위해 선형 회귀 분석을 수행하였다. 플라스미드 전달을 위해, 오프-타겟 편집은 온-타겟 편집 비율와 밀접하게 연관되었지만 (비-0 기울기, F-시험에 대해 R^2 = 0.95, p = 0.0012), 단백질 전달을 사용한 오프-타겟과 온-타겟 편집 사이에 유의한 연관이 없었다 (비-0 기울기, F-시험에 대해 R^2 = 0.078, p = 0.59).

[0853] 이들 데이터는 염기 편집제의 단백질 전달이 투여량과는 독립적인 고유한 특이성 이점을 제공한다는 것을 나타낸다. 이전 관찰과 함께[29,34], 이러한 발견은 DNA 전달과 비교하여 단백질 전달로부터의 염기 편집의 보다 높은 DNA 특이성이 염기 편집제에 대한 게놈의 연장된 노출을 피함으로써, 온-타겟 유전자좌가 이미 변형된 후에 오프-타겟 유전자좌를 프로세싱하는 염기 편집제의 기회를 최소화하는 단백질 전달의 능력으로부터 발생한다는 모델을 지지한다.

[0854] 제브라피쉬 및 마우스에서의 DNA-무함유 염기 편집

[0855] 상기 관찰은 검출가능한 오프-타겟 염기 편집을 제거하면서 온-타겟 염기 편집을 유지하기 위한 BE3의 단백질 전달의 가능성을 제안하였다. 따라서, BE3의 단백질 전달이 수정된 제브라피쉬 배아 내로 티로시나제 유전자좌를 표적화하는 BE3:sgRNA 복합체를 주사함으로써 제브라피쉬 내의 특정 점 돌연변이를 생성하는 데 사용될 수 있는지 시험하였다. 주사 4일 후에 생성된 제브라피쉬 유충으로부터 게놈 DNA를 수거하고, 고처리량 서열분석에 의해 염기 편집 및 indel 빈도를 측정하였다 (도 132a). 시험된 3개의 BE3:sgRNA 복합체 중 2개는 생체내에서 실질적인 점 돌연변이를 유도하였다 (TYR1: $C_3 \rightarrow T_3$ 5.3±1.8%, TYR2: $C_4 \rightarrow T_4$ 4.3±2.1%; n = 3개의 주사된 배아의 평균 ± s.d.; 도 132a). 제브라피쉬 유전자좌의 서열이 표 17에 열거되어 있다.

HEK 세포 샘플

샘플 설명	반복 1	2	3	앰플리콘
단백질, BE3	237256	295609	159391	VEGFA 온 타겟
단백질, HF-BE3	383480	389874	383467	VEGFA 온 타겟
플라스미드, BE3	315213	280891	335668	VEGFA 온 타겟
플라스미드, HF-BE3	196373	251965	369201	VEGFA 온 타겟
대조군	390748	395523	353614	VEGFA 온 타겟
단백질, BE3	192805	26472	24799	FANCF 오프 타겟 부위 #1
단백질, HF-BE3	36383	30007	39193	FANCF 오프 타겟 부위 #1
플라스미드, BE3	35580	29557	22243	FANCF 오프 타겟 부위 #1
플라스미드, HF-BE3	40371	27187	28248	FANCF 오프 타겟 부위 #1
대조군	35106	37274	34939	FANCF 오프 타겟 부위 #1
단백질, BE3	82978	124689	83840	VEGFA 오프 타겟 부위 #1
단백질, HF-BE3	142404	140482	142220	VEGFA 오프 타겟 부위 #1
플라스미드, BE3	112027	117071	100894	VEGFA 오프 타겟 부위 #1
플라스미드, HF-BE3	114187	98876	122553	VEGFA 오프 타겟 부위 #1
대조군	76854	90547	89271	VEGFA 오프 타겟 부위 #1
단백질, BE3	14514	25515	19325	VEGFA 오프 타겟 부위 #2
단백질, HF-BE3	24678	24363	25312	VEGFA 오프 타겟 부위 #2
플라스미드, BE3	16945	19918	10225	VEGFA 오프 타겟 부위 #2
플라스미드, HF-BE3	12200	14769	17797	VEGFA 오프 타겟 부위 #2
대조군	8739	13648	7818	VEGFA 오프 타겟 부위 #2
단백질, BE3	179245	111693	173909	VEGFA 오프 타겟 부위 #3
단백질, HF-BE3	243899	300503	276139	VEGFA 오프 타겟 부위 #3
플라스미드, BE3	208476	291370	155430	VEGFA 오프 타겟 부위 #3
플라스미드, HF-BE3	117174	154033	199152	VEGFA 오프 타겟 부위 #3
대조군	119263	170436	121686	VEGFA 오프 타겟 부위 #3
단백질, BE3	237799	262947	185371	VEGFA 오프 타겟 부위 #4
단백질, HF-BE3	313253	233699	244922	VEGFA 오프 타겟 부위 #4
플라스미드, BE3	243094	230316	234421	VEGFA 오프 타겟 부위 #4
플라스미드, HF-BE3	170958	160091	140603	VEGFA 오프 타겟 부위 #4
대조군	158691	148720	137270	VEGFA 오프 타겟 부위 #4
단백질, BE3	28684	28237	43315	HEK3 온 타겟
단백질, HF-BE3	49300	42576	57690	HEK3 온 타겟
플라스미드, BE3	55008	55813	54310	HEK3 온 타겟
플라스미드, HF-BE3	55199	11384	7659	HEK3 온 타겟
대조군	63741	42878	48524	HEK3 온 타겟
단백질, BE3	104827	181792	161090	HEK3 오프 타겟 부위 #1
단백질, HF-BE3	204580	175561	177303	HEK3 오프 타겟 부위 #1
플라스미드, BE3	178694	152264	206863	HEK3 오프 타겟 부위 #1
플라스미드, HF-BE3	191257	138423	160789	HEK3 오프 타겟 부위 #1
대조군	190303	190061	190516	HEK3 오프 타겟 부위 #1
단백질, BE3	146089	95113	135015	HEK3 오프 타겟 부위 #2
단백질, HF-BE3	155947	136541	157991	HEK3 오프 타겟 부위 #2
플라스미드, BE3	150036	128438	158905	HEK3 오프 타겟 부위 #2
플라스미드, HF-BE3	371077	123642	142562	HEK3 오프 타겟 부위 #2
대조군	130322	134545	141833	HEK3 오프 타겟 부위 #2
단백질, BE3	145058	175338	161837	HEK3 오프 타겟 부위 #3
단백질, HF-BE3	212337	178893	179887	HEK3 오프 타겟 부위 #3
플라스미드, BE3	186452	166500	80441	HEK3 오프 타겟 부위 #3
플라스미드, HF-BE3	163732	118453	134719	HEK3 오프 타겟 부위 #3
대조군	131461	134470	155608	HEK3 오프 타겟 부위 #3
단백질, BE3	41986	61678	67890	FANCF 온 타겟
단백질, HF-BE3	41057	55850	86411	FANCF 온 타겟
플라스미드, BE3	39114	48575	70074	FANCF 온 타겟
플라스미드, HF-BE3	41617	53638	75718	FANCF 온 타겟
대조군	68852	59422	81265	FANCF 온 타겟
단백질, BE3	113462	80529	191344	FANCF 오프 타겟 부위 #1
단백질, HF-BE3	202662	233981	203024	FANCF 오프 타겟 부위 #1
플라스미드, BE3	208912	202044	107234	FANCF 오프 타겟 부위 #1
플라스미드, HF-BE3	86494	113989	86807	FANCF 오프 타겟 부위 #1
대조군	92255	72386	56661	FANCF 오프 타겟 부위 #1
단백질, BE3	96271	117442	84374	FANCF 오프 타겟 부위 #2
단백질, HF-BE3	105624	102312	105343	FANCF 오프 타겟 부위 #2
플라스미드, BE3	101002	98747	70052	FANCF 오프 타겟 부위 #2
플라스미드, HF-BE3	108966	69787	83184	FANCF 오프 타겟 부위 #2
대조군	99986	100344	100659	FANCF 오프 타겟 부위 #2
단백질, BE3	255748	182451	65388	FANCF 오프 타겟 부위 #3
단백질, HF-BE3	71858	75553	71785	FANCF 오프 타겟 부위 #3
플라스미드, BE3	60980	57360	78166	FANCF 오프 타겟 부위 #3
플라스미드, HF-BE3	68316	34659	85718	FANCF 오프 타겟 부위 #3
대조군	49685	57388	60418	FANCF 오프 타겟 부위 #3
단백질, BE3	46981	90793	79439	EMX1 온 타겟
단백질, HF-BE3	58629	71186	61575	EMX1 온 타겟
플라스미드, BE3	70817	82736	75706	EMX1 온 타겟
플라스미드, HF-BE3	77038	71123	78511	EMX1 온 타겟
대조군	62183	48574	68439	EMX1 온 타겟
단백질, BE3	165905	257565	142888	EMX 1 오프 타겟 부위 # 1
단백질, HF-BE3	148339	151300	130712	EMX 1 오프 타겟 부위 # 1
플라스미드, BE3	101950	103226	203004	EMX 1 오프 타겟 부위 # 1
플라스미드, HF-BE3	167969	175193	97010	EMX 1 오프 타겟 부위 # 1
대조군	101476	150435	102327	EMX 1 오프 타겟 부위 # 1
단백질, BE3	136738	213438	118711	EMX 1 오프 타겟 부위 #2
단백질, HF-BE3	123413	126114	109375	EMX 1 오프 타겟 부위 #2
플라스미드, BE3	85576	86600	169592	EMX 1 오프 타겟 부위 #2
플라스미드, HF-BE3	140317	145738	137050	EMX 1 오프 타겟 부위 #2
대조군	84818	125139	85454	EMX 1 오프 타겟 부위 #2
단백질, BE3	11940	36593	24946	EMX1 오프 타겟 부위 #3
단백질, HF-BE3	26762	31566	36377	EMX1 오프 타겟 부위 #3
플라스미드, BE3	32420	21547	14659	EMX1 오프 타겟 부위 #3
플라스미드, HF-BE3	31427	16592	17385	EMX1 오프 타겟 부위 #3
대조군	17385	28128	32717	EMX1 오프 타겟 부위 #3

NIH 3T3 세포 처리로부터의 유전 샘플

샘플 설명	반복 1	2	3	앰플리콘
플라스미드, BE3	16641	102216	463611	온 타겟 VEGFA (Mus)
플라스미드, HF-BE3	89330	1126543	100993	온 타겟 VEGFA (Mus)
단백질, BE3	88998	81697	511242	온 타겟 VEGFA (Mus)
단백질, HF-BE3	1282181	29193	1315115	온 타겟 VEGFA (Mus)
대조군	187677	38866	58985	온 타겟 VEGFA (Mus)
플라스미드, BE3	174782	1675041	1825653	CFD 오프 타겟 1
플라스미드, HF-BE3	1671201	1825201	1923895	CFD 오프 타겟 1
단백질, BE3	2305691	2126051	1381441	CFD 오프 타겟 1
단백질, HF-BE3	2286681	2114571	1833701	CFD 오프 타겟 1
대조군	1717381	1911171	2087951	CFD 오프 타겟 1
플라스미드, BE3	206475	227332	2060891	CFD 오프 타겟 2
플라스미드, HF-BE3	2138091	2030281	1990781	CFD 오프 타겟 2
단백질, BE3	2159951	2757541	2499691	CFD 오프 타겟 2
단백질, HF-BE3	2509181	2720631	2410591	CFD 오프 타겟 2
대조군	1937601	1759591	2469631	CFD 오프 타겟 2
플라스미드, BE3	603881	1262781	732851	CFD 오프 타겟 3
플라스미드, HF-BE3	890451	1285081	517851	CFD 오프 타겟 3
단백질, BE3	1671951	3300461	1116311	CFD 오프 타겟 3
단백질, HF-BE3	821201	3093521	1039331	CFD 오프 타겟 3
대조군	832041	1769391	566131	CFD 오프 타겟 3
플라스미드, BE3	1928461	1137091	1710781	CFD 오프 타겟 4
플라스미드, HF-BE3	2096011	1514341	1888941	CFD 오프 타겟 4
단백질, BE3	2718941	1819931	2083981	CFD 오프 타겟 4
단백질, HF-BE3	2119661	1489761	1868381	CFD 오프 타겟 4
대조군	1839331	1303181	1974761	CFD 오프 타겟 4

마우스 와우 샘플

샘플 설명	반복 1	2	3	앰플리콘
혈관조	378801	205706	620911	온 타겟 (VEGFA)
코르티 기관	1484471	1750041	2907531	온 타겟 (VEGFA)
와우축	1828061	1813821	612691	온 타겟 (VEGFA)
주사하지 않은 대조군	2282221	2419791	2727591	온 타겟 (VEGFA)
혈관조	44457	244487	2446461	CFD 오프 타겟 1
코르티 기관	1363351	1183181	347471	CFD 오프 타겟 1
와우축	677176	2095431	686991	CFD 오프 타겟 1
주사하지 않은 대조군	3431001	3427171	3790151	CFD 오프 타겟 1
혈관조	72962	1319881	2657931	CFD 오프 타겟 2
코르티 기관	1984561	1314301	600301	CFD 오프 타겟 2
와우축	920141	2511091	814131	CFD 오프 타겟 2
주사하지 않은 대조군	3991381	3459651	4839201	CFD 오프 타겟 2
혈관조	83251	803221	1425561	CFD 오프 타겟 3
코르티 기관	810141	459761	181031	CFD 오프 타겟 3
와우축	99281	755551	113411	CFD 오프 타겟 3
주사하지 않은 대조군	3991381	3459651	4839201	CFD 오프 타겟 3
혈관조	2321941	3977701	5540541	CFD 오프 타겟 4
코르티 기관	3134721	2853021	1768721	CFD 오프 타겟 4
와우축	2301051	3713991	2581421	CFD 오프 타겟 4
주사하지 않은 대조군	5245031	6379461	6247091	CFD 오프 타겟 4

제브라피쉬 샘플

	1	2	3	앰플리콘
처리된 제브라피쉬	723551	459581	810611	TYR 1
스크램블링된 sgRNA	1079191	985021	924291	TYR 1
처리된 제브라피쉬	514341	480141	415471	TYR 2
스크램블링된 sgRNA	614661	623741	667651	TYR 2
처리된 제브라피쉬	64871	572471	758831	TYR 3
스크램블링된 sgRNA	645961	712341	756241	TYR 3

마우스 와우 샘플 - 비공인된 sgRNA로 처리됨

샘플			앰플리콘
혈관조	코르티	와우축	
537459	249767	389274	온 타겟 (VEGFA)

[0856]

[0857] 표 17: 참조 서열에 정렬되고 물질 및 방법에 기재된 품질 필터를 통과하는 HTS 판독물의 수.

[0858] 마지막으로, 이러한 개발을 적용하여 마우스에서 DNA-무함유, 고-특이성 염기 편집을 달성하였다. 생체내에서 온- 및 오프-타겟 염기 편집을 관찰할 가능성을 최대화하기 위해, VEGFA 부위 2를 표적화하는 고도로 반복적인 sgRNA를 사용하였고; 편리하게는, 뮤린 및 인간 게놈은 이 표적 부위에서 동일하다.

[0859] 배양된 뮤린 NIH/3T3 세포를 사용하여, BE3 단백질 전달이 이 유전자좌에서 효율적인 온-타겟 염기 편집 34±11%를 생성하였다는 것이 확인되었다 (도 133a; 본 단락에서의 모든 편집 백분율은 n = 3개의 생물학적 반복에 대해 평균 ± s.d.를 나타냄). 절단 빈도 결정인자 (CFD) 알고리즘[29,34]을 사용하여 VEGFA 부위 2 sgRNA와 연관된 마우스 게놈 내의 오프-타겟 유전자좌를 예측하였다 (표 18). 배양된 NIH/3T3 세포를 사용하여, 상위 4개의 예측된 오프-타겟 유전자좌 중 2개가 실제로 배양된 뮤린 세포에서의 BE3의 플라스미드 전달에 의해 변형된 것으로 확인되었다 (CFD 오프-타겟 유전자좌 1, 9±5% 편집; 및 CFD 오프-타겟 유전자좌 4, 3±2% 편집; 도 133b-133e). 인간 세포로부터의 결과와 일치하게, BE3의 단백질 전달은 오프-타겟 편집을 음성 대조군과 유사한 수

준으로 감소시켰다 (도 133c 및 133e). CFD 오프-타겟 유전자좌 1 및 4에 대한 평균 염기 편집 특이성 비는 BE3의 플라스미드 전달의 경우 28±13에서 BE3의 단백질 전달의 경우 ≥ 780±300으로 증가하였다 (값은 평균 ± s.e.m.; n = 3개의 생물학적 반복을 나타냄).

[0860]

부위	서열	SEQ ID NO	CFD 점수	유전자좌의 설명
온-타겟	GACCCCCTCCACCCCGCCTCCGG	497		VEGFA 부위 2
오프-타겟 1	TCCCCCCTCCACCCCACCTCCGG	498	0.7857	유전자간:mmu-mir-21c-Nrp1/Mir1903
오프-타겟 2	TGCCCACCTCACCCCGCCTCTGG	499	0.65	인트론:Vipr1
오프-타겟 3	GCCCCTCCCAACCCCACCTCTGG	500	0.6323	인트론:Nos1ap
오프-타겟 4	CACCCCCCTCACCCCGCCTCAGG	501	0.625	유전자간:Unc5b-mmu-mir-6408

[0861] 표 18: VEGFA 부위 2 sgRNA와 연관된 마우스 게놈 내의 예측된 오프-타겟 유전자좌에 대한 프로토스페이서 및 PAM 서열. CFD 점수[46]는 CRISPOR를 사용하여 계산하였다[47]. 온-타겟 서열과 상이한 오프-타겟 프로토스페이서 내의 위치는 밑줄표시된다.

[0862] 마우스에서 DNA-무함유 염기 편집을 확립하기 위해, BE3:sgRNA 복합체를 리포펙타민 2000과 조합하고 (도 132b), 와우내 주사를 P1-P2에서 마우스 새끼에게 수행하였다. 주사된 와우 조직을 주사 3-4일 후에 수거하고, 와우 영역당 5-7개의 샘플로 미세 해부하였다. 주사되지 않은 마우스로부터의 대조군 와우를 동시에 수거하였다. 수거된 조직으로부터 게놈 DNA를 추출하고, qPCR에 의해 후기-지수기로 증폭시키고, C→T 전환을 측정하기 위해 고처리량 DNA 서열분석에 적용하였다. 염기 편집제 단백질에 노출된 세포로부터 독점적으로 DNA를 회수하는 것이 가능하지 않기 때문에 처리된 세포 중에서 염기 편집 효율을 정량화하는 것은 불가능하지만, 와우의 3개 영역에서의 조직으로부터 명백한 염기 편집이 관찰되었다: 코르티 기관의 기저 말단, 혈관조 및 와우축 (도 132c-132d). 처리된 조직 샘플에서 유의한 indel 형성이 검출되지 않았다 (< 0.1% indel; 도 134b).

[0863] 표적 C→T 전환을 함유하는 와우 세포의 백분율 (도 132c)은 배양물 내의 처리된 NIH/3T3 세포에서 관찰된 것보다 유의하게 더 낮았고 (도 133a), 이는 지질-기반 단백질 전달의 고도로 국재화된 성질 및 염기 편집제에 노출된 세포로부터 독점적으로 DNA를 단리할 수 없다는 것과 일치한다. 그럼에도 불구하고, 국부 전달은 세포 유형이 편집되는 제어, 및 제조 및 투여의 용이성을 포함한, 접근가능한 용도에 대한 주요 이점을 제공한다.

[0864] 마지막으로, BE3:sgRNA:지질 복합체의 와우내 주사 후 오프-타겟 편집을 분석하였다. BE3:VEGFA 부위 2 sgRNA:지질 복합체를 주사한 마우스의 와우 조직으로부터의 게놈 DNA에서 확인된 오프-타겟 부위 CFD 유전자좌 1 및 CFD 유전자좌 4를 포함한 모든 4개의 예측된 오프-타겟 유전자좌의 분석은 시험된 오프-타겟 유전자좌 중 임의의 것에 대해 비처리 대조군 샘플에서 관찰된 것 초과의 검출가능한 C→T 전환 또는 indel 형성을 나타내지 않았다 (도 134a).

[0865] 함께, 이들 생체내 염기 편집 결과는 높은 DNA 서열 특이성을 갖는 동물의 게놈 DNA에서의 개별 뉴클레오티드의 정확한 전환을 위한 바이러스-무함유, DNA-무함유 전략을 확립한다.

[0866] 논의

[0867] 본 연구에서 개발되고 실시되는 전략은 오프-타겟 염기 편집을 제거하거나 감소시키고, 생체내 염기 편집을 지지하는 DNA-무함유 전달 방법을 확립함으로써 염기 편집의 유용성 및 적용가능성을 확장시킨다. 단백질 전달은 동일한 구축물의 플라스미드 전달과 비교하여 인간 및 뮤린 세포에서 염기 편집 특이성을 개선시키고 (도 137,130, 및 133), 제브라피시 및 마우스 와우에서 특이적 염기 편집을 가능하게 한다 (도 132).

[0868] Cas9의 DNA 특이성을 증진시키는 것으로 공지된 BE3 돌연변이를 설치함으로써 고충실도 염기 편집제가 생성되었다[20]. 이러한 돌연변이의 Cas9 내로의 설치는, 여기서 조사된 EMX1 유전자좌를 포함한 비-반복적 sgRNA와 연관된 오프 타겟 유전자좌에에서 검출불가능한 indel 형성을 발생시키는 것으로 보고되었다 (도 127a)[20]. HF-BE3에서 관찰된 특이성 증진은 실질적이지만, 보다 보통이었고; HF-BE3은 플라스미드 DNA로서 포유동물 세포 내로 전달되는 경우에 반복적 및 비-반복적 유전자좌 둘 다에서 검출가능한 오프-타겟 염기 편집을 나타내었다 (도 127a, 127d, 133c 및 133e). 이러한 특이성 증진 차이는 염기 편집이 Cas9-매개 indel 형성과 달리 DNA 절단을 필요로 하지 않고, 단지 DNA-결합 및 R-루프 형성을 필요로 한다는 사실로부터 발생할 수 있고[14], HF-Cas9의 증진된 특이성 중 일부가 이미-결합된 오프-타겟 유전자좌에서의 손상된 DNA 절단으로부터 발생할 수 있는 것으

로 추측되는 것이 그럴듯하다.

[0869] 오프-타겟 염기 편집을 감소시키기 위한 제2 시도에서, 염기 편집제의 RNP 전달은 온- 및 오프-타겟 편집의 탈커플링을 야기하는 것으로 입증되었다 (도 128b-128c). RNP 전달은 온-타겟 편집을 플라스미드 전달과 대등하게 유지하면서 비-반복적 부위에서의 오프-타겟 편집을 제거하였고 (도 130a-130c 및 128a), 고도로 반복적인 VEGFA 부위 2에서도 오프-타겟 편집을 크게 감소시켰다 (도 130d). 염기 편집제의 RNP 전달은 특히, 세포 투여량이 전형적으로 제어 또는 특징화되기 어려운 생체내 편집 용도에 유용할 수 있다.

[0870] 공여자 DNA 주형의 전달과 커플링된 Cas9의 RNP 전달은 이전에 포유동물 세포에서 HDR-기반 게놈 편집을 수행하는 데 사용되었다. 그러나, 이들 접근법은 낮은 효율, 세포-상태 의존성, 및 전형적으로 특히 점 돌연변이 교정에 대해 목적하는 HDR 결과의 것을 초과하는 indel 형성 효율에 의해 제한되어 있다[29,30,32,35]. DNA-무함유 염기 편집은 대조적으로 생체내 및 세포 둘 다에서 확률적 indel에 비해 실질적인 과량의 편집된 생성물을 생성한다 (도 132a, 134a, 및 134b). 본 발명자들의 지식이 미치는 한, 염기 편집제의 RNP 전달은 외인성 DNA를 필요로 하지 않으면서 게놈 DNA에 대한 특이적이고 정확한 변형을 생성하기 위한 제1 전략을 나타낸다.

[0871] 방법

[0872] 플라스미드의 클로닝

[0873] 본 연구에서 플라스미드는 USER 클로닝에 의해 생성되었다. 퓨션 U 핫 스타트 폴리머라제 (써모 피셔)를 사용하여 점 돌연변이를 설치하고, 이전에 보고된 구축물로부터 단백질 발현 플라스미드를 구축하였다[36]. 단백질 서열은 보충 정보에 열거되어 있고, BE3 및 HF-BE3의 발현을 위한 플라스미드는 애드진으로부터 입수가능하다.

[0874] BE3 및 HF-BE3의 발현 및 정제

[0875] BL21 스타 (DE3)-적격 이. 콜라이 세포를 His_6 N-말단 정제 태그를 갖는 박테리아 코돈 최적화된 염기 편집제를 코딩하는 플라스미드로 형질전환시켰다. 단일 콜로니를 밤새 50 μg mL^{-1} 카나마이신을 함유하는 루리아-베르타니 (LB) 브로쓰에서 37℃에서 성장시켰다. 세포를 2L의 동일 배지 내로 1:200 희석하고, OD_{600} = 0.70-0.75가 될 때까지 37℃에서 성장시켰다. 배양물을 얼음 상에서 60분 동안 인큐베이션하고, 단백질 발현을 0.5 mM 이소프로필-β-D-1-티오갈락토피라노시드 (IPTG, 골드바이오(GoldBio))로 유도하였다. 18℃에서 진탕시키면서 14-16시간 동안 발현을 지속하였다. 후속 정제 단계를 4℃에서 수행하였다. 세포를 6,000 g에서 20분 동안 원심분리에 의해 수집하고, 세포 수집 완충제 (100 mM 트리스(히드록시메틸)-아미노메탄 (트리스)-HCl, pH 8.0, 1M NaCl, 20% 글리세롤, 5 mM 트리스(2-카르복시에틸)포스핀 (TCEP; 골드바이오), 0.4 mM 페닐메탄 술포닐 플루오라이드 (PMSF; 시그마 알드리치) 및 사용된 50 mL 완충제당 1개의 컴플리트(cOmplete) EDTA-무함유 프로테아제 억제제 펠릿 (로슈(Roche)) 중에 재현탁시켰다. 세포를 초음파처리에 의해 용해시키고 (총 6분, 3초 온, 3초 오프), 용해물을 25,000 g (20분)에서 원심분리에 의해 정화하였다.

[0876] 정화한 용해물을 His-Pur 니켈 니트릴로아세트산 (니켈-NTA) 수지 (배양물 리터당 1 mL 수지, 써모 피셔)와 함께 60-90분 동안 4℃에서 회전시키면서 인큐베이션하였다. 수지를 20 칼럼 부피의 세포 수집 완충제로 세척한 후에 결합된 단백질을 용리 완충제 (100 mM 트리스(히드록시메틸)-아미노메탄 (트리스)-HCl, pH 8.0, 0.5 M NaCl, 20% 글리세롤, 5 mM 트리스(2-카르복시에틸) 포스핀 (TCEP; 골드바이오), 200 mM 이미다졸)로 용리시켰다. 생성된 단백질 분획을 Akta 퓨어 FPLC를 사용하여 5 mL 하이-트랩(Hi-Trap) HP SP (지이 헬스케어 (GE Healthcare)) 양이온 교환 칼럼 상에서 추가로 정제하였다. 단백질-함유 분획을 3,000 g에서 원심분리되는 100,000 kDa 컷오프를 갖는 칼럼 (밀리포어(Millipore))을 사용하여 농축시키고, 농축된 용액을 .22 μm PVDF 막 (밀리포어)을 통해 멸균 여과하였다.

[0877] 멸균 여과 후, 단백질을 환원제 상용성 비신코닌산 (BCA) 검정 (피어스 바이오테크놀로지(Pierce Biotechnology))으로 정량화하고, 액체 질소에서 급속 동결시키고, -80℃에서 분취액으로 저장하였다. 발현된 단백질의 서열은 보충 노트 2에 열거된다.

[0878] sgRNA의 시험관내 전사

[0879] sgRNA 백본 및 목적하는 20 bp sgRNA 프로토스페이서 상류에 T7 RNA 폴리머라제 프로모터 서열을 함유하는 선형 DNA 단편을 보충 정보에 열거된 바와 같은 프라이머를 사용하여 PCR (Q5 핫 스타트 마스터믹스, 뉴 잉글랜드 바이오랩스)에 의해 생성하고, 민일루트(minelute) 칼럼 (퀴아젠) 상에서 농축시켰다. sgRNA를 20 μL 반응물당 1 μg의 선형 주형을 사용하여 14-16시간 동안 하이스크라이브(HiScribe) T7 고수율 RNA 합성 키트 (뉴 잉글랜

드 바이오랩스)로 16℃에서 전사하였다. sgRNA를 제조업체의 지침에 따라 메가클리어(MEGAClear) 전사 클린 업 키트 (써모 피셔)를 사용하여 정제하였다. 정제된 sgRNA를 -80℃에서 분취액으로 저장하였다.

[0880] 시험관내 탈아미노화 검정

[0881] 시험관내 탈아미노화 검정을 위한 주형으로서 사용된 DNA 올리고뉴클레오티드의 서열은 보충 노트 3에 제시된다. 모든 올리고뉴클레오티드는 IDT로부터 구입하였다. 상보적 서열을 사용하여 합성된 단일-가닥 올리고뉴클레오티드를 트리스 완충제 pH 8.0 내에서 합하고 (100 μM 용액 5 μL), 95℃로 5분 동안 가열함으로써 어닐링하고, 이어서 0.1℃/초$^{-1}$의 속도로 37℃로 서서히 냉각시켜 79개 염기 쌍 (bp) dsDNA 기질을 생성하였다. 새로 해동된 염기-편집 단백질 (10 μL 반응 부피 중 2 μM 최종 농도)을 실온에서 5분 동안 반응 완충제 (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, 10 mM $MgCl_2$)에서 지시된 sgRNA (2.2 μM 최종 농도)와 복합체화하였다[37]. 이어서, 어닐링된 dsDNA 기질을 250 nM의 최종 농도로 첨가하였다. 반응을 37℃에서 30분 동안 진행시킨 후, 단백질 변성을 99℃에서 5분 동안 가열함으로써 수행하였다. PB 완충제 (퀴아젠, 100 μL) 및 이소프로판올 (25 μL)을 첨가하여 단백질이 기질 DNA로부터 해리되도록 하였다. DNA를 민엘루트 칼럼 (퀴아젠)으로 정제하고, 생성된 생성물을 U-내성 폴리머라제 (퓨션 U 핫 스타트, 써모피셔) 및 보충 정보에 열거된 바와 같은 프라이머를 사용하여 15 사이클의 qPCR (12ng 투입 DNA, 50 μL 반응 부피)로 선형 범위의 상단으로 증폭시켰다. 증폭된 DNA를 래피드팁2 (디피니티 게노믹스)를 사용하여 정제하고, 하기 기재된 바와 같이 일루미나 MiSeq 상에서의 서열분석을 위해 제조하기 전에 제2 라운드의 PCR (8 사이클, 5 ng 투입)로 바코딩하였다.

[0882] 게놈 DNA의 정제 및 서열분석

[0883] 제조업체의 지침에 따라 아젠코트 DNAdvance 게놈 DNA 단리 키트 (베크만 쿨터)를 사용하여 게놈 DNA를 단리하였다. 제1 PCR을 위해, DNA를 제조업체의 지침에 따라, 그러나 3% DMSO 및 SYBR 골드 핵산 스테인 (써모 피셔)을 첨가하여 Q5 핫 스타트 DNA 폴리머라제 (NEB)를 사용하여 선형 범위의 상단으로 증폭시켰다. 모든 앰플리콘에 대해, 사용된 PCR 프로토콜은 98℃에서 2분의 초기 가열 단계에 이어서 최적화된 수의 증폭 사이클 (98℃에서 12초, 61℃에서 25초, 72℃에서 30초)이었다. 제브라피쉬 및 형질감염된 세포 샘플에 대해, 30 ng의 투입 DNA를 50 μL 반응물에 사용하였고, 와우 샘플에 대해 20 ng을 25 μL 반응물에 사용하였다. 각각의 앰플리콘에 대한 최적 사이클 수를 결정하기 위해 qPCR을 수행하였다. 증폭된 DNA를 래피드팁2 (디피니티 게노믹스)를 사용하여 정제하고, 추가의 PCR (8 사이클, 5ng 투입)로 바코딩하였다. 제1 라운드 PCR에 사용된 고유한 정방향 및 역방향 프라이머는 어닐링 영역의 5'에 불변 영역을 함유하였고 (정방향: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN-3' (서열식별번호: 502), 역방향: 5'-TGGAGTTCAGACGTGTGCTCTTCCGATCT-3' (서열식별번호: 503)), 이는 제2 라운드 PCR을 위한 바코딩 프라이머의 증폭된 DNA에 대한 결합을 용이하게 한다.

[0884] 제2 라운드 PCR은 3개 영역을 갖는 프라이머를 사용하였다: 앰플리콘이 일루미나 유동 세포 (이탤릭체)에 결합하도록 하는 5' 불변 영역, 8-염기 바코딩 영역 (X), 및 바코딩 프라이머가 제1 라운드 PCR 앰플리콘 (볼드체)에 결합하도록 하는 3' 불변 영역. 프라이머 서열의 예는 하기와 같다:

정방향: 5'-

*AATGATACGGCGACCACCGAGATCTACAC*XXXXXXXX**ACACTCTTTCCCTACACGAC**

-3' (SEQ ID NO: 504)

역방향: 5'-

CAAGCAGAAGACGGCATACGAGATXXXXXXXX**GTGACTGGAGTTCAGACGTGTGC**

[0885] **TCTTC-3'** (SEQ ID NO: 505)

[0886] 서열분석 어댑터 및 이중-바코딩 서열은 TruSeq 인덱싱 어댑터 (일루미나)에 기초한다. 바코딩된 샘플을 풀링하고, 겔 추출 (퀴아젠)에 의해 정제한 다음, 암퓨어(Ampure) 비드 (베크만 쿨터)를 사용하여 정제한 후, 큐비트(Qubit) dsDNA HS 키트 (써모 피셔) 및 qPCR (카파 바이오시스템즈)를 사용하여 제조업체의 지침에 따라 정량화하였다. 제조업자의 지침에 따라 MiSeq (일루미나) 상에서 180-250개 염기 (앰플리콘 크기에 따름)로부터의 단일-말단 판독물을 사용하여 풀링된 샘플의 서열분석을 수행하였다.

[0887] PCR 증폭에 사용된 올리고뉴클레오티드의 서열은 보충 노트 3에 제시된다. 모든 올리고뉴클레오티드는 IDT로부터 입수하였다. 본 연구에서 각각의 앰플리콘에 대한 최적화된 수의 PCR 사이클은 하기와 같다: VEGFA 부위 2

인간 게놈 DNA (어닐링 온도는 모든 연장 단계에 대해 25초 동안 61℃임): 온-타겟: 29 사이클, 오프-타겟 #1: 32 사이클, 오프-타겟 #2: 28 사이클, 오프-타겟 #3: 27 사이클, 오프-타겟 #4: 27 사이클, VEGFA 부위 2 뮤린 게놈 DNA: 온-타겟: 31 사이클, 오프-타겟 #1, #2, #3 및 #4: 31 사이클. HEK293 부위 3: 오프-타겟 #1: 29 사이클, 오프-타겟 #2: 28 사이클, 오프-타겟 #3: 28 사이클. FANCF 오프-타겟 #1: 29 사이클, 오프-타겟 #2: 28 사이클, 오프-타겟 3: 28 사이클. EMX1 오프-타겟 #1, #2 및 #3: 28 사이클. 제브라피쉬 DNA의 증폭을 위한 TYR1, TYR2 및 TYR3 sgRNA: 32 사이클. EMX1, FANCF 및 HEK293 부위 3 유전자좌의 온-타겟 증폭을 위한 최적화된 프로토콜은 이전에 기재된 바와 같이 이어졌다[14].

[0888] 게놈 DNA 서열분석 판독물의 분석 및 정렬

[0889] 서열분석 판독물을 이전에 기재된 바와 같이 분석하였다[14]. 간략하게, 서열분석 판독물을 MiSeq 리포터 (일루미나)를 사용하여 디멀티플렉싱하고, 개별 FASTQ 파일을 이전에 보고된 주문제작 매트랩 스크립트로 분석하였다[14]. 스미스-워터만 알고리즘을 사용하여 판독물을 참조 서열에 정렬시켰다. 30 미만의 Q-점수를 갖는 염기 호출을 플레이스홀더 뉴클레오티드 (N)로 대체하였다. 이러한 품질 역치는 1,000 중 1의 예상 오차율을 갖는 뉴클레오티드 빈도를 발생시킨다. indel 빈도는 nCas9 절단 부위 주위의 30-염기 윈도우에서 발생하고 최소 2-염기 삽입 또는 결실인 indel을 카운팅하는 이전에 공개된 주문제작 매트랩 스크립트로 정량화하였다[14]. indel은 처리된 샘플 및 비처리 대조군에서의 indel 형성 사이의 유의한 차이 (스튜던트 양측 t-검정, $p < 0.05$)가 존재하는 경우 검출가능한 것으로 정의하였다.

[0890] 뮤린 게놈 내의 VEGFA 부위 2 sgRNA와 연관된 서열분석된 앰플리콘 중 하나인 CFD 오프-타겟 #3의 경우, indel 형성을 정확하게 측정하는 것이 가능하지 않았다. 이 유전자좌에서 프로토스페이서는 12개의 구아닌 염기에 직전에 선행하고, 이는 이 부위의 PCR 및 고처리량 서열분석을 무작위 삽입 또는 결실에 취약하게 만들고; 서열분석 판독물의 20%만큼 높은 결실 비율은 다중 독립적 비처리 대조군 샘플에서 관찰되었다. 어떠한 유의한 염기 편집도 임의의 처리 조건 하에 이 오프-타겟 유전자좌에서 검출되지 않았기 때문에 (도 132 및 133), indel 형성이 또한 이 유전자좌에서 무시할만한 것으로 의심된다.

[0891] phred.II Q30 점수는 염기호출에서 추정 99.9% 정확도에 상응한다[38]. 주어진 위치에서의 부정확한 염기 호출의 0.1% 확률은 염기 호출 오류가 4개의 염기에 걸쳐 무작위로 분포되는 것으로 가정되는 경우 0.1/4 = 0.025%의 염기 호출에 대한 하한에 상응한다. 이러한 역치 미만에 속하는 C→T 편집 백분율은 검출불가능한 것으로 분류되었다. PCR 동안의 자발적 탈아미노화[39] 또는 폴리머라제 오차는 또한 인공 C→T 편집을 도입할 수 있다. 염기 편집제-유도된 C→T 편집을 인공 C→T 편집 비율로부터 구별하기 위해, 비처리 대조군 세포를 각각의 앰플리콘에 대해 서열분석하였고, 특정한 조건 하에 C→T 편집이 역치로서 $p < 0.05$를 갖는 스튜던트 양측 t-검정을 사용하여 통계적으로 유의한지 계산하였다. 통계적으로 유의한 편집 비율이 >0.025%인 오프-타겟 부위를 측정가능한 것으로 간주하였다. 각각의 샘플에 대한 정렬된 및 품질 여과된 판독물의 수를 표 17에 포함시켰다.

[0892] 게놈 DNA 서열 정렬의 통계적 분석

[0893] 달리 언급되지 않는 한, 본문 전체에 걸쳐 언급된 평균 값은 $n \geq 3$개의 독립적인 생물학적 반복을 대표하고, 평균 ± 표준 편차가 언급되어 있다.

[0894] 처리된 샘플에서 온- 및 오프-타겟 편집 백분율을 음성 대조군 샘플 (비처리)에서 측정된 임의의 편집과 비교함으로써 도 2 및 3에 나타낸 고처리량 서열분석 데이터의 통계적 분석을 수행하였다. 스튜던트 양측 t 검정이 사용되었고, 개별 p-값은 표 16에 제시된다. * $p \leq 0.05$, ** $p \leq 0.01$ 및 *** $p \leq 0.001$. 편집이 검출 한계 (0.025%) 미만인 경우, 유의성은 계산되지 않았고; 모든 비처리 대조군 샘플은 검출불가능한 편집을 보여주었다.

[0895] 도 128a의 경우, 평균 온-타겟 염기 편집은 염기 편집 활성 윈도우 내의 시토신의 편집을 평균냄으로써 계산하였다 (HEK293 부위 3 및 EMX1에 대해 C_4-C_8, FANCF 및 VEGFA 부위 2에 대해 C_4-C_9).

[0896] 염기 편집 활성에서의 sgRNA-의존성 차이를 설명하기 위해, 염기 편집:indel 비를 계산하였다 (도 130b). 이 비는 C→T 전환을 갖는 HTS 판독물의 백분율 (각각의 부위에 대해 염기 편집 윈도우에 걸쳐 평균냄)을 indel을 함유하는 HTS 판독물의 백분율로 나눔으로써 생성하였다. 상기 기재된 바와 같이, 특정한 유전자좌에 대한 오프-타겟 편집이 검출 한계 미만인 경우, 특이성 비율을 계산하기 위한 목적으로 검출 방법의 추정 상한

(0.025%)을 보존적으로 가정하였다.

[0897] 시험관내 편집된 DNA의 데이터 분석

[0898] 서열분석 판독물을 MiSeq 리포터 (일루미나)를 사용하여 자동으로 디멀티플렉싱하였다. 온라인 패키지 usegalaxy.org를 사용하여 품질 필터링을 수행하였다[40]. 30이하의 일루미나 품질 점수를 갖는 개별 염기를 FASTQ 그루머에 이어서 FASTA 마스커를 사용하여 플레이스홀더 뉴클레오티드 'N'으로 전환시켰다[41]. 생성된 품질-필터링된 FASTQ 파일을 후속적으로 보충 노트 1에 제공된 주문제작 피톤 스크립트로 분석하였다. 서열분석 판독물을 표적 DNA 서열의 양쪽 면에 플랭킹하는 2개의 14-염기 서열에 대한 정확한 매치에 대해 스캐닝하였다. 정확한 매치가 발견되지 않는 경우, 판독물을 분석에서 배제하였다. 14-염기 서열 둘 다가 위치하고 이들 사이의 서열 길이가 예상된 프로토스페이서 길이 (20개 염기)와 동일한 경우, 플랭킹 영역 사이에서 발견된 프로토스페이서 서열을 저장하고, 프로토스페이서 내 각각의 부위에서 고처리량 서열분석에 의해 호출된 염기를 집계하였다.

[0899] 세포 배양

[0900] HEK293T (ATCC CRL-3216) 및 NIH/3T3 (ATCC CRL-1658) 둘 다를 10% (v/v) 소 태아 혈청 (FBS)이 보충된 둘베코 변형 이글 배지 + 글루타맥스 (써모피셔) 중에서 37℃에서 5% CO_2 하에 유지시켰다. 세포는 ATCC로부터 입수하였고, 구매시 ATCC에 의해 미코플라스마가 없는 것으로 인증 및 검증되었다.

[0901] 염기 편집제의 HEK293T 세포로의 플라스미드 형질감염

[0902] HEK293T 세포를 항생제 무함유 배지에서 48-웰 콜라겐-코팅된 바이오코트 플레이트 (코닝) 상에 시딩하고, 대략 70% 전면생장률로 형질감염시켰다. 달리 언급되지 않는 한, 750 ng의 BE 및 250 ng의 sgRNA 발현 플라스미드를 제조업체의 프로토콜에 따라 웰당 1.5 μl의 리포펙타민 2000 (써모 피셔)을 사용하여 형질감염시켰다.

[0903] 염기 편집제의 HEK293T 세포로의 단백질 형질감염

[0904] HEK293T 세포를 250 μL 항생제 무함유 배지 중에서 48-웰 콜라겐-코팅된 바이오코트 플레이트 (코닝) 상에 시딩하고, 대략 70% 전면성장률로 형질감염시켰다. 염기 편집제 단백질을 실온에서 5분 동안 1.1X 몰 과량의 필요한 sgRNA와 함께 인큐베이션하였다. 이어서, 복합체를 1.5 μL 리포펙타민 2000 (써모 피셔)과 함께 인큐베이션하고, 플라스미드 전달을 위해 제조업체의 프로토콜에 따라 형질감염시켰다. 달리 언급되지 않는 한, BE 단백질을 200 nM의 최종 농도로 첨가하였다 (275 μL의 총 웰 부피에 기초함).

[0905] 염기 편집제의 NIH/3T3 세포로의 플라스미드 형질감염

[0906] NIH/3T3 세포를 항생제-무함유 DMEM 배지에서 48-웰 콜라겐-코팅된 바이오코트 플레이트 (코닝) 상에 시딩하고, 대략 75% 전면생장률로 형질감염시켰다. 달리 언급되지 않는 한, 600 ng의 BE 및 200 ng의 sgRNA 발현 플라스미드를 제조업체의 프로토콜에 따라 웰당 1.4 μL의 리포펙타민 3000과 1 μL의 P3000 시약 (써모 피셔)을 사용하여 형질감염시켰다.

[0907] 염기 편집제의 NIH/3T3 세포로의 단백질 형질감염

[0908] NIH/3T3 세포를 항생제 무함유 DMEM 배지에서 48-웰 콜라겐-코팅된 바이오코트 플레이트 (코닝) 상에 시딩하고, 대략 75% 전면성장률로 형질감염시켰다. 염기 편집제 단백질을 1.1배 몰 과량의 지시된 sgRNA와 함께 25℃에서 5분 동안 인큐베이션하였다. 이어서, 복합체를 1.4 μL 리포펙타민 3000 (써모 피셔)과 함께 인큐베이션하고, 플라스미드 전달을 위해 제조업체의 프로토콜에 따라 형질감염시켰다. P3000 시약은 그의 첨가가 단백질 침전 및 염기 편집 효율의 감소를 야기하기 때문에 사용되지 않았다. 달리 언급되지 않는 한, BE 단백질을 400 nM의 최종 농도로 첨가하였다 (275 μL의 총 웰 부피에 기초함).

[0909] 양이온성 지질 내에 캡슐화된 BE3 단백질:가이드 RNA의 와우내 전달

[0910] 모든 동물 실험은 매사추세츠 안과 및 이과 병원의 동물 실험 윤리 위원회에 의해 승인되었다. 이전에 기재된 바와 같이 혼합 유전자 배경의 P1-P2 마우스에서 와우내 전달을 수행하였다[42]. 마우스를 수술 절차 전에 체온을 낮춤으로써 마취시켰다. 오른쪽 귀 근처에서 이후부 절개를 수행하고, 낭포를 들어올려 와우를 노출시켰다. BE3 단백질 (57.7 μM)을 1:1.1 몰비로 sgRNA (100 μM)와 사전-복합체화한 다음, 리포펙타민 2000 (써모 피셔)과 1:1 부피 비로 혼합하였다. 생성된 용액 (1.2-1.5 μL)을, 나노리터 마이크로펌프 (WPI, UMP3 + 마이

크로4 + 나노필)에 250 nL/분$^{-1}$의 속도로 부착된 와우 기저 회전에서, 와우 캡슐을 통해 유리 피펫 (말단 직경, 5 μm)을 사용하여 와우관 내로 주입하였다. 주사 후, 절개부를 닫고, 마우스를 가열 패드에 넣어 회복시켰다. 3-4일 후, 마우스의 와우를 코르티 기관, 혈관조 및 와우축으로 해부하였다. 각각의 조직을 5 내지 7개의 별개의 조각으로 추가로 미세-해부하고, DNA 추출을 각각의 샘플에 대해 개별적으로 수행하고, 이어서 상기 기재된 바와 같이 고처리량 서열분석을 수행하였다. 도 132 및 도 134에 제시된 데이터는 각각의 와우 영역에 대한 하나의 미세-해부된 샘플의 추출로부터 생성된 서열분석 데이터를 보여준다.

[0911] BE3 단백질:가이드 RNA의 제브라피쉬 배아 내로의 미세주사

[0912] 제브라피쉬 (튀빙겐 계통)를 보스톤 아동 병원의 내부 규제 검토에 준하여 표준 조건 하에 유지하였다. 1-세포 단계 제브라피쉬 배아에 적절한 sgRNA 또는 비관련된 sgRNA 대조군과 사전-복합체화된 대략 2 nL의 BE3 단백질을 1:1 몰비 (4.5 μM 최종 농도)로 주사하였다. 수정 4일 후, DNA를 50 mM NaOH 중에서 30분 동안 95℃에서 이전에 기재된 바와 같은 유충으로부터 추출하고[43], 생성된 용액을 트리스-HCl로 중화시켰다. 게놈 DNA를 상기 기재된 바와 같이 정량화하고, PCR에 의해 증폭시키고, 서열분석하였다.

[0913] 단백질 겔 분석

[0914] 제시된 모든 단백질 겔은 사전 성형된 4-12% 폴리아크릴아미드 비스-트리스 플러스 (써모 피셔)였다. 이들을 MOPS 완충제 (써모 피셔) 중에서 180 V에서 50분 동안 구동시켰다. 샘플을 변성을 위한 100mM DTT 및 1X 리튬 도데실 술페이트 (LDS) 샘플 완충제 중에서 99℃로 10분 동안 가열함으로써 로딩을 위해 제조하였다. 제조업체의 지침에 따라 인스턴트 블루 단백질 염색 (엑스페디온(Expedion))을 사용하여 겔을 염색하였다.

[0915] 세포 용해물 분석을 위해, 2 mL의 유도후 밤샘 배양물을 15,000 g에서 펠릿화한 후, 제조업체의 지침에 따라 100 μL B-PER (써모 피셔)에서 용해시켰다.

[0916] 데이터 이용가능성

[0917] 본 연구의 발견을 지지하는 고처리량 서열분석 데이터를 수탁 번호 SRP097884 하에 NCBI 서열 판독 아카이브 데이터베이스에 기탁하였다. 단백질 발현을 위한 HF-BE3 및 BE3, 뿐만 아니라 포유동물 발현을 위한 HF-BE3을 코딩하는 플라스미드는 애드진으로부터 수탁 ID 87439 (pCMV-HF-BE3), 87438 (pET42b-HF-BE3), 87437 (pET42b-BE3)으로 입수가능하다.

[0918] 참고문헌

1 Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* **21**, 121-131, doi:10.1038/nm.3793 (2015).
2 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).
3 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20-36, doi:10.1016/j.cell.2016.10.044 (2017).
4 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* **32**, 347-355, doi:10.1038/nbt.2842 (2014).
5 Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome Res* **25**, 1442-1455, doi:10.1101/gr.190124.115 (2015).
6 Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. *Annual review of genetics* **40**, 363-383, doi:10.1146/annurev.genet.40.110405.090451 (2006).
7 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
8 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
9 Bodles-Brakhop, A. M., Heller, R. & Draghia-Akli, R. Electroporation for the Delivery of DNA-based Vaccines and Immunotherapeutics: Current Clinical Developments. *Mol Ther* **17**, 585-592, doi:10.1038/mt.2009.5 (2009).
10 Kay, M. A., Glorioso, J. C. & Naldini, L. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nature medicine* **7**, 33-40, doi:Doi 10.1038/83324 (2001).
11 Midoux, P., Pichon, C., Yaouanc, J. J. & Jaffres, P. A. Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers. *Brit J Pharmacol* **157**, 166-178, doi:10.1111/j.1476-5381.2009.00288.x (2009).
12 Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature biotechnology* **32**, 551-553, doi:10.1038/nbt0914-952d (2014).

[0919]

13 Hu, J. H., Davis, K. M. & Liu, D. R. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. *Cell chemical biology* **23**, 57-73, doi:10.1016/j.chembiol.2015.12.009 (2016).
14 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, doi:10.1038/nature17946 (2016).
15 Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature methods* **13**, 1036-1042, doi:10.1038/nmeth.4038 (2016).
16 Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature methods* **13**, 1029-1035, doi:10.1038/nmeth.4027 (2016).
17 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729, doi:10.1126/science.aaf8729 (2016).
18 Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol Plant* **In Press**, doi:10.1016/j.molp.2016.12.001 (2016).
19 Koyama, S., Ishii, K. J., Coban, C. & Akira, S. Innate immune response to viral infection. *Cytokine* **43**, 336-341, doi:10.1016/j.cyto.2008.07.009 (2008).
20 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).
21 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
22 Fu, Y. F. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
23 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
24 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187-197, doi:10.1038/nbt.3117 (2015).
25 Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature methods* **11**, 429-435, doi:10.1038/nmeth.2845 (2014).
26 Singh, R., Kuscu, C., Quinlan, A., Qi, Y. J. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research* **43**, 677-683, doi:ARTN e118
10.1093/nar/gkv575 (2015).
27 Liang, X. Q. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44-53, doi:10.1016/j.jbiotec.2015.04.024 (2015).
28 Wang, M. *et al.* Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2868-2873, doi:10.1073/pnas.1520244113 (2016).
29 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33**, 73-80, doi:10.1038/nbt.3081 (2015).

[0920]

30 DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med* **8**, 360ra134, doi:10.1126/scitranslmed.aaf9336 (2016).
31 Kim, K. *et al.* Highly efficient RNA-guided base editing in mouse embryos. *Nature biotechnology*, doi:10.1038/nbt.3816 (2017).
32 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. *Nature biotechnology* **34**, 339-344, doi:10.1038/nbt.3481 (2016).
33 Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature communications* **5**, 4767, doi:10.1038/ncomms5767 (2014).
34 Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **11**, 316-318, doi:10.1038/Nchembio.1793 (2015).
35 Lin, S., Staahl, B., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*, e04766, doi:10.7554/eLife.04766 (2014).
36 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34**, 184-191, doi:10.1038/nbt.3437 (2016).
37 Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
38 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194 (1998).
39 Liu, M. & Schatz, D. G. Balancing AID and DNA repair during somatic hypermutation. *Trends in immunology* **30**, 173-181, doi:10.1016/j.it.2009.01.007 (2009).
40 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research* **44**, W3-W10, doi:10.1093/nar/gkw343 (2016).
41 Blankenberg, D. *et al.* Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-1785, doi:10.1093/bioinformatics/btq281 (2010).
42 Hu, L. X. *et al.* Diphtheria Toxin-Induced Cell Death Triggers Wnt-Dependent Hair Cell Regeneration in Neonatal Mice. *J Neurosci* **36**, 9479-9489, doi:10.1523/Jneurosci.2447-15.2016 (2016).
43 Meeker, N. D., Hutchinson, S. A., Ho, L. & Trede, N. S. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques* **43**, 610, 612, 614 (2007).
44 Zuris, J.A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nature biotechnology 33, 73-80 (2015).
45 Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature biotechnology 33, 187-197 (2015).
46 Doench, J.G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature biotechnology 34, 184-191 (2016).
47 Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol 17, 148 (2016).

[0921]

[0922] 보충 정보

[0923] 보충 노트 1: 품질-필터링된 시험관내-편집된 DNA를 분석하는 데 사용된 피톤 스크립트.

```
from __future__ import print_function
from __future__ import division

import Bio #This will import the BioPython suite
from Bio import SeqIO #Necessary to read/write sequence handles
from Bio.Seq import Seq
import os
import collections
import csv

inputfile = "please_specify_your_input_file_here_containing_filtered_reads" #specify th
e filenames that contain sequences
filenames = []

for file in os.listdir(inputfile):
    if file.endswith(".fastqsanger"):
        filenames.append(file)

spacer = []
list_of_filenames = []

for file in filenames:
    site = {}
    output = open(file + ".txt", "w")
    list_of_filenames.append(file + ".txt") #allows calling of the txt files that come
from fastq files later
    for rec in SeqIO.parse(file, "fastq"):
        split1=rec.seq.tostring().split("GTTCGCGGCGATCG") #14-
base pair constant_region_before_protospacer
        if len(split1)>=2:
            split2=split1[1].split("TGGATCGCCTGGCA") #14-
base pairc constant_region_after_protospacer
            site=split2[0]
            if len(site)==20:
                output.write(site + "\n")

BASES = 'ATGCN'
UNRECOGNIZED = 'X'
BASE_SEPERATOR = dict(zip(BASES, ',,,,\n'))
a_index = 0
t_index = 1
g_index = 2
c_index = 3
n_index = 4

def get_counts_by_column(base, count, library):
    current_count = library[count]
    if base == 'A':
        current_count[a_index] += 1
    elif base == 'T':
        current_count[t_index] += 1
    elif base == 'G':
        current_count[g_index] += 1
    elif base == 'C':
        current_count[c_index] += 1
    elif base == 'N':
        current_count[n_index] += 1

```

[0924]

```
def dna_counts(list_of_sequences, sample):
    first_oligo = list_of_sequences[0]
    for i in range (len(first_oligo)):
        sample.append([0,0,0,0,0])
    for j in range(len(first_oligo)):
        for i in range(len(list_of_sequences)):
            get_counts_by_column(list_of_sequences[i][j], j, libname)


for file in list_of_filenames:
    spacer_list = open(file).read().splitlines()
    output2=[]
    dna_counts(spacer_list, output2)
    with open(file  + ".csv", "wb") as f:
        writer = csv.writer(f)
        writer.writerows(output2)
```

[0925]

[0926] 보충 노트 2: 본 연구에 사용된 단백질의 서열

[0927] 발현된 BE3의 단백질 서열

MGSSHHHHHHSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQ
LTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKK
KRKV (SEQ ID NO: 185)

[0928]

[0929] 발현된 HF-BE3의 단백질 서열

MGSSHHHHHHSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQ
LTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSK
KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL

[0930]

EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKL
KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKK
KRKV (SEQ ID NO: 186)

[0931]

[0932] 보충 노트 3: 본 연구에 사용된 올리고뉴클레오티드의 서열

인간 세포에서의 HTS에 대한 오프 타겟 게놈 DNA를 증폭시키는 데 사용된 비공개된 프라이머 (SEQ ID NOS: 506-513)

fwd_VEGFA_부위2_오프_타겟_1_인간	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTACAAGTAACAGTCCAAGAA
rev_VEGFA_부위2_오프_타겟_1_인간	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTCTGCAACTTAACTTACGTGAAA
fwd_VEGFA_부위2_오프_타겟_2_인간	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCAAGCCCATTTGTCCAGG
rev_VEGFA_부위2_오프_타겟_2_인간	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTCTTTTTGAGCTTTGGGC
fwd_VEGFA_부위2_오프_타겟_3_인간	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCATACCAGCAGCAGTTCC
rev_VEGFA_부위2_오프_타겟_3_인간	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCACCTCAGCTCCTGCAC
fwd_VEGFA_부위2_오프_타겟_4_인간	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCACTGATTCTACACCATGGT
rev_VEGFA_부위2_오프_타겟_4_인간	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGTTCCCAACCTTTTTGACA

다른 프라이머 (HEK_3, EMX1, FANCF와 연관된 오프 타겟 부위에 대한 것)는 이전에 공개되었다

뮤린 세포에서의 HTS에 대한 오프 타겟 게놈 DNA를 증폭시키는 데 사용된 프라이머 (SEQ ID NOS: 514-521)

fwd_VEGFA_부위2_오프_타겟_1_뮤린	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGCTGGAGATTCAGAGACAC
rev_VEGFA_부위2_오프_타겟_1_뮤린	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGCCCCTTCTGACACACATAC
fwd_VEGFA_부위2_오프_타겟_2_뮤린	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCCCTCAAGGCTTGACATTTC
rev_VEGFA_부위2_오프_타겟_2_뮤린	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAAAGTTGGGAGAGGGGATG
fwd_VEGFA_부위2_오프_타겟_3_뮤린	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGTACCCCAGTCCCCTCATC
rev_VEGFA_부위2_오프_타겟_3_뮤린	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAGTTACGGGGATGTCACTTG
fwd_VEGFA_부위2_오프_타겟_4_뮤린	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTAACATCCAGTCTCCCAAACACA
rev_VEGFA_부위2_오프_타겟_4_뮤린	TGGAGTTCAGACGTGTGCTCTTCCGATCTACACACACACACTACTAGGACA

뮤린 세포에서의 HTS에 대한 온 타겟 게놈 DNA를 증폭시키는 데 사용된 프라이머 (SEQ ID NOS: 522-523)

fwd_VEGFA_부위2_온_타겟_뮤린	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGCTACTACGGAGCGAGAAG
rev_VEGFA_부위2_온_타겟_뮤린	TGGAGTTCAGACGTGTGCTCTTCCGATCTACAGGGGCAAAGTGAGTGAC

sgRNA의 T7 전사를 위한 기질로서의 역할을 하는 PCR 산물을 생성하는 데 사용된 프라이머 (SEQ ID NOS: 524-529)

rev_sgRNA_T7: 모든 경우에 사용됨	AAAAAAAGCACCGACTCGGTGCCAC
fwd_sgRNA_T7_EMX1	TAATACGACTCACTATAGGGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_FANCF	TAATACGACTCACTATAGGGGAATCCCTTCTGCAGCACCGTTTTAGAGCTAGAAATAGCA

[0933]

fwd_sgRNA_T7_HEK_부위_3	TAATACGACTCACTATAGGGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_VEGFA_부위_2	TAATACGACTCACTATAG GACCCCCTCCACCCCGCCTCGTTTTAGAGCTAGAAATAGCA
fwd_sgRNA_T7_TC_반복부_시험관내	TAATACGACTCACTATAGGTCTCTCTCTCTCTCTCTCTCGTTTTAGAGCTAGAAATAGCA

sgRNA 형질감염 플라스미드를 생성하는 데 사용된 프라이머 (SEQ ID NOS: 530-531)

pFYF1320 플라스미드는 이전에 기재된 바와 같이 (Komor et al.) 주형으로서 사용되었다. 다른 sgRNA 플라스미드의 서열은 이전에 보고되었다

rev_sgRNA_플라스미드	GGTGTTTCGTCCTTTCCACAAG
fwd_VEGFA_부위_2	GACCCCCTCCACCCCGCCTCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC

시험관내 데아미나제 검정에 사용된 ssDNA 기질의 서열 (SEQ ID NOS: 532-533)

fwd_TC_반복부_기질	ACGTAAACGGCCACAAGTTCGCGGCGATCGTCTCTCTCTCTCTCTCTCTCTGGATCGCCTGGCATCTTC TTCAAGGACG
rev_TC_반복부_기질	CGTCCTTGAAGAAGATGCCAGGCGATCCAGAGAGAGAGAGAGAGAGAGACGATCGCCGCGAACTTGT GGCCGTTTACGT

인간 세포에서의 HTS에 대한 오프 타겟 게놈 DNA를 증폭시키는 데 사용된 이전에 공개된 프라이머 (SEQ ID NOS: 534-557)

fwd_EMX1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_EMX1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_FANCF_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCATTGCAGAGAGGCGTATCA
rev_FANCF_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_HEK293_부위3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGTGGGCTGCCTAGAAAGG
rev_HEK293_부위3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGCCAAACTTGTCAACC
fwd_EMX1_오프1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAGCCTCTTTCTCAATGTGC
rev_EMX1_오프1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTTCACAAGGATGCAGTCT
fwd_EMX1_오프2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGAGCTAGACTCCGAGGGGA
rev_EMX1_오프2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTCGTCCTGCTCTCACTT
fwd_EMX1_오프3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAGGCTGAAGAGGAAGACCA
rev_EMX1_오프3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCAGCTGTGCATTCTAT
fwd_FANCF_오프1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCCACTGAAGAAGCAGGG
rev_FANCF_오프1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCTTAATCCGGCTCCAT
fwd_FANCF_오프2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCAGTGTTTCCATCCCGAA
rev_FANCF_오프2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTGACCTCCACAACTCT

[0934]

fwd_FANCF_오프3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGGTACAGTTCTGCGTGT
rev_FANCF_오프3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCACTCTGAGCATCGCCAAG
fwd_HEK293_부위3_오프1_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_부위3_오프1_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK293_부위3_오프2_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTGGTGTTGACAGGGAGCAA
rev_HEK293_부위3_오프2_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_부위3_오프3_HTS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_부위3_오프3_HTS	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCCAAGAACCT

제브라피쉬에서의 HTS에 대한 온 타겟 게놈 DNA를 증폭시키는 데 사용된 프라이머 (SEQ ID NOS: 558-563)

fwd_TYR1_제브라피쉬	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTCCCCCGAGTCTGCACCT
rev_TYR1_제브라피쉬	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGAACTTGCATTCGCCGCAA
fwd_TYR2_제브라피쉬	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTCTGCCTTGGCATCGGGTG
rev_TYR2_제브라피쉬	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCATACCGCCCCTAGAACTAACATTC
fwd_TYR3_제브라피쉬	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACAACTGCTTTCCATGGTGTGT
rev_TYR3_제브라피쉬	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCCAGGGCTTTCGTGGAGA

[0935]

부위	서열	SEQ ID NO
TYR1	GTC$_3$AGGTC$_8$GAGGGTTCTGTCAGG	564
TYR2	CTTC$_4$C$_5$AGGATGAGAACACAGAGG	565
TYR3	CAAC$_4$C$_5$AC$_7$TGCTCAAAGATGCTGG	566

[0936]

[0937] 보충 표 1. 본 연구에서 연구된 제브라피쉬 게놈 유전자좌에 대한 프로토스페이서 및 PAM 서열.

[0938] 실시예 17: 보다 높은 효율 및 생성물 순도를 갖는 C:G에서 T:A로의 염기 편집제

[0939] 염기 편집은 이중-가닥 DNA 파괴를 유도하거나 또는 Cas9 변이체 및 시티딘 데아미나제의 조작된 융합을 사용한 상동성-지정 복구를 필요로 하지 않으면서 표적 C:G 염기 쌍의 T:A 염기 쌍으로의 프로그램가능한 전환이다 (1). 제3 세대 염기 편집제 (BE3) 및 관련 기술은 광범위한 유기체에서 많은 연구자들에 의해 성공적으로 사용되었다 (2-13). 염기 편집의 생성물 분포-표적 C:G 염기 쌍이 목적하는 T:A 생성물과 함께 목적하지 않은 부산물의 혼합물로 전환되는 빈도-는 표적 부위-의존성 방식으로 달라진다 (2, 3, 6-8). 여기서 본 발명자들은 인간 세포에서의 염기 편집 결과의 결정인자를 특징화하고, 목적하지 않은 생성물의 형성이 우라실 N-글리코실라제 (UNG)에 의존성이고 염기 편집 활성 윈도우 내의 단일 C만을 함유하는 표적 부위에서 발생할 가능성이 더 크다는 것을 확립한다. 일부 서열에 대해 염기 편집 효율을 증가시키는 시티딘 데아미나제 상동체를 사용하는 구축물 CDA1-BE3 및 AID-BE3을 조작하였다. 추가적으로, 염기 편집된 중간체에 대한 UNG의 접근을 보다 효율적으로 차단하는 제4 세대 에스. 피오게네스 Cas9-유래 염기 편집제 (BE4)를 또한 조작하였다. BE3과 비교하여, BE4는 목적하지 않은 부산물의 빈도를 반으로 줄이면서 C:G에서 T:A로의 염기 편집의 효율을 대략 50%만큼 증가시킨다. 이러한 개선을 또한 적용하여, 에스. 아우레우스 Cas9-유래 BE4 (SaBE4)를 생성하였고, 이는 BE4보다 실질적으로 더 작고 대안적 표적화 범주를 갖는다.

[0940] 도입

[0941] 전통적인 게놈 편집 방법은 게놈 표적 유전자좌에서 이중 가닥 DNA 파괴 (DSB)를 도입한다 (14). DSB 병변에 대한 세포 반응은 주로 비-상동 말단 연결 (NHEJ) 및 관련 과정을 통해 진행된다 (15). NHEJ는 통상적으로 DSB에 플랭킹된 2개의 말단을 재연결하지만, 전형적인 게놈 편집 조건 하에 DSB는 연속적으로 재도입되어 결국 삽입 및 결실 (indel)의 축적 또는 DSB 부위에서의 전위 및 상응하는 게놈 유전자좌의 파괴를 발생시킨다 (16). 활발하게 분열하는 세포는 또한 DSB를 둘러싸는 영역에 대한 상동성을 함유하는 공여자 DNA 주형의 존재 하에 상동성-지정 복구 (HDR)를 개시함으로써 DSB에 반응할 수 있고, 이는 연구자가 NHEJ를 통해 가능한 것보다 더

정확하고 예측가능하게 게놈을 조작할 수 있게 한다 (17). HDR-의존성 게놈 편집은 NHEJ 결과와의 경쟁, 및 유사분열에 대한 HDR의 의존성으로부터 발생하는 낮은 효율에 의해 제한된다 (18).

[0942] HDR 또는 DSB의 도입을 필요로 하지 않으면서 프로그램가능한 방식으로 C:G 염기 쌍의 T:A 염기 쌍으로의 직접적인 비가역적 전환을 가능하게 하는 염기 편집의 개발이 보고되었다 (1). 염기 편집제는 촉매적으로 손상된 Cas9 단백질에 테더링된 단일-가닥 DNA-특이적 시티딘 데아미나제 효소 및 염기 절제 복구 억제제로 이루어진다 (1, 4, 9, 10). Cas9 변이체는 상응하는 가이드 RNA에 의해 프로그램된 관심 게놈 유전자좌에 결합한다. 단백질:RNA:DNA 3원 "R-루프" 복합체의 형성 (19)은 테더링된 시티딘 데아미나제 효소에 대한 기질로서의 역할을 하는 단일 가닥 DNA의 작은 (~5-nt) 윈도우를 노출시킨다. 이 윈도우 내의 임의의 시티딘은 우라실로 가수분해성으로 탈아미노화되어 G:U 중간체를 생성한다.

[0943] 염기 절제 복구 (BER)는 G:U 미스매치에 대한 세포의 일차 반응이고 우라실 N-글리코실라제 (UNG)에 의한 우라실의 절제에 의해 개시된다 (20). UNG에 의한 절제로부터 편집된 G:U 중간체를 보호하기 위한 노력으로, 83-아미노산 우라실 글리코실라제 억제제 (UGI)를 촉매적 데드 Cas9 (dCas9)의 C-말단에 직접 융합시켰다 (1). 세포 DNA 미스매치 복구 시스템을 G:U 미스매치 내의 G를 A로 우선적으로 대체하도록 조작하기 위해, dCas9 내의 Ala 840 아미노산을 His로 복귀시키고, 이로써 Cas9 단백질이 새로 형성된 우라실에 대향하는 DNA 가닥을 니킹하도록 하여, G:U 중간체의 목적하는 A:U 및 A:T 생성물로의 훨씬 더 효율적인 전환을 발생시켰다 (1). 이들 2가지 조작 노력을 조합하여 16-아미노산 링커를 통해 에스. 피오게네스 dCas9 (A840H)에 테더링된 APOBEC1 시티딘 데아미나제 효소의 3-부분 융합체로 이루어진 단일 단백질인 BE3을 생성하였고, 이는 4-아미노산 링커를 통해 UGI에 공유 연결된다 (1). BE3 및 관련 염기 편집제는 현재 식물 게놈 편집, 생체내 포유동물 게놈 편집, 표적화된 돌연변이유발 및 녹아웃 연구를 포함한 다양한 용도에 사용되어 왔다 (2-13). 염기 편집의 범주는 변경된 PAM 요건 (4), 좁아진 편집 윈도우 (4), 감소된 오프-타겟 편집 (10), 및 소분자 의존성 (21)을 갖는 BE3 변이체를 보고함으로써 최근에 확장되었다.

[0944] 일부 유전자좌에서, 염기 편집제 예컨대 BE3은 표적 C:G 염기 쌍이 목적하는 T:A 생성물보다는 G:C 또는 A:T 염기 쌍으로 전환되는 목적하지 않은 부산물을 생성한다 (2, 3, 6-8). 여기서, 본 발명자들은 염기 편집 생성물 순도의 결정인자를 조명하고, UNG 활성이 목적하지 않은 부산물의 형성에 요구된다는 것을 확립한다. 우라실 중간체에 대한 UNG 접근을 차단하는 것은 목적하지 않은 생성물을 최소화하기 위해 단일 C가 편집 윈도우 내에 있는 표적 유전자좌에 특히 중요한 것으로 결정되었다. BE3을 포함한 이전에 기재된 염기 편집제와 비교하여 더 높은 효율 및 크게 개선된 생성물 순도를 갖는 염기 편집을 수행하는 제4-세대 염기 편집제인 BE4 및 SaBE4를조작하였다.

[0945] 결과

[0946] UNG 활성은 부산물 형성에 요구된다

[0947] 목적하지 않은 염기 편집 부산물은 우라실-함유 DNA 가닥 내의 무염기성 부위의 형성 및 오류-유발 분해로 인한 염기 절제 복구 동안 발생할 수 있다. 우라실 N-글리코실라제 (UNG)가 결여된 세포에서의 염기 편집의 생성물 순도가 개선되는지 결정하기 위해, HAP1 세포 (반수체 인간 세포주) 및 HAP1 UNG^{-} 세포를 EMX1, FANCF, HEK2, HEK3, HEK4 또는 RNF2 유전자좌를 표적화하는 BE3 및 sgRNA를 코딩하는 플라스미드로 뉴클레오펙션하였다 (표적 서열에 대해서는 도 135b 참조). 뉴클레오펙션 3일 후, 게놈 DNA를 추출하고, 표적 유전자좌를 PCR에 의해 증폭시키고, 고처리량 DNA 서열분석 (HTS)에 의해 분석하였다. 염기 편집 생성물 순도는, 표적 C가 T로 편집된, 편집된 서열분석 판독물 (표적 C가 A, G 또는 T로 전환된 판독물)의 퍼센트로서 정의된다. BE3-처리된 HAP1 세포의 염기 편집 생성물 순도는 6개의 유전자좌에서 12개의 표적 C를 걸쳐 68 ± 6%로 평균되었다 (n = 3개의 생물학적 반복에 대해 평균 ± S.D.). HAP1 UNG^{-} 세포에서, 시험된 모든 12개의 표적 C는 생성물 순도 > 98%로 염기 편집되었다 (도 135a). 또한, 모든 6개의 시험된 유전자좌에서의 indel 빈도는 UNG 녹아웃시 7 내지 100배 감소하였다 (도 135c). 이들 데이터는 UNG 활성이 염기 편집 동안 목적하지 않은 생성물 형성에 필요하다는 것을 강하게 암시하며, 이는 무염기성 부위 형성 및 오류-유발 폴리머라제를 사용한 후속 염기 절제 복구가 표적 뉴클레오티드의 무작위화 및 때때로 indel을 생성하는 가닥 파괴를 야기하는 모델과 일치한다.

[0948] 다중 편집가능한 C를 갖는 표적은 보다 높은 생성물 순도를 나타낸다

[0949] BE3에 의한 염기 편집 효율은 G의 바로 하류인 일부 (전부는 아님) 표적 C에 대해 더 낮을 수 있고 (1), 이는 APOBEC1의 공지된 서열 선호도와 일치한다 (22) (도 136a). 이러한 표적을 효율적으로 편집하기 위한 노력으로, APOBEC1 데아미나제를 대체한 BE3 변이체를 CDA1 (CDA1-BE3을 생성하기 위함), AID (AID-BE3을 생성

하기 위함), 또는 APOBEC3G (APOBEC3G-BE3을 생성하기 위함)로 대체하여, 상이한 서열 선호도를 갖는 3개의 단일 가닥 DNA-특이적 시티딘 데아미나제 효소를 생성하였다 (23). HEK293T 세포를 이들 BE3 변이체 및 EMX1, FANCF, HEK2, HEK3, HEK4, 또는 RNF2 유전자좌를 표적화하는 sgRNA를 코딩하는 플라스미드로 형질감염시켰다. 형질감염 3일 후, 게놈 DNA를 추출하고, 표적 유전자좌를 PCR에 의해 증폭시키고, HTS를 사용하여 염기 편집에 대해 평가하였다. G에 바로 이어지는 표적 C의 보다 효율적인 편집이 BE3과 비교하여 CDA1-BE3 및 AID-BE3에서 관찰되었다 (도 136a, 도 140a-d, 및 도 146-151). 일반적으로, CDA1-BE3 및 AID-BE3은 G에 이어지지 않는 표적 C에서 BE3보다 더 낮은 편집 효율을 나타내었다 (도 140a-d). 대조적으로, APOBEC3G-BE3은 예측가능하지 않은 서열 선호도를 나타내었으며, 이때 전체적으로 BE3과 비교하여 더 낮은 C에서 T로의 편집의 수율을 가졌다. 이러한 발견은 CDA1-BE3 및 AID-BE3이 일부 표적 5'-GC-3' 서열에 대해 BE3에 비해 더 높은 편집 효율을 제공할 수 있다는 것을 시사한다.

[0950] 이들 데이터를 분석하면서, CDA1-BE3 및 AID-BE3의 생성물 순도는 전형적으로 CDA1-BE3 및 AID-BE3이 BE3보다 더 많은 C를 편집하는 그러한 부위에서 BE3의 생성물 순도보다 더 높다는 것에 주목하였다 (도 136a-d). 예를 들어, HEK4 유전자좌에서, BE3은 단지 단일 C만을 효율적으로 편집하지만 (G가 C에 선행하지 않음), CDA1-BE3 및 AID-BE3 둘 다는 3개의 C를 편집한다 (도 140a-c). 이러한 유전자좌에서의 BE3의 생성물 순도는 50 ± 7% (n = 3개의 생물학적 반복에 대해 평균 ± S.D.)인 반면, CDA1-BE3 및 AID-BE3의 생성물 순도는 각각 97 ± 2% 및 93 ± 2%이다. 더욱이, 각각 84 ± 3% 및 91 ± 2%의 생성물 순도를 갖는, BE3에 의해 편집된 EMX1 및 FANCF는 대등한 효율로 편집된 다중 C를 함유하는 반면 (도 S2), 각각 28 ± 3% 및 64 ± 3%의 훨씬 더 낮은 생성물 순도를 갖는, BE3에 의해 편집된 HEK2 및 RNF2는 동일하지 않은 효율로 편집된 다중 C를 함유한다 (도 141a-c). 대등한 효율을 갖는, HEK2 유전자좌 내에서 두 C를 편집하는 CDA1-BE3 및 AID-BE3은 이 유전자좌에서 훨씬 더 높은 생성물 순도를 나타낸다 (각각 85 ± 5% 및 81 ± 4%) (도 136a-d 및 도 140c). HEK2 및 RNF2 부위에서 다중 C가 초기에 대등한 효율로 BE3에 의해 U로 전환된 다음, DNA 복구 시스템에 의해 상이한 효율로 프로세싱될 가능성이 배제되었다. 이를 고려할 때, 이들 부위가 CDA1-BE3 또는 AID-BE3에 비해 BE3으로 처리되는 경우에 상이한 생성물 분포가 관찰되기 보다는 유사한 생성물 분포가 예상될 것이다 (도 136b 및 도 146-151). 대신에, 단리된 G:U는 G:U 병변의 클러스터보다 UNG에 의해 보다 용이하게 프로세싱될 수 있다. BE3에서의 시티딘 데아미나제 도메인의 프로세싱성 (1, 24)은 다중 편집가능한 C를 함유하는 유전자좌에서의 BE3의 체류 시간을 증가시킴으로써, 단일 편집가능한 C를 함유하는 유전자좌에서보다 더 효과적으로 UNG에 의한 접근을 차단할 수 있다는 것이 가능하다.

[0951] 생성물 순도, 개별 서열분석 판독물에서의 편집된 C의 수, 및 UNG 활성 사이의 관계를 추가로 분석하였다. 명백한 UNG 억제의 부재 하에 염기 편집된 DNA의 운명을 밝혀내기 위해, BE3의 UGI 성분을 제거하여 BE3B를 생성하였다. HEK293T 세포를 BE3 또는 BE3B 및 EMX1, FANCF, HEK2, HEK3, HEK4 또는 RNF2 유전자좌를 표적화하는 sgRNA를 코딩하는 플라스미드로 형질감염시켰다. 상기 확립된 염기 편집 결과를 다양화하는 데 있어서의 UNG의 역할을 고려할 때 예상되는 바와 같이, 모든 표적 C에서의 생성물 순도는 BE3-처리된 DNA와 비교하여 BE3B-처리된 DNA에서 크게 감소하였고, 비-T를 함유하는 편집 생성물의 분율은 평균 1.8 ± 0.4배만큼 증가하였다 (도 142b).

[0952] 다중-C 부위 HEK2, HEK3, 및 RNF2를 표적화하는 sgRNA 및 BE3 또는 BE3B로 처리된 HEK293T 세포로부터의 개별 DNA 서열분석 판독물을 분석하였다. 각각의 부위에 대해, 1차 표적 C가 가장 효율적으로 변형된 뉴클레오티드로서 지정되었다. 모든 3개의 부위에 걸쳐, 1차 표적 C의 목적하지 않은 C에서 비-T로의 편집을 함유하는 서열분석 판독물의 평균 80 ± 10%는 단일 염기 편집 사건만을 나타내었다 (도 142a-d 및 도 143). 대조적으로, 동일한 3개의 다중-C 부위에 걸쳐, 1차 표적 C의 명확한 C에서 T로의 편집을 함유하는 서열분석 판독물의 훨씬 더 낮은 평균 32 ± 4%는 단일 명확한 염기 편집 사건만을 나타내었다 (도 142a-d 및 도 143). 또한, 편집 윈도우 내에서 오직 1개의 C를 함유하는 부위인 BE3B-처리된 HEK4 DNA에 대한 생성물의 분포는 대략적으로 A:G:T에 대해 1:3:1의 비를 따른다 (도 143d). 이러한 관찰은, 주어진 표적 내의 단일 시티딘이 UGI의 부재 하에 U로 전환되는 경우, 이는 UNG-개시된 BER에 의해 효율적으로 프로세싱되어 생성물의 혼합물을 제공한다는 것을 집합적으로 나타낸다.

[0953] 이들 데이터는 클러스터링된 G:U 미스매치가 단리된 G:U 미스매치와는 상이하게 프로세싱되고, 명확한 C에서 T로의 편집을 생성할 가능성이 더 크다는 모델과 일치한다. 단지 단일 C에서 T로의 편집 사건을 목적하는 경우, 상기 관찰은 UNG 억제가 목적하지 않은 부산물을 최소화하는 데 중요하다는 것을 시사한다. 그러나, dCas9-데아미나제 융합체, 예컨대 TAM(8) 및 CRISPR-X(2)를 사용하여 표적화된 무작위 돌연변이유발을 수행하는 경우, 상기 관찰은 단지 단일 편집가능한 C를 갖는 표적 부위가 생성물 혼합물을 최대화할 것임을 시사한다.

[0954] 개선된 생성물 순도를 위한 BE3 아키텍처의 최적화

[0955] SSB-BE3을 생성하기 위해 BE3의 UGI 성분을 단일-가닥 DNA 결합 (SSB) 단백질로 대체하여, SSB가 R-루프의 우라실-함유 ssDNA 부분에 UNG가 접근하는 것을 차단할 수 있도록 하였다. SSB-BE3에 의한 염기 편집 효율의 큰 감소가 관찰되었고, 여기서 4개의 부위에 걸쳐 모든 7개의 C가 단지 평균 1.9 ± 0.5% C에서 T로의 전환을 나타내었다 (도 137c).

[0956] 염기 편집 결과를 결정하는 단계 동안 APOBEC, UGI, 및 UNG의 상대적인 배치가 공지되어 있지 않기 때문에, UGI는 UNG 억제를 개선시키기 위한 노력으로 BE3의 N-말단으로 재배치되었다 (N-UGI-BE3). UGI를 BE3의 N-말단으로 이동시켜, BE3과 비교하여 2.3 ± 0.6배의 모든 7개의 시험된 표적 C에 걸쳐 C에서 T로의 편집 백분율의 평균 감소 (도 137c), 및 BE3과 비교하여 평균 2.2 ± 0.5배의 4개의 모든 부위에서의 전체 생성물 순도의 감소가 생성되었다 (도 137b).

[0957] 대조적으로, BE3의 C-말단에 UGI의 추가의 카피를 부가하여 (BE3-2xUGI), BE3에 비해 생성물 순도의 큰 증가 및 BE3의 것과 대등한 C에서 T로의 편집 백분율을 발생시켰다. 비-T 편집 생성물은 시험된 4개의 유전자좌에 걸쳐 평균 2.2 ± 0.8배 감소하였다 (도 137b). 이들 관찰은 UGI의 제2 카피의 부가가 G:U 염기 편집 중간체에 대한 UNG의 접근을 실질적으로 감소시킴으로써 생성물 순도를 크게 개선시킨다는 것을 시사한다.

[0958] 상기 실험이 또한 성분의 아키텍처에 대한 염기 편집 결과의 감수성을 밝혀냈기 때문에, 다음으로 본 발명자들은 생성물 순도 및 편집 효율을 추가로 증가시키기 위해 BE3 성분 사이의 링커를 최적화하였다. 본 발명자들은 rAPOBEC1-dCas9(A840H) 링커를 16개의 아미노산 (BE3)에서 32개의 아미노산(BE3C)으로 다양화하고, dCas9(A840H)-UGI 링커를 4개 (BE3)에서 9개 (BE3D)에서 16개의 아미노산 (BE3E, 도 138a)으로 다양화하였다. dCas9(A840H)-UGI 링커가 4개 아미노산 (BE3) 대신에 9개 아미노산 잔기 길이 (BE3D)인 경우에 비-T 생성물 형성은 평균적으로 1.3 ± 0.1배 감소하였고 (도 138d), C에서 T로의 편집 효율에서의 명백한 차이는 없었다 (도 138c). rAPOBEC1-dCas9 (A840H) 링커를 16개의 아미노산 (BE3)에서 32개의 아미노산 (BE3C)으로 증가시키는 것은 C에서 T로의 편집 효율을 HEK2 유전자좌에서 평균 1.2 ± 0.1배 상승시켰다 (도 138c). 이러한 유전자좌는 이전에 가장 불균일하게 편집된 시험된 다중-C 부위였고 (도 141a-c), 이 링커를 확장시켜 C4에 비한 C6의 우선적인 편집을 2.6 ± 0.2배에서 1.8 ± 0.1배로 감소시켰다 (C6에서 편집된 서열분석 판독물의 백분율 대 C4의 것의 비). 본 발명자들은 이러한 보다 긴 링커가 데아미나제가 R-루프 내의 ssDNA에 보다 잘 접근할 수 있게 하고, 다중 표적 C가 염기 편집 윈도우 내에 존재하는 경우에 보다 균일한 탈아미노화를 발생시킬 수 있다는 것을 추론하였다. BE3C는 또한 시험된 다른 유전자좌에서 대등하거나 개선된 염기 편집 효율 및 생성물 순도를 나타내었다 (도 138c-d).

[0959] 증진된 효율 및 생성물 순도를 갖는 C:G에서 T:A로의 염기 편집제인 BE4

[0960] 염기 편집제 구축물 BE4는 모든 3가지 개선을 조합함으로써-rAPOBEC1-dCas9 링커를 32개의 아미노산으로 연장시키고, dCas9-UGI 링커를 9개의 아미노산으로 연장시키고, UGI의 제2 카피를 또 다른 9-아미노산 링커로 구축물의 C-말단에 부가함으로써 조작하였다. 문헌 [Nishida et al.]에 의해 보고된 대안적 염기 편집제 구축물인 표적-AID (9)를 또한 BE4와 동일한 플라스미드 백본 내로 클로닝하였다. HEK293T 세포를 BE3, BE4, 또는 표적-AID 및 EMX1, FANCF, HEK2, HEK3, HEK4, 또는 RNF2 유전자좌를 표적화하는 sgRNA를 코딩하는 플라스미드로 형질감염시켰다. 형질감염 3일 후, 게놈 DNA를 추출하고, 표적 유전자좌를 PCR에 의해 증폭시키고, HTS에 의해 분석하였다. BE3에 비해 BE4에 대한 모든 12개의 편집된 C에 걸쳐 1.5 ± 0.3배의 C에서 T로의 편집 효율의 평균 증가가 관찰되었다 (도 139c). 분석된 동일한 위치에서의 표적-AID에 대한 C에서 T로의 편집의 평균 효율은 BE3의 것보다 1.5 ± 0.5배 더 낮았고, BE4의 것보다 2.1 ± 0.5배 더 낮았지만, CDA1 데아미나제를 사용하는 표적-AID는, 위치 C3 및 C4에 대해 최적의 편집을 가지면서 편집 윈도우가 BE3 및 BE4에 비해 이동하는 것으로 보이는 것으로 보인다는 것에 주목하는 것이 중요하다 (도 139c). 이러한 이동된 편집 윈도우는 표적-AID와 BE3 또는 BE4 사이의 효율 및 생성물 순도의 비교를 어렵게 하는데, 이는 주어진 표적 C가 동일한 가이드 RNA를 사용하는 경우에도 상이한 편집 윈도우 내의 더 최적인 또는 덜 최적인 위치에 놓일 수 있기 때문이다.

[0961] 더 큰 C에서 T로의 편집 효율 이외에도, BE4는 또한 시험된 모든 게놈 유전자좌에서 BE3에 비해 실질적으로 개선된 생성물 순도를 나타내었고, 2.3 ± 0.3배의 비-T 생성물 형성에서의 평균 감소를 나타내었다 (도 139d). indel을 야기할 수 있는 염기 절제 복구를 추가로 방해하는 것에서 예상되는 바와 같이 (25), BE3과 비교하여 BE4 처리 후 모든 6개의 유전자좌에 걸쳐 평균 2.3 ± 1.1배의 indel 비율의 감소가 또한 관찰되었다 (도 144a-c). 종합하면, 이들 결과는 BE4가 시험된 모든 유전자좌에서 C에서 T로의 편집의 높은 효율, 높은 생성물

순도, 및 낮은 indel 형성 비율을 제공한다는 것을 나타낸다.

[0962] BE4 개선을 에스. 아우레우스 Cas9와 통합하여 (26) SaBE4를 생성하였고, 이는 에스. 피오게네스 dCas9(A840H)를 더 작은 에스. 아우레우스 dCas9(A580N)로 대체하고, 그의 대안적 PAM 요건으로 인해 상이한 표적에 접근할 수 있다. HEK293T 세포를 SaBE3 (4) 또는 SaBE4 및 FANCF, HEK3, 또는 HEK4 유전자좌를 표적화하는 sgRNA를 코딩하는 플라스미드로 형질감염시켰다. BE4 및 BE3을 비교하는 결과와 일치하게, 본 발명자들은 SaBE3에 비해 SaBE4에 대한 모든 10개의 편집된 C에 걸쳐 1.4 ± 0.2배의 C에서 T로의 편집 효율의 평균 증가를 관찰하였고 (도 145a), 목적하지 않은 비-T 편집 생성물에서는 1.8 ± 0.5배 평균 감소를 관찰하였다 (도 145b). 이들 결과는 BE4 증진으로부터 발생하는 염기 편집 효율 및 생성물 순도의 이득이 또한 다른 Cas9 상동체로부터 유래된 염기 편집제에도 적용된다는 것을 나타낸다.

[0963] 물질 및 방법

[0964] 플라스미드의 클로닝

[0965] 본 연구에서 모든 플라스미드는 퓨션 U 핫 스타트 폴리머라제 (써모 피셔)를 사용하여 USER 클로닝에 의해 생성하였다. 데아미나제 및 SSB 유전자를 g블록스 유전자 단편 (인테그레이티드 DNA 테크놀로지스)으로서 합성하고, 표적-AID를 애드진으로부터 입수하였다 (플라스미드 # 79620). 단백질 서열은 보충 노트에 열거된다.

[0966] 세포 배양

[0967] HEK293T (ATCC CRL-3216) 세포를 10% (v/v) 소 태아 혈청 (FBS)이 보충된 둘베코 변형 이글 배지 + 글루타맥스 (써모피셔) 중에서 37℃에서 5% CO_2 하에 유지하였다. HAP1 (호리즌 디스커버리(Horizon Discovery) C631) 및 HAP1 UNG^- (호리즌 디스커버리 HZGHC001531c012)를 10% (v/v) 소 태아 혈청 (FBS)이 보충된 이스코브 변형 둘베코 배지 + 글루타맥스 (써모피셔 사이언티픽) 중에서 37℃에서 5% CO_2 하에 유지하였다.

[0968] 형질감염

[0969] HEK293T 세포를 48-웰 콜라겐-코팅된 바이오코트 플레이트 (코닝) 상에 시딩하고, 대략 75% 전면생장률로 형질감염시켰다. 간략하게, 750 ng의 BE 및 250 ng의 sgRNA 발현 플라스미드를 제조업체의 프로토콜에 따라 웰당 1.5 μL의 리포펙타민 2000 (써모피셔 사이언티픽)을 사용하여 형질감염시켰다.

[0970] HAP1 및 HAP1 UNG^- 세포를 SE 세포주 4D뉴클레오펙터™ X 키트 S (론자(Lonza))를 사용하여 제조업체의 프로토콜에 따라 뉴클레오펙션시켰다. 간략하게, 4 x 10^5개의 세포를 4D뉴클레오펙터™ 프로그램 DZ-113을 사용하여 300 ng의 BE 및 100 ng의 sgRNA 발현 플라스미드로 뉴클레오펙션시켰다.

[0971] 게놈 DNA 샘플의 고처리량 DNA 서열분석

[0972] 형질감염된 세포를 3일 후에 수거하고, 세포를 용해 완충제 (10 mM 트리스-HCl pH 8.0, 0.05% SDS, 25 μg/mL 프로테이나제 K) 중에서 37℃에서 1시간 동안, 이어서 80℃에서 30분 동안 인큐베이션함으로써 게놈 DNA를 단리하였다. 관심 게놈 영역을 이전에 기재된 바와 같이 플랭킹 HTS 프라이머 쌍을 사용하여 PCR에 의해 증폭시켰다 (6, 1). PCR 증폭을 제조업체의 지침에 따라, 및 이전에 기재된 바와 같이 퓨션 고충실도 DNA 폴리머라제 (써모피셔)를 사용하여 수행하였다. 정제된 DNA를 서열분석 어댑터를 함유하는 프라이머를 사용하여 PCR에 의해 증폭시켰다. 생성물을 겔-정제하고, 퀀티T™ 피코그린 dsDNA 검정 키트 (써모피셔) 및 카파 라이브러리 정량화 키트-일루미나 (카파 바이오시스템즈)를 사용하여 정량화하였다. 샘플을 이전에 기재된 바와 같이 일루미나 MiSeq 상에서 서열분석하였다.

[0973] 데이터 분석

[0974] 서열분석 판독물을 MiSeq 리포터 (일루미나)를 사용하여 자동으로 디멀티플렉싱하고, 개별 FASTQ 파일을 이전에 기재된 바와 같이 주문제작 매트랩 스크립트로 분석하였다 (1). 각각의 판독물을 스미스-워터만 알고리즘을 사용하여 적절한 참조 서열에 쌍별 정렬시켰다. 31 미만의 Q-점수를 갖는 염기 호출을 N으로 대체하고, 따라서 뉴클레오티드 빈도를 계산하는 것에서 제외하였다. 이러한 처리는 대략 1,000 중 1의 예상된 MiSeq 염기-호출 오류율을 생성한다. 판독 및 참조 서열이 갭을 함유하지 않는 정렬된 서열을, 염기 빈도가 각각의 유전자좌에 대해 표로 제시될 수 있는 정렬 표에 저장하였다.

[0975] indel 빈도는 이전에 기재된 매트랩 스크립트로 정량화하였다 (5, 6, 1). 간략하게, 서열분석 판독물을 indel이 발생할 수 있는 윈도우의 양 측에 플랭킹하는 2개의 10-bp 서열에 대한 정확한 매치에 대해 스캐닝하였다. 정확한 매치가 위치하지 않으면, 판독물을 분석으로부터 제외하였다. 이러한 indel 윈도우의 길이가 참조 서열과 정확하게 매칭되는 경우, 판독물을 indel을 함유하지 않는 것으로 분류하였다. indel 윈도우가 참조 서열보다 2개 이상의 염기가 더 길거나 또는 더 짧은 경우, 서열분석 판독물은 각각 삽입 또는 결실로서 분류되었다.

[0976] 편집 윈도우 내의 다중 표적 시티딘에서의 염기 편집 결과 사이의 상호의존성 (연결 불균형)을 평가하기 위해, BE 처리된 세포로부터의 표적 부위 서열을 주문제작 피톤 스크립트에 의해 분석하였다 (보충 노트 1). 간략하게, 서열분석 판독물을 프로토스페이서의 각 측에 플랭킹하는 2개의 7-bp 서열에 대한 정확한 매치에 대해 스캐닝하였다. 개재 영역이 정확히 20-bp가 아닌 경우, 이를 추가의 분석에서 제외하였다. 프로토스페이서 서열을 가장 비-T 편집 결과 (1차 표적 C)를 갖는 위치에서 뉴클레오티드의 정체에 기초하여 4개의 군으로 추가로 필터링하였다. 이들 4개의 군 각각 뿐만 아니라 전체 풀에 대해, 프로토스페이서 내의 20개의 위치 각각에서 뉴클레오티드 존재비를 집계하였다.

[0977] 실시예 18: LbCpf1 (뉴클레아제 데드, 뉴클레아제 활성 및 니카제)을 포함하는 염기 편집제.

[0978] 상기 논의된 바와 같이, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능한 DNA 결합 단백질 (napDNAbp)은 LbCpf1 단백질일 수 있다. 일부 실시양태에서, LbCpf1 단백질은 뉴클레아제 불활성, 뉴클레아제 활성, 또는 LbCpf1 니카제이다. LbCpf1의 형태를 포함하는 융합 단백질의 여러 구축물을 상이한 표적 서열에서 C에서 T로의 편집을 생성하는 그의 능력에 대해 시험하였다. 시험된 구축물의 개략적 표현이 도 152에 나타나 있다. 구축물 10은 [Apobec]-[LbCpf1]-[UGI]-[UGI]의 도메인 배열을 갖고; 구축물 11은 [Apobec]-[LbCpf1]-[UGI]의 도메인 배열을 갖고; 구축물 12는 [UGI]-[Apobec]-[LbCpf1]의 도메인 배열을 갖고; 구축물 13은 [Apobec]-[UGI]-[LbCpf1]의 도메인 배열을 갖고; 구축물 14는 [LbCpf1]-[UGI]-[Apobec]의 도메인 배열을 갖고; 구축물 15는 [LbCpf1]-[Apobec]-[UGI]의 도메인 배열을 갖는다. 각각의 구축물에 대해, 3개의 상이한 LbCpf1 단백질 (D/N/A, 이는 뉴클레아제 데드 LbCpf1 (D); LbCpf1 니카제 (N) 및 뉴클레아제 활성 LbCpf1 (A)을 지칭함)을 사용하였다. 각각의 이들 구축물에 대해, 도메인을 연결하는 링커가 하기에 제시되고, 여기서 XTEN은 서열 SGSETPGTSESATPES (서열식별번호: 604)를 갖는 XTEN 링커를 지칭하고, BPNLS는 서열 KRTADGSEFEPKKKRKV (서열식별번호: 740)를 갖는 핵 국재화 서열을 지칭한다. 구축물은 N-말단 (좌측)에서 C-말단 (우측)으로 제시된다.

구축물 10: Apobec-SGGSSGGS*XTEN*SGGSSGGS-LbCpf1-SGGSGGSGGS-UGI-SGGSGGSGGS-UGI-SGGS-BPNLS

구축물 11: Apobec-SGGSSGGS*XTEN*SGGSSGGS-LbCpf1-SGGSGGSGGS-UGI-SGGS-BPNLS

구축물 12: UGI-SGGSGGSGGS-Apobec-SGGSSGGS*XTEN*SGGSSGGS-LbCpf1-SGGS-BPNLS

구축물 13: Apobec-SGGSGGSGGS-UGI-SGGSSGGS*XTEN*SGGSSGGS-LbCpf1-SGGS-BPNLS

구축물 14: LbCpf1-SGGSGGSGGS-UGI-SGGSGGSGGS-Apobec-SGGS-BPNLS

구축물 15: LbCpf1-SGGSGGSGGS-Apobec-SGGSGGSGGS-UGI-SGGS-BPNLS

[0979]

[0980] 실험에서 사용된 공통 가이드 백본 (sgRNA)은 GTAATTTCTACTAAGTGTAGAT (서열식별번호: 741)[가이드 서열]TTTTTTT이고, 여기서 서열식별번호: 741의 각각의 T는 우라실 (U)이고, 구축물을 특정 뉴클레오티드 서열로 표적화하는 가이드 서열은 괄호 사이에 제시된다. 일부 실시양태에서, 본원에 제공된 구축물 중 임의의 것은 GTAATTTCTACTAAGTGTAGAT (서열식별번호: 741)의 백본 서열을 포함하는 sgRNA와 복합체화되고, 여기서 서열식별번호: 741의 각각의 T는 우라실 (U)이다. 일부 실시양태에서, 본원에 제공된 가이드 RNA 중 임의의 것은 서열, 예를 들어 표적 DNA 서열에 완벽하게 상보적인 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개 뉴클레오티드를 포함하는 가이드 서열을 포함한다. 수행된 실험에서, 시험된 가이드 서열이 하기에 제시된다:

EMX23: TACTTTGTCCTCCGGTTCTGGAA (SEQ ID NO: 742)

EMX20: TACTTTGTCCTCCGGTTCTG (SEQ ID NO: 743)

EMX19: TACTTTGTCCTCCGGTTCT (SEQ ID NO: 744)

EMX18: TACTTTGTCCTCCGGTTC (SEQ ID NO: 745)

EMX17: TACTTTGTCCTCCGGTT (SEQ ID NO: 746)

Hek2_23: CAGCCCGCTGGCCCTGTAAAGGA (SEQ ID NO: 747)

Hek2_20: CAGCCCGCTGGCCCTGTAAA (SEQ ID NO: 748)

Hek2_19: CAGCCCGCTGGCCCTGTAA (SEQ ID NO: 749)

Hek2_18: CAGCCCGCTGGCCCTGTA (SEQ ID NO: 750)

[0981] Hek2_17: CAGCCCGCTGGCCCTGT (SEQ ID NO: 751)

[0982] 다양한 구축물 및 표적 서열을 사용한 C에서 T로의 염기 쌍 편집 백분율을 입증하는 데이터가 도 153-159에 제시되고, 편집 백분율 값 (indel 카운트에 기초한 조정 후), 및 실험을 위한 indel의 백분율이 도 160-166에 제시된다.

[0983] 보충 서열

[0984] CDA1-BE3, AID-BE3, BE4, 및 SaBE4 융합 단백질의 아미노산 서열.

[0985] *CDA1*-**XTEN**-dCas9-**UGI**-*NLS* 1차 서열 (CDA1-BE3):

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTE
RGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACK
LYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKR
*RSELSIMIQVKILHTTKSPAV***SGSETPGTSESATPES**DKKYSIGLAIGTNSVGWAVITDEYK

VPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGV
DAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGT
EELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRY
TSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESILMLP**
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIK
MLSGGS*PKKKRKV* (SEQ ID NO: 165)

AID-**XTEN**-dCas9-**UGI**-*NLS* 1차 서열 (AID-BE3):

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPE
GLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVD
*DLRDAFRTLGL***SGSETPGTSESATPES**DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKV
LGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSA
RLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY
EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER
LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY

[0986]

LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGDSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN**
KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGS*PKKK*
RKV (SEQ ID NO: 166)

rAPOBEC1-링커-dCas9-**UGI**-***UGI***-*NLS* 1차 서열 (BE4):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVE
VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNR
QGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC
*LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK***SGGSSGGSSGSETPGTSESATPESS**
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL
AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGDSGGSGGSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTA**
YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGS***TNLSDIIEK***
ETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLSGGS*PKKKRK* (SEQ ID NO: 167)

[0987]

rAPOBEC1-링커-SaCas9d-**UGI**-***UGI***-NLS 1차 서열 (SaBE4):

*MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVE VNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNR QGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK***SGGSSGGSSGSETPGTSESATPESS GGSSGGSG**KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLA KRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFK TSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYE MLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQ KKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIA KILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQ IAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL ARENKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSK ISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMN LLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVD KKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDP QTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLD ITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEA KKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMND KRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGGSPKKKRKVSSDYKDHDG DYKDHDIDYKDDDDKSGGSGGSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIG NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML**SGGSGG SGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLL TSDAPEYKPWALVIQDSNGENKIKML***SGGS*PKKKRKV* (SEQ ID NO: 168)

[0988]

[0989] 실시예 17에 대한 참고문헌

1. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424 (2016).

2. G. T. Hess et al., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods 13, 1036-1042 (2016).

3. K. Kim et al., Highly efficient RNA-guided base editing in mouse embryos. Nat Biotechnol 35, 435-437 (2017).

4. Y. B. Kim et al., Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol 35, 371-376 (2017).

[0990]

5. C. Kuscu et al., CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat Meth advance online publication, (2017).

6. J. Li, Y. Sun, J. Du, Y. Zhao, L. Xia, Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System. Mol Plant 10, 526-529 (2017).

7. Y. Lu, J. K. Zhu, Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol Plant 10, 523-525 (2017).

8. Y. Ma et al., Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat Methods 13, 1029-1035 (2016).

9. K. Nishida et al., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, (2016).

10. H. A. Rees et al., Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun 8, 15790 (2017).

11. L. Yang et al., Engineering and optimising deaminase fusions for genome editing. Nat Commun 7, 13330 (2016).

12. Y. Zong et al., Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol 35, 438-440 (2017).

13. Z. Shimatani et al., Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat Biotech 35, 441-443 (2017).

14. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20-36 (2017).

15. A. J. Davis, D. J. Chen, DNA double strand break repair via non-homologous end-joining. Translational cancer research 2, 130-143 (2013).

16. M. M. Vilenchik, A. G. Knudson, Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. Proceedings of the National Academy of Sciences 100, 12871-12876 (2003).

17. F. Liang, M. Han, P. J. Romanienko, M. Jasin, Homology-directed repair is a major doublestrand break repair pathway in mammalian cells. Proceedings of the National Academy of Sciences 95, 5172-5177 (1998).

18. Y. Miyaoka et al., Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. Scientific Reports 6, 23549 (2016).

[0991]

19. M. M. Jore et al., Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol 18, 529-536 (2011).
20. L. H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily. Mutation Research/DNA Repair 460, 165-181 (2000).
21. W. Tang, J. H. Hu, D. R. Liu, Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. 8, 15939 (2017).
22. G. Saraconi, F. Severi, C. Sala, G. Mattiuz, S. G. Conticello, The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. Genome Biology 15, 417 (2014).
23. R. M. Kohli et al., Local Sequence Targeting in the AID/APOBEC Family Differentially Impacts Retroviral Restriction and Antibody Diversification. Journal of Biological Chemistry 285, 40956-40964 (2010).
24. L. Chelico, P. Pham, M. F. Goodman, Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 583-593 (2009).
25. E. A. Kouzminova, A. Kuzminov, Patterns of chromosomal fragmentation due to uracil-DNA incorporation reveal a novel mechanism of replication-dependent double-stranded breaks. Molecular Microbiology 68, 202-215 (2008).
26. F. A. Ran et al., In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186-191 (2015).
27. F. d. A. di Fagagna, G. R. Weller, A. J. Doherty, S. P. Jackson, The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. EMBO Reports 4, 47-52 (2003).
28. C. Shee et al., Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. eLife 2, e01222 (2013).

[0992]

[0993] 등가물 및 범위

[0994] 관련 기술분야의 통상의 기술자는 상용 실험을 사용하여 본원에 기재된 실시양태의 많은 등가물을 인식하거나, 확인할 수 있을 것이다. 본 개시내용의 범주는 상기 설명에 제한되는 것으로 의도되지 않지만, 오히려 첨부된 청구범위에 제시된 바와 같다.

[0995] 단수 형태는 달리 나타내거나 또는 달리 문맥으로부터 명백하지 않는 한 1개 또는 1개 초과를 의미할 수 있다. 군의 2개 이상의 구성원 사이에 "또는"을 포함하는 청구항 또는 설명은 달리 나타내거나 또는 달리 문맥으로부터 명백하지 않는 한 군 구성원 중 1개, 1개 초과, 또는 전부가 존재하는 경우 충족되는 것으로 간주된다. 2개 이상의 군 구성원 사이에 "또는"을 포함하는 군의 개시내용은 군의 정확하게 1개의 구성원이 존재하는 실시양태, 군의 1개 초과의 구성원이 존재하는 실시양태, 및 군 구성원의 전부가 존재하는 실시양태를 제공한다. 간결성의 목적으로 이들 실시양태는 본원에 개별적으로 설명되지 않았지만, 이들 실시양태 각각은 본원에 제공되며, 구체적으로 청구되거나 비청구될 수 있는 것으로 이해될 것이다.

[0996] 본 발명은 청구항 중 하나 이상으로부터 또는 설명의 하나 이상의 관련 부분으로부터, 하나 이상의 제한, 요소, 조항, 또는 설명적 용어가 또 다른 청구항 내로 도입되는 모든 변형, 조합, 및 순열을 포괄하는 것으로 이해되어야 한다. 예를 들어, 또 다른 청구항에 종속되는 청구항은 동일한 기초 청구항에 종속되는 임의의 다른 청구항에서 발견되는 제한 중 하나 이상을 포함하도록 변형될 수 있다. 게다가, 청구항이 조성물을 인용하는 경우, 달리 나타내거나 또는 모순 또는 불일치가 발생할 것이 관련 기술분야의 통상의 기술자에게 명백하지 않는 한, 본원에 개시된 제조 또는 사용 방법 중 임의의 것에 따라 또는 존재하는 경우 관련 기술분야에 공지된 방법에 따라 조성물을 제조하거나 사용하는 방법이 포함되는 것으로 이해되어야 한다.

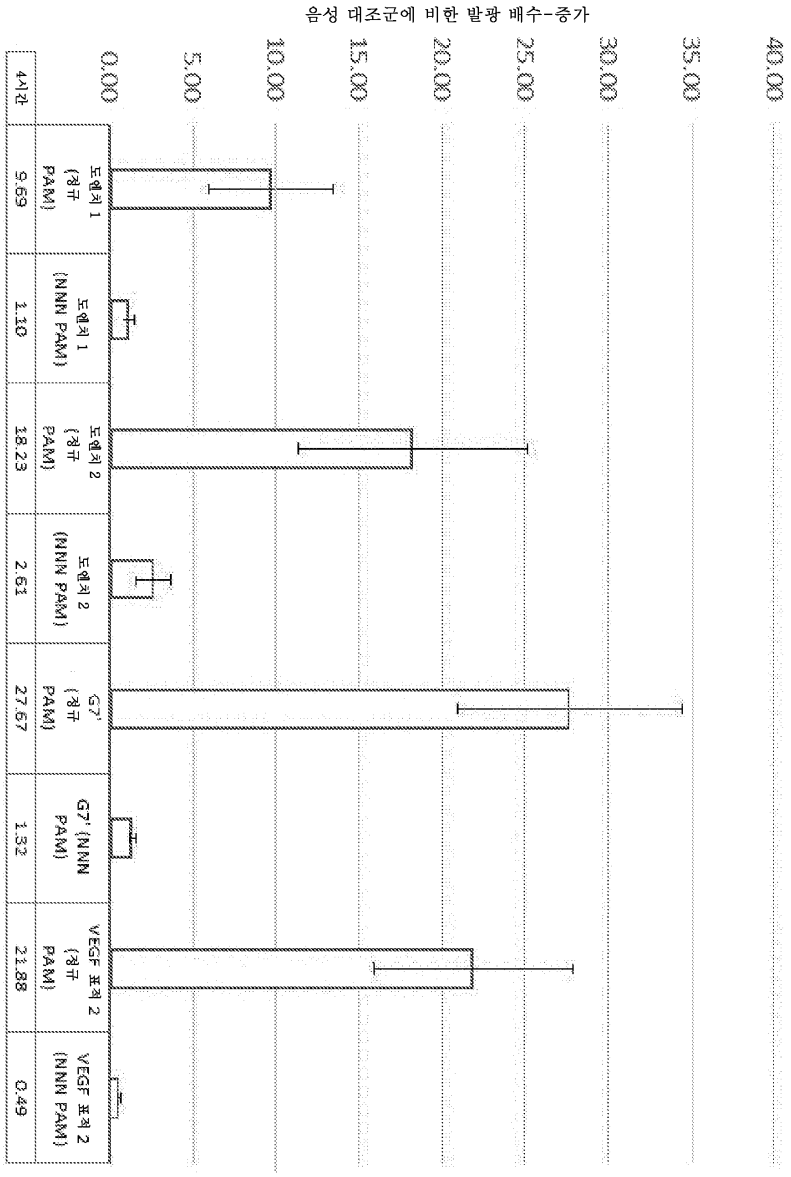
[0997] 요소가 목록으로서, 예를 들어 마쿠쉬 군 포맷으로 제공되는 경우, 요소의 모든 가능한 하위군이 또한 개시되고, 임의의 요소 또는 요소의 하위군은 군으로부터 제거될 수 있는 것으로 이해되어야 한다. 또한, 용어 "포함하는"은 개방적이며, 추가의 요소 또는 단계의 포함다는 것을 허용하는 것으로 의도되는 것에 주목한다. 일반적으로, 실시양태, 생성물, 또는 방법이 특정한 요소를 포함하는 것으로 언급되는 경우, 이러한 요소,

특색, 또는 단계로 이루어지거나, 이로 본질적으로 이루어지는 특정한 요소, 특색, 또는 단계, 실시양태, 생성물, 또는 방법이 또한 제공되는 것으로 이해되어야 한다. 간결성의 목적으로 이들 실시양태는 본원에 개별적으로 설명되지 않았지만, 이들 실시양태 각각은 본원에 제공되며, 구체적으로 청구되거나 비청구될 수 있는 것으로 이해될 것이다.

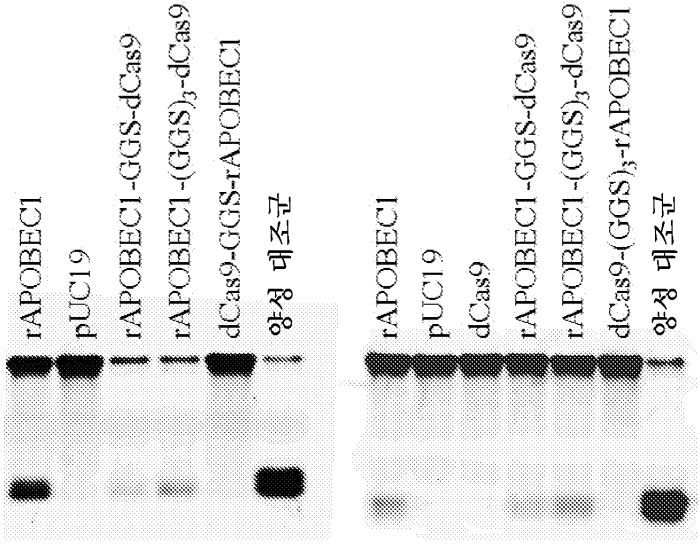
[0998] 범위가 주어지는 경우, 종점이 포함된다. 게다가, 달리 나타내거나 또는 달리 문맥 및/또는 관련 기술분야의 통상의 기술자의 이해로부터 명백하지 않는 한, 범위로서 표현된 값은 일부 실시양태에서 언급된 범위 내의 임의의 특정 값을, 문맥이 달리 명백하게 나타내지 않는 한, 범위의 하한의 단위의 1/10까지 가정할 수 있다는 것이 이해되어야 한다. 간결성의 목적으로 각각의 범위 중의 값은 본원에 개별적으로 설명되지 않았지만, 이들 값 각각은 본원에 제공되며, 구체적으로 청구되거나 비청구될 수 있는 것으로 이해될 것이다. 또한, 달리 나타내거나 또는 달리 문맥 및/또는 관련 기술분야의 통상의 기술자의 이해로부터 명백하지 않는 한, 범위로서 표현된 값은 주어진 범위 내의 임의의 하위범위를 가정할 수 있으며, 여기서, 하위범위의 종점은 범위의 하한의 단위의 1/10만큼 동일한 정도의 정확성으로 표현되는 것으로 이해되어야 한다. 또한, 본 발명의 임의의 특정한 실시양태는 청구항 중 어느 하나 이상으로부터 명백히 제외될 수 있는 것으로 이해되어야 한다. 범위가 주어지는 경우, 범위 내의 임의의 값은 청구항 중 어느 하나 이상으로부터 명백히 제외될 수 있다. 본 발명의 조성물 및/또는 방법의 임의의 실시양태, 요소, 특색, 적용, 또는 측면은 임의의 하나 이상의 청구항으로부터 제외될 수 있다. 간결성의 목적으로, 하나 이상의 요소, 특색, 목적, 또는 측면이 제외되는 모든 실시양태는 본원에 명백하게 제시되지 않는다.

도면

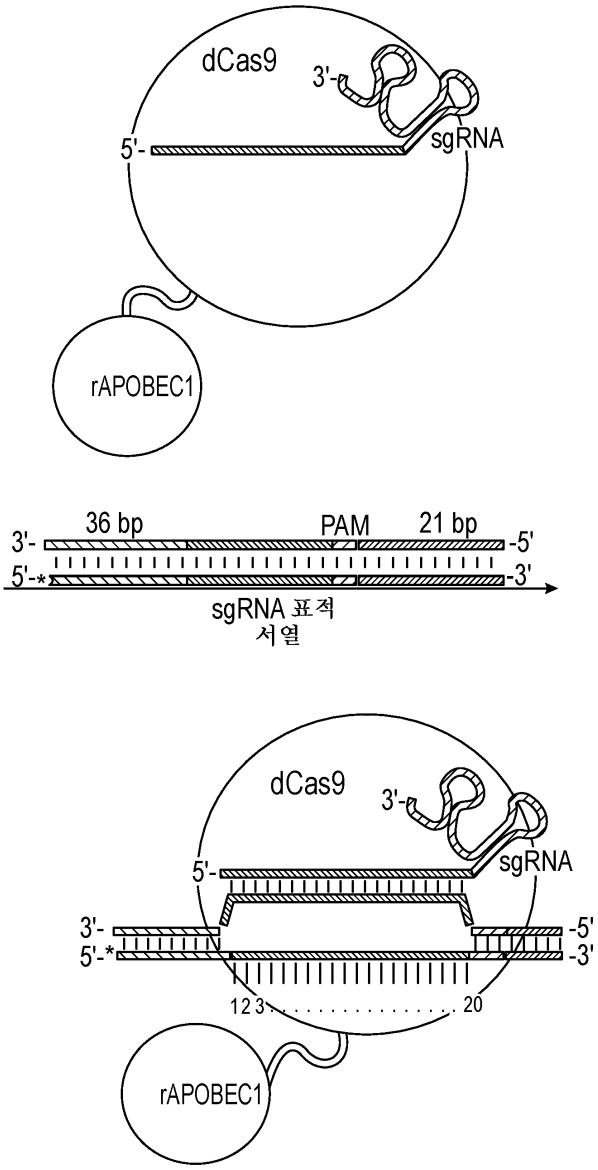
도면1


음성 대조군에 비한 발광 배수-증가
0.00
5.00
10.00
15.00
20.00
25.00
30.00
35.00
40.00
4시간
9.69
1.10
18.23
2.61
27.67
1.32
21.88
0.49
도엠치 1 (정규 PAM)
도엠치 1 (NNN PAM)
도엠치 2 (정규 PAM)
도엠치 2 (NNN PAM)
G7' (정규 PAM)
G7' (NNN PAM)
VEGF 표적 2 (정규 PAM)
VEGF 표적 2 (NNN PAM)

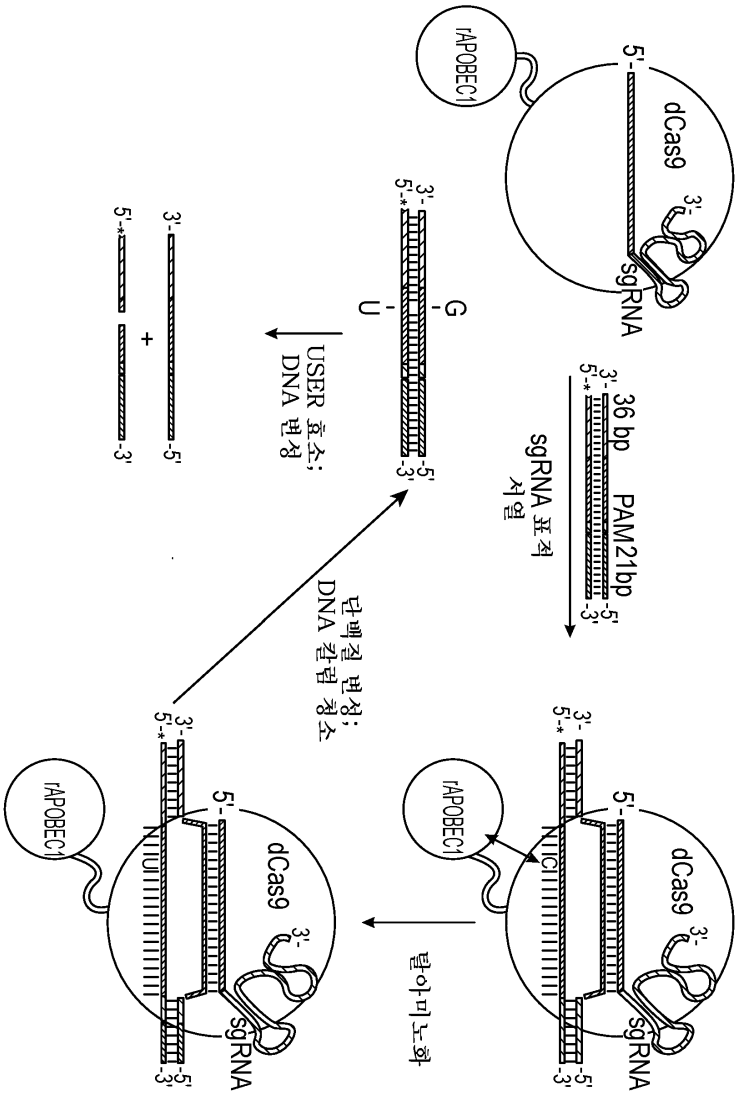
도면2


rAPOBEC1
pUC19
rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
dCas9-GGS-rAPOBEC1
양성 대조군
rAPOBEC1
pUC19
dCas9
rAPOBEC1-GGS-dCas9
rAPOBEC1-(GGS)3-dCas9
dCas9-(GGS)3-rAPOBEC1
양성 대조군

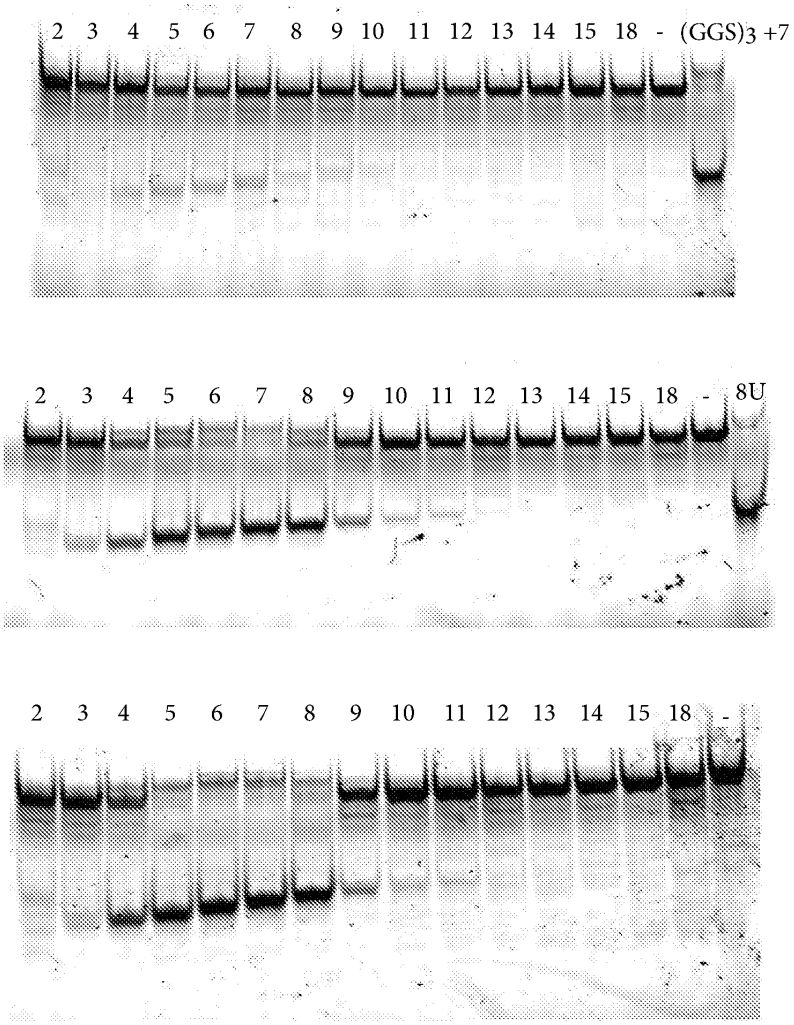
도면3


dCas9
3'-
5'-
sgRNA
rAPOBEC1
36 bp
PAM
21 bp
3'-
-5'
5'-*
-3'
sgRNA 표적
서열
dCas9
3'-
5'-
sgRNA
3'-
-5'
5'-*
-3'
1 2 3
20
rAPOBEC1

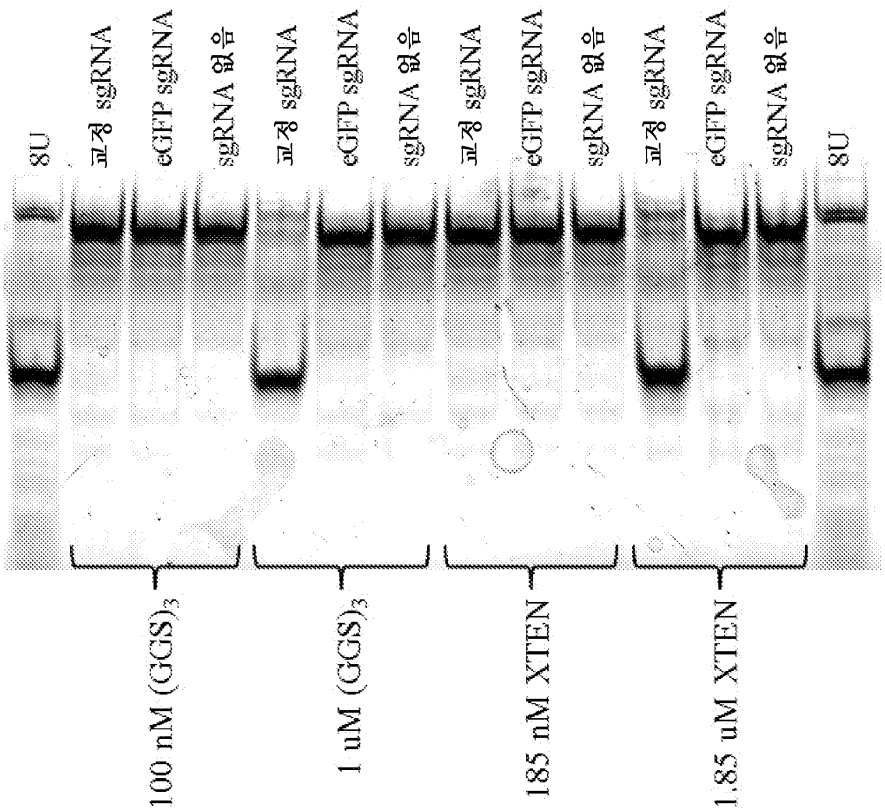
도면4


rAPOBEC1
dCas9
sgRNA
36 bp
PAM21bp
sgRNA 표적 서열
USER 효소; DNA 변성
단백질 변성; DNA 칼럼 정소
탈아미노화

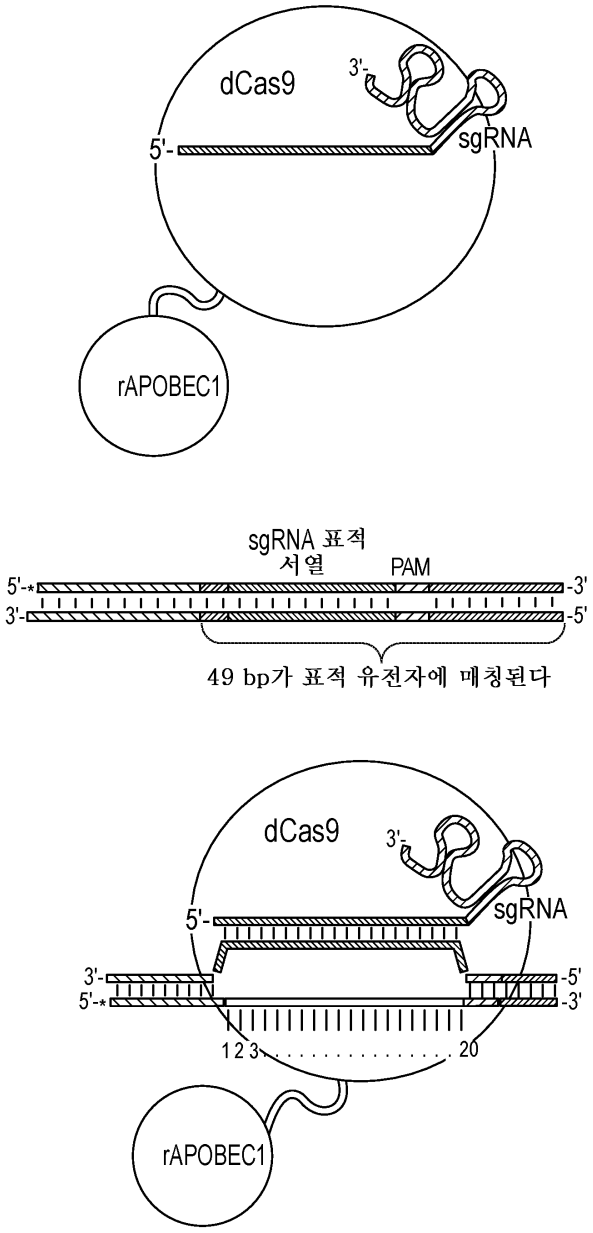
도면5


2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 - (GGS)3 +7
2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 - 8U
2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 -

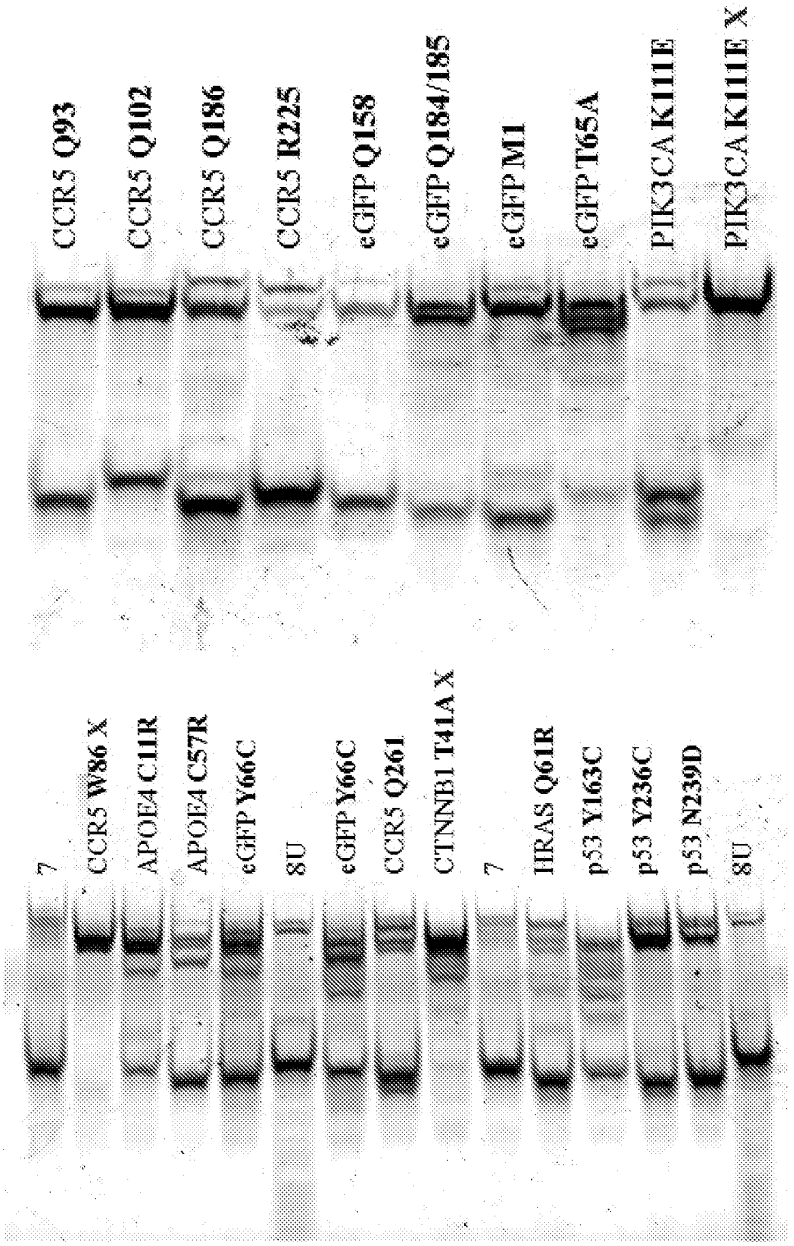
도면6


8U
교정 sgRNA
eGFP sgRNA
sgRNA 없음
교정 sgRNA
eGFP sgRNA
sgRNA 없음
교정 sgRNA
eGFP sgRNA
sgRNA 없음
교정 sgRNA
eGFP sgRNA
sgRNA 없음
8U
100 nM (GGS)3
1 uM (GGS)3
185 nM XTEN
1.85 uM XTEN

도면7


dCas9
3'-
5'-
sgRNA
rAPOBEC1
sgRNA 표적
서열
PAM
5'-*
-3'
3'-
-5'
49 bp가 표적 유전자에 매칭된다
dCas9
3'-
5'-
sgRNA
3'-
-5'
5'-*
-3'
1 2 3 20
rAPOBEC1

도면8


7
CCR5 W86 X
APOE4 C11R
APOE4 C57R
eGFP Y66C
8U
eGFP Y66C
CCR5 Q261
CTNNB1 T41A X
7
HRAS Q61R
p53 Y163C
p53 Y236C
p53 N239D
8U
CCR5 Q93
CCR5 Q102
CCR5 Q186
CCR5 R225
eGFP Q158
eGFP Q184/185
eGFP M1
eGFP T65A
PIK3CA K111E
PIK3CA K111E X

도면9

EMX1		C_5	C_6	C_{10}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.8%
	G	0.0%	0.0%	0.1%
	T	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.0%	0.1%
	C	60.4%	61.0%	99.1%
	G	0.0%	0.0%	0.1%
	T	39.5%	39.0%	0.7%

FANCF		C_6	C_7	C_8	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	99.8%	99.8%	99.9%	99.9%
	G	0.0%	0.1%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	63.9%	64.7%	65.0%	72.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	36.0%	35.1%	34.9%	27.3%

HEK293 부위 2		C_4	C_6	C_{11}
0 uM XTEN	A	0.1%	0.1%	0.1%
	C	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%
1.85 uM XTEN	A	0.1%	0.1%	0.1%
	C	80.6%	76.9%	99.6%
	G	0.0%	0.0%	0.0%
	T	19.3%	22.9%	0.3%

HEK293 부위 3		C_3	C_4	C_5	C_9
0 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	99.8%	99.9%	99.9%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.0%	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.0%	0.1%
	C	92.2%	74.8%	71.5%	96.6%
	G	0.0%	0.0%	0.0%	0.0%
	T	7.7%	25.1%	28.5%	3.3%

HEK293 부위 4		C_3	C_5	C_8	C_{11}
0 uM XTEN	A	0.1%	0.0%	0.1%	0.0%
	C	99.8%	99.9%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.1%	0.0%
1.85 uM XTEN	A	0.1%	0.1%	0.1%	0.1%
	C	98.8%	60.1%	97.0%	99.4%
	G	0.0%	0.0%	0.0%	0.0%
	T	1.1%	39.8%	2.9%	0.5%

RNF2		C_3	C_6
0 uM XTEN	A	0.1%	0.0%
	C	99.9%	99.9%
	G	0.0%	0.0%
	T	0.0%	0.0%
1.85 uM XTEN	A	0.1%	0.0%
	C	59.1%	57.8%
	G	0.0%	0.0%
	T	40.8%	42.1%

도면10

EMX1		C_5	C_6	C_{10}
비처리	A	0.0%	0.0%	0.0%
	C	99.5%	99.7%	100.0%
	G	0.0%	0.1%	0.0%
	T	0.5%	0.2%	0.0%
XTEN	A	0.7%	0.5%	0.0%
	C	93.5%	95.8%	100.0%
	G	2.1%	0.3%	0.0%
	T	3.6%	3.3%	0.0%
XTEN-UGI	A	0.2%	0.0%	0.0%
	C	81.8%	82.5%	100.0%
	G	0.6%	0.3%	0.0%
	T	17.4%	17.1%	0.0%

FANCF		C_6	C_7	C_8	C_{11}
비처리	A	0.0%	0.0%	0.2%	0.1%
	C	99.9%	99.8%	99.8%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.1%	0.1%	0.0%	0.0%
XTEN	A	0.3%	0.1%	0.0%	0.0%
	C	98.1%	99.2%	99.0%	99.8%
	G	0.4%	0.0%	0.0%	0.0%
	T	1.2%	0.7%	1.0%	0.2%
XTEN-UGI	A	0.0%	0.0%	0.1%	0.0%
	C	93.2%	93.5%	93.4%	98.2%
	G	0.0%	0.0%	0.0%	0.0%
	T	6.7%	6.5%	6.5%	1.8%

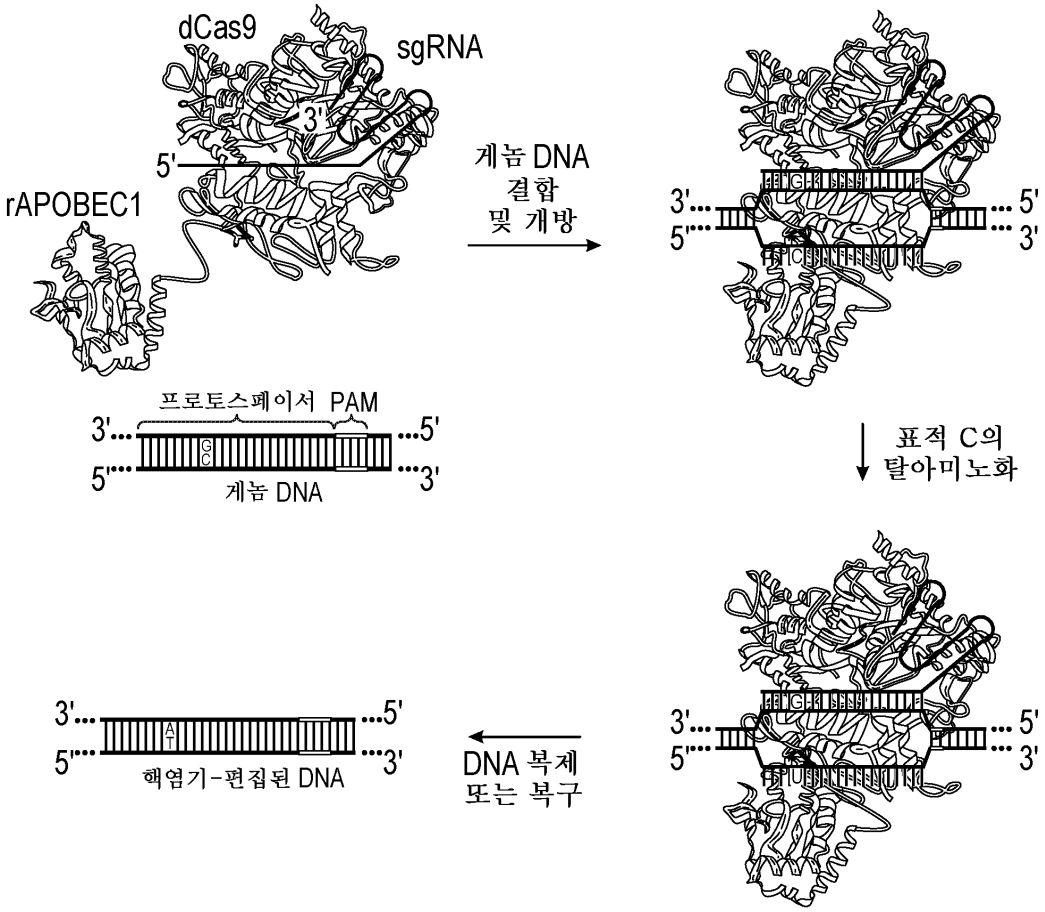
HEK293부위2		C_4	C_6	C_{11}
비처리	A	0.3%	0.2%	0.2%
	C	99.7%	99.7%	99.7%
	G	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%
XTEN	A	0.3%	0.3%	0.3%
	C	99.7%	99.4%	99.7%
	G	0.0%	0.3%	0.0%
	T	0.0%	0.0%	0.0%
XTEN-UGI	A	0.3%	0.2%	0.2%
	C	98.8%	98.2%	99.8%
	G	0.0%	0.3%	0.0%
	T	0.9%	1.3%	0.0%

HEK293부위2		C_3	C_4	C_5	C_9
비처리	A	0.0%	0.0%	0.0%	0.0%
	C	100.0%	100.0%	100.0%	99.9%
	G	0.0%	0.0%	0.0%	0.0%
	T	0.0%	0.0%	0.0%	0.1%
XTEN	A	0.0%	0.6%	0.3%	0.1%
	C	100.0%	95.8%	95.8%	99.2%
	G	0.0%	0.2%	0.7%	0.4%
	T	0.0%	3.4%	3.2%	0.3%
XTEN-UGI	A	0.0%	0.3%	0.3%	0.0%
	C	96.8%	83.0%	79.2%	98.5%
	G	0.0%	0.0%	1.1%	0.2%
	T	3.2%	16.8%	19.4%	1.3%

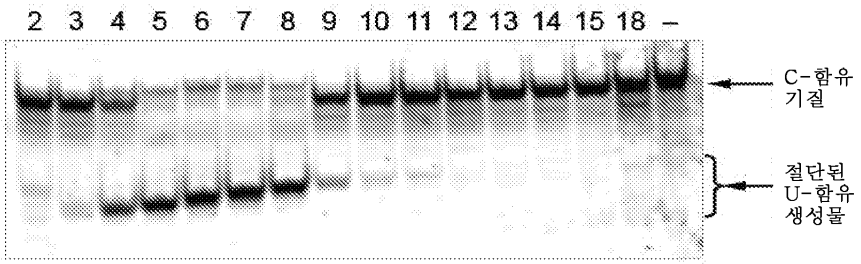
HEK293부위4		C_3	C_5	C_8	C_{11}
비처리	A	0.0%	0.4%	0.0%	0.0%
	C	99.8%	97.6%	99.9%	100.0%
	G	0.0%	1.0%	0.0%	0.0%
	T	0.2%	1.0%	0.0%	0.0%
XTEN	A	0.0%	1.1%	0.0%	0.0%
	C	99.6%	92.2%	99.9%	100.0%
	G	0.0%	2.2%	0.0%	0.0%
	T	0.4%	4.5%	0.0%	0.0%
XTEN-UGI	A	0.0%	0.5%	0.0%	0.0%
	C	99.4%	86.7%	99.1%	100.0%
	G	0.0%	1.8%	0.0%	0.0%
	T	0.6%	11.0%	0.9%	0.0%

RNF2		C_3	C_6
비처리	A	0.0%	0.0%
	C	99.9%	99.5%
	G	0.0%	0.2%
	T	0.0%	0.3%
XTEN	A	0.0%	0.0%
	C	99.8%	99.3%
	G	0.0%	0.2%
	T	0.2%	0.5%
XTEN-UGI	A	0.0%	0.0%
	C	99.6%	99.1%
	G	0.0%	0.4%
	T	0.4%	0.5%

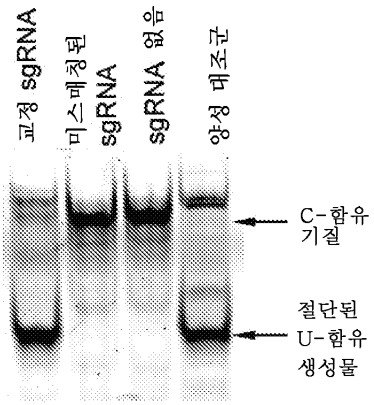
도면11a


dCas9
sgRNA
3'
5'
rAPOBEC1
게놈 DNA
결합
및 개방
프로토스페이서 PAM
3'
5'
게놈 DNA
표적 C의
탈아미노화
핵염기-편집된 DNA
DNA 복제
또는 복구

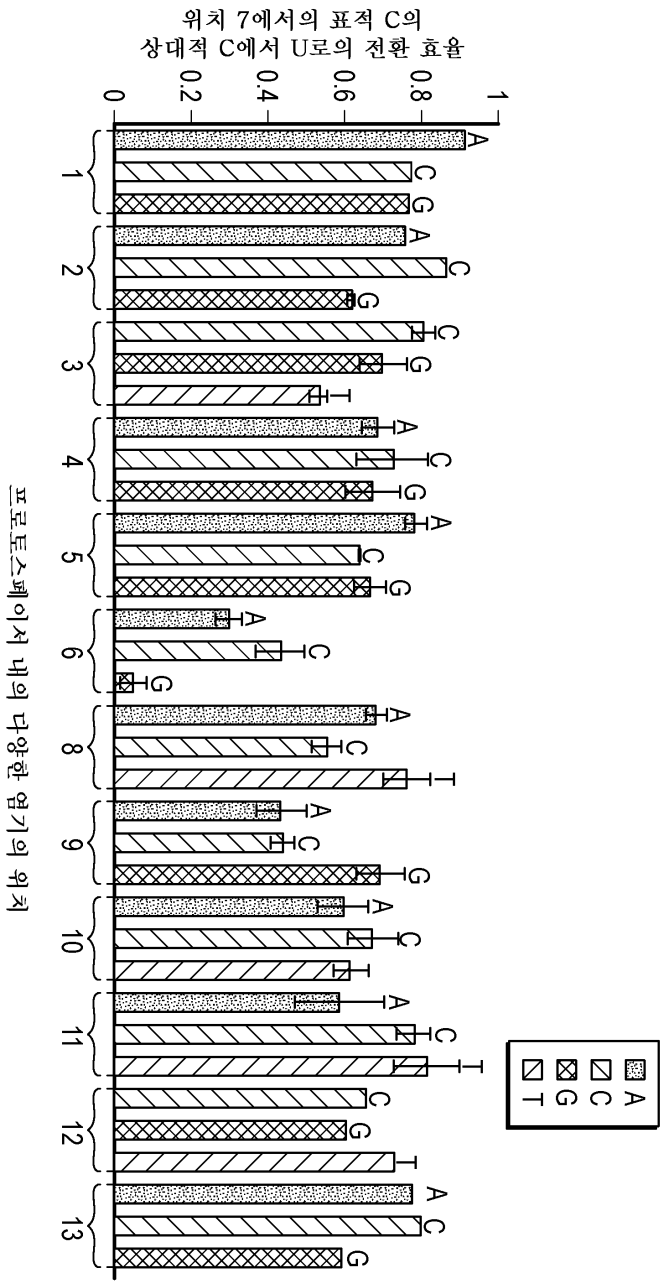
도면11b


2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 –
C-함유
기질
절단된
U-함유
생성물

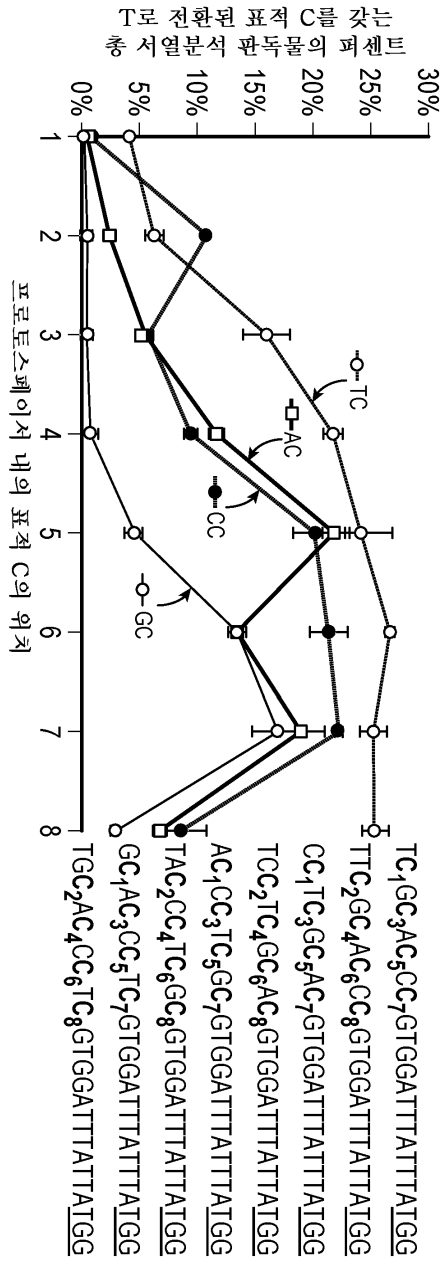
도면11c



도면12a


위치 7에서의 표적 C의
상대적 C에서 U로의 전환 효율
0
0.2
0.4
0.6
0.8
1
프로토스페이서 내의 다양한 염기의 위치
1
2
3
4
5
6
8
9
10
11
12
13
A
C
G
T

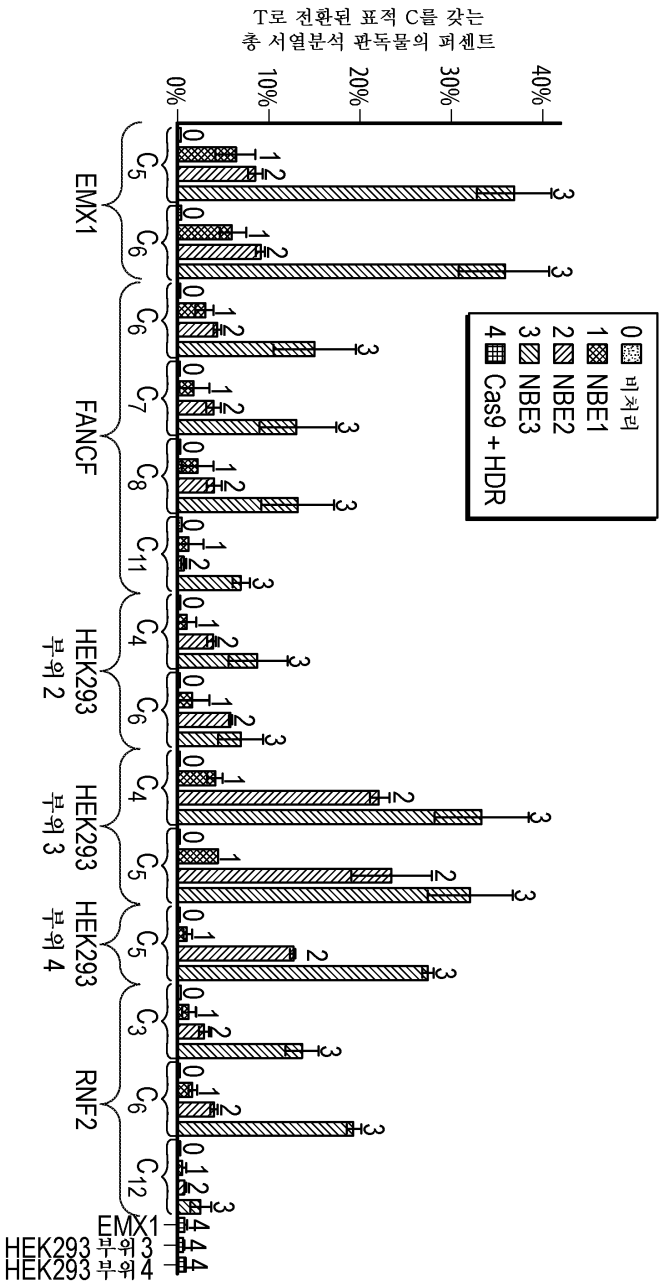
도면12b



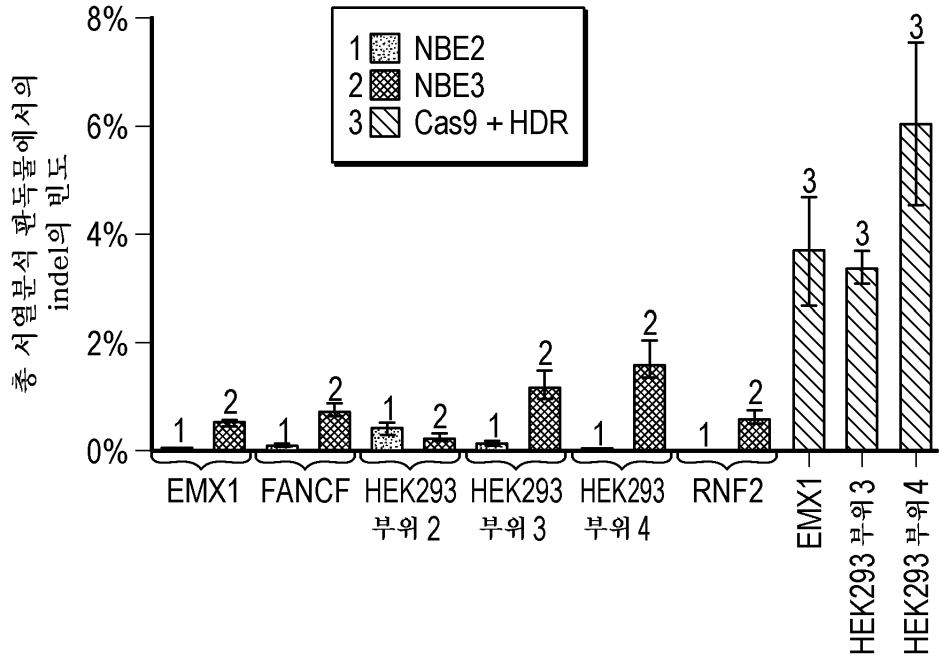
도면13a

EMX1:	GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG
FANCF:	GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG
HEK293 부위 2:	GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK293 부위 3:	GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK293 부위 4:	GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2:	GTC$_3$ATC$_6$TTAGTC$_{12}$ATTACCTGAGG

도면13b


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0%
10%
20%
30%
40%
0 비처리
1 NBE1
2 NBE2
3 NBE3
4 Cas9 + HDR
EMX1
FANCF
HEK293 부위 2
HEK293 부위 3
HEK293 부위 4
RNF2
C5
C6
C7
C8
C11
C4
C3
C12

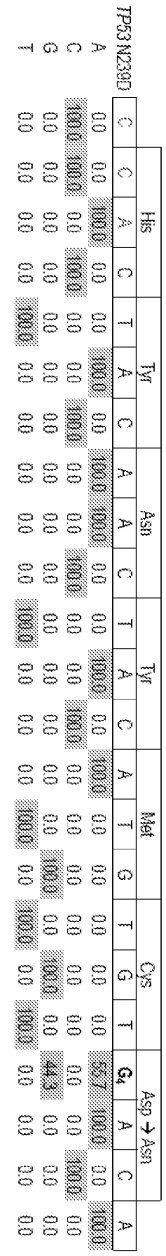
도면13c


총 서열분석 판독물에서의 indel의 빈도
8%
6%
4%
2%
0%
1 NBE2
2 NBE3
3 Cas9 + HDR
EMX1
FANCF
HEK293 부위 2
HEK293 부위 3
HEK293 부위 4
RNF2
EMX1
HEK293 부위 3
HEK293 부위 4

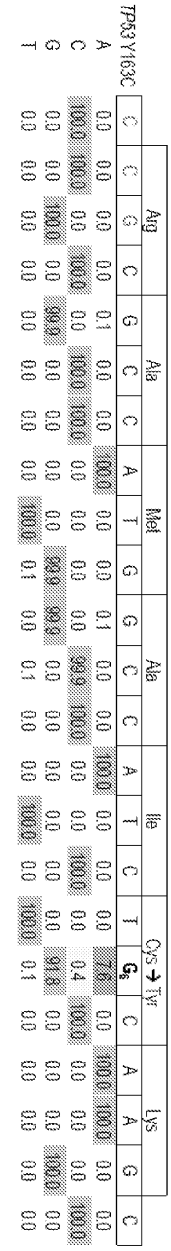
도면14a

APOE4 C158R	G	Lys			Arg→Cys			Leu→Leu			Ala			Val			Tyr			Gln		
		A	A	G	G	C	C	T	G	G	C	A	G	T	G	T	A	C	C	A	G	G
A	0.0	100.0	100.0	0.0	0.2	0.0	0.1	0.3	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	99.7	0.0	91.3	91.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0
G	100.0	0.0	0.0	100.0	0.2	100.0	0.4	0.1	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0
T	0.0	0.0	0.0	0.0	100.0	0.0	8.2	8.4	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

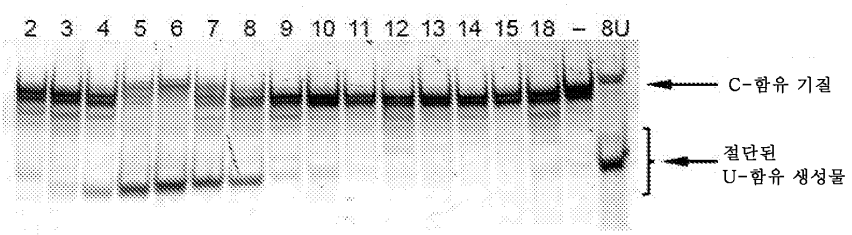
도면14b


TP53 NM_000546.6 c.1239T>A
A
C
G
T
His
Tyr
Asn
Tyr
Met
Cys
Asp → Asn
G4

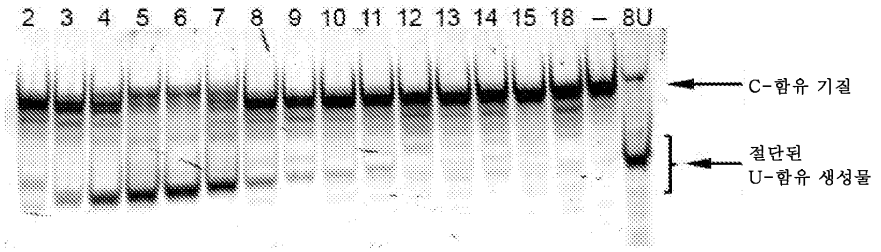
도면14c


TP53 Y163C
A
C
G
T
Arg
Ala
Met
Ala
Ile
Ala
Cys → Tyr
Lys

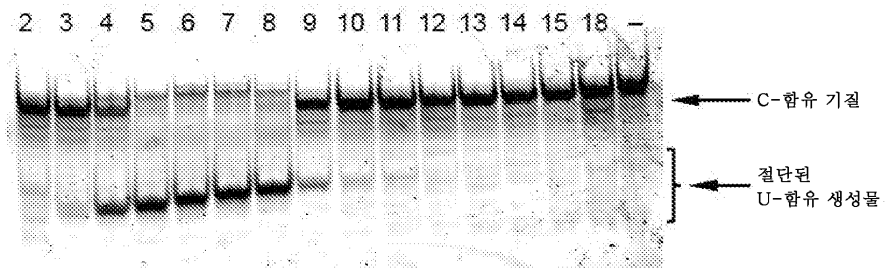
도면15a


2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 – 8U
C-함유 기질
절단된
U-함유 생성물

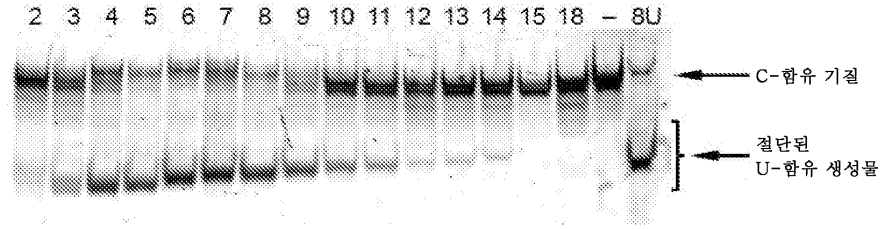
도면15b



도면15c



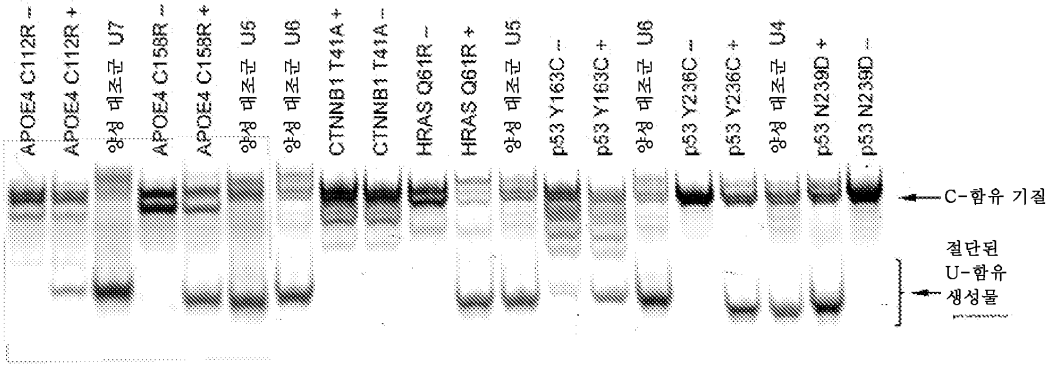
도면15d



도면16a

APOE4 Cys112Arg:	5'-GGAGGACGTGC$_{11}$GCGGCCGCCTGG
APOE4 Cys158Arg:	5'-GAAGC$_5$GCCTGGCAGTGTACCAGG
CTNNB1 Thr41Ala:	5'-CTGTGGC$_7$AGTGGCACCAGAATGG
HRAS Gln61Arg:	5'-CCTCCC$_6$GGCCGGCGGTATCCAGG
p53 Tyr163Cys:	5'-GCTTGC$_6$AGATGGCCATGGCGCGG
p53 Tyr236Cys:	5'-ACACATGC$_8$AGTTGTAGTGGATGG
p53 Asn239Asp:	5'-TGTC$_4$ACACATGTAGTTGTAGTGG

도면16b



도면17

프로토스페이서 및 PAM 서열: 5'-TTCCCCCCCCGATTTATTTATGG-3'

서열	총 판독물의 %
CCCCCCCC	62.4
TTTTTTCC	18.2
TTTTTTTC	13.4
TTTTTTTT	3.3
TCCCCCCC	0.8
CCCCTTCC	0.3
CCCTTTCC	0.3
TTTTTCCC	0.3
CCCCTCCC	0.3

도면18a

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG
FANCF: GGAATC$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG
HEK293 부위 2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG
HEK293 부위 3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG
HEK293 부위 4: GGCAC$_5$TGCGGCTGGAGGTCCGGG
RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

도면18b

EMX1	C_5	C_6
NBE1	6.2%	6.5%
NBE1 + UGI	9.7%	10.1%
NBE2	8.0%	8.7%

도면18c

FANCF	C_6	C_7	C_8	C_{10}
NBE1	3.7%	3.2%	3.4%	2.4%
NBE1 + UGI	7.5%	7.6%	7.5%	1.6%
NBE2	4.7%	4.6%	4.6%	0.8%

도면18d

HEK293 부위 2	C_4	C_6
NBE1	0.4%	0.4%
NBE1 + UGI	1.6%	2.6%
NBE2	3.4%	5.9%

도면18e

HEK293 부위 3	C_4	C_5
NBE1	2.0%	1.9%
NBE1 + UGI	6.5%	6.7%
NBE2	10.0%	12.5%

도면18f

HEK293 부위 4	C_5
NBE1	1.4%
NBE1 + UGI	5.4%
NBE2	8.2%

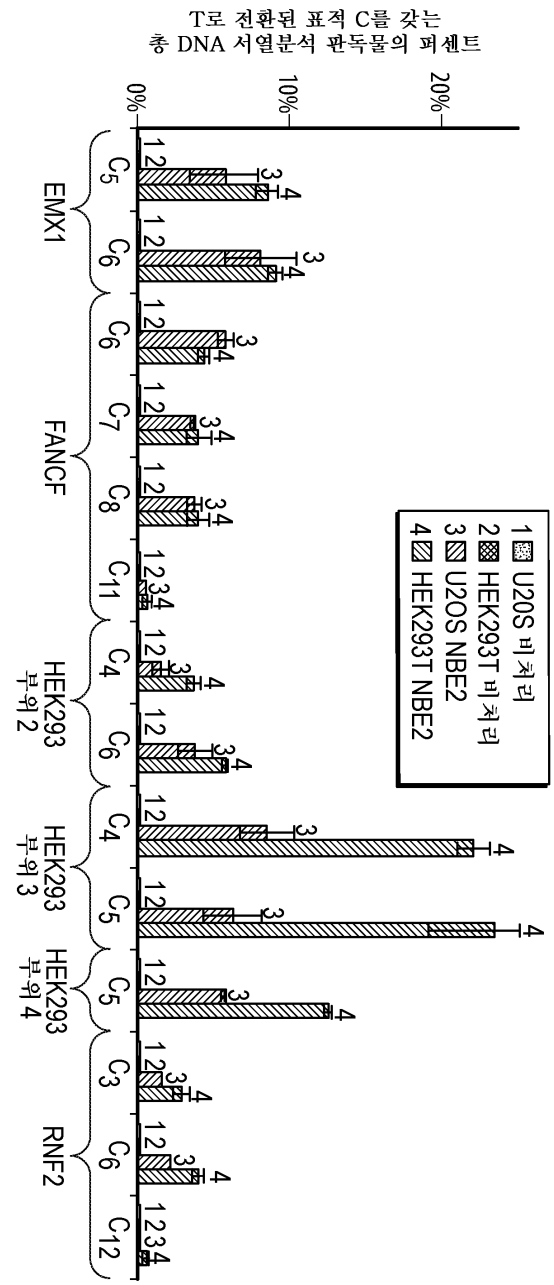
도면18g

RNF2	C_3	C_8
NBE1	0.7%	1.4%
NBE1 + UGI	3.4%	3.9%
NBE2	2.5%	3.7%

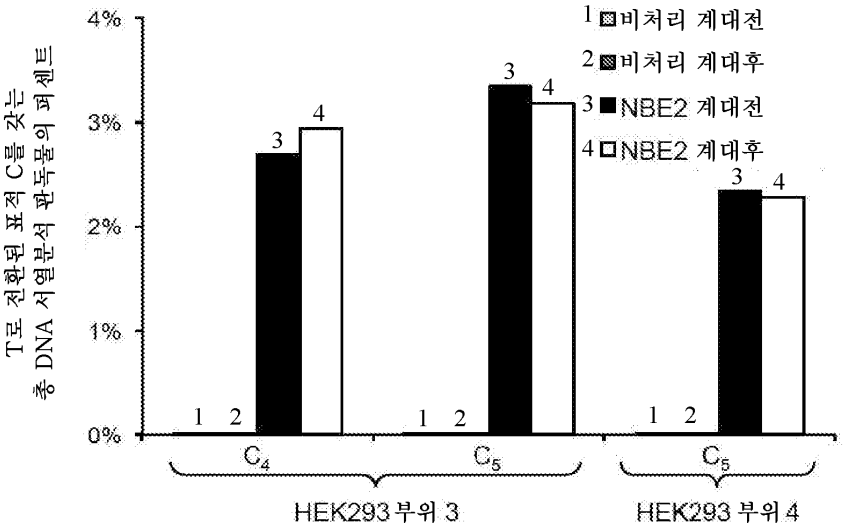
도면18h

비-프로토스페이서	C	C	T
비처리		99.93%	0.03%
NBE1		99.95%	0.03%
NBE1 + UGI		99.91%	0.06%
NBE2		99.92%	0.04%

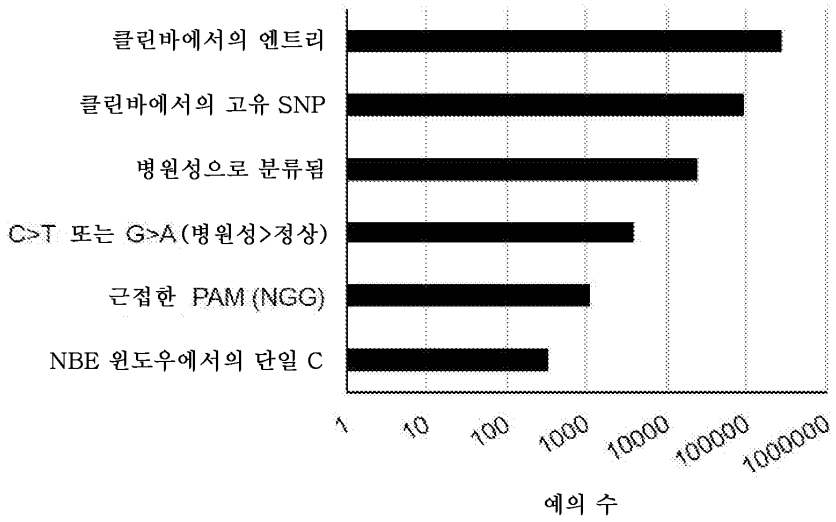
도면19


T로 전환된 표적 C를 갖는
총 DNA 서열분석 판독물의 퍼센트
0%
10%
20%
1 U2OS 비처리
2 HEK293T 비처리
3 U2OS NBE2
4 HEK293T NBE2
EMX1
FANCF
HEK293 부위 2
HEK293 부위 3
HEK293 부위 4
RNF2
C5
C6
C6
C7
C8
C11
C4
C6
C4
C5
C5
C3
C6
C12

도면20


4%
3%
2%
1%
0%
T로 전환된 표적 C를 갖는
총 DNA 서열분석 판독물의 퍼센트
1 비처리 계대전
2 비처리 계대후
3 NBE2 계대전
4 NBE2 계대후
1 2 3 4
1 2 3 4
1 2 3 4
C4
C5
C5
HEK293 부위 3
HEK293 부위 4

도면21


클린바에서의 엔트리
클린바에서의 고유 SNP
병원성으로 분류됨
C>T 또는 G>A (병원성>정상)
근접한 PAM (NGG)
NBE 윈도우에서의 단일 C
1
10
100
1000
10000
100000
1000000
예의 수

도면22i

EMX1 1.85 uM NBE1

	G	A	G	T	C5	C6	G	A	G	C	A	G	A	A	G	A	A	G	G	G	G	G	G	G	G
A	0.2	99.9	0.1	0.1	0.1	0.0	0.3	100.0	0.1	0.1	99.9	0.2	99.9	99.7	0.3	100.0	99.9	0.2	99.9	97.9	0.1	0.2	0.2	0.2	0.2
C	0.0	0.0	0.0	0.0	60.4	61.0	0.0	0.0	0.0	99.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
G	99.7	0.0	99.9	0.0	0.0	0.0	99.7	0.0	99.0	0.1	0.0	99.7	0.0	0.0	99.7	0.0	0.0	0.0	99.8	0.0	0.0	99.8	99.8	99.8	99.8
T	0.1	0.0	99.9	99.9	39.5	39.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

FANCF 1.85 uM NBE1

	G	G	A	A	T	C6	C7	C8	T	T	C11	T	G	A	C	A	A	G	G	A	C	C	C	T	G	G	G
A	0.1	0.2	99.8	99.9	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	99.9	0.3	100.0	99.9	0.2	0.1	0.1	0.1	0.1	0.1	0.3	0.4	
C	0.0	0.0	0.1	0.1	0.0	63.9	64.7	65.0	0.0	0.0	72.0	0.0	0.0	99.4	0.1	0.0	99.7	4.9	99.8	99.9	0.0	0.0	0.0	0.0	0.0	0.0	
G	99.8	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.2	0.0	99.7	0.2	0.0	0.1	0.0	0.0	0.0	99.9	0.0	99.6	99.5		
T	0.1	0.1	0.0	0.0	99.9	36.0	35.1	34.9	99.9	99.9	27.3	99.8	0.1	0.3	0.0	0.1	0.0	0.0	0.0	0.0	99.9	0.1	0.1				

HEK293 부위2 1.85 uM NBE1

	G	A	A	C4	A	C6	A	A	A	G	C	A	T	A	G	A	C	C	T	G	C	G	A	C	G	T	G	A	T	G	G
A	0.1	100.0	99.8	0.1	100.0	0.1	99.9	100.0	99.9	0.1	0.1	100.0	99.9	0.2	100.0	0.1	0.0	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
C	0.0	0.0	0.1	80.6	0.0	76.9	0.0	0.0	0.0	99.8	0.0	0.0	0.1	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	99.8	0.0	0.0	19.3	0.0	22.9	0.0	0.0	0.0	0.1	0.3	0.0	99.9	0.0	0.1	0.0	0.0	99.9	0.0	0.1	0.0	0.1	0.0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
T	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

HEK293 부위3 1.85 uM NBE1

	G	G	C	C4	C5	A	G	A	C	T	G	A	G	C	A	C	G	T	G	A	T	G	G
A	0.2	0.1	0.1	0.1	0.0	99.9	0.2	99.9	0.1	0.0	0.2	99.9	0.1	0.1	100.0	0.1	0.1	0.0	0.2	96.0	0.0	0.1	0.1				
C	0.0	0.0	92.2	74.0	71.5	0.0	0.0	0.0	96.0	0.0	0.0	0.0	0.0	99.8	0.0	99.7	0.0	0.0	0.0	1.3	0.0	0.0	0.0				
G	99.7	99.8	0.0	0.0	0.0	99.8	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	99.0	0.0	99.8	0.0	0.0	0.0	99.8	99.8				
T	0.1	0.1	7.7	25.1	28.5	0.0	0.0	0.0	3.3	99.9	0.1	0.0	0.0	0.1	0.0	0.2	0.0	99.9	0.0	0.0	100.0	0.0	0.0				

도면22ii

HEK293 부위3

1.85 uM NBE1

	G	G	C	C4	C5	A	G	A	C	T	G	A	G	C	T	G	A	T	G
A	0.2	0.1	0.1	0.1	0.0	99.9	0.2	99.9	0.1	0.0	0.2	99.9	0.1	0.1	100.0	0.1	0.1	0.0	0.2	99.6	0.0	0.1	0.1
C	0.0	0.0	92.2	74.9	71.5	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0
G	99.7	99.8	0.0	0.0	0.0	0.0	99.8	0.0	0.0	99.9	0.0	0.0	0.0	99.8	0.0	99.8	0.0	99.8	0.0	0.0	99.9	99.8	99.8
T	0.1	0.1	7.7	25.1	28.5	0.0	0.0	0.0	3.3	99.9	0.1	0.0	0.0	0.0	0.0	0.2	0.0	99.9	0.0	0.0	100.0	0.0	0.0

HEK293 부위4

1.85 uM NBE1

	G	G	C	A	C5	T	G	C	G	G	C	T	G	G	A	G	G	T	G	G	T	G	G
A	0.3	0.2	0.1	100.0	0.1	0.0	0.2	0.1	0.2	0.2	0.1	0.0	0.3	0.5	100.0	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.2	0.2
C	0.0	0.0	99.8	0.0	99.1	0.0	0.0	97.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	99.7	99.7	0.0	0.0	0.0	99.8	99.8	0.0	99.8	99.8	0.0	100.0	99.7	99.4	0.0	99.8	99.8	0.1	99.8	99.8	99.8	99.8	99.8	99.7
T	0.0	0.1	1.1	0.0	0.0	100.0	0.1	2.9	0.0	0.1	0.5	100.0	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.1	0.1	0.0	0.0	0.0

RNF2

1.85 uM NBE1

	G	T	C3	A	T	C6	T	T	A	G	T	C12	A	T	T	A	C	C	T	G	A	G	G
A	0.1	0.0	0.1	99.9	0.0	0.0	0.0	0.0	99.9	0.1	0.0	0.1	99.9	0.0	0.0	99.9	0.1	0.0	0.0	0.1	98.6	0.0	0.0
C	0.0	0.0	99.1	0.0	0.0	57.8	0.0	0.0	0.1	0.0	0.0	99.4	0.0	0.0	0.0	0.1	99.9	99.9	0.0	0.0	1.4	0.0	0.0
G	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	99.9	99.9
T	0.0	99.9	0.8	0.0	100.0	42.1	100.0	100.0	0.0	0.1	99.9	13.4	0.0	100.0	99.9	0.0	0.0	0.0	99.9	0.1	0.0	0.1	0.0

도면23i

EMX1 온 타겟
비처리
NBE1
NBE2
NBE3
0.1 0.0
4.2 4.4
4.1 9.3
20.0 20.0
0.0
0.0
0.0
GUIDE-seq 카운트
4521
EMX1 오프 타겟1
비처리
NBE1
NBE2
NBE3
0.0
0.0
0.0
0.0
3123
EMX1 오프 타겟2
비처리
NBE1
NBE2
NBE3
0.0
1.0
3.0
8.5
0.0
0.0
0.0
1445

도면23ii

| EMX1 오프 타겟3 | | |

도면23iii

EMX1 오프 타겟7																					143
	G	A	G	C	C	G	G	A	G	C	A	G	A	A	G	A	A	G	G	G	G
비처리				0.0	0.0					0.0											
NBE1				0.0	0.0					0.0											
NBE2				0.1	0.0					0.0											
NBE3				0.1	0.1					0.0											

EMX1 오프 타겟8																					102
	A	A	G	T	C	C	G	A	G	G	A	G	G	A	G	A	A	A	G	G	G
비처리				0.0	0.0																
NBE1				0.0	0.0																
NBE2				0.0	0.0																
NBE3				0.1	0.1																

EMX1 오프 타겟9																					67
	G	A	A	T	C	C	A	A	G	C	A	G	G	A	G	A	A	G	A	G	A
비처리				0.0	0.1					0.0											
NBE1				0.0	0.1					0.0											
NBE2				0.0	0.1					0.0											
NBE3				0.3	0.3					0.0											

EMX1 오프 타겟10																					38
	A	C	G	T	C	T	G	A	G	C	A	G	A	A	G	A	A	G	A	T	G	G
비처리				0.1						0.0											
NBE1				0.3						0.0											
NBE2				0.5						0.0											
NBE3				2.1						0.0											

도면24i

FANCF 온 타겟
비처리
NBE1
NBE2
NBE3
GUIDE-seq 카운트
4616
FANCF 오프 타겟1
2859
FANCF 오프 타겟2
524
FANCF 오프 타겟3
150
FANCF 오프 타겟4
125

도면24ii

FANCF 오프 타겟 5

	G	G	A	T	T	G	C	C	A	T	C	G	C	A	G	C	A	C	C	T	G	G
비처리						0.1	0.0			0.0	0.0	0.0	0.0		0.0	0.0		0.0	0.0			
NBE1						0.1	0.0			0.0	0.1	0.0			0.0			0.0	0.0			
NBE2						0.1	0.0			0.0	0.0	0.0			0.0			0.0	0.0			
NBE3						0.1	0.0			0.0	0.0	0.0			0.0			0.0	0.0			

101

FANCF 오프 타겟 6

	T	G	A	A	T	C	C	C	A	T	C	T	C	C	A	G	C	A	C	C	A	G	G
비처리						0.0	0.0	0.0		0.0		0.1	0.1			0.0			0.0	0.0			
NBE1						0.0	0.0	0.0			0.0	0.0	0.0			0.0			0.0	0.0			
NBE2						0.2	0.2	0.2			0.2	0.1	0.0			0.0			0.0	0.0			
NBE3						0.3	0.3	0.3			0.3	0.1	0.0			0.0			0.0	0.0			

78

FANCF 오프 타겟 7

	G	G	A	G	T	C	C	C	T	C	G	T	A	C	A	G	C	A	C	C	A	G	G
비처리						0.0	0.0	0.0		0.0	0.0		0.0			0.0			0.0	0.0			
NBE1						0.0	0.0	0.0		0.0	0.0		0.0			0.0			0.0	0.0			
NBE2						0.0	0.0	0.0		0.0	0.0		0.0			0.0			0.0	0.0			
NBE3						0.5	0.5	0.5		0.3	0.1		0.0			0.0			0.0	0.0			

77

FANCF 오프 타겟 8

	G	G	A	G	T	C	C	C	T	C	C	T	G	C	A	G	C	A	C	C	T	G	A
비처리						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0			
NBE1						0.0	0.0	0.0		0.0	0.0			0.0			0.0		0.0	0.0			
NBE2						1.4	0.5	0.5		0.4	0.5			0.0			0.0		0.0	0.0			
NBE3						0.1	0.1	0.0		0.0	0.0			0.0			0.0		0.0	0.0			

4

도면25

HEK부위2 온 타겟																					GUIDE-Seq 카운트
	G	A	A	C	A	C	A	A	A	G	C	A	T	G	A	C	T	G	C	G	G	1670
비처리																					
NBE1	0.0	1.8	2.9																		
NBE2		6.7	10.4																		
NBE3		17.6	14.6																		

HEK부위2 오프 타겟1																					
	G	A	A	C	A	C	A	A	T	G	C	A	T	A	G	A	C	T	G	C	G	G	99
비처리																					
NBE1	0.0	0.0	0.0																		
NBE2	0.0	0.0	0.4																		
NBE3	0.2	0.6																			

HEK부위2 오프 타겟2																					
	A	A	A	C	A	T	A	A	A	G	C	A	T	A	G	A	C	T	G	C	A	A	4
비처리																					
NBE1	0.0	0.0																			
NBE2	0.0	0.0																			
NBE3	0.0	0.0																			

도면26

HEK 부위 3 온 타겟

	G	G	C	C	C	A	G	A	C	T	G	A	C	A	C	G	T	G	A	T	G	G
비처리		0.1	0.1	0.1			0.3				0.3		0.0		0.0							
NBE1		0.1	4.8	4.3			0.9				0.9		0.0		0.0							
NBE2		3.6	41.3	20.4			0.3				0.3		0.3		0.0							
NBE3		0.5	39.4	43.1			0.5				0.5		0.0		0.2							

GUIDE-seq 카운트 2074

HEK 부위 3 오프 타겟 1

	C	A	C	C	C	A	G	A	C	T	G	A	G	C	A	C	G	T	G	C	T	G	G
비처리	0.0		0.0	0.0	0.0			0.0				0.0		0.0		0.0				0.0			
NBE1	0.0		0.0	0.0	0.0			0.0				0.0		0.0		0.0				0.0			
NBE2	0.0		0.0	0.0	0.0			0.0				0.0		0.0		0.0				0.0			
NBE3	0.0		0.0	0.0	0.0			0.0				0.0		0.0		0.0				0.0			

327

HEK 부위 3 오프 타겟 2

	G	A	C	A	C	A	G	A	C	T	G	G	G	C	A	C	G	T	G	A	G	G	G
비처리		0.0		0.0				0.0						0.0		0.0							
NBE1		0.0		0.0				0.0						0.0		0.0							
NBE2		0.0		0.0				0.0						0.0		0.0							
NBE3		0.0		0.0				0.0						0.0		0.0							

306

HEK 부위 3 오프 타겟 3

	A	G	C	T	C	A	G	A	C	T	G	A	G	C	A	A	G	T	G	A	G	G	G
비처리			0.0		0.0				0.0						0.0								
NBE1			0.0		0.0				0.0						0.0								
NBE2			0.0		0.0				0.0						0.0								
NBE3			0.0		0.9				0.0						0.0								

136

HEK 부위 3 오프 타겟 4

	A	G	A	C	C	A	G	A	C	T	G	A	G	C	A	A	G	T	G	A	G	G	G
비처리			0.0	0.0				0.0						0.0									
NBE1			0.0	0.0				0.0						0.1									
NBE2			0.0	0.0				0.0						0.0									
NBE3			0.0	0.0				0.0						0.0									

10

HEK 부위 3 오프 타겟 5

	G	A	G	C	C	A	G	A	T	G	A	G	C	A	C	G	T	G	A	G	G	G
비처리			0.0	0.0									0.0	0.0								
NBE1			0.0	0.0									0.0	0.0								
NBE2			0.0	0.0									0.0	0.0								
NBE3			0.0	0.0									0.0	0.0								

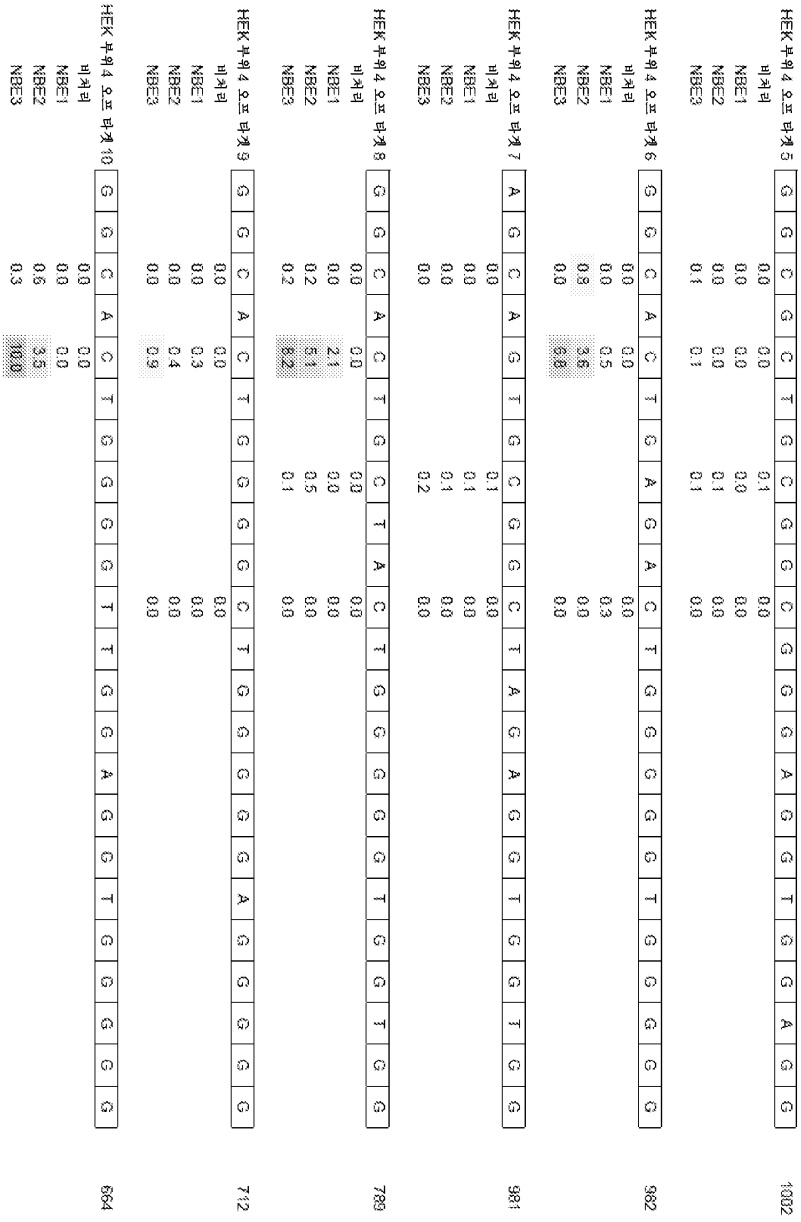
2

도면27i

HEK 부위 4 온 타겟	G	G	C	A	C	T	G	C	G	G	C	T	G	G	A	G	T	G	G	G	G	1654
비처리			0.0		0.0		0.1				0.0											
NBE1			0.0		0.4		0.1				0.0											
NBE2			0.1		12.7		0.1				0.0											
NBE3			1.5		22.3		1.0				0.0											
HEK 부위 4 오프 타겟 1	T	G	C	A	C	T	G	C	G	G	C	C	G	G	A	G	A	G	G	T	G	2475
비처리			0.0		0.0		0.0				0.0	0.0										
NBE1			0.0		0.6		0.1				0.1	0.0										
NBE2			0.3		3.3		0.3				0.0	0.0										
NBE3			0.2		10.2		0.3				0.0	0.0										
HEK 부위 4 오프 타겟 2	G	G	C	T	C	T	G	C	G	G	C	T	G	G	A	G	G	G	G	T	G	2020
비처리			0.0		0.0		0.1				0.0											
NBE1			0.0		0.6		0.1				0.0											
NBE2			0.0		0.4		0.1				0.1											
NBE3			0.0		3.3		0.1				0.0											
HEK 부위 4 오프 타겟 3	G	G	C	A	C	G	A	C	G	G	C	T	G	G	A	G	G	T	G	G	G	1375
비처리			0.0		0.0		0.0				0.0											
NBE1			0.1		11.8		4.0				0.0											
NBE2			0.0		9.5		3.9				0.0											
NBE3			0.1		19.0		1.9				0.1											
HEK 부위 4 오프 타겟 4	G	G	C	A	T	C	A	C	G	G	C	T	G	G	A	G	G	T	G	G	A	1097
비처리			0.0		0.0		0.0				0.0											
NBE1			0.0		0.2		0.3				0.0											
NBE2			0.5		5.3		2.7				0.0											
NBE3			0.2		49.1		3.1				0.0											

GUIDE-seq 카운트

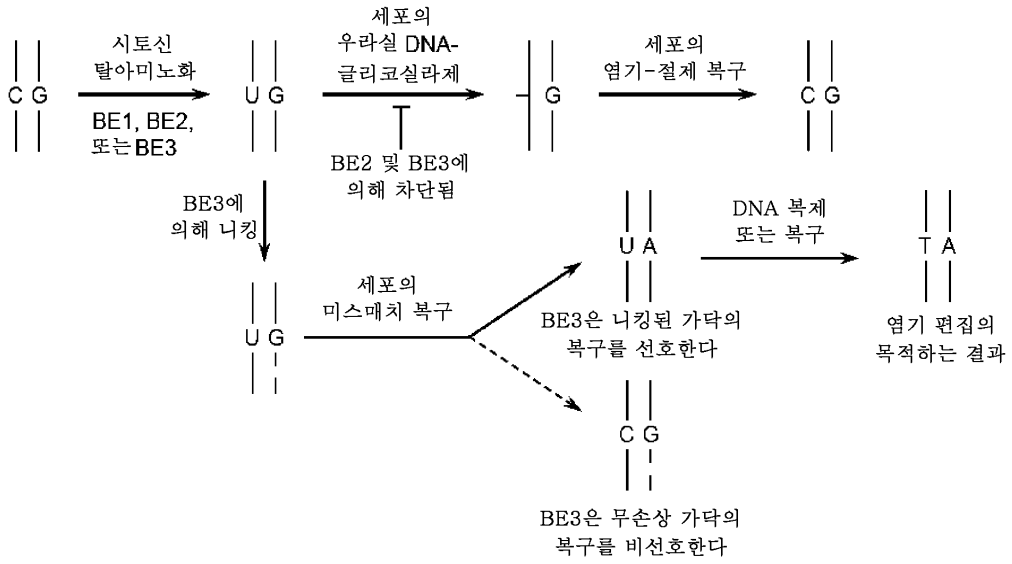
도면27ii



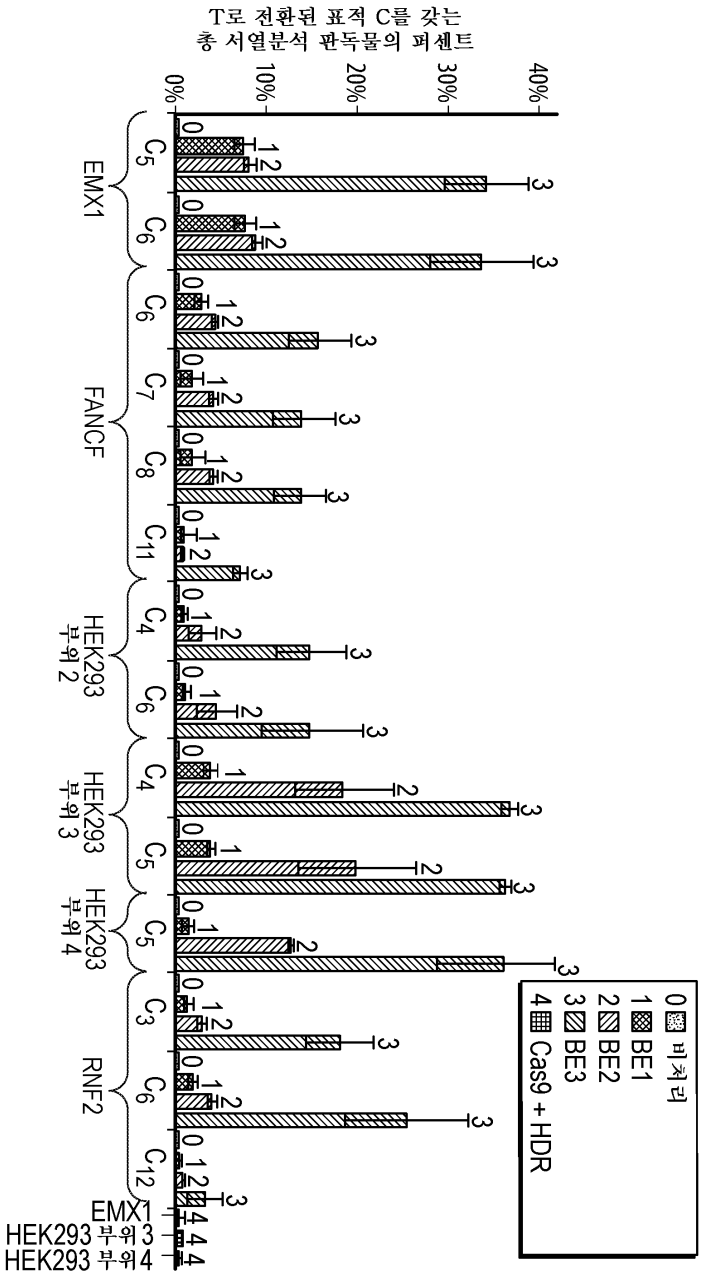
도면28

비-프로토스페이서 C	C (%)	T (%)
비처리	99.94	0.04
NBE1	99.92	0.05
NBE2	99.92	0.05
NBE3	99.94	0.03

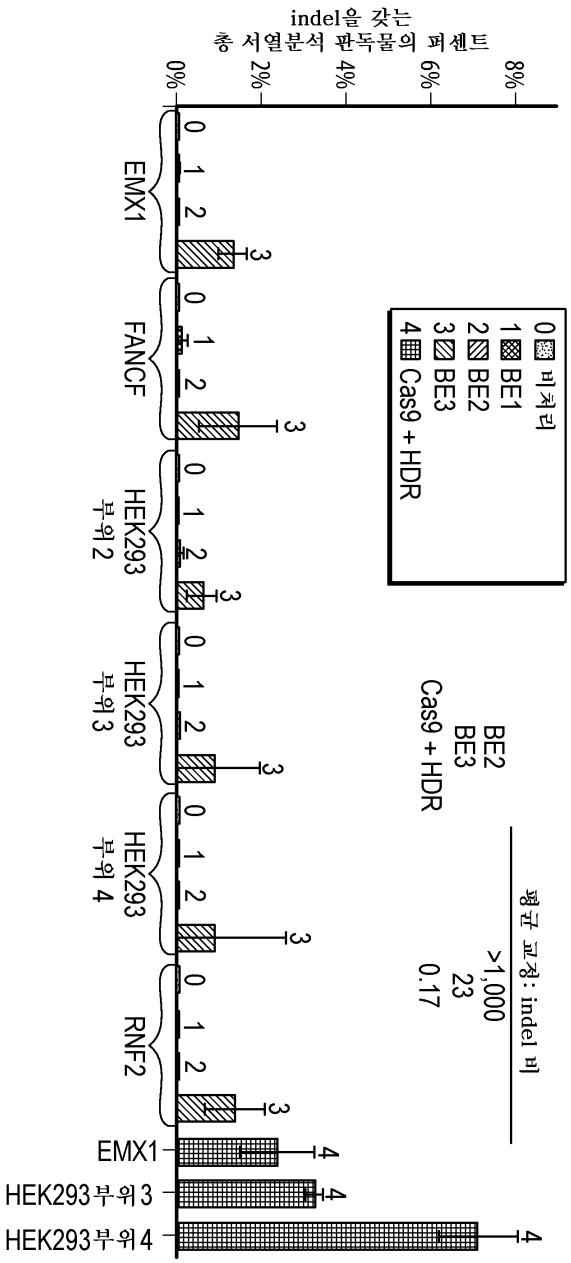
도면29a


C G
시토신
탈아미노화
BE1, BE2,
또는 BE3
U G
세포의
우라실 DNA-
글리코실라제
BE2 및 BE3에
의해 차단됨
G
세포의
염기-절제 복구
C G
BE3에
의해 니킹
U G
세포의
미스매치 복구
U A
BE3은 니킹된 가닥의
복구를 선호한다
DNA 복제
또는 복구
T A
염기 편집의
목적하는 결과
C G
BE3은 무손상 가닥의
복구를 비선호한다

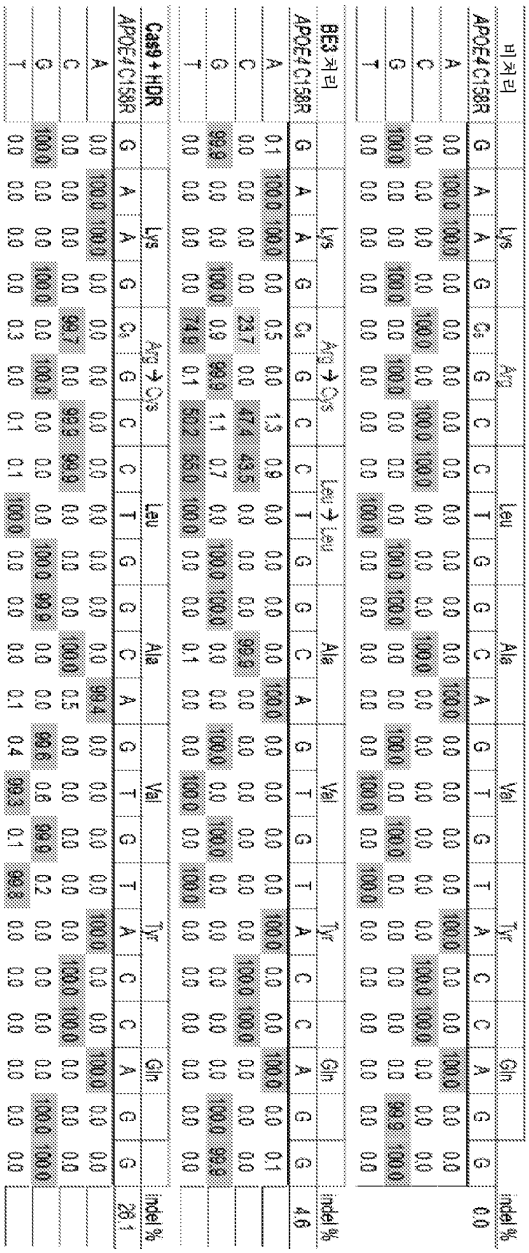
도면29b


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0%
10%
20%
30%
40%
EMX1
FANCF
HEK293 부위 2
HEK293 부위 3
HEK293 부위 4
RNF2
EMX1
HEK293 부위 3
HEK293 부위 4
0 비처리
1 BE1
2 BE2
3 BE3
4 Cas9 + HDR

도면29c


indel을 갖는
총 서열분석 판독물의 퍼센트
0%
2%
4%
6%
8%
0 비처리
1 BE1
2 BE2
3 BE3
4 Cas9 + HDR
EMX1
FANCF
HEK293 부위2
HEK293 부위3
HEK293 부위4
RNF2
EMX1
HEK293부위3
HEK293부위4
BE2
BE3
Cas9 + HDR
평균 교정: indel 비
>1,000
23
0.17

도면30a



비처리

APOE4 C158R	Lys			Arg		Leu			Ala				Val				Tyr			Gln			indel %
	G	A	A	G	C	C	T	G	G	C	A	G	G	T	G	T	A	C	C	A	G	G	
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	90.9	100.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0		

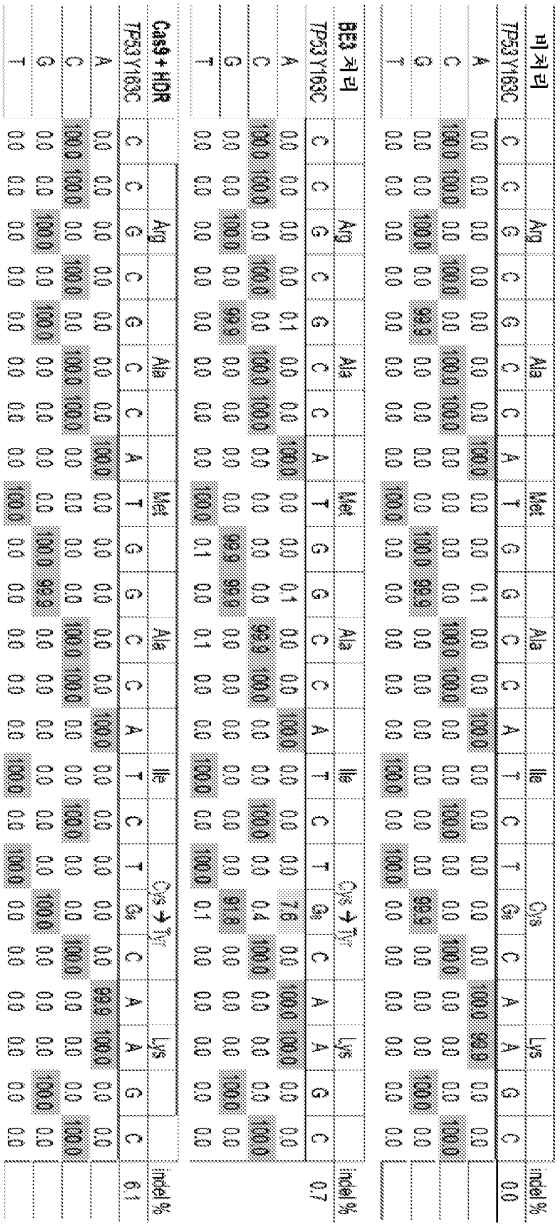
BE3 처리

APOE4 C158R	Lys			Arg → Cys		Leu → Leu			Ala				Val				Tyr			Gln			indel %
	G	A	A	G	C	C	T	G	G	C	A	G	G	T	G	T	A	C	C	A	G	G	
A	0.1	100.0	100.0	0.0	0.5	0.0	1.3	0.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.1	4.6
C	0.0	0.0	0.0	0.0	23.7	0.0	47.4	43.5	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	100.0	0.9	99.8	1.1	0.7	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	99.8	
T	0.0	0.0	0.0	0.0	74.9	0.1	50.2	55.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Cas9 + HDR

APOE4 C158R	Lys			Arg → Cys		Leu			Ala				Val				Tyr			Gln			indel %
	G	A	A	G	C	C	T	G	G	C	A	G	G	T	G	T	A	C	C	A	G	G	
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.5	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	28.1
C	0.0	0.0	0.0	0.0	99.7	0.0	99.3	99.0	0.0	0.0	0.0	99.6	0.6	99.0	0.1	99.3	0.0	0.0	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	99.8	0.0	0.0	0.4	99.3	0.1	99.7	0.2	0.0	0.0	0.0	0.0	0.0	0.0	
T	0.0	0.0	0.0	0.0	0.3	0.0	0.1	0.1	100.0	0.3	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

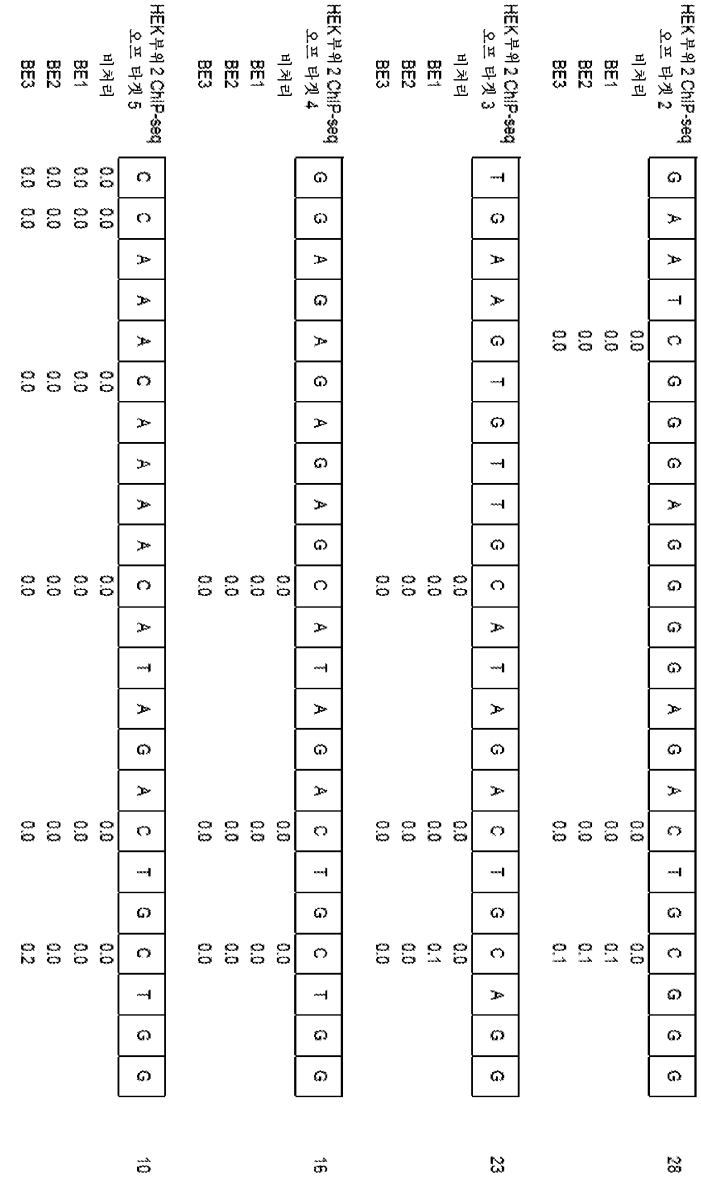
도면30b


비처리
TP53Y163C
Arg
Ala
Met
Cys
Lys
indel%
BE3 처리
Cys → Trp
Cas9 + HDR

도면31i

HEK부위 2 온 타겟
GUIDE-seq 카운트/ ChIP-seq 강도
1670/ 117
G A A C A C A A G C A T A G T C C G G G
비처리 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
BE1 1.8 2.9 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
BE2 8.5 10.5 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
BE3 17.9 14.9 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
HEK부위 2 GUIDE-seq 오프 타겟 1
99
G A A C A C A A T G C A T A G T C C G G G
비처리 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.0 0.0 0.0
BE1 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.2 0.0 0.0
BE2 0.0 0.4 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.0 0.0 0.0
BE3 0.2 0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.0 0.0 0.0
HEK부위 2 GUIDE-seq 오프 타겟 2
4
A A A C A T A A G C A T A G T C C A A A
비처리 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
BE1 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
BE2 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
BE3 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
HEK부위 2 ChIP-seq 오프 타겟 1
31
T C A G G G T G A G A T A G C A T C G G G
비처리 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
BE1 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.0 0.0 0.0
BE2 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.0 0.0 0.0
BE3 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.1 0.0 0.0 0.0

도면31ii


HEK부위 2 ChIP-seq
오프 타겟 2
비처리
BE1
BE2
BE3
28
HEK부위 2 ChIP-seq
오프 타겟 3
23
HEK부위 2 ChIP-seq
오프 타겟 4
16
HEK부위 2 ChIP-seq
오프 타겟 5
10

도면32i

HEK부위3 온 타겟
GUIDE-seq 카운트/ ChIP-seq 강도
2374/ 163
비처리
BE1
BE2
BE3
HEK부위3 GUIDE-seq 오프 타겟 1
327
HEK부위3 GUIDE-seq 오프 타겟 2
305
HEK부위3 GUIDE-seq 오프 타겟 3
136
HEK부위3 GUIDE-seq 오프 타겟 4
10

도면32ii

HEK부위3 GUIDE-seq 오프 타겟 5
비처리
BE1
BE2
BE3
2
HEK부위3 ChIP-seq 오프 타겟 1
47
HEK부위3 ChIP-seq 오프 타겟 2
45
HEK부위3 ChIP-seq 오프 타겟 3
33
HEK부위3 ChIP-seq 오프 타겟 4
31
HEK부위3 ChIP-seq 오프 타겟 5
31

도면33i

HEK 부위 4 온 타겟																						GUIDE-seq 카운트 / ChIP-seq 강도
	G	G	C	A	C	T	G	C	G	G	C	T	G	G	A	G	T	G	G	G	G	1654 / 49
비처리					0.0		0.1				0.0											
BE1					0.0	0.4					0.0											
BE2					0.1	12.7					0.0											
BE3					1.5	27.0					0.0											

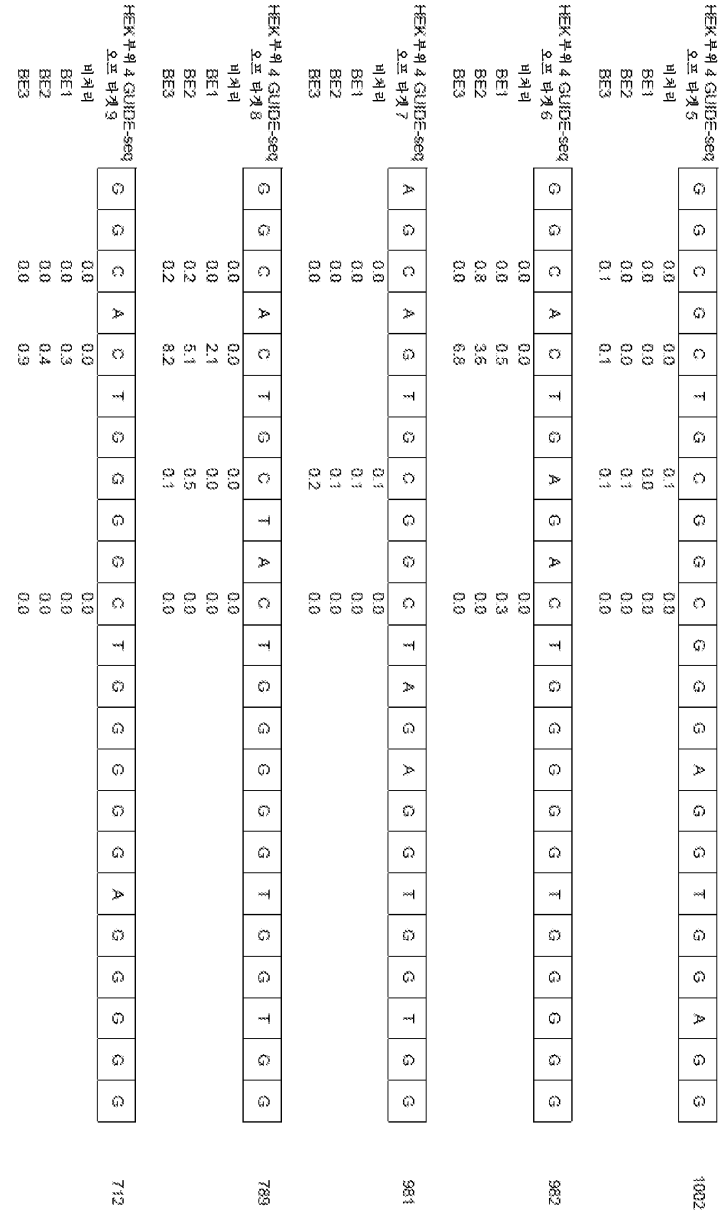
HEK 부위 4 GUIDE-seq 오프 타겟 1																						2475
	T	G	C	A	C	T	G	C	G	G	C	C	G	G	A	G	G	A	G	T	G	G	
비처리				0.0		0.0		0.0				0.0										
BE1				0.0		0.6		0.1				0.1	0.0									
BE2				0.3		3.3		0.3				0.0	0.0									
BE3				0.2		10.2		0.3				0.0	0.0									

HEK 부위 4 GUIDE-seq 오프 타겟 2																						2020
	G	G	C	T	C	T	G	C	G	G	C	T	G	G	A	G	G	G	G	T	G	G	
비처리				0.0		0.0		0.1				0.0										
BE1				0.0		0.6		0.1				0.0										
BE2				0.0		0.4		0.1				0.1										
BE3				0.0		3.3		0.1				0.0										

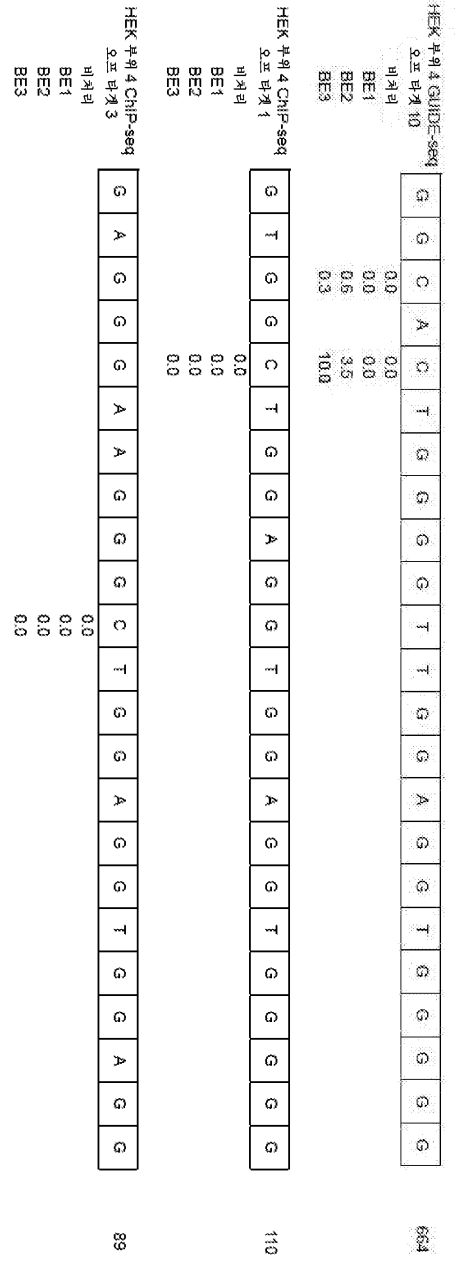
HEK 부위 4 GUIDE-seq 오프 타겟 3																						1375
	G	G	C	A	C	A	G	C	G	G	C	T	G	G	A	G	G	T	G	G	A	G	
비처리				0.0		0.0		0.0				0.0										
BE1				0.1		10.8		4.0				0.0										
BE2				0.0		9.6		3.9				0.0										
BE3				0.1		99.0		1.9				0.1										

HEK 부위 4 GUIDE-seq 오프 타겟 4																						1097
	G	G	C	A	T	C	A	C	G	G	C	T	G	G	A	G	G	T	G	G	A	G	
비처리				0.0		0.0		0.0				0.0										
BE1				0.0		0.2		0.3				0.0										
BE2				0.5		5.3		2.7				0.0										
BE3				0.2		90.1		3.1				0.0										

도면33ii


HEK부위 4 GUIDE-seq
오프 타겟 5
비처리
BE1
BE2
BE3
1002
HEK부위 4 GUIDE-seq
오프 타겟 6
982
HEK부위 4 GUIDE-seq
오프 타겟 7
981
HEK부위 4 GUIDE-seq
오프 타겟 8
789
HEK부위 4 GUIDE-seq
오프 타겟 9
712

도면33iii


HEK 부위 4 GUIDE-seq
오프 타겟 10
664
비처리
BE1
BE2
BE3
0.0
0.0
0.6
0.3
0.0
0.0
3.5
10.0
HEK 부위 4 ChIP-seq
오프 타겟 1
110
비처리
BE1
BE2
BE3
0.0
0.0
0.0
0.0
HEK 부위 4 ChIP-seq
오프 타겟 3
89
비처리
BE1
BE2
BE3
0.0
0.0
0.0
0.0

도면34i

도면34ii

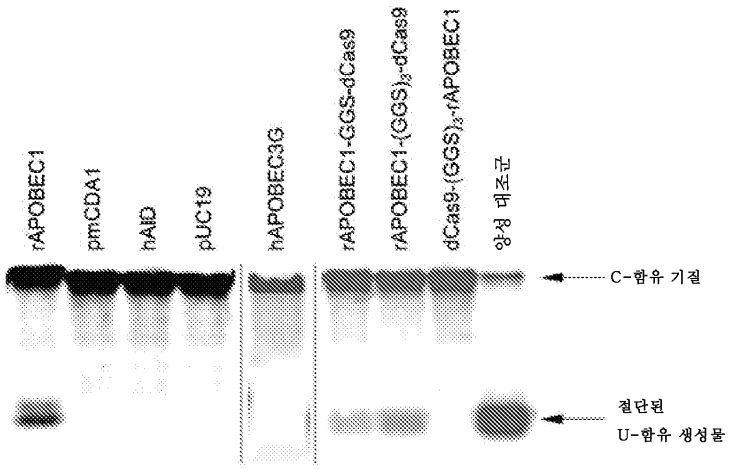
도면35i

TP63
비처리
C-50 C-49 T-48 G-47 T-46 G-45 C-44 A-43 G-42 C-41 T-40 G-39 T-38 G-37 G-36 G-35 T-34 T-33 G-32 A-31 T-30 T-29 C-28 C-27 A-26
BE3+온-타겟 sgRNA
BE3+오프-타겟 sgRNA
TP63
비처리
C-25 A-24 C-23 C-22 C-21 C-20 C-19 G-18 C-17 C-16 C-15 G-14 G-13 C-12 A-11 C-10 C-9 C-8 G-7 C-6 G-5 T-4 C-3 C-2 G-1
BE3+온-타겟 sgRNA
BE3+오프-타겟 sgRNA

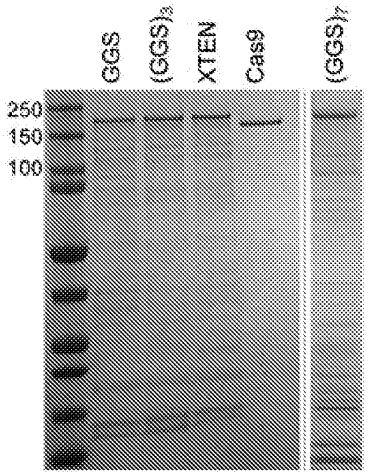
도면35ii

TP53
비처리
BE3 + 온-타겟 sgRNA
BE3 + 오프-타겟 sgRNA
TP53
비처리
BE3 + 온-타겟 sgRNA
BE3 + 오프-타겟 sgRNA

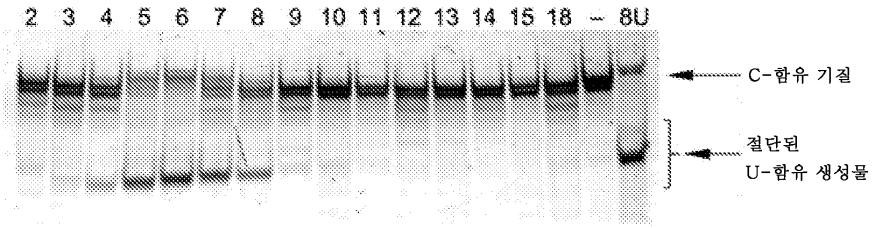
도면36a



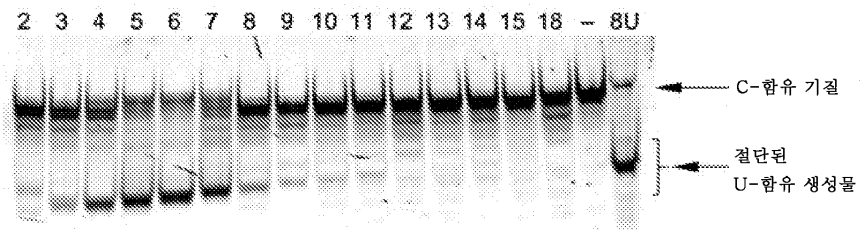
도면36b



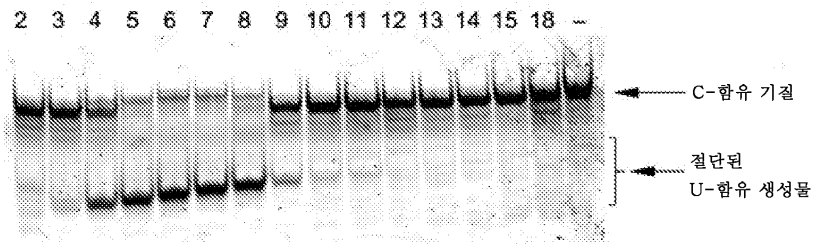
도면36c



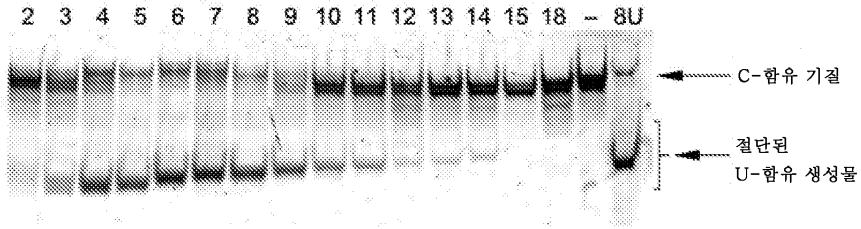
도면36d



도면36e



도면36f



도면37a

EMX1: GAGTC_5C_6GAGCAGAAGAAGAAGGG

FANCF: GGAAT$C_6C_7C_8$TTC_{11}TGCAGCACCTGG

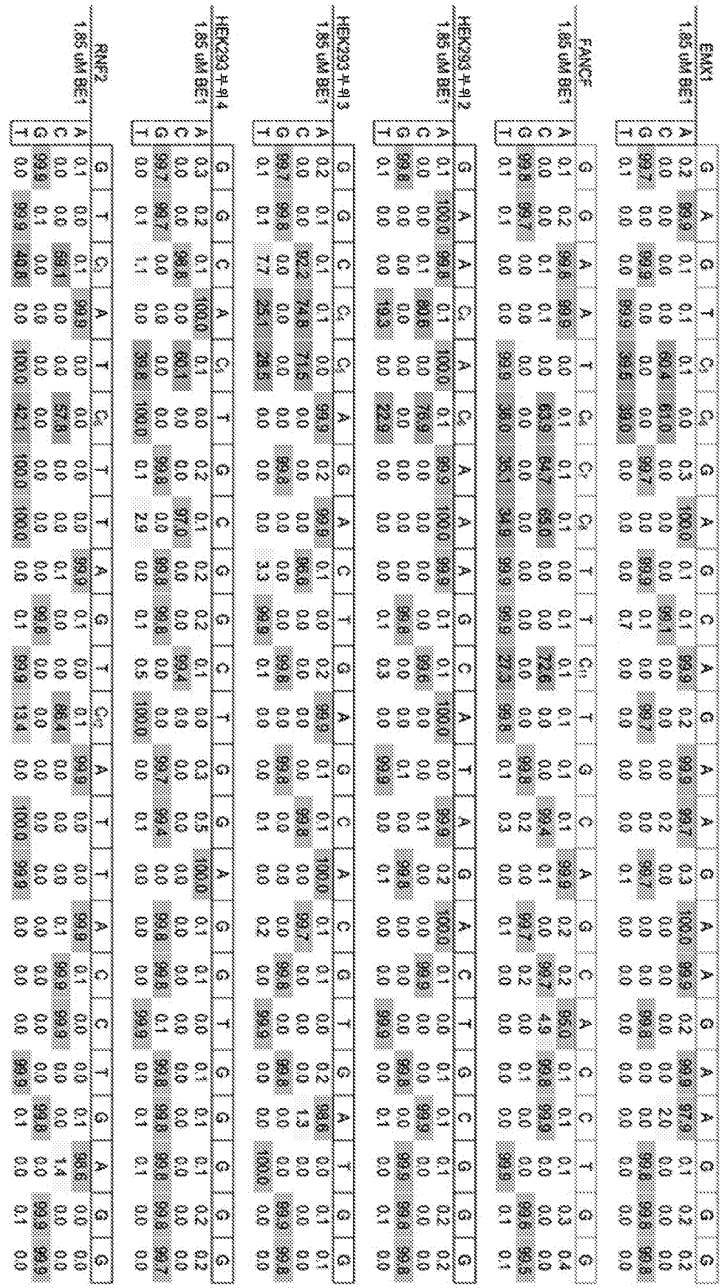
HEK293 부위 2: GAAC_4AC_6AAAGCATAGACTGCGGG

HEK293 부위 3: GGCC_4C_5AGACTGAGCACGTGATGG

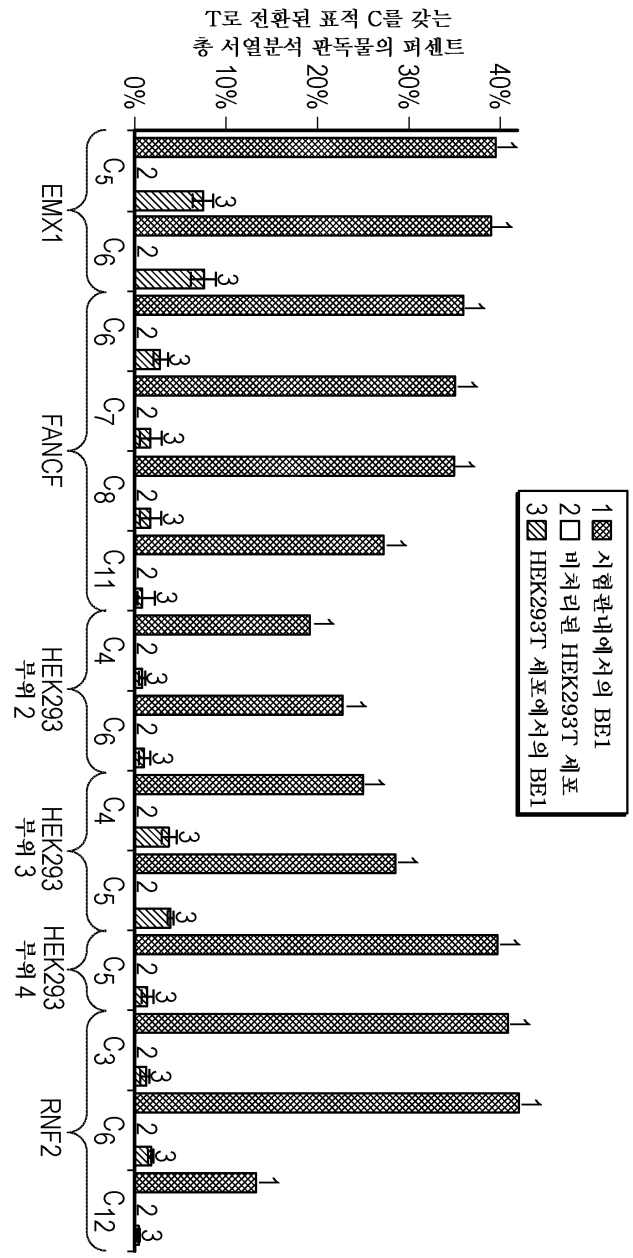
HEK293 부위 4: GGCAC_5TGCGGCTGGAGGTCCGGG

RNF2: GTC_3ATC_6TTAGTC_{12}ATTACCTGAGG

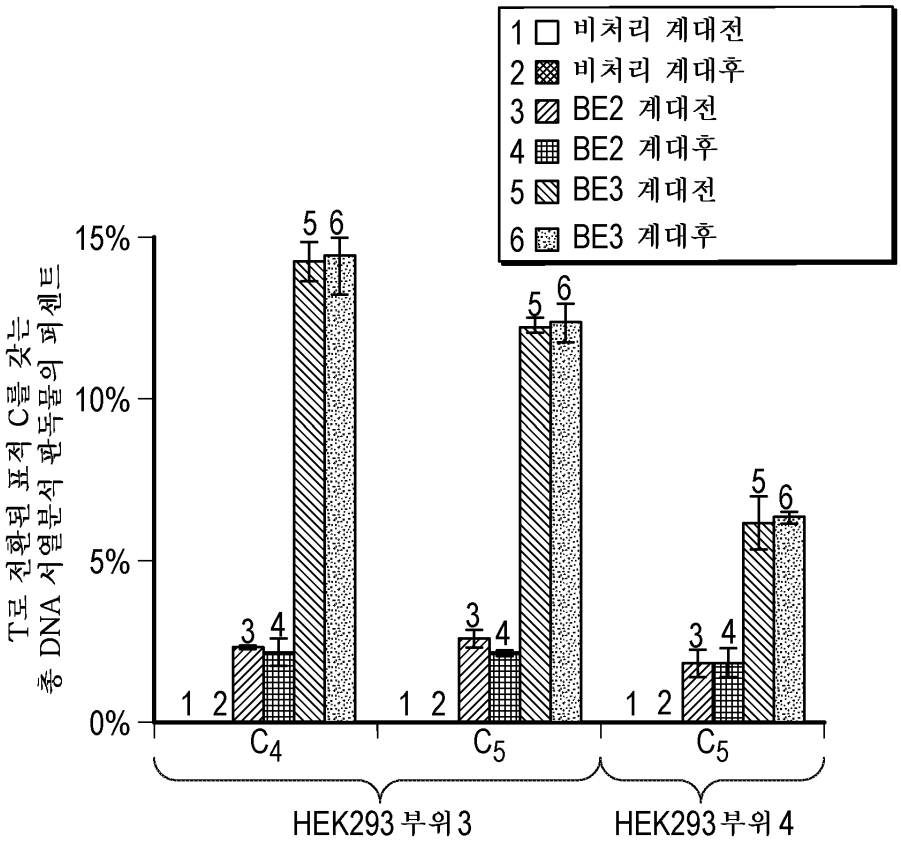
도면37b


EMX1
1.85 uM BE1
FANCF
1.85 uM BE1
HEK293 +#2
1.85 uM BE1
HEK293 +#3
1.85 uM BE1
HEK293 +#4
1.85 uM BE1
RNF2
1.85 uM BE1

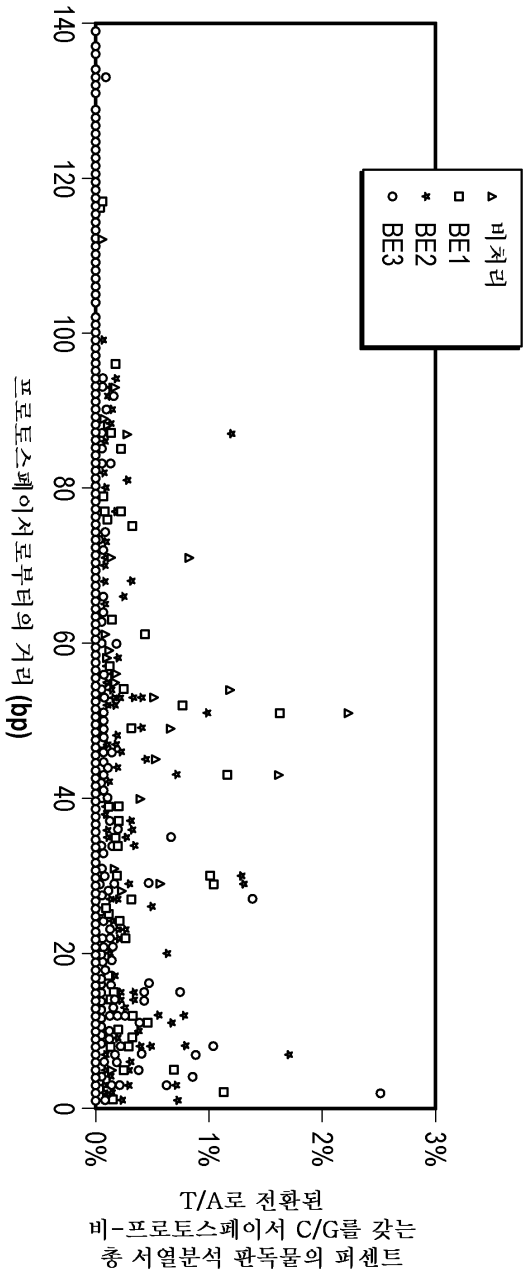
도면37c


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0%
10%
20%
30%
40%
1 시험관내에서의 BE1
2 비처리된 HEK293T 세포
3 HEK293T 세포에서의 BE1
EMX1
C5
C6
FANCF
C6
C7
C8
C11
HEK293 부위 2
C4
C6
HEK293 부위 3
C4
C5
HEK293 부위 4
C5
C3
RNF2
C6
C12

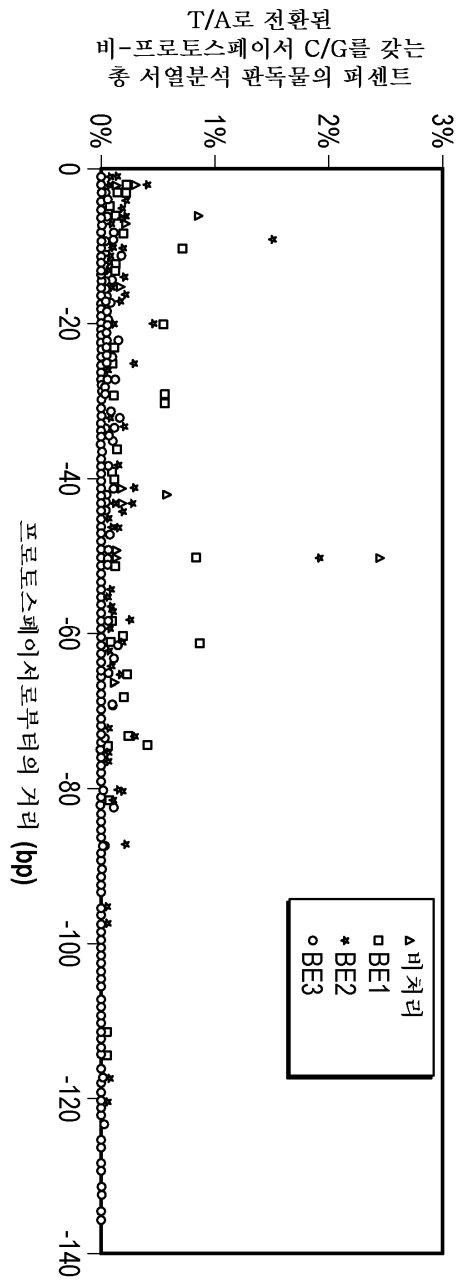
도면38


1 비처리 계대전
2 비처리 계대후
3 BE2 계대전
4 BE2 계대후
5 BE3 계대전
6 BE3 계대후
T로 전환된 표적 C를 갖는 총 DNA 서열분석 판독물의 퍼센트
15%
10%
5%
0%
C4
C5
C5
HEK293 부위 3
HEK293 부위 4

도면39a


비치리
BE1
BE2
BE3
프로토스페이서로부터의 거리 (bp)
0
20
40
60
80
100
120
140
0%
1%
2%
3%
T/A로 전환된
비-프로토스페이서 C/G를 갖는
총 서열분석 판독물의 퍼센트

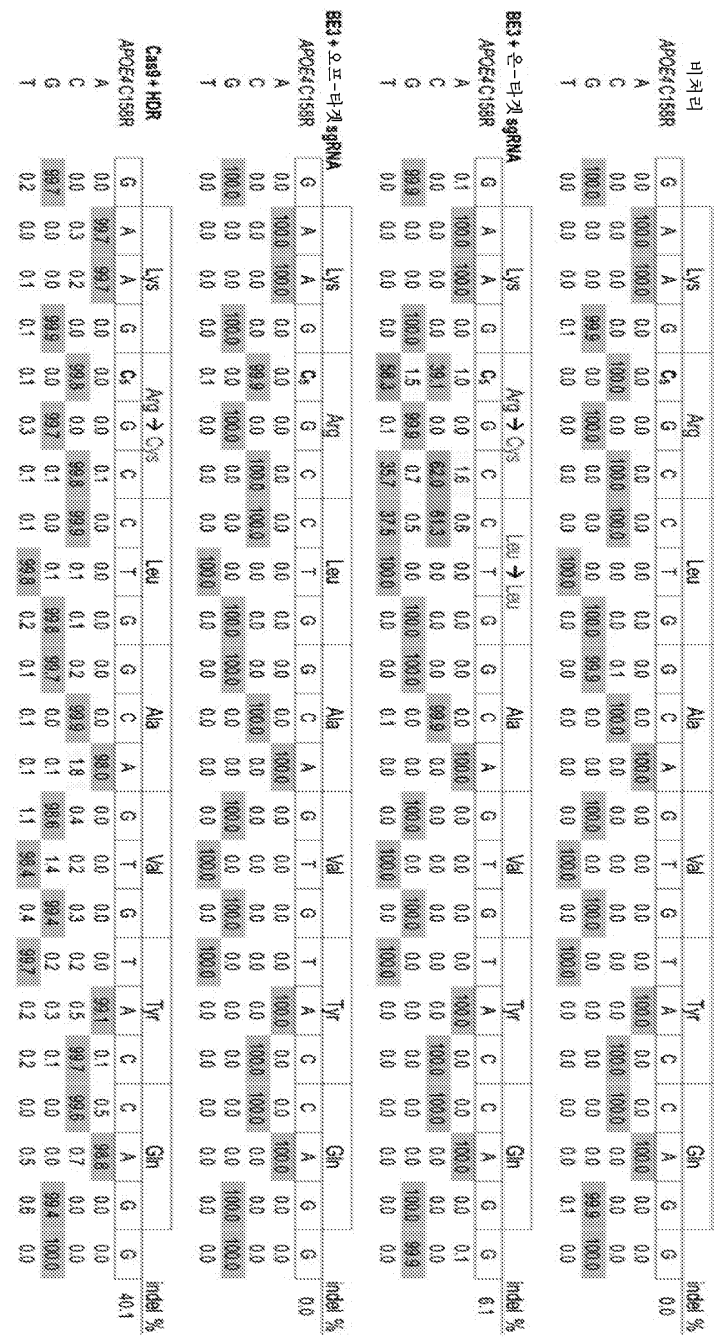
도면39b


T/A로 전환된
비-프로토스페이서 C/G를 갖는
총 서열분석 판독물의 퍼센트
0%
1%
2%
3%
0
-20
-40
-60
-80
-100
-120
-140
프로토스페이서로부터의 거리 (bp)
비처리
BE1
BE2
BE3

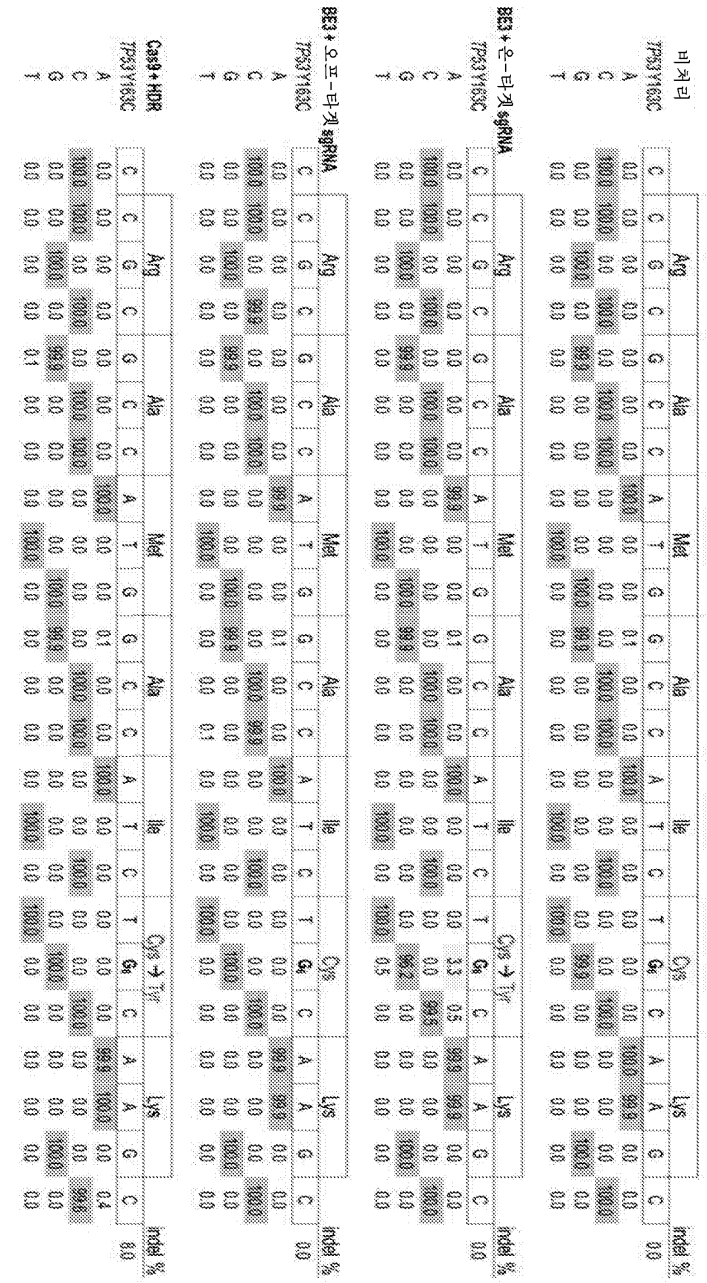
도면39c

비-프로토스페이서 C/G	평균 C/G (%)	평균 T/A (%)	최저 T/A (%)	최고 T/A (%)
비처리	99.95 ± 0.14	0.02 ± 0.02	0.00	2.44
BE1	99.95 ± 0.24	0.03 ± 0.03	0.00	1.64
BE2	99.95 ± 0.13	0.03 ± 0.03	0.00	1.92
BE3	99.97 ± 0.09	0.02 ± 0.02	0.00	2.52

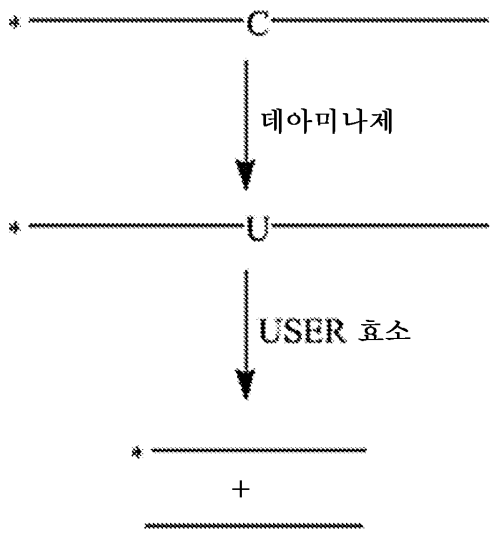
도면40a


비처리
APOE4-C158R
Lys
Arg
Ala
Leu
Val
Tyr
Gln
indel %
BE3 + 온-타겟 sgRNA
Arg → Cys
Leu → Leu
BE3 + 오프-타겟 sgRNA
Cas9 + HDR
Arg → Cys

도면40b


비처리
TP53Y163C
Arg
Ala
Met
Ile
Cys
Lys
indel %
BE3 + 온-타겟 sgRNA
BE3 + 오프-타겟 sgRNA
Cas9 + HDR
Arg
Ala
Met
Ile
Cys → Tyr
Lys
indel %

도면41



도면42

CDA

	P	A	M	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	-1	-2	-3
	T	G	C	A	T	C	A	C	G	T	G	C	T	C	A	A	T	C	G	G	G	G	C	C	C	C	A
A	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	99.9%	0.0%	0.0%	0.5%	0.0%	0.0%	0.1%	100.0%	5.8%	0.0%	0.0%	0.0%	33.9%	33.9%	33.1%	0.0%	0.0%	0.0%	0.0%	100.0%
C	0.0%	0.0%	99.9%	99.9%	0.0%	99.9%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	99.9%	0.0%	1.0%	0.0%	99.9%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	99.9%	99.9%	0.0%
G	0.0%	99.9%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	0.0%	99.5%	0.0%	0.0%	0.0%	0.0%	91.8%	0.0%	0.0%	65.7%	65.0%	65.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
T	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	1.5%	100.0%	0.0%	0.3%	1.1%	1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

rAPOBEC1

	A	C	G	T
A	0.0%	0.0%	0.0%	100.0%
C	0.0%	0.0%	99.9%	0.0%
G	0.0%	99.9%	0.0%	0.0%
T	100.0%	0.0%	0.0%	0.0%

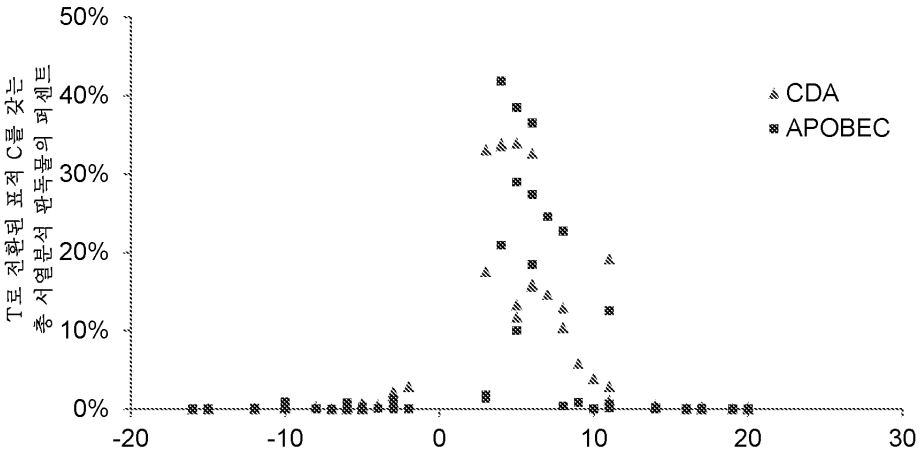
비처리

	A	C	G	T
A	0.0%	0.0%	0.0%	100.0%
C	0.0%	0.0%	99.9%	0.0%
G	0.0%	99.9%	0.0%	0.0%
T	100.0%	0.0%	0.0%	0.0%

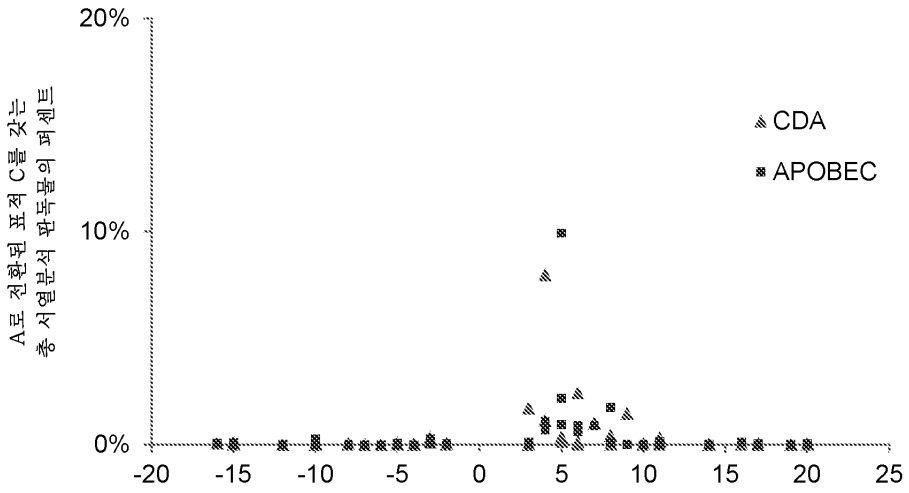
도면43

Hek-2부위, 나탄네이점: 코딩 가닥 (G→A), 표적: 주형 가닥 (C→T)
pmCDA1
nCas9
UGI
NLS
P A M 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 -1 -2 -3
C G C C C G C A G T C T A T G C T T T T G T T T T C C A G
CDA
A 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0% 19.2% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 32.6% 0.0% 33.6% 0.0% 0.0% 0.0% 99.8% 99.9% 0.8%
C 99.9% 0.0% 100.0% 99.7% 99.9% 99.5% 0.0% 99.9% 0.0% 0.0% 0.0% 0.5% 100.0% 0.0% 0.0% 0.0% 0.0% 4.5% 0.0% 0.1% 0.0% 0.0% 0.0% 99.8% 0.0% 0.3%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 60.4% 0.0% 58.3% 0.0% 0.0% 0.0% 0.0% 99.8% 98.8%
T 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 0.0% 80.0% 0.0% 100.0% 100.0% 100.0% 2.4% 100.0% 8.0% 100.0% 100.0% 100.0% 0.2% 0.4% 0.0% 0.1%
rAPOBEC1
A 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.7% 0.0% 0.0% 0.0% 0.0% 0.0% 18.4% 0.0% 20.9% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0%
C 99.9% 0.0% 100.0% 100.0% 99.9% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 38.6% 0.0% 0.4% 0.0% 0.0% 0.0% 99.9% 99.9% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 99.3% 0.0% 0.0% 0.0% 0.0% 0.0% 42.3% 0.0% 77.6% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.7%
T 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 0.0% 100.0% 0.0% 100.0% 100.0% 100.0% 0.7% 100.0% 1.1% 100.0% 100.0% 100.0% 0.0% 0.0% 0.0% 0.3%
비처리
A 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.1% 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0%
C 99.9% 0.0% 100.0% 100.0% 99.9% 99.9% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9% 0.0%
G 0.0% 99.9% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 99.9%
T 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 100.0% 0.0% 100.0% 0.0% 100.0% 0.0% 100.0% 100.0% 100.0% 0.0% 100.0% 0.0% 100.0% 100.0% 100.0% 0.0% 0.0% 0.0% 0.0%

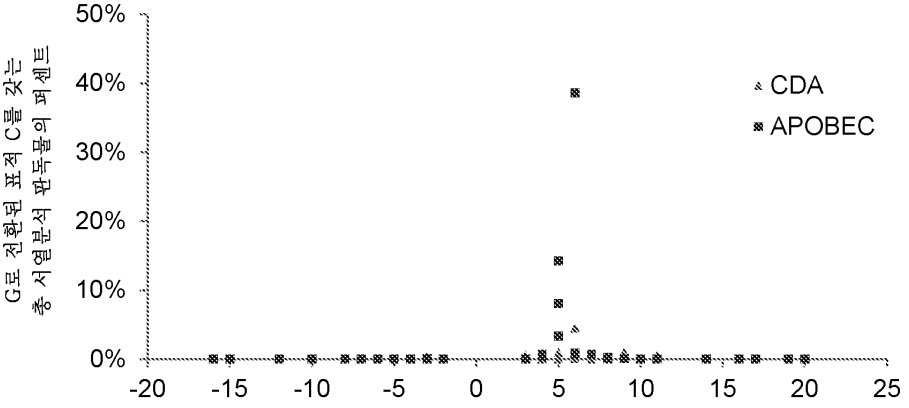
도면44



도면45



도면46



도면47

rAPOBEC1

	P	A	M	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
	C	C	C	G	C	A	G	T	C	T	A	T	G	C	T	T	T	G	T	G	T	T	C	
A	0.02%	0.02%	0.01%	0.02%	0.02%	0.05%	0.01%	99.96%	0.09%	0.00%	0.01%	0.00%	99.96%	0.01%	0.02%	0.01%	0.00%	0.00%	0.00%	9.30%	0.01%	10.87%	0.00%	0.01%	0.01%
C	99.96%	0.00%	99.88%	99.97%	99.97%	99.97%	0.00%	99.88%	0.02%	0.01%	0.01%	99.98%	0.01%	0.00%	0.01%	0.01%	0.01%	18.14%	0.01%	0.54%	0.01%	0.01%	99.95%		
G	0.00%	99.96%	0.00%	0.00%	0.00%	0.00%	99.93%	0.00%	0.01%	99.86%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%	72.40%	0.00%	88.10%	0.00%	0.00%	0.01%		
T	0.02%	0.02%	0.01%	0.01%	0.01%	0.01%	0.02%	0.01%	0.03%	99.96%	0.01%	99.96%	0.02%	99.96%	0.01%	99.99%	99.98%	99.99%	0.17%	99.98%	0.49%	99.98%	99.98%	0.03%	

APOBEC3G

A	0.01%	0.01%	0.01%	0.01%	0.01%	0.02%	99.97%	0.01%	0.00%	0.00%	0.00%	99.97%	0.01%	0.03%	0.00%	0.03%	2.18%	0.01%	0.34%	0.00%	0.00%	0.00%	0.01%
C	0.00%	99.97%	0.00%	0.00%	0.00%	0.00%	0.00%	99.98%	0.00%	0.00%	0.00%	0.00%	0.00%	99.96%	0.01%	0.01%	5.34%	0.01%	0.15%	0.01%	0.02%	99.98%	0.01%
G	0.02%	0.01%	0.01%	0.01%	0.02%	0.01%	0.00%	0.00%	0.00%	99.99%	0.00%	0.00%	99.98%	0.00%	0.00%	0.00%	92.44%	0.00%	99.46%	0.00%	0.00%	0.00%	0.00%
T																							

APOBEC3G_RR																							
A	0.01%	0.02%	0.01%	0.01%	0.02%	0.01%	0.01%	0.01%	0.00%	0.00%	0.00%	0.04%	0.01%	0.00%	0.01%	0.00%	2.18%	0.01%	6.12%	0.00%	0.00%	0.00%	0.01%
C	99.97%	0.00%	99.98%	99.99%	99.97%	99.97%	0.00%	99.98%	0.02%	0.01%	0.01%	0.01%	99.98%	0.01%	0.01%	0.01%	4.98%	0.01%	2.49%	0.01%	0.02%	99.99%	
G	0.00%	99.96%	0.00%	0.00%	0.00%	0.00%	99.92%	0.00%	0.01%	99.98%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	92.80%	0.00%	91.04%	0.00%	0.00%	0.00%	0.00%
T	0.02%	0.02%	0.01%	0.01%	0.01%	0.02%	0.02%	0.01%	0.00%	0.01%	99.98%	0.01%	0.01%	99.98%	0.00%	0.00%	0.05%	99.98%	0.35%	99.98%	99.98%	0.01%	

비커리

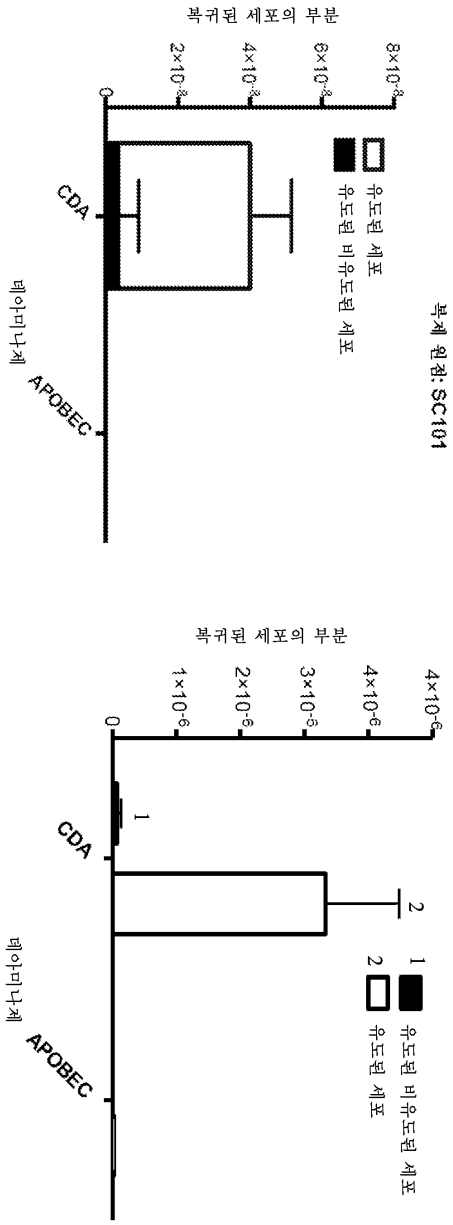
A	0.02%	0.02%	0.00%	0.01%	0.02%	0.01%	0.05%	0.01%	99.98%	0.00%	0.00%	0.01%	0.00%	99.97%	0.01%	0.00%	0.00%	0.00%	0.02%	0.01%	0.01%	0.00%	0.01%
C	99.96%	0.00%	99.99%	99.98%	99.97%	99.97%	0.00%	99.98%	0.01%	0.01%	0.01%	0.00%	99.98%	0.01%	0.00%	0.00%	0.00%	0.00%	0.02%	0.01%	0.01%	0.01%	0.01%
G	0.00%	99.96%	0.00%	0.00%	0.00%	0.00%	99.91%	0.00%	0.01%	99.98%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.96%	0.00%	99.96%	0.00%	0.00%	0.00%
T	0.02%	0.02%	0.01%	0.01%	0.01%	0.01%	0.03%	0.01%	0.00%	0.01%	99.99%	0.02%	99.98%	0.01%	99.99%	99.99%	99.99%	0.02%	99.98%	0.02%	99.98%	99.98%	0.01%

도면48

F 플라스미드 상에 코딩된 lacZ를 갖는 박테리아 균주를 이용함
중요한 글루탐산 잔기가 돌연변이된다. 복귀는 lacZ 활성을 복원한다.

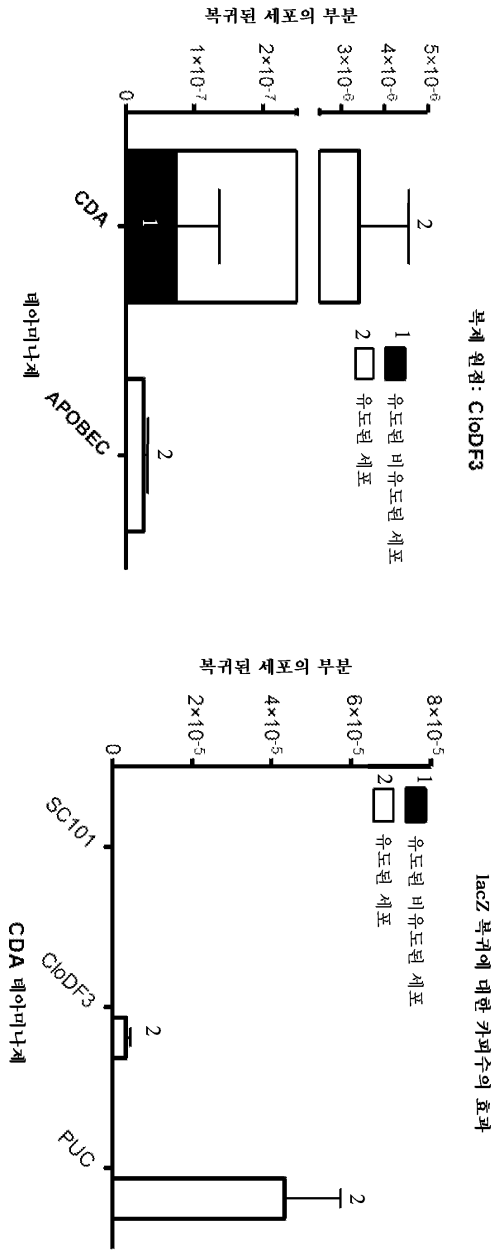
선택 검정을 개발하는 데 사용된 균주
사용된 구축물: APOBEC1 및 CDA

도면49i


복귀된 세포의 부분
0
2×10-8
4×10-8
6×10-8
8×10-8
유도된 세포
유도된 비유도된 세포
CDA
데아미나제
APOBEC
복제 원점: SC101

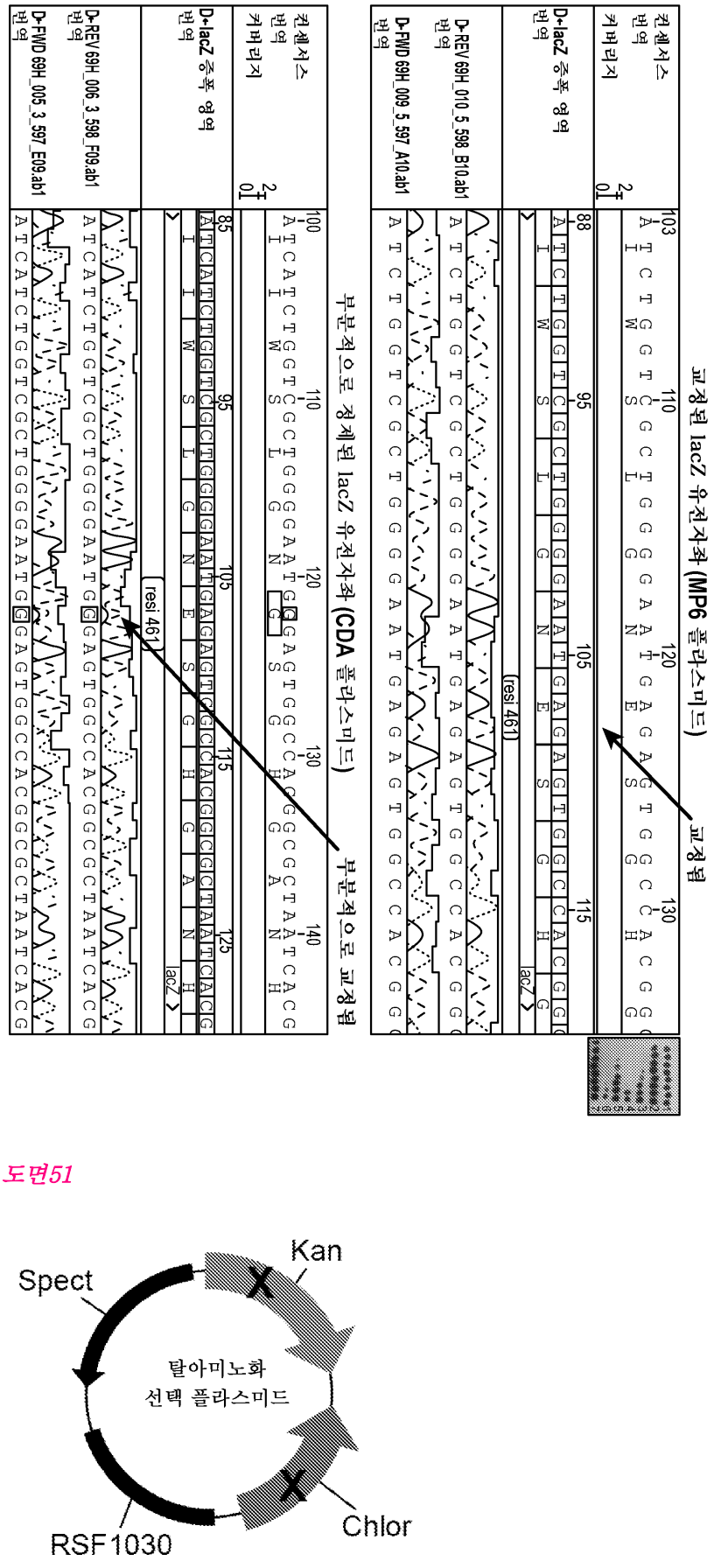
복귀된 세포의 부분
0
1×10-6
2×10-6
3×10-6
4×10-6
4×10-6
1
2
유도된 비유도된 세포
유도된 세포
CDA
데아미나제
APOBEC

도면49ii


복귀된 세포의 부분
5×10⁻⁶
4×10⁻⁶
3×10⁻⁶
2×10⁻⁷
1×10⁻⁷
0
CDA
APOBEC
1
2
2
2
1 유도된 비유도된 세포
2 유도된 세포
복제 원점: ClODF3
테아미나제

복귀된 세포의 부분
8×10⁻⁵
6×10⁻⁵
4×10⁻⁵
2×10⁻⁵
0
SC101
ClODF3
PUC
1
2
2
2
1 유도된 비유도된 세포
2 유도된 세포
lacZ 복귀에 대한 카피수의 효과
CDA 테아미나제

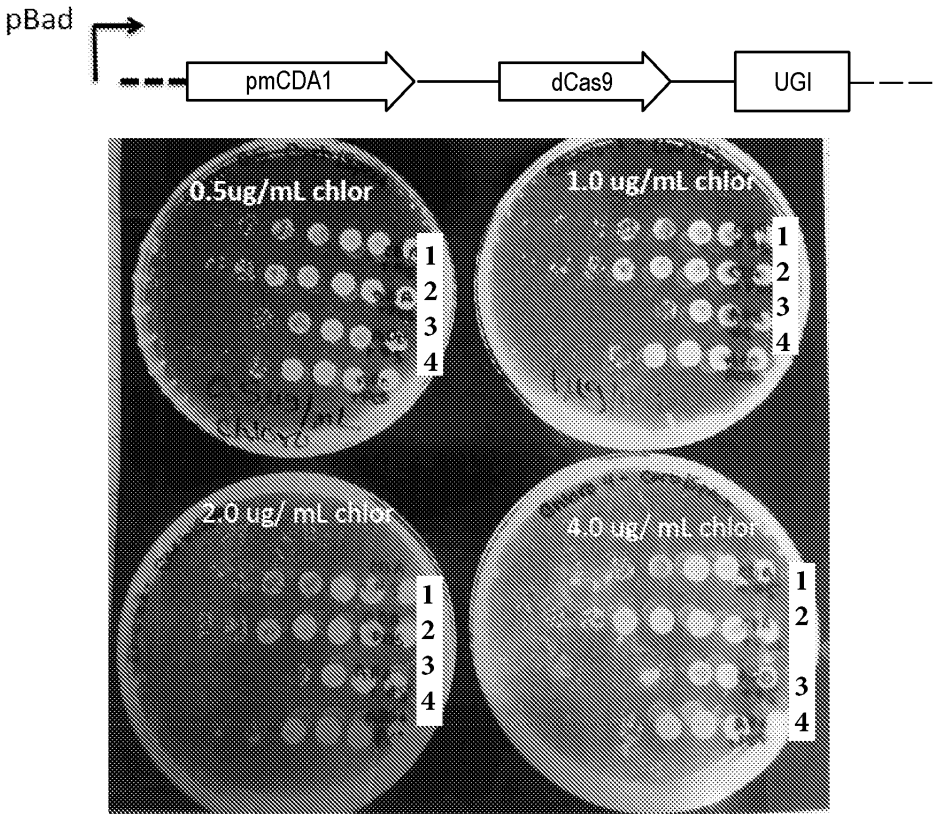
도면50


교정된 lacZ 유전자좌 (MP6 플라스미드)
교정됨
부분적으로 정제된 lacZ 유전자좌 (CDA 플라스미드)
부분적으로 교정됨
컨센서스
번역
커버리지
D-lacZ 증폭 영역
번역
D-REV 69H_010_5_598_B10.ab1
번역
D-FWD 69H_009_5_597_A10.ab1
번역
D-REV 69H_006_3_598_F09.ab1
번역
D-FWD 69H_005_3_597_E09.ab1
번역
(resi 461)
lacZ

도면51

Kan
Spect
탈아미노화
선택 플라스미드
Chlor
RSF1030

도면52



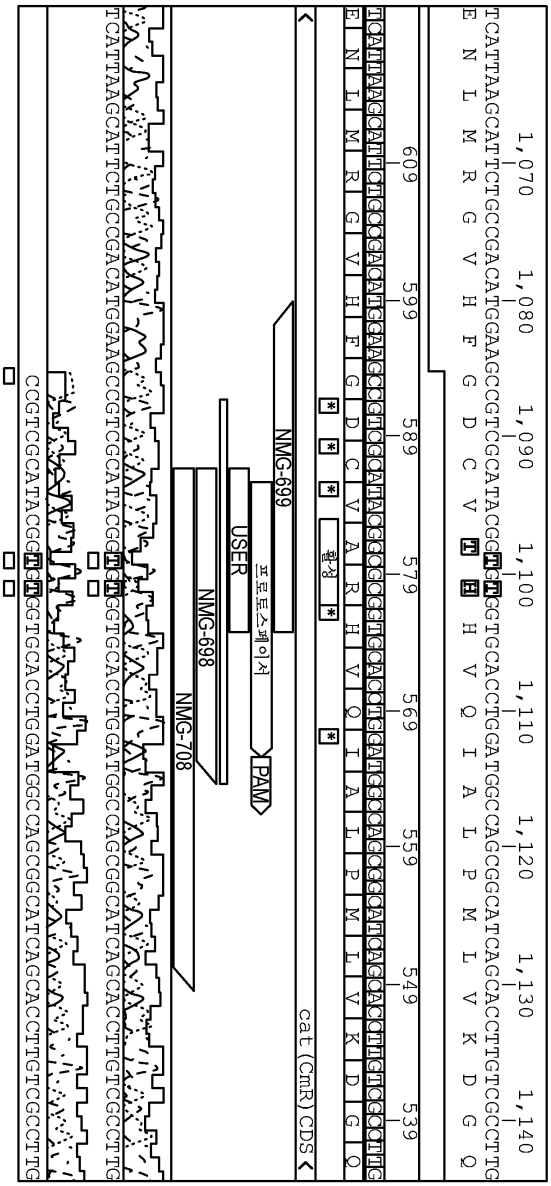
제1열: CDA-dCas9 + 선택 플라스미드 (chlor^{S})

제2열: CDA-dCas9 + 양성 대조군 선택 (chlor^{R})

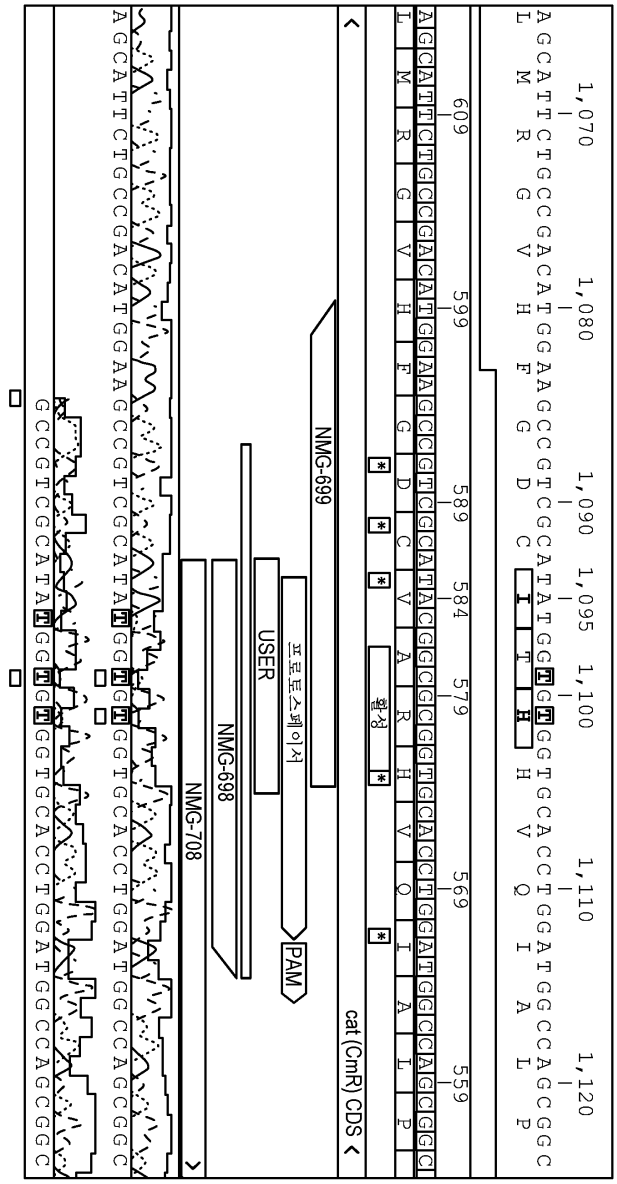
제3열: rAPOBEC-dCas9 + 선택 플라스미드 (chlor^{S})

제4열: rAPOBEC-dCas9 + 양성 대조군 선택 (chlor^{R})

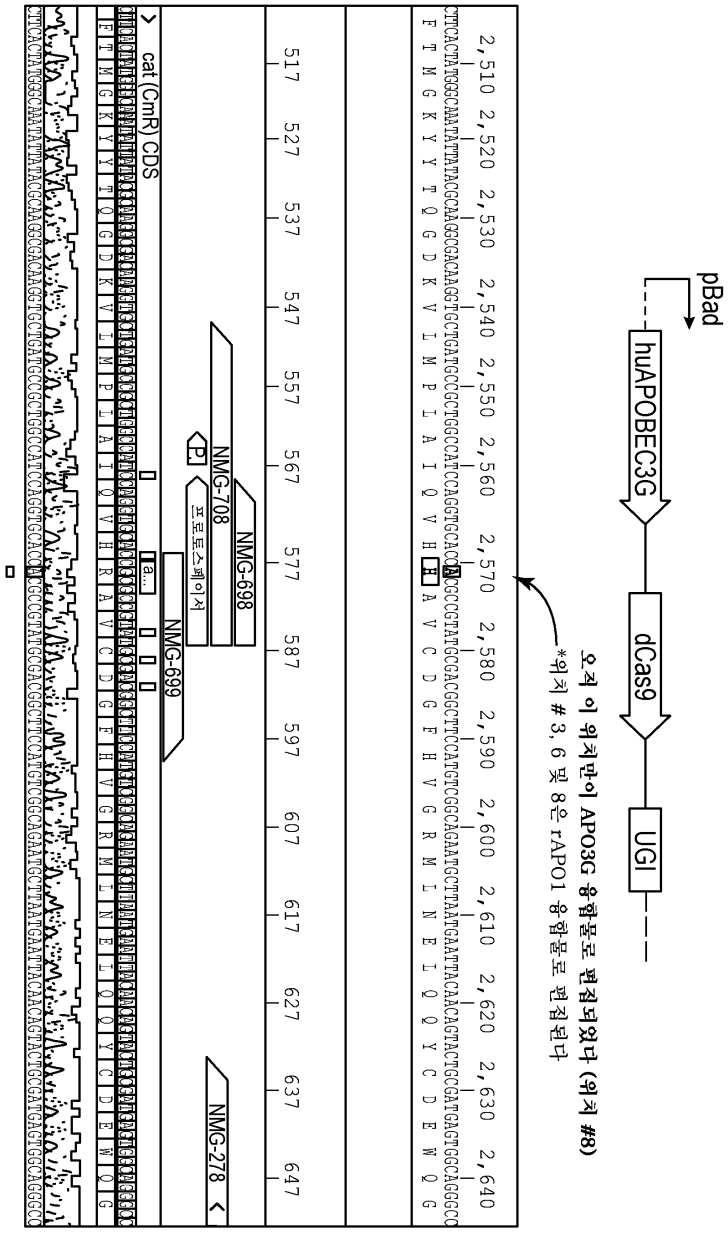
도면53a


8 ug/mL chlor 상의 rAPOBEC-XTEN-dCas9-UGI
NMG-699
프로토스페이서
PAM
USER
NMG-698
NMG-708
활성
cat(CmR) CDS

도면53b


CDA-XTEN-dCas9-UGI: 8 ug/mL 클로람페니콜 상의 생존
NMG-699
프로토스페이서
PAM
USER
NMG-698
NMG-708
cat (CmR) CDS

도면54


pBad
huAPOBEC3G
dCas9
UGI
오직 이 위치만이 APO3G 융합물로 편집되었다 (위치 #8)
*위치 # 3, 6 및 8은 rAPO1 융합물로 편집된다
NMG-698
NMG-708
프로모터
NMG-699
NMG-278
cat (CmR) CDS

도면55i

타겟X1 온 타겟
비처리
BE3
HF-BE3
타겟X1 오프 타겟1
비처리
BE3
HF-BE3
타겟X1 오프 타겟2
비처리
BE3
HF-BE3
타겟X1 오프 타겟3
비처리
BE3
HF-BE3
타겟X1 오프 타겟4
비처리
BE3
HF-BE3

도면55ii

EMX1 오프 타겟 6
비커리
BE3
HF-BE3
EMX1 오프 타겟 7
비커리
BE3
HF-BE3
EMX1 오프 타겟 8
비커리
BE3
HF-BE3
EMX1 오프 타겟 9
비커리
BE3
HF-BE3
EMX1 오프 타겟 10
비커리
BE3
HF-BE3

도면56

EMX1		C_3	C_6	C_{12}	indel %
비처리	A	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	
	T	0±0	0±0	0±0	
BE3	A	1.3±0.1	0.7±0.2	0±0	2.6±0.3
	C	51.7±1	55.2±0.1	99.9±0	
	G	4±0.7	1.9±0.4	0±0	
	T	42.8±0.2	41.9±0.4	0±0	
HF BE3	A	1.5±0.7	0.5±0.1	0±0	1±0.1
	C	44.8±8.5	49.9±6.6	99.9±0	
	G	4.2±1.2	0.6±0.2	0±0	
	T	49.3±6.5	48.7±6.3	0±0	

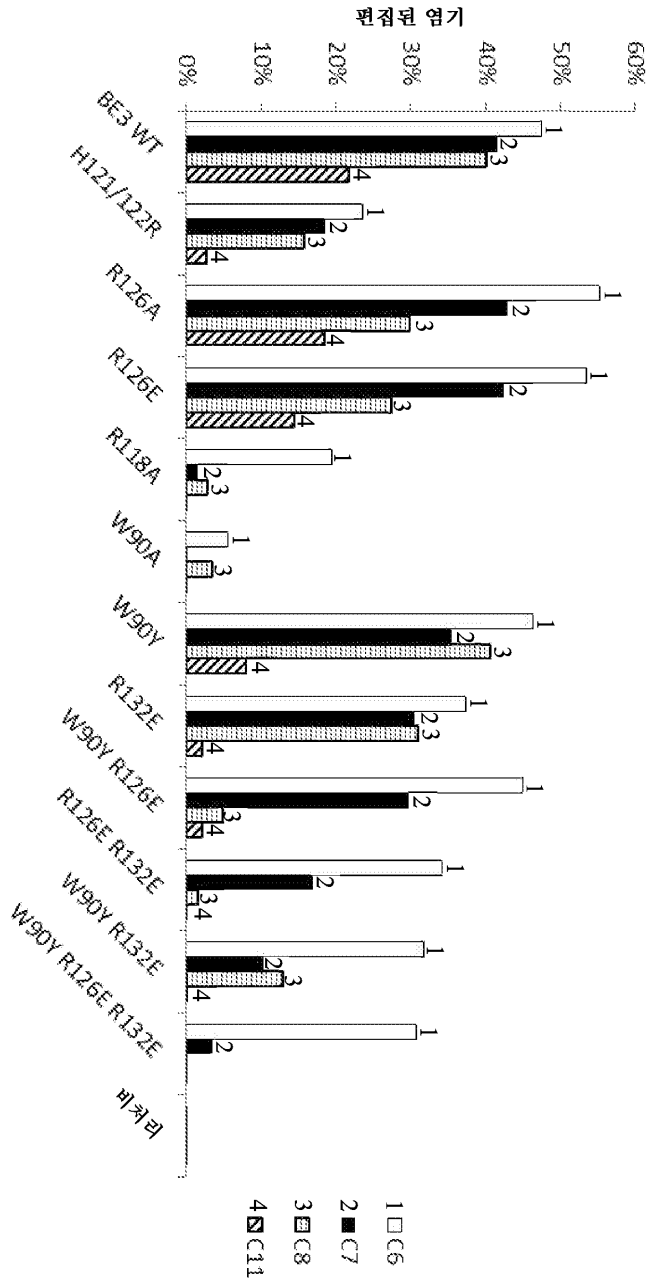
RNF2		C_3	C_6	C_{12}	indel %
비처리	A	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	
	T	0±0	0±0	0±0	
BE3	A	0.4±0	1.5±0.2	0.1±0	2.3±0.3
	C	70.9±2.7	45±0	95±0.2	
	G	0±0	12.6±2.4	0.1±0	
	T	28.5±2.7	40.7±2.7	4.6±0.2	
HF BE3	A	0±0	0.2±0	0±0	0.5±0.1
	C	76.8±1.8	64.5±3.3	95.4±0.1	
	G	0±0	2.4±0.5	0±0	
	T	23±1.8	32.7±2.6	4.5±0.1	

FANCF		C_6	C_7	C_8	C_{11}	indel %
비처리	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0.8±0.1	1.2±0.3	1.2±0.1	0.2±0.1	5.8±0.9
	C	64.7±5.9	67.9±5.1	68.9±5.2	85.4±2.7	
	G	0.6±0	0.7±0	0.7±0.1	0.1±0	
	T	33.8±5.7	29.9±4.8	29±5	14.1±2.6	
HF BE3	A	0.9±0.2	2±0.4	0.9±0.2	0.4±0	5.9±0.7
	C	52.5±8.4	57.7±7.6	61.9±6.8	84.7±2.4	
	G	1.4±0.2	0.6±0	0.5±0.1	0.3±0	
	T	45±8	39.5±7.2	36.5±6.5	14.4±2.2	

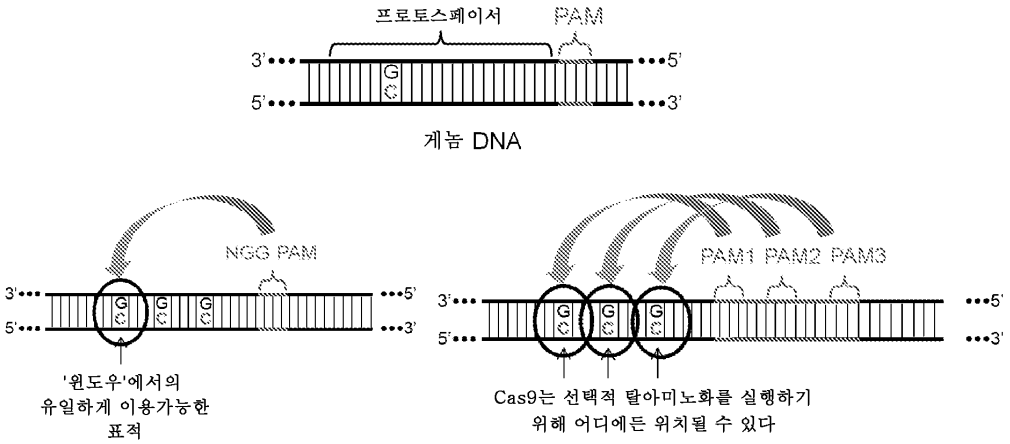
HEK3		C_6	C_7	C_8	C_{11}	indel %
비처리	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0±0	1.6±0.2	1.5±0.3	0.2±0	2.7±0.4
	C	98.7±0.2	48.6±5.4	40.8±6.7	98.6±0.1	
	G	0±0	1.4±0.2	10.5±1.2	0.3±0	
	T	1.1±0.2	48.2±4.9	47±5.2	0.7±0.1	
HF BE3	A	0±0	0.8±0.2	1.1±0.3	0.4±0	4.1±0.7
	C	97.6±0.4	65±3.8	40.4±6.3	95.9±0.5	
	G	0±0	0.9±0.2	6.1±1.1	0.7±0.2	
	T	2±0.4	33±3.4	52.2±5.2	2.7±0.2	

HEK4		C_6	C_7	C_8	C_{11}	indel %
비처리	A	0±0	0±0	0±0	0±0	0±0
	C	99.9±0	99.9±0	99.9±0	99.9±0	
	G	0±0	0±0	0±0	0±0	
	T	0±0	0±0	0±0	0±0	
BE3	A	0±0	5.4±1.3	0±0	0±0	3±0.7
	C	98.7±0	44.5±8.2	98.9±0	99.6±0	
	G	0±0	18.1±2.3	0±0	0±0	
	T	1.1±0.1	31.9±4.8	0.9±0.1	0.1±0	
HF BE3	A	0±0	5.2±0.8	0±0	0±0	1.3±0
	C	99.6±0	36.6±9	98.5±0.1	99.9±0	
	G	0±0	16±2.8	0±0	0±0	
	T	0.2±0	42.1±5.3	1.3±0.1	0±0	

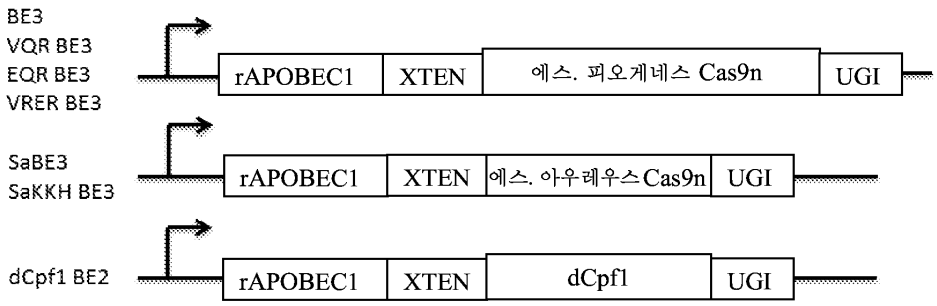
도면57


편집된 염기
0%
10%
20%
30%
40%
50%
60%
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
비처리
1 C6
2 C7
3 C8
4 C11

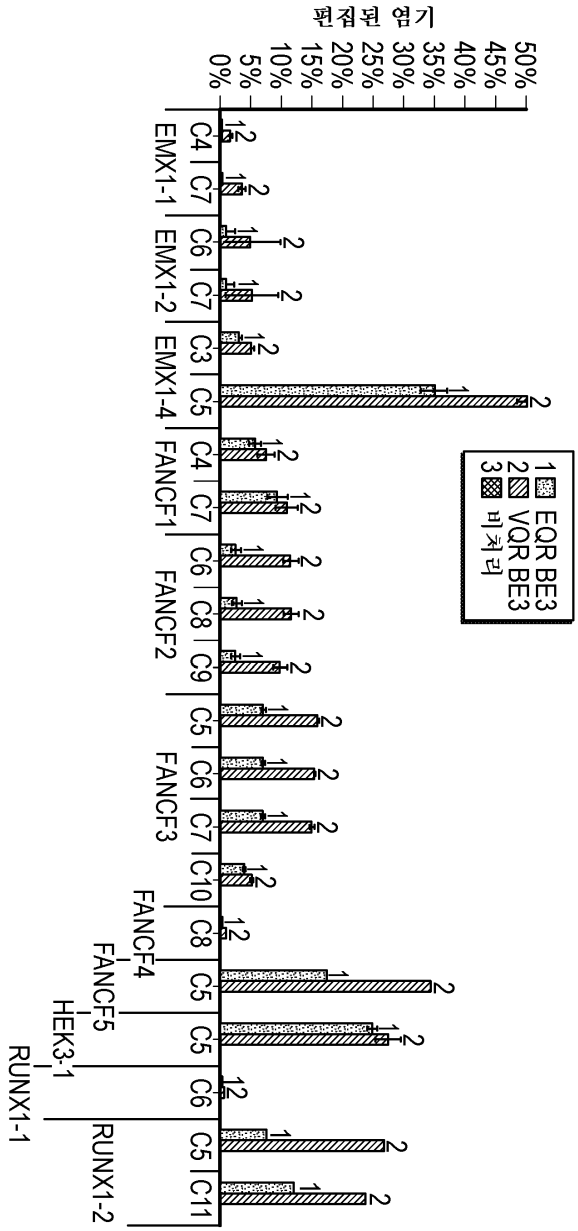
도면58


프로토스페이서
PAM
3'
5'
5'
3'
게놈 DNA
NGG PAM
'윈도우'에서의
유일하게 이용가능한
표적
PAM1 PAM2 PAM3
Cas9는 선택적 탈아미노화를 실행하기
위해 어디에든 위치될 수 있다

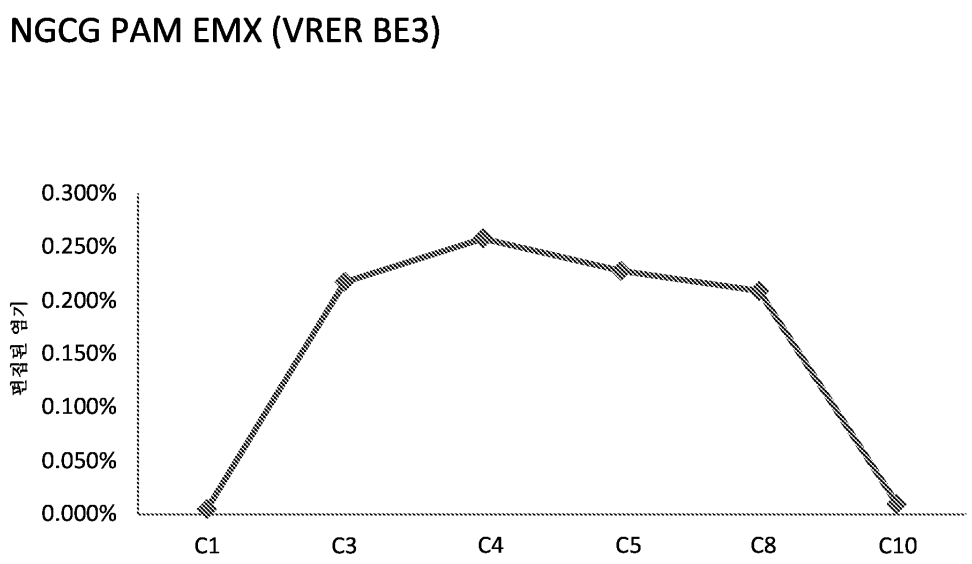
도면59


BE3
VQR BE3
EQR BE3
VRER BE3
rAPOBEC1
XTEN
에스. 피오게네스 Cas9n
UGI
SaBE3
SaKKH BE3
rAPOBEC1
XTEN
에스. 아우레우스 Cas9n
UGI
dCpf1 BE2
rAPOBEC1
XTEN
dCpf1
UGI

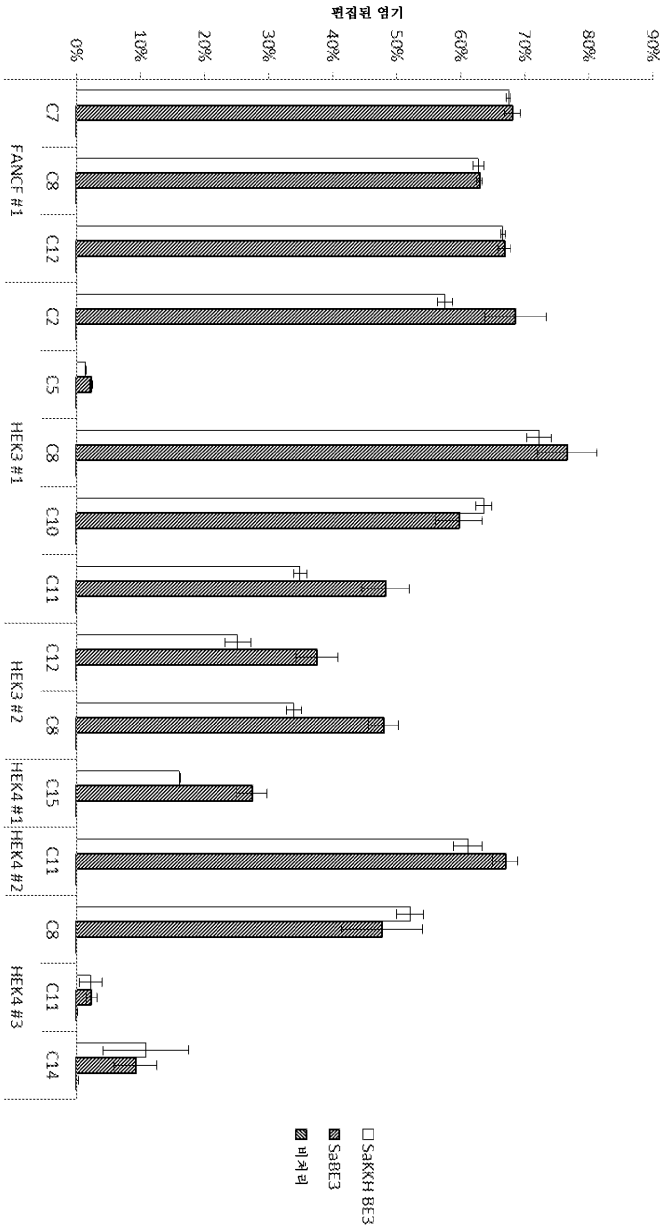
도면60


편집된 염기
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
1 EQR BE3
2 VQR BE3
3 비처리
EMX1-1
EMX1-2
EMX1-4
FANCF1
FANCF2
FANCF3
FANCF4
FANCF5
HEK3-1
RUNX1-1
RUNX1-2
C4
C7
C6
C7
C3
C5
C4
C7
C6
C8
C9
C5
C6
C7
C10
C8
C5
C5
C6
C5
C11

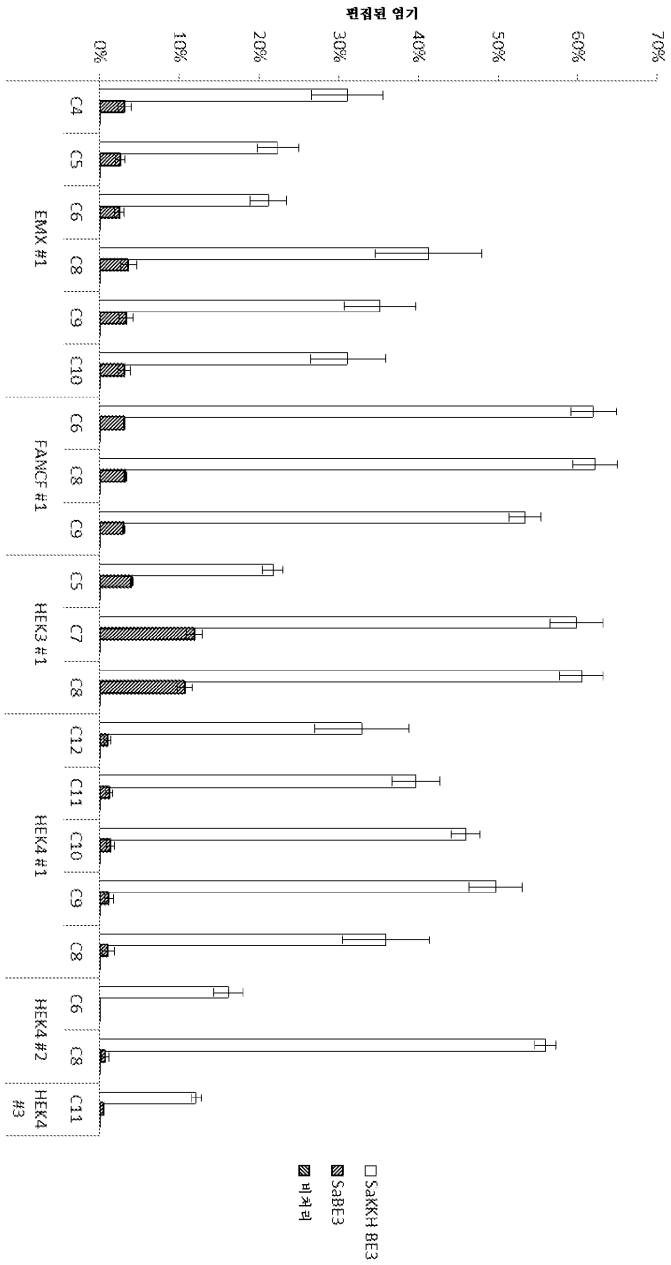
도면61


NGCG PAM EMX (VRER BE3)
0.300%
0.250%
0.200%
0.150%
0.100%
0.050%
0.000%
편집된 염기
C1
C3
C4
C5
C8
C10

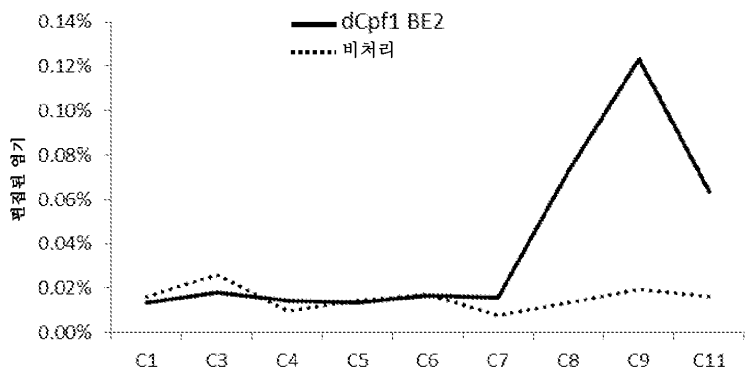
도면62


편집된 염기
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
FANCF #1
C7
C8
C12
HEK3 #1
C2
C5
C8
C10
C11
HEK3 #2
C12
C8
HEK4 #1
C15
HEK4 #2
C11
HEK4 #3
C8
C11
C14
비처리
SaBE3
SaKKH BE3

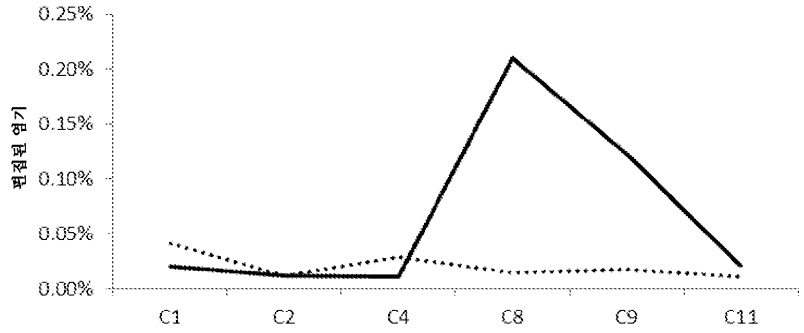
도면63


편집된 염기
0%
10%
20%
30%
40%
50%
60%
70%
EMX #1
C4
C5
C6
C8
C9
C10
FANCF #1
C6
C8
C9
HEK3 #1
C5
C7
C8
HEK4 #1
C12
C11
C10
C9
C8
HEK4 #2
C6
C8
HEK4 #3
C11
비처리
SaBE3
SaKKH BE3

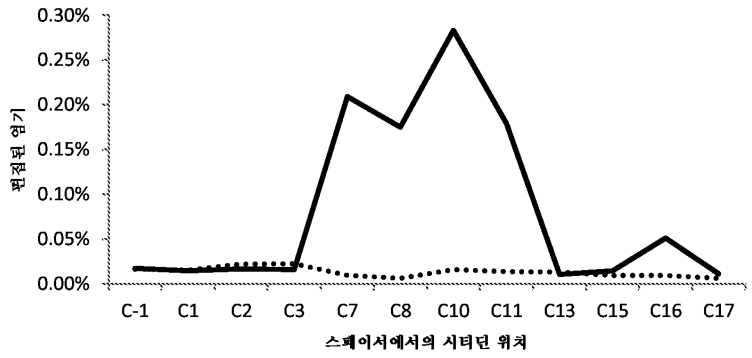
도면64a


dCpf1 BE2
비처리
편집된 염기
0.14%
0.12%
0.10%
0.08%
0.06%
0.04%
0.02%
0.00%
C1
C3
C4
C5
C6
C7
C8
C9
C11

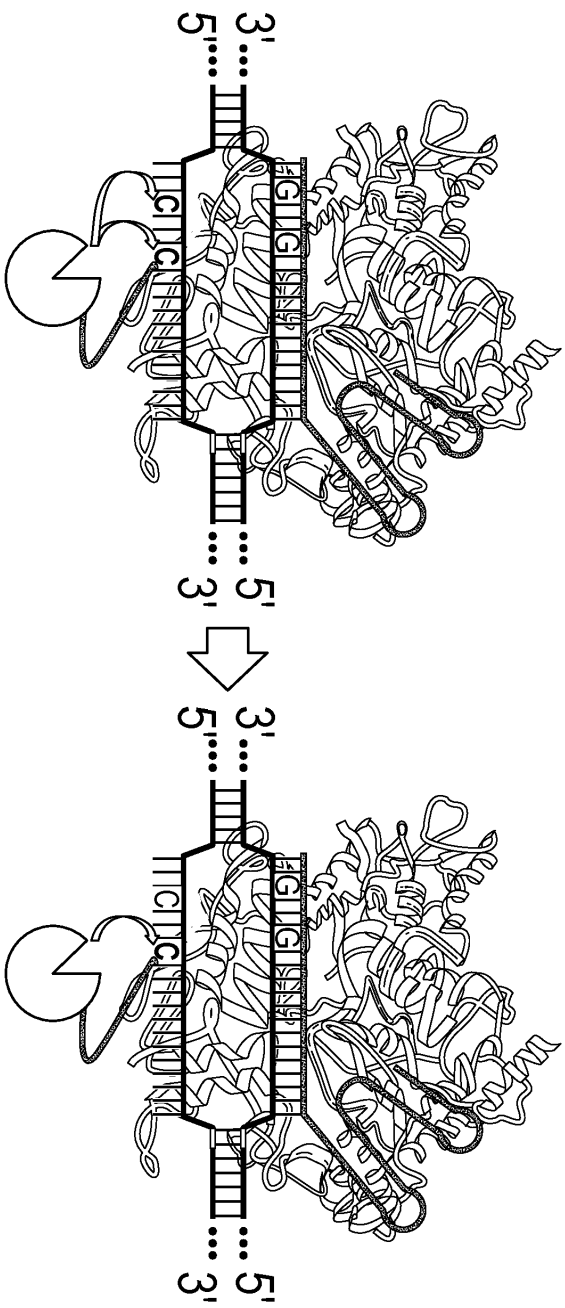
도면64b


0.25%
0.20%
0.15%
0.10%
0.05%
0.00%
편집된 염기
C1
C2
C4
C8
C9
C11

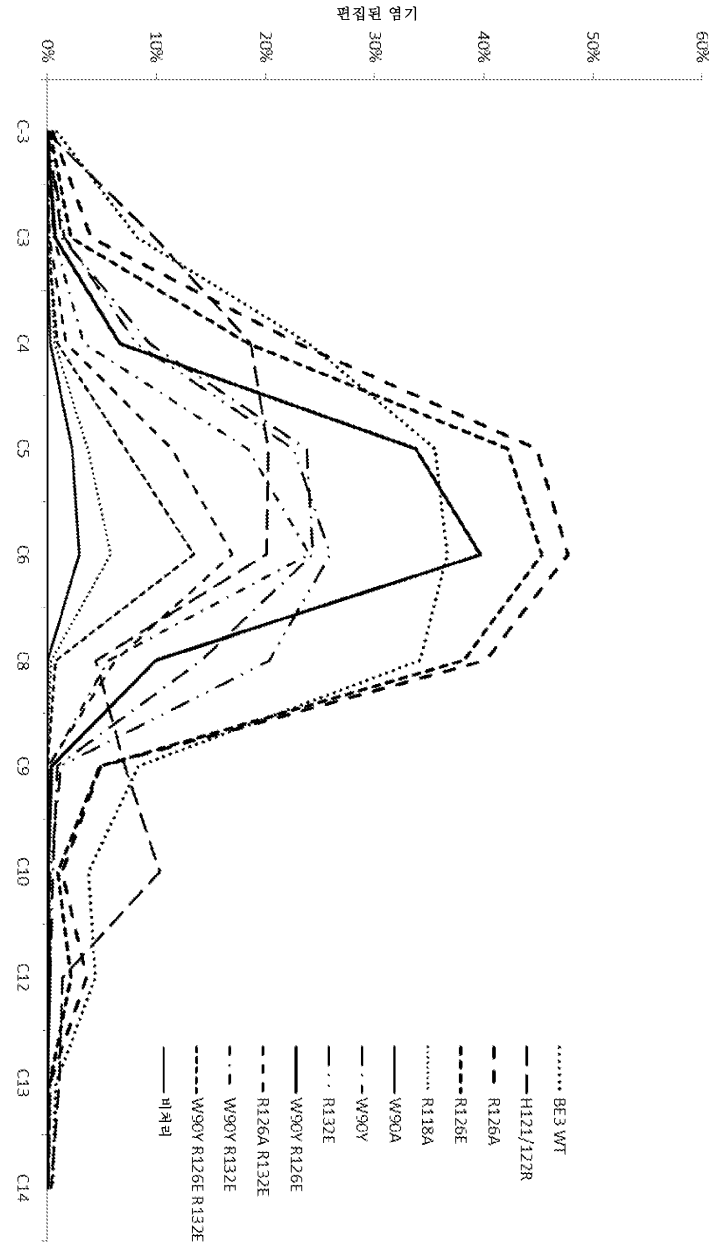
도면64c


0.30%
0.25%
0.20%
0.15%
0.10%
0.05%
0.00%
편집된 염기
C-1
C1
C2
C3
C7
C8
C10
C11
C13
C15
C16
C17
스페이서에서의 시티딘 위치

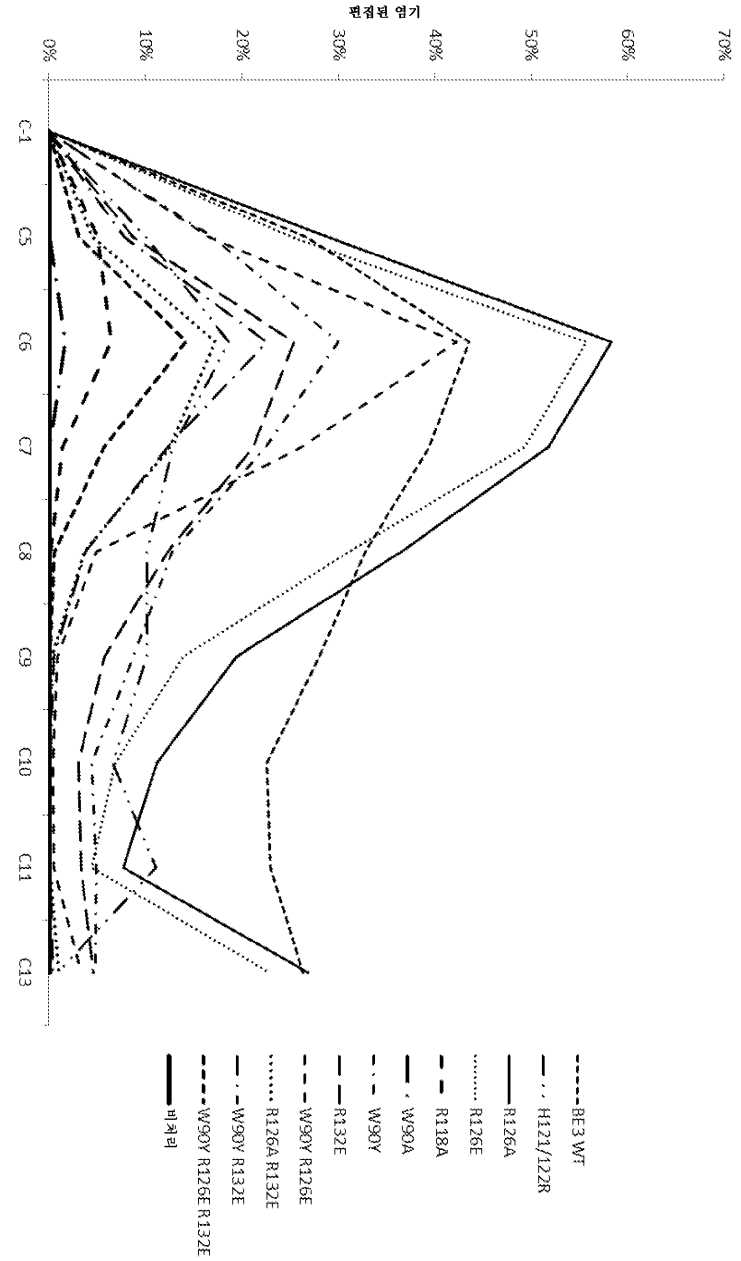
도면65



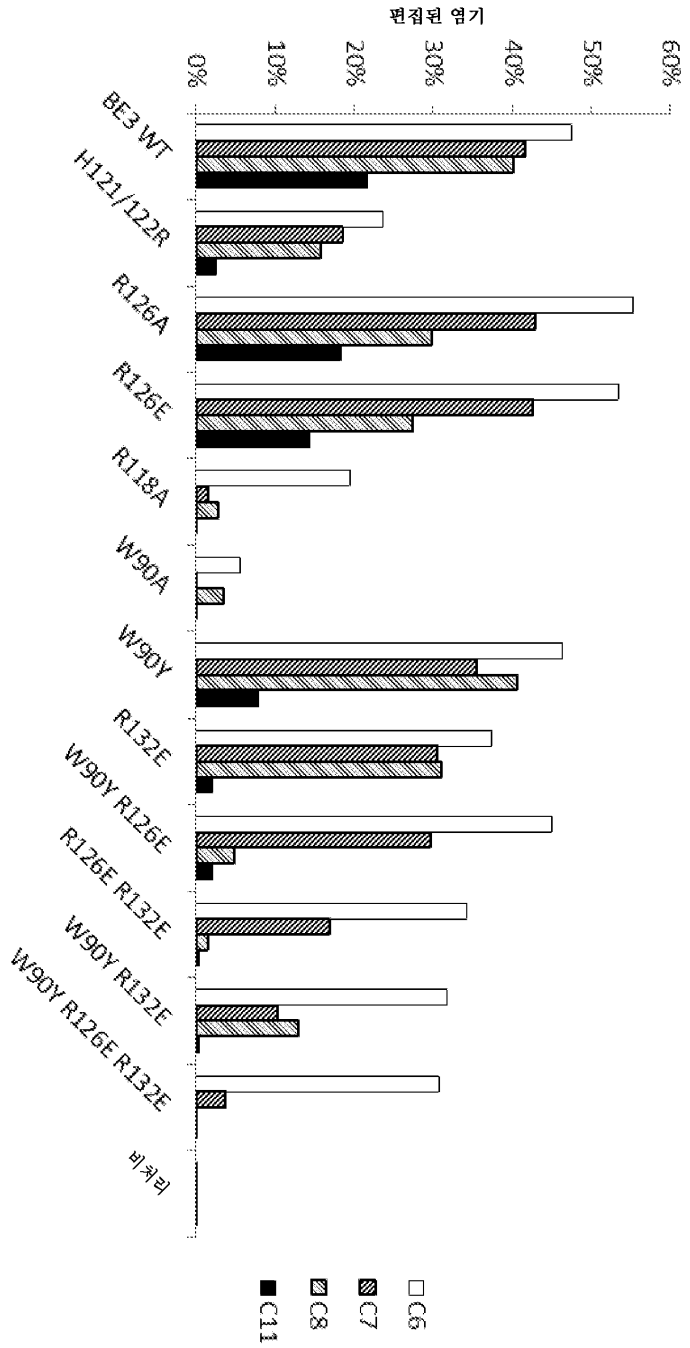
도면66


편집된 염기
0%
10%
20%
30%
40%
50%
60%
C-3
C3
C4
C5
C6
C8
C9
C10
C12
C13
C14
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126A R132E
W90Y R132E
W90Y R126E R132E
비처리

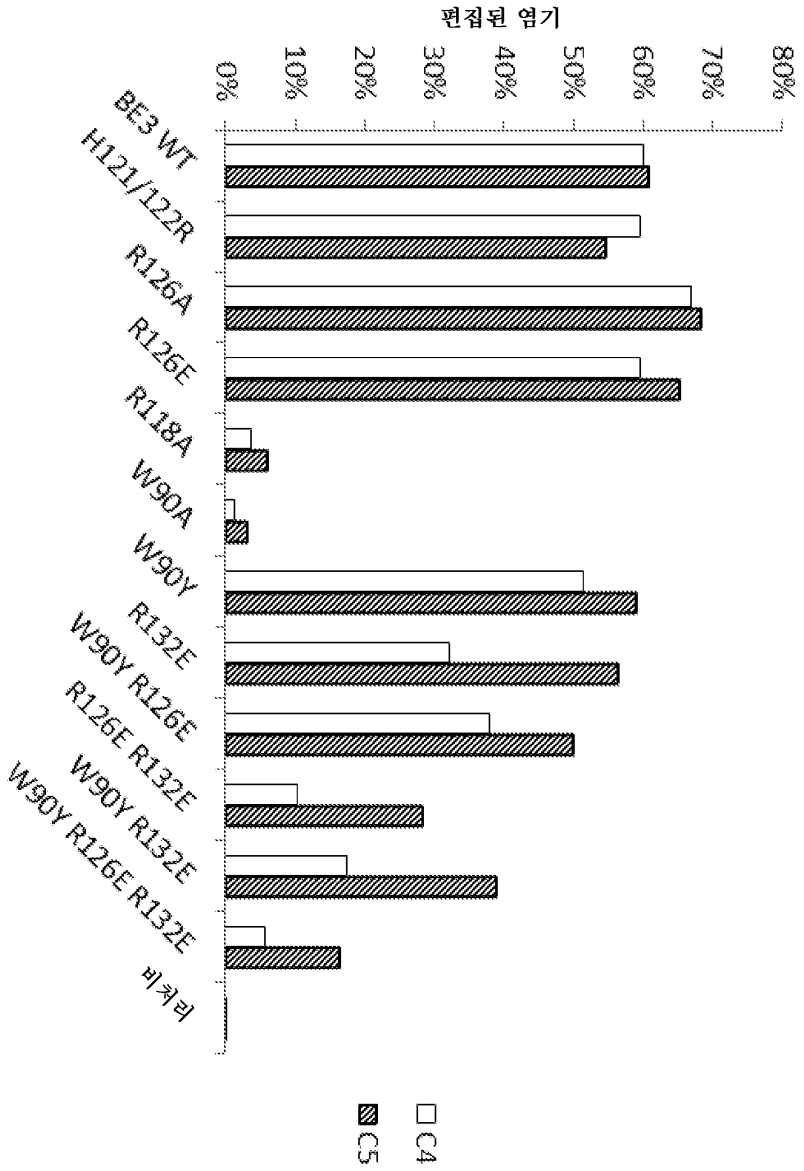
도면67


편집된 염기
0%
10%
20%
30%
40%
50%
60%
70%
C-1
C5
C6
C7
C8
C9
C10
C11
C13
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126A R132E
W90Y R132E
W90Y R126E R132E
비처리

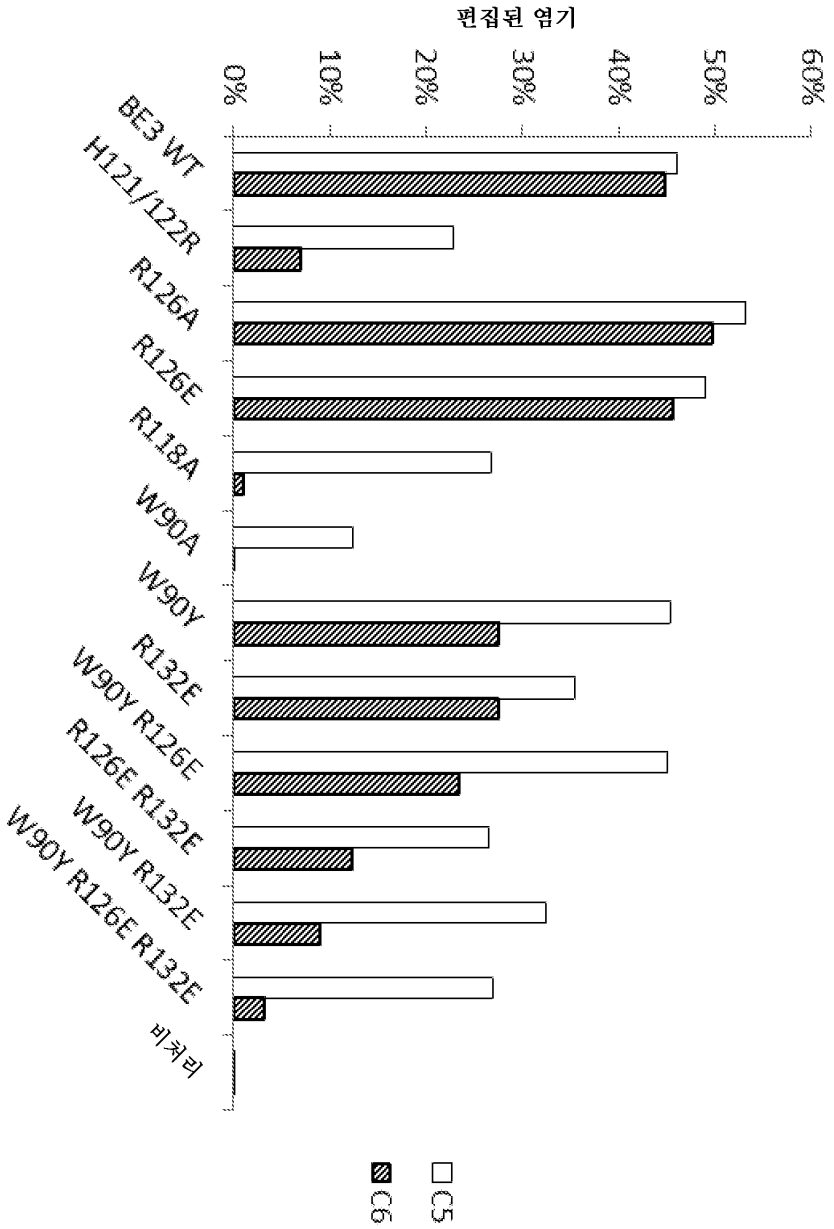
도면68


편집된 염기
0%
10%
20%
30%
40%
50%
60%
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
비처리
C11
C8
C7
C6

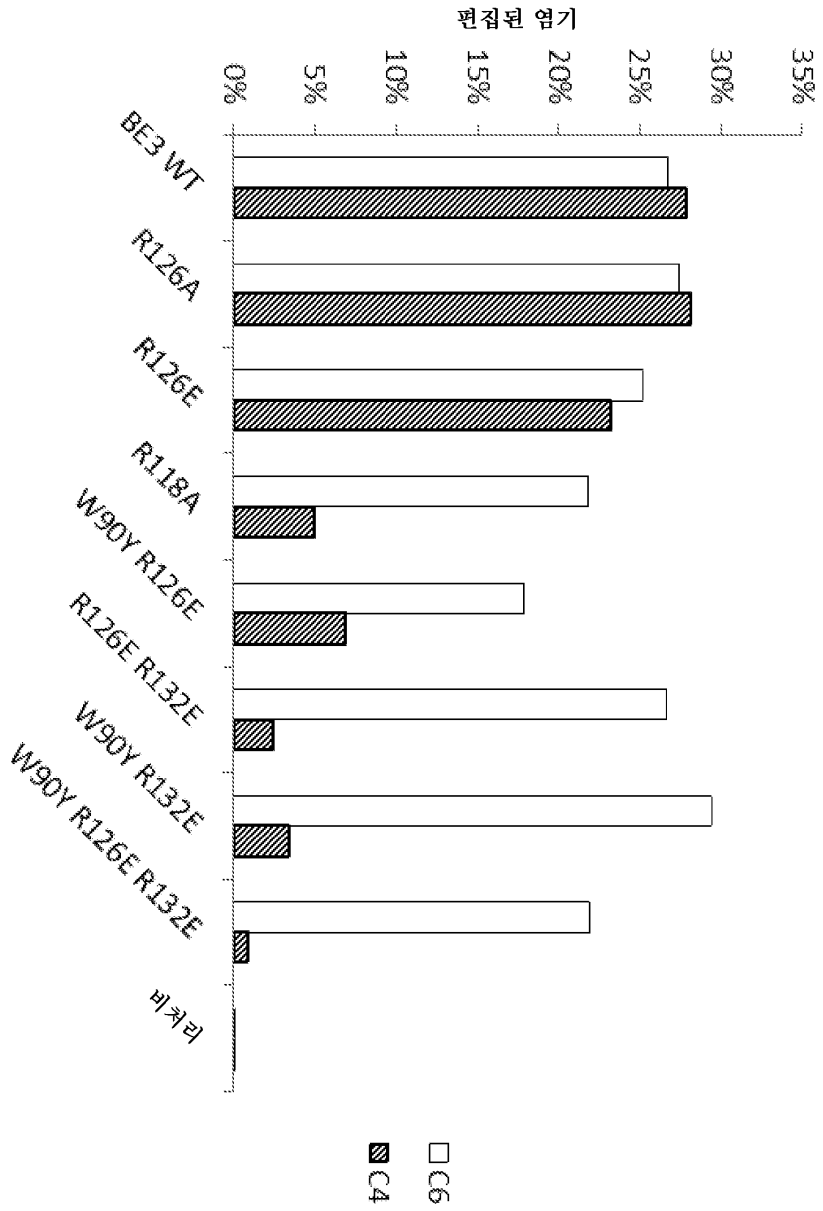
도면69


편집된 염기
0%
10%
20%
30%
40%
50%
60%
70%
80%
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
비커리
C5
C4

도면70


편집된 염기
0%
10%
20%
30%
40%
50%
60%
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
비처리
C6
C5

도면71


편집된 염기
0%
5%
10%
15%
20%
25%
30%
35%
BE3 WT
R126A
R126E
R118A
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
비처리
C4
C6

도면72

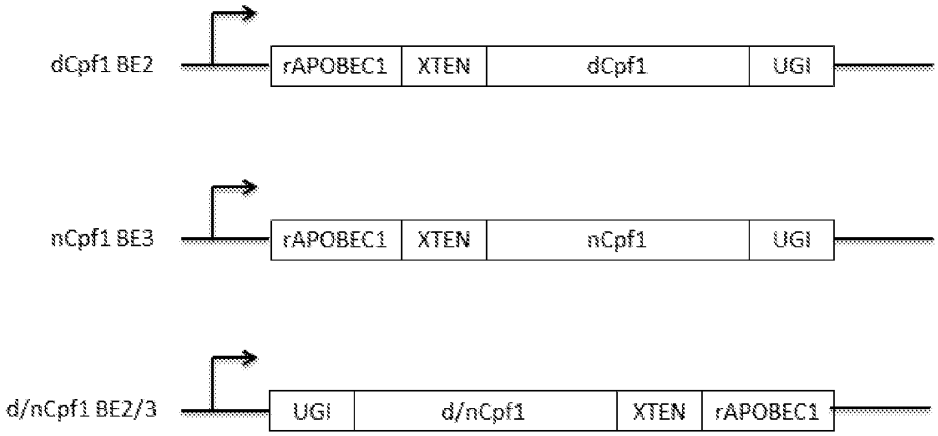
BE3

염기	Lys				Arg → Cys				Leu → Leu				Ala				Tyr				Gln			
	G	A	A	G	C	G	C	C	C	T	G	G	C	A	G	T	G	T	A	C	C
A	0.2	99.9	99.8	0.0	1.0	0.1	1.8	1.2	0.0	0.0	0.0	0.1	0.1	99.8	0.0	0.0	0.1	0.1	99.9	0.1	0.0
C	0.0	0.0	0.1	0.0	38.1	0.0	49.8	52.3	0.0	0.0	0.0	0.0	99.7	0.1	0.0	0.1	0.0	0.0	0.0	99.9	99.9
G	99.8	0.0	0.0	99.9	1.8	99.8	1.3	0.7	0.1	99.9	99.8	0.0	0.1	99.9	0.1	99.9	0.1	0.0	0.0	0.0	0.0
T	0.0	0.0	0.1	0.1	59.2	0.1	47.0	45.8	99.9	0.1	0.1	0.2	0.0	0.0	99.8	0.0	99.8	0.0	0.1	0.1	0.1

BE3 W90Y R132E

염기	Lys				Arg → Cys				Leu → Leu				Ala				Tyr				Gln			
	G	A	A	G	C	G	C	C	C	T	G	G	C	A	G	T	G	T	A	C	C
A	0.0	99.9	99.9	0.0	0.5	0.1	0.3	0.1	0.0	0.0	0.0	0.0	0.1	99.8	0.0	0.0	0.0	0.1	99.9	0.1	0.0
C	0.0	0.0	0.0	0.1	78.0	0.0	94.8	98.9	0.0	0.0	0.0	0.0	99.8	0.1	0.0	0.0	0.0	0.0	0.0	99.9	99.9
G	100.0	0.0	0.0	99.9	0.5	99.9	0.1	0.1	0.0	99.9	99.9	0.0	0.1	100.0	0.1	99.9	0.0	0.0	0.0	0.0	0.0
T	0.0	0.0	0.0	0.0	21.0	0.0	4.8	0.9	99.9	0.0	0.0	0.1	0.0	0.0	99.8	0.0	99.9	0.0	0.0	0.0	0.0

도면73


dCpf1 BE2
rAPOBEC1
XTEN
dCpf1
UGI
nCpf1 BE3
rAPOBEC1
XTEN
nCpf1
UGI
d/nCpf1 BE2/3
UGI
d/nCpf1
XTEN
rAPOBEC1

도면74

EMX1	C_3	C_6	C_{12}	indel%
비처리				
A	0±0	0±0	0±0	0±0
C	99.9±0	99.9±0	99.9±0	
G	0±0	0±0	0±0	
T	0±0	0±0	0±0	

FANCF	C_6	C_7	C_8	C_{11}	indel%
비처리					
A	0±0	0±0	0±0	0±0	0±0
C	99.9±0	99.9±0	99.9±0	99.9±0	
G	0±0	0±0	0±0	0±0	
T	0±0	0±0	0±0	0±0	

RNF2	C_3	C_6	C_{12}	indel%
비처리				
A	0±0	0±0	0±0	0±0
C	99.9±0	99.9±0	99.9±0	
G	0±0	0±0	0±0	
T	0±0	0±0	0±0	

BE3	C_3	C_6	C_{12}	indel%
A	1.3±0.1	0.7±0.2	0±0	2.6±0.3
C	51.7±1	55.2±0.1	99.9±0	
G	4±0.7	1.9±0.4	0±0	
T	42.8±0.2	41.9±0.4	0±0	

BE3	C_6	C_7	C_8	C_{11}	indel%
A	0.8±0.1	1.2±0.3	1.2±0.1	0.2±0.1	5.8±0.9
C	64.7±5.9	67.9±5.1	68.9±5.2	85.4±2.7	
G	0.6±0	0.7±0	0.7±0.1	0.1±0	
T	33.8±5.7	29.9±4.8	29±5	14.1±2.6	

BE3	C_3	C_6	C_{12}	indel%
A	0.4±0	1.5±0.2	0.1±0	2.3±0.3
C	70.9±2.7	45±0	95±0.2	
G	0±0	12.6±2.4	0.1±0	
T	28.5±2.7	40.7±2.7	4.6±0.2	

HF BE3	A	C	G	T	1±0.1
A	1.5±0.7	0.5±0.1	0±0		
C	44.8±8.5	49.9±6.6	99.9±0		
G	4.2±1.2	0.6±0.2	0±0		
T	49.3±6.5	48.7±6.3	0±0		

HF BE3	A	C	G	T	5.9±0.7
A	0.9±0.2	2±0.4	0.9±0.2	0.4±0	
C	52.5±8.4	57.7±7.6	61.9±6.8	84.7±2.4	
G	1.4±0.2	0.6±0	0.5±0.1	0.3±0	
T	45±8	39.5±7.2	36.5±6.5	14.4±2.2	

HF BE3	A	C	G	T	0.5±0.1
A	0±0	0.2±0	0±0		
C	76.8±1.8	64.5±3.3	95.4±0.1		
G	0±0	2.4±0.5	0±0		
T	23±1.8	32.7±2.6	4.5±0.1		

수는 3회의 독립적 반복으로부터의 μ±σ이다

도면75

HEK3		C_6	C_7	C_8	C_{11}	indel %
비처리	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	

HEK4		C_6	C_7	C_8	C_{11}	indel %
비처리	A	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C	99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G	0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T	0 ± 0	0 ± 0	0 ± 0	0 ± 0	

BE3		C_6	C_7	C_8	C_{11}	indel %
	A	0 ± 0	1.6 ± 0.2	1.5 ± 0.3	0.2 ± 0	2.7 ± 0.4
	C	98.7 ± 0.2	48.6 ± 5.4	40.8 ± 6.7	98.6 ± 0.1	
	G	0 ± 0	1.4 ± 0.2	10.5 ± 1.2	0.3 ± 0	
	T	1.1 ± 0.2	48.2 ± 4.9	47 ± 5.2	0.7 ± 0.1	

BE3		C_6	C_7	C_8	C_{11}	indel %
	A	0 ± 0	5.4 ± 1.3	0 ± 0	0 ± 0	3 ± 0.7
	C	98.7 ± 0	44.5 ± 8.2	98.9 ± 0	99.8 ± 0	
	G	0 ± 0	18.1 ± 2.3	0 ± 0	0 ± 0	
	T	1.1 ± 0.1	31.9 ± 4.8	0.9 ± 0.1	0.1 ± 0	

HF BE3		C_6	C_7	C_8	C_{11}	indel %
	A	0 ± 0	0.8 ± 0.2	1.1 ± 0.3	0.4 ± 0	4.1 ± 0.7
	C	97.8 ± 0.4	65 ± 3.8	40.4 ± 6.3	95.9 ± 0.5	
	G	0 ± 0	0.9 ± 0.2	6.1 ± 1.1	0.7 ± 0.2	
	T	2 ± 0.4	33 ± 3.4	52.2 ± 5.2	2.7 ± 0.2	

HF BE3		C_6	C_7	C_8	C_{11}	indel %
	A	0 ± 0	5.2 ± 0.8	0 ± 0	0 ± 0	1.3 ± 0
	C	99.6 ± 0	36.6 ± 9	98.5 ± 0.1	99.9 ± 0	
	G	0 ± 0	16 ± 2.8	0 ± 0	0 ± 0	
	T	0.2 ± 0	42.1 ± 5.3	1.3 ± 0.1	0 ± 0	

수는 3회의 독립적 반복으로부터의 μ±σ이다

도면76i

EMX1 온 타겟	G	A	G	T	C	C	G	A	G	A	G	G	A	A	A	A	G	A	A	G	G	G
비처리																						
BE3					0±0	0±0						0±0										
HF-BE3					42.8±5.3	41.9±5.5						0±0										
EMX1 오프 타겟1	G	A	G	T	T	A	G	A	G	C	A	G	A	A	A	A	G	A	A	A	G	G
비처리																						
BE3					49.3±5.5	48.7±5.6						0±0										
HF-BE3												0±0										
EMX1 오프 타겟2	G	A	G	T	C	T	A	A	G	C	A	G	A	A	A	A	G	A	A	G	A	G
비처리					0±0							0±0										
BE3					12.1±0.9							0±0										
HF-BE3					1.7±0.3							0±0										
EMX1 오프 타겟3	G	A	G	G	C	C	G	A	G	C	A	G	A	A	A	A	G	A	A	C	G	G
비처리					0±0	0±0						0±0										
BE3					6.8±2	8.1±2.5						0.2±0										
HF-BE3					0±0	0±0						0±0										
EMX1 오프 타겟4	G	A	G	T	C	C	T	A	G	C	A	G	A	G	A	A	G	A	A	G	A	G
비처리					0±0	0±0						0±0										
BE3					1.3±0.1	1.3±0.1						0±0										
HF-BE3					0.1±0	0.2±0						0±0										

도면76ii

EMX1 오프 타겟6	G	A	G	T	C	C	G	G	G	A	A	G	A	G	A	A	A	G	G
비처리					0±0	0±0													
BE3					0.4±0	0.4±0													
HF-BE3					0±0	0±0													

EMX1 오프 타겟7	G	A	G	C	C	G	G	A	G	C	A	G	A	A	G	A	G	G	G
비처리				0±0	0±0					0±0									
BE3				0±0	0±0					0±0									
HF-BE3				0±0	0±0					0±0									

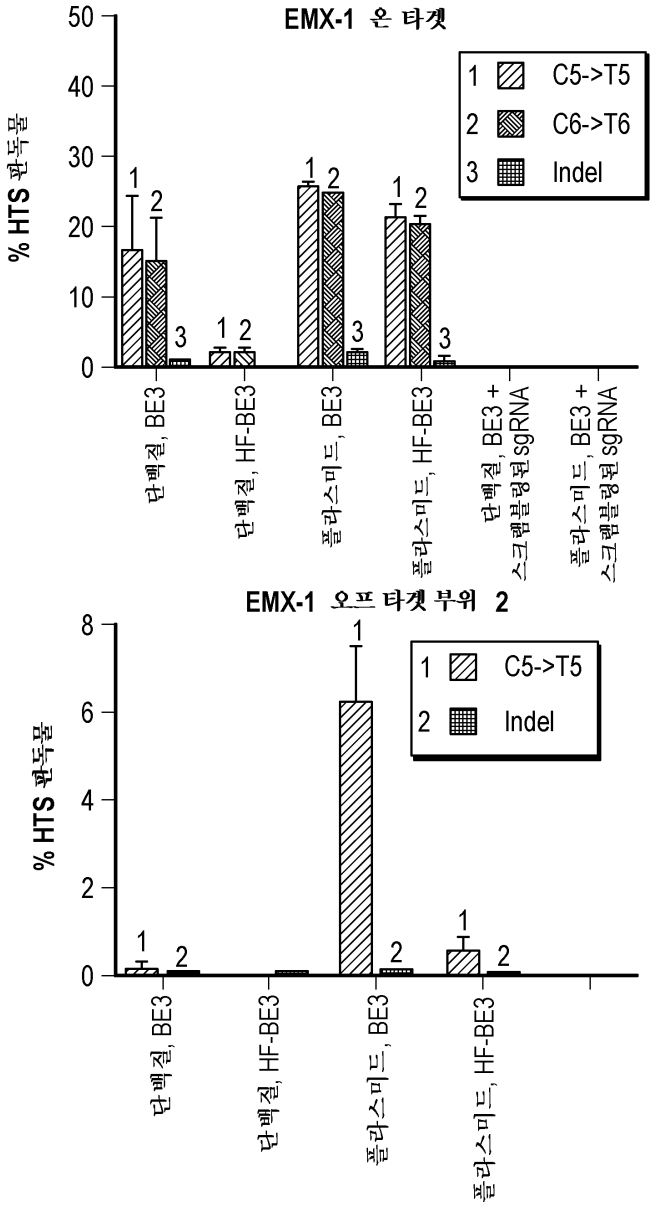
EMX1 오프 타겟8	A	A	G	T	C	C	G	A	G	G	A	G	A	G	A	A	A	G	G
비처리				0±0	0±0														
BE3				0±0	0±0														
HF-BE3				0±0	0±0														

EMX1 오프 타겟9	G	A	A	T	C	C	A	A	G	C	A	G	A	A	G	A	A	G	A
비처리				0±0	0±0					0±0									
BE3				0.1±0	0.1±0					0±0									
HF-BE3				0±0	0±0					0±0									

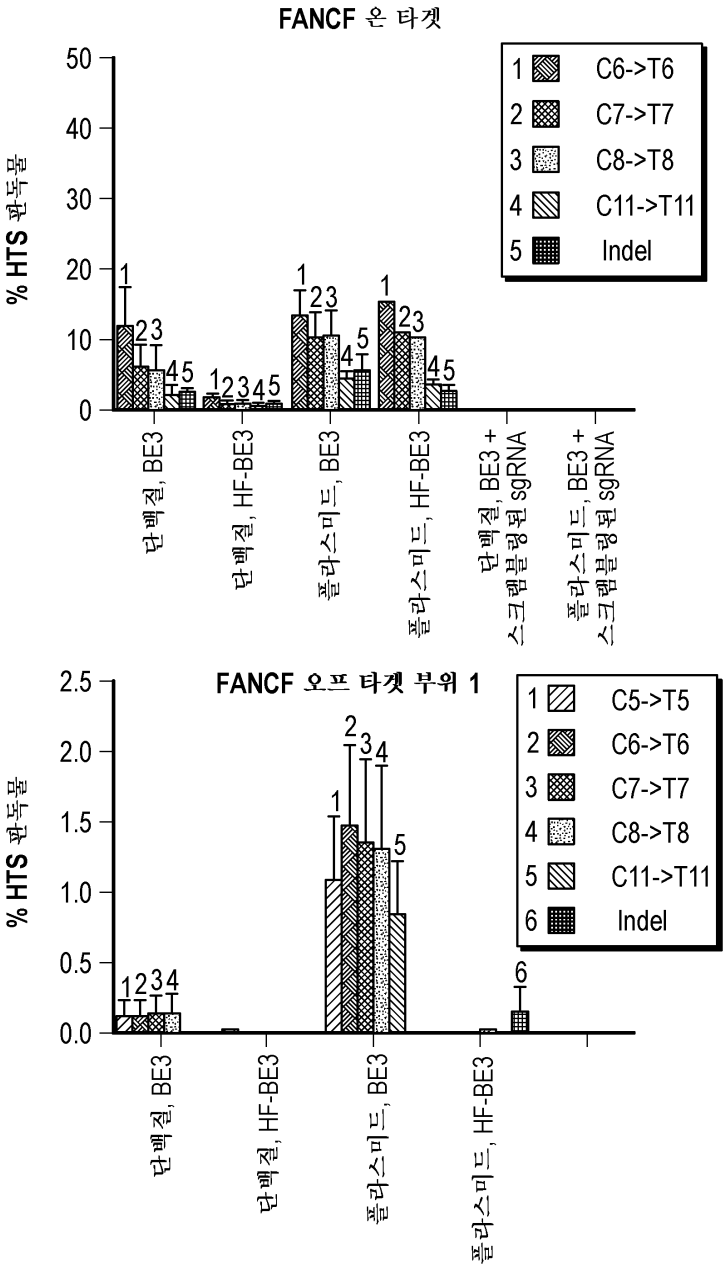
EMX1 오프 타겟 10	A	C	G	T	C	T	G	A	G	C	A	G	A	A	G	A	A	G	A	T	G	G
비처리	0±0			0±0						0±0												
BE3	0±0			1.5±0.1						0±0												
HF-BE3	0±0			1.1±0.2						0±0												

수는 3회의 독립적 반복으로부터의 μ ± σ이다

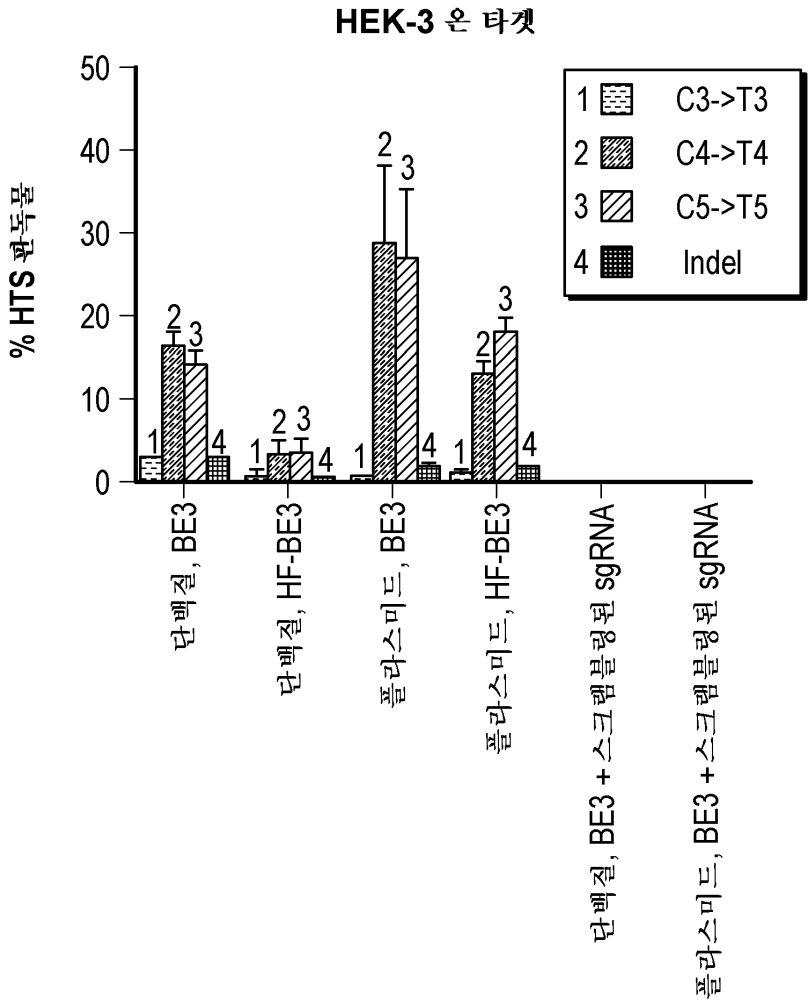
도면77


EMX-1 온 타겟
% HTS 편독물
50
40
30
20
10
0
1 C5->T5
2 C6->T6
3 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3
단백질, BE3 + 스크램블링된sgRNA
플라스미드, BE3 + 스크램블링된 sgRNA
EMX-1 오프 타겟 부위 2
% HTS 편독물
8
6
4
2
0
1 C5->T5
2 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3

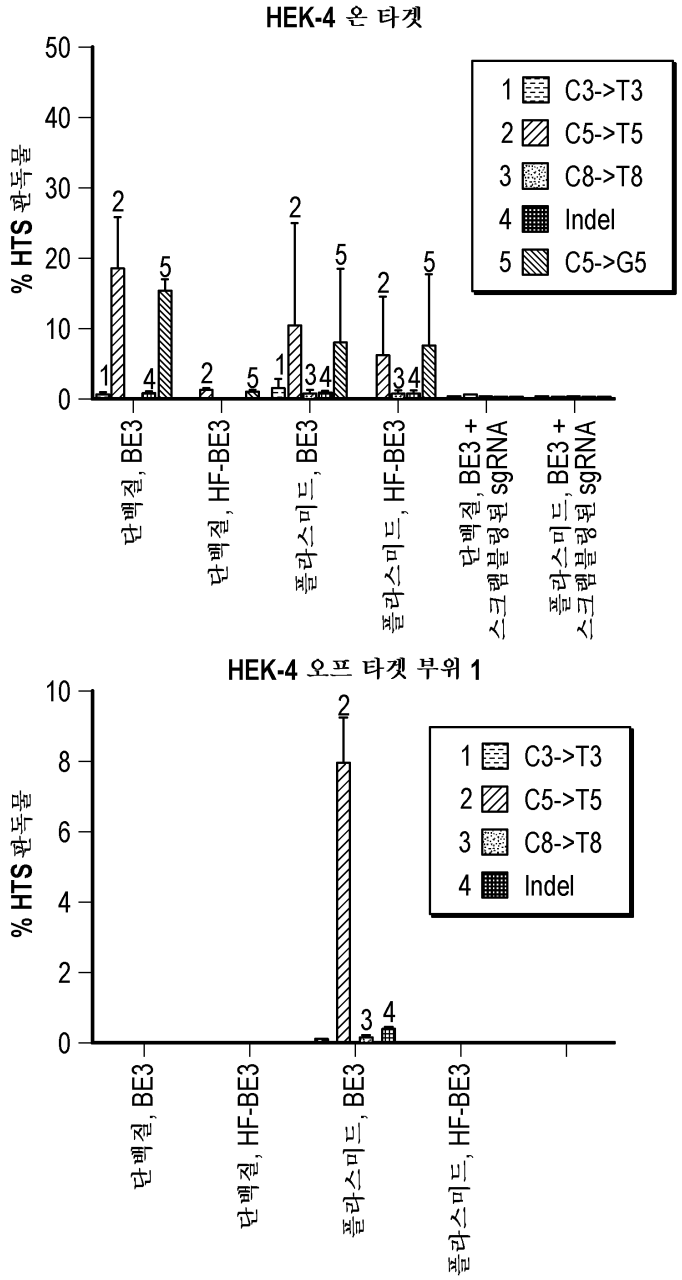
도면78


FANCF 온 타겟
% HTS 판독물
50
40
30
20
10
0
1 C6->T6
2 C7->T7
3 C8->T8
4 C11->T11
5 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3
단백질, BE3 + 스크램블링된 sgRNA
플라스미드, BE3 + 스크램블링된 sgRNA
FANCF 오프 타겟 부위 1
% HTS 판독물
2.5
2.0
1.5
1.0
0.5
0.0
1 C5->T5
2 C6->T6
3 C7->T7
4 C8->T8
5 C11->T11
6 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3

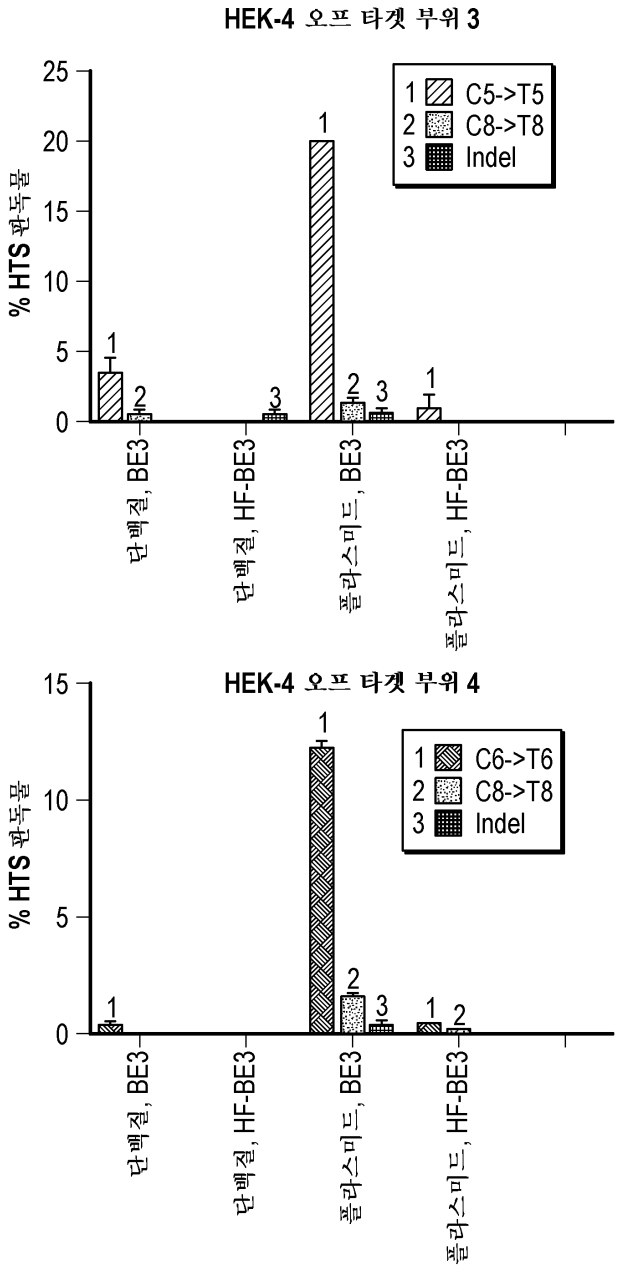
도면79


HEK-3 온 타겟
% HTS 판독물
50
40
30
20
10
0
1 C3->T3
2 C4->T4
3 C5->T5
4 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3
단백질, BE3 + 스크램블링된 sgRNA
플라스미드, BE3 + 스크램블링된 sgRNA

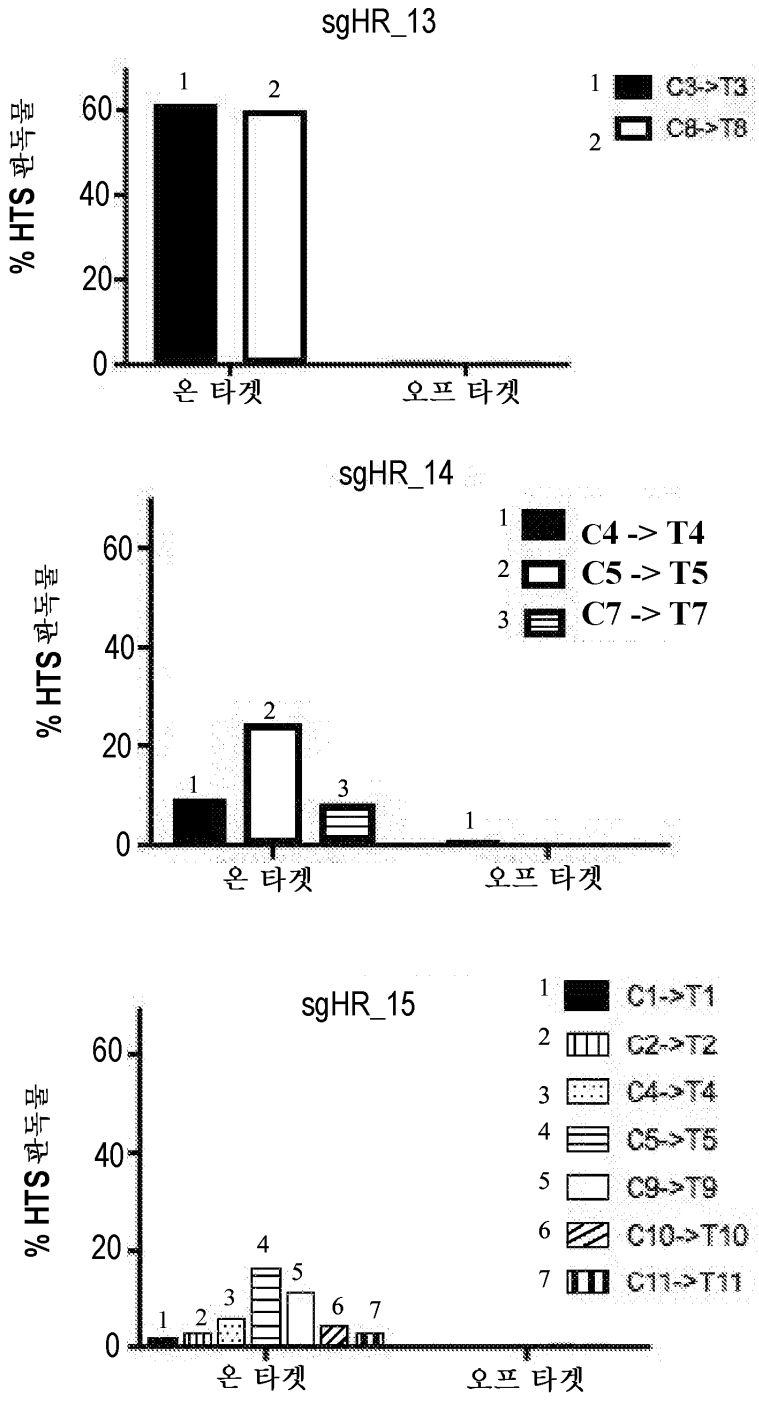
도면80i


HEK-4 온 타겟
% HTS 판독물
50
40
30
20
10
0
1 C3->T3
2 C5->T5
3 C8->T8
4 Indel
5 C5->G5
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3
단백질, BE3 + 스크램블링된 sgRNA
플라스미드, BE3 + 스크램블링된 sgRNA
HEK-4 오프 타겟 부위 1
% HTS 판독물
10
8
6
4
2
0
1 C3->T3
2 C5->T5
3 C8->T8
4 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3

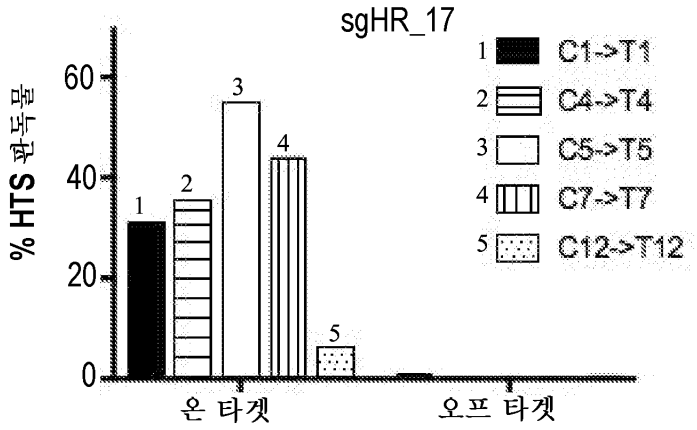
도면80ii

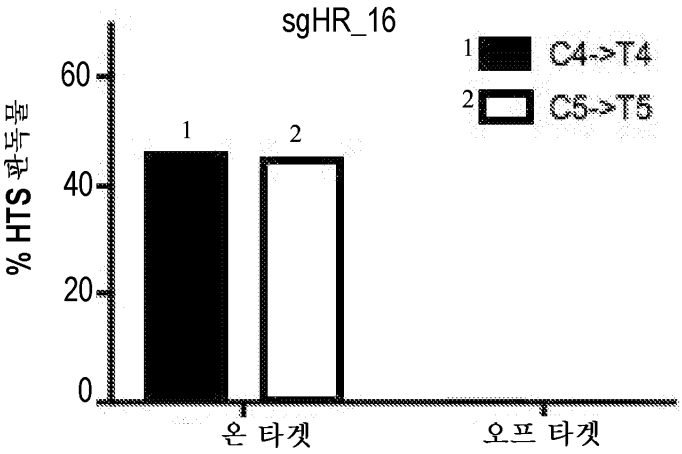

HEK-4 오프 타겟 부위 3
% HTS 판독물
25
20
15
10
5
0
1 C5->T5
2 C8->T8
3 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3
HEK-4 오프 타겟 부위 4
% HTS 판독물
15
10
5
0
1 C6->T6
2 C8->T8
3 Indel
단백질, BE3
단백질, HF-BE3
플라스미드, BE3
플라스미드, HF-BE3

도면81

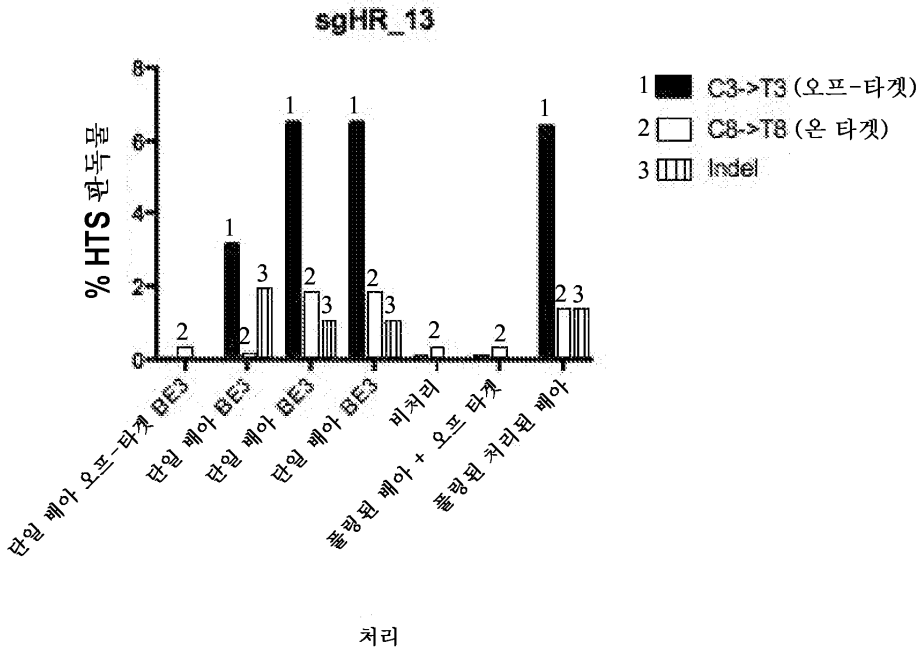

sgHR_13
% HTS 판독물
60
40
20
0
1
2
온 타겟
오프 타겟
1 C3->T3
2 C8->T8
sgHR_14
% HTS 판독물
60
40
20
0
1
2
3
1
온 타겟
오프 타겟
1 C4 -> T4
2 C5 -> T5
3 C7 -> T7
sgHR_15
% HTS 판독물
60
40
20
0
1
2
3
4
5
6
7
온 타겟
오프 타겟
1 C1->T1
2 C2->T2
3 C4->T4
4 C5->T5
5 C9->T9
6 C10->T10
7 C11->T11

도면82

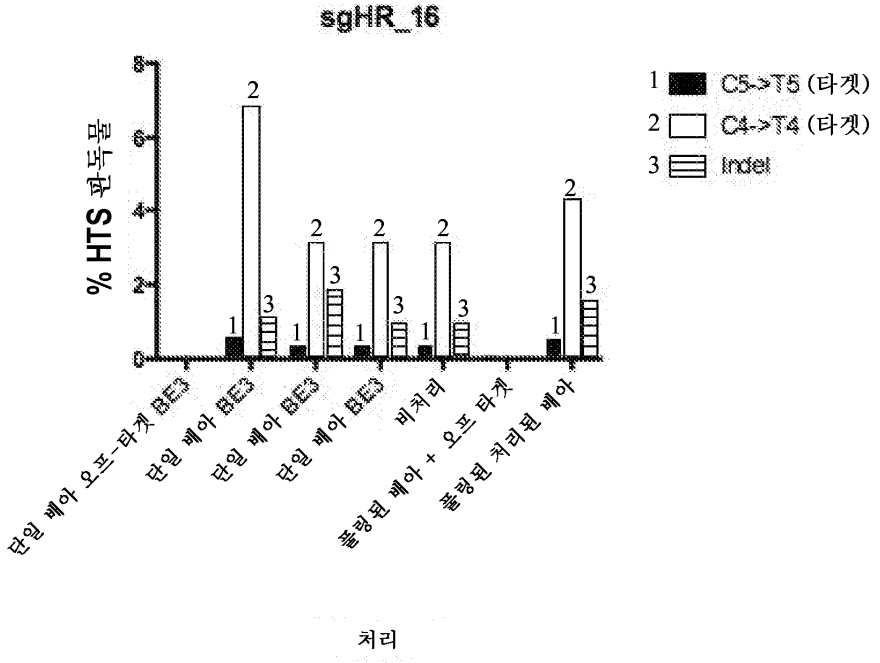

sgHR_17
% HTS 판독물
60
40
20
0
1
2
3
4
5
온 타겟
오프 타겟
1 C1->T1
2 C4->T4
3 C5->T5
4 C7->T7
5 C12->T12


sgHR_16
% HTS 판독물
60
40
20
0
1
2
온 타겟
오프 타겟
1 C4->T4
2 C5->T5

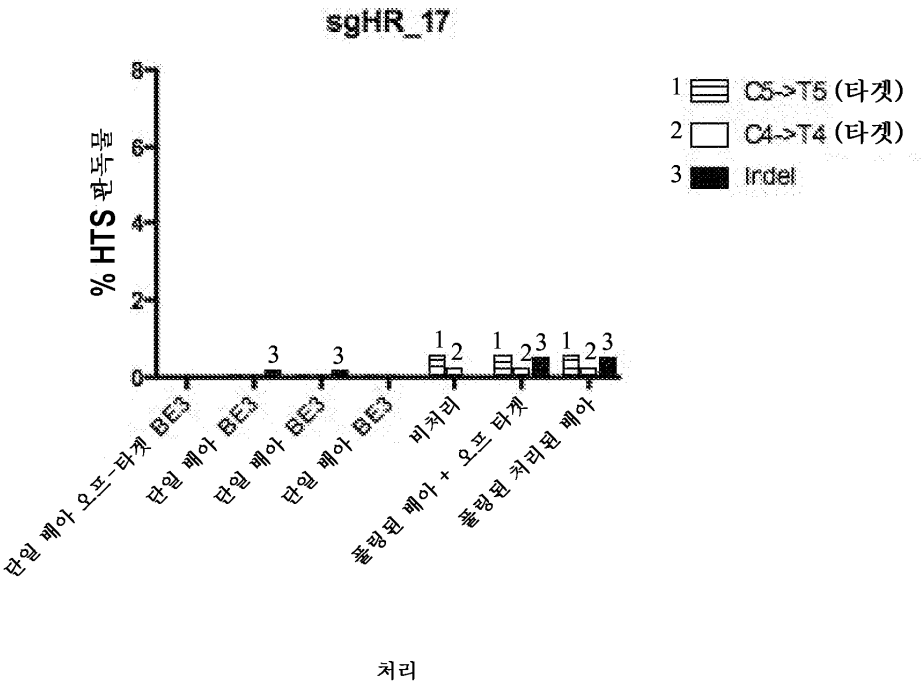
도면83


sgHR_13
% HTS 판독물
8
6
4
2
0
단일 베이 오프-타겟 BE3
단일 베이 BE3
단일 베이 BE3
단일 베이 BE3
비처리
폴링된 베이 + 오프 타겟
폴링된 처리된 베이
처리
1 C3->T3 (오프-타겟)
2 C8->T8 (온 타겟)
3 Indel

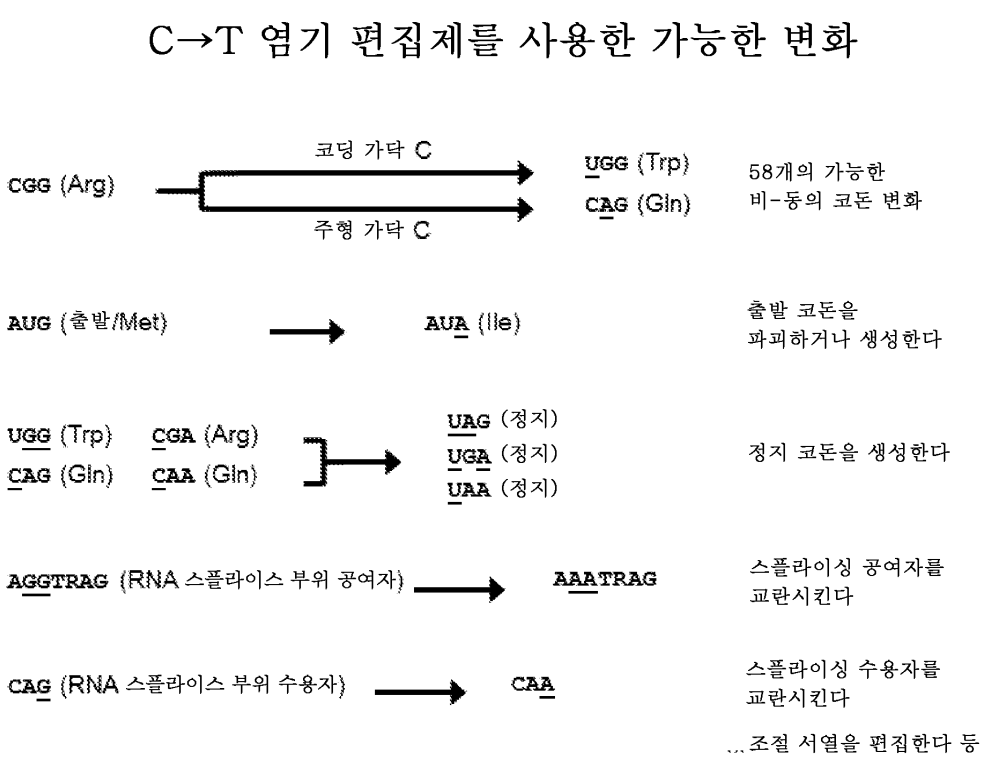
도면84


sgHR_16
% HTS 판독물
1 C5->T5 (타겟)
2 C4->T4 (타겟)
3 Indel
단일 베이 오프-타겟 BE3
단일 베이 BE3
단일 베이 BE3
단일 베이 BE3
비처리
풀링된 베이 + 오프 타겟
풀링된 처리된 베이
처리

도면85


sgHR_17
% HTS 판독물
1 C5->T5 (타겟)
2 C4->T4 (타겟)
3 Indel
단일 베이 오프-타겟 BE3
단일 베이 BE3
단일 베이 BE3
단일 베이 BE3
비처리
풀링된 베이 + 오프 타겟
풀링된 처리된 베이
처리

도면86


C→T 염기 편집제를 사용한 가능한 변화
CGG (Arg)
코딩 가닥 C
주형 가닥 C
UGG (Trp)
CAG (Gln)
58개의 가능한
비-동의 코돈 변화
AUG (출발/Met)
AUA (Ile)
출발 코돈을
파괴하거나 생성한다
UGG (Trp)
CGA (Arg)
CAG (Gln)
CAA (Gln)
UAG (정지)
UGA (정지)
UAA (정지)
정지 코돈을 생성한다
AGGTRAG (RNA 스플라이스 부위 공여자)
AAATRAG
스플라이싱 공여자를
교란시킨다
CAG (RNA 스플라이스 부위 수용자)
CAA
스플라이싱 수용자를
교란시킨다
... 조절 서열을 편집한다 등

도면87

*APOE4*에서 *APOE3r*로의 염기 편집

- 아포지단백질 E는 아미노산 112 및 158에서 상이한 4가지 이소형을 갖는다
- *APOE4*는 후기-발병 알츠하이머병에 대한 가장 크고 가장 통상적인 유전적 위험 인자이다

대립유전자	잔기 112	잔기 158	AD 위험 (공지되거나 추정됨)
APOE2	Cys TGC	Cys TGC	낮음
APOE3	Cys TGC	Arg CGC	중간
APOE3r	Arg CGC	Cys TGC	중간
APOE4	Arg CGC	Arg CGC	높음

BE3

도면88

비처리

APOE4 C158R	Lys			Arg 158	Leu		Ala		Val		Tyr		Gln		indel %
	G	A	G	C_5	C	T	G	C	G	T	T	A	G	G	**0.0**
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	
G	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	
T	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	

Cas9 + HDR

APOE4 C158R	Lys			Arg 158 → Cys		Leu		Ala		Val		Tyr		Gln		indel%
	G	A	G	C_5	G	C	T	G	C	A	G	T	T	A	G	G	**26.1**
A	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	0.0	99.7	0.0	99.9	99.9	0.0	0.0	100.0	0.5	0.0	0.0	0.0	0.0	
G	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	99.9	0.0	0.0	99.5	0.6	99.9	0.2	0.0	0.0	0.0	100.0	100.0
T	0.0	0.0	0.0	**0.3**	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.1	0.4	99.3	0.1	99.8	0.0	0.0	0.0	0.0	0.0

BE3-커리큘

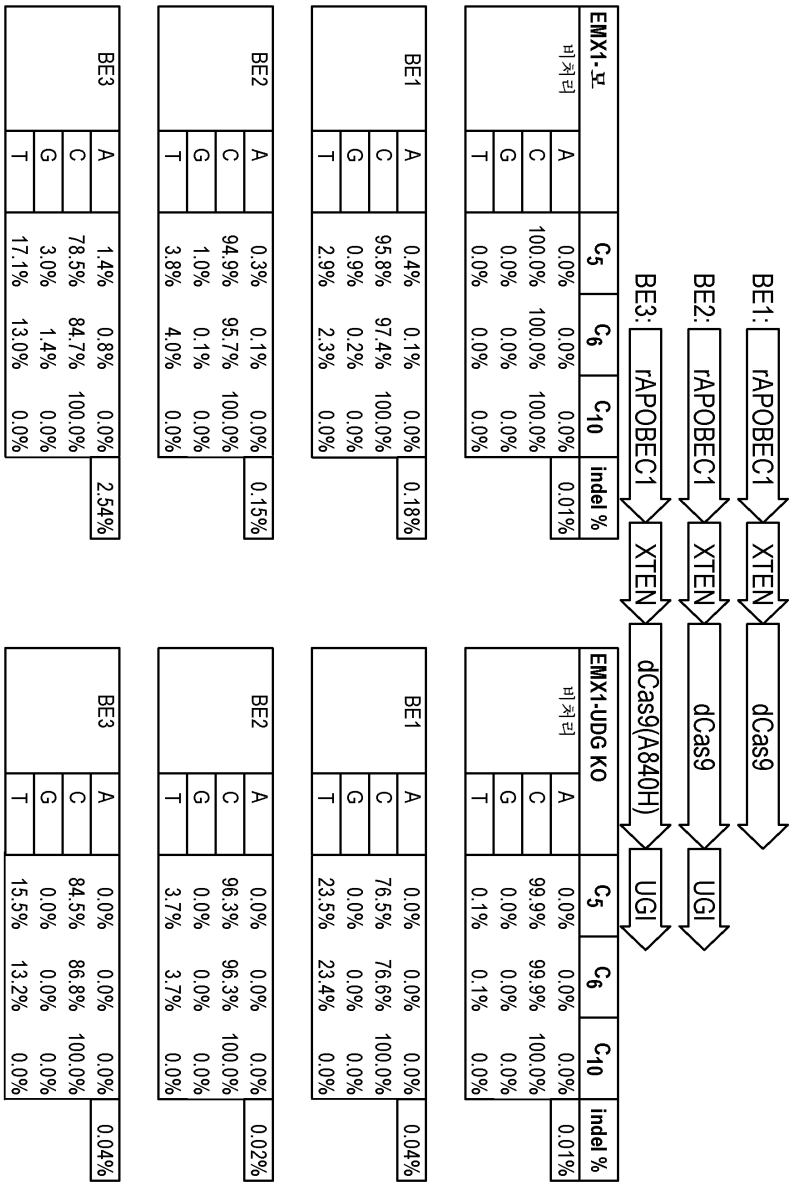
APOE4 C158R	Lys			Arg 158 → Cys		Leu → Leu		Ala		Val		Tyr		Gln		indel%
	G	A	A	G	C_5	G	C_7	C_8	T	G	G	C	A	G	T	T	A	C	C	A	G	G	**4.6**
A	0.1	100.0	100.0	0.0	0.5	0.0	1.3	0.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.1	
C	0.0	0.0	0.0	0.0	23.7	0.0	47.4	43.5	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	
G	99.9	0.0	0.0	100.0	0.9	99.9	1.1	0.7	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	99.9	
T	0.0	0.0	0.0	0.0	**74.9**	0.1	**50.2**	**55.0**	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	

도면89

BE3-처리된 *PRNP* R37X

	Gly			Ser → Ser		Arg 37 → Stop			Tyr		Pro			Gly			Glu		
	G	G	C	A	G	C_7	C_8	G	A	T	A	C	C	G	G	G	C	A	G	G	G
A	0.0	0.0	0.1	100.0	0.0	1.1	0.3	0.1	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
C	0.0	0.0	98.9	0.0	0.0	54.9	57.8	0.0	0.0	0.0	0.0	99.8	99.8	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
G	100.0	100.0	0.0	0.0	100.0	2.1	0.6	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	100.0
T	0.0	0.0	1.1	0.0	0.0	41.9	41.2	0.0	0.0	100.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

도면90i



EMX1-포 비처리	C_5	C_6	C_{10}	indel %
A	0.0%	0.0%	0.0%	0.01%
C	100.0%	100.0%	100.0%	
G	0.0%	0.0%	0.0%	
T	0.0%	0.0%	0.0%	

EMX1-UDG KO 비처리	C_5	C_6	C_{10}	indel %
A	0.0%	0.0%	0.0%	0.01%
C	99.9%	99.9%	100.0%	
G	0.0%	0.0%	0.0%	
T	0.1%	0.1%	0.0%	

BE1	C_5	C_6	C_{10}	indel %
A	0.4%	0.1%	0.0%	0.18%
C	95.8%	97.4%	100.0%	
G	0.9%	0.2%	0.0%	
T	2.9%	2.3%	0.0%	

BE1	C_5	C_6	C_{10}	indel %
A	0.0%	0.0%	0.0%	0.04%
C	76.5%	76.6%	100.0%	
G	0.0%	0.0%	0.0%	
T	23.5%	23.4%	0.0%	

BE2	C_5	C_6	C_{10}	indel %
A	0.3%	0.1%	0.0%	0.15%
C	94.9%	95.7%	100.0%	
G	1.0%	0.1%	0.0%	
T	3.8%	4.0%	0.0%	

BE2	C_5	C_6	C_{10}	indel %
A	0.0%	0.0%	0.0%	0.02%
C	96.3%	96.3%	100.0%	
G	0.0%	0.0%	0.0%	
T	3.7%	3.7%	0.0%	

BE3	C_5	C_6	C_{10}	indel %
A	1.4%	0.8%	0.0%	2.54%
C	78.5%	84.7%	100.0%	
G	3.0%	1.4%	0.0%	
T	17.1%	13.0%	0.0%	

BE3	C_5	C_6	C_{10}	indel %
A	0.0%	0.0%	0.0%	0.04%
C	84.5%	86.8%	100.0%	
G	0.0%	0.0%	0.0%	
T	15.5%	13.2%	0.0%	

도면90ii

FANCF-모		C_6	C_7	C_8	C_{11}	indel %
비처리	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	99.9%	99.9%	99.9%	99.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.1%	0.1%	0.0%	
BE1	A	0.5%	0.5%	0.3%	0.0%	0.13%
	C	94.2%	97.8%	97.9%	99.8%	
	G	0.7%	0.0%	0.0%	0.0%	
	T	4.7%	1.6%	1.8%	0.2%	
BE2	A	0.3%	0.4%	0.2%	0.0%	0.25%
	C	95.3%	97.3%	97.6%	99.8%	
	G	0.4%	0.1%	0.0%	0.0%	
	T	4.0%	2.3%	2.1%	0.1%	
BE3	A	2.4%	3.2%	2.2%	0.9%	18.88%
	C	60.3%	72.6%	73.8%	86.6%	
	G	1.2%	0.6%	0.4%	0.3%	
	T	36.2%	23.6%	23.5%	12.3%	

FANCF-UDG KO		C_6	C_7	C_8	C_{11}	indel %
비처리	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	99.9%	99.9%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.1%	0.0%	0.0%	0.0%	
BE1	A	0.0%	0.1%	0.0%	0.0%	0.09%
	C	60.9%	61.1%	61.3%	94.7%	
	G	0.1%	0.1%	0.1%	0.0%	
	T	39.0%	38.8%	38.6%	5.2%	
BE2	A	0.0%	0.0%	0.0%	0.0%	0.02%
	C	88.6%	88.6%	88.7%	98.9%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	11.4%	11.4%	11.3%	1.1%	
BE3	A	0.0%	0.1%	0.1%	0.0%	0.35%
	C	47.9%	48.5%	48.8%	55.8%	
	G	0.0%	0.1%	0.3%	0.2%	
	T	52.1%	51.3%	50.8%	44.0%	

도면90iii

HEK3-모 비처리		C_3	C_4	C_5	C_9	indel %
	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	99.9%	99.9%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	

BE1						
	A	0.0%	0.4%	0.3%	0.1%	0.07%
	C	99.9%	96.3%	94.4%	99.8%	
	G	0.0%	0.1%	1.9%	0.1%	
	T	0.1%	3.2%	3.4%	0.1%	

BE2						
	A	0.0%	0.2%	0.3%	0.1%	0.05%
	C	100.0%	97.7%	96.1%	99.8%	
	G	0.0%	0.0%	1.2%	0.1%	
	T	0.0%	2.1%	2.4%	0.1%	

BE3						
	A	0.1%	2.9%	2.4%	0.3%	3.27%
	C	99.4%	62.6%	55.9%	98.8%	
	G	0.0%	1.5%	10.4%	0.3%	
	T	0.6%	33.0%	31.3%	0.6%	

HEK3-UDG KO 비처리		C_3	C_4	C_5	C_9	indel %
	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	100.0%	100.0%	100.0%	100.0%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	0.0%	0.0%	0.0%	0.0%	

BE1						
	A	0.0%	0.0%	0.0%	0.0%	0.03%
	C	96.0%	50.3%	41.1%	96.3%	
	G	0.0%	0.0%	0.1%	0.1%	
	T	4.0%	49.6%	58.8%	3.6%	

BE2						
	A	0.0%	0.0%	0.0%	0.0%	0.00%
	C	99.0%	80.2%	73.2%	98.8%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.0%	19.7%	26.8%	1.1%	

BE3						
	A	0.0%	0.0%	0.0%	0.0%	0.09%
	C	98.5%	55.6%	40.3%	99.4%	
	G	0.0%	0.0%	0.0%	0.0%	
	T	1.4%	44.3%	59.7%	0.6%	

도면90iv

HEK4-포	C_3	C_5	C_8	C_{11}	indel %
비처리					
A	0.0%	0.0%	0.0%	0.0%	0.00%
C	100.0%	100.0%	99.9%	100.0%	
G	0.0%	0.0%	0.0%	0.0%	
T	0.0%	0.0%	0.1%	0.0%	

BE1	C_3	C_5	C_8	C_{11}	indel %
A	0.0%	1.4%	0.0%	0.0%	0.08%
C	100.0%	95.5%	99.9%	100.0%	
G	0.0%	2.2%	0.0%	0.0%	
T	0.0%	0.9%	0.1%	0.0%	

BE2	C_3	C_5	C_8	C_{11}	indel %
A	0.0%	0.4%	0.0%	0.0%	0.07%
C	100.0%	97.7%	99.9%	100.0%	
G	0.0%	1.1%	0.0%	0.0%	
T	0.0%	0.7%	0.1%	0.0%	

BE3	C_3	C_5	C_8	C_{11}	indel %
A	0.0%	7.1%	0.0%	0.0%	6.05%
C	99.9%	58.5%	99.7%	100.0%	
G	0.0%	17.4%	0.0%	0.0%	
T	0.1%	17.0%	0.2%	0.0%	

HEK4-UDG KO	C_3	C_5	C_8	C_{11}	indel %
비처리					
A	0.0%	0.0%	0.0%	0.0%	0.00%
C	100.0%	100.0%	99.9%	100.0%	
G	0.0%	0.0%	0.0%	0.0%	
T	0.0%	0.0%	0.1%	0.0%	

BE1	C_3	C_5	C_8	C_{11}	indel %
A	0.0%	0.0%	0.0%	0.0%	0.00%
C	99.9%	67.5%	99.0%	100.0%	
G	0.0%	0.0%	0.1%	0.0%	
T	0.1%	32.5%	1.0%	0.0%	

BE2	C_3	C_5	C_8	C_{11}	indel %
A	0.0%	0.0%	0.0%	0.0%	0.00%
C	100.0%	91.9%	99.7%	100.0%	
G	0.0%	0.0%	0.0%	0.0%	
T	0.0%	8.1%	0.2%	0.0%	

BE3	C_3	C_5	C_8	C_{11}	indel %
A	0.0%	0.0%	0.0%	0.0%	0.03%
C	99.9%	65.9%	99.7%	100.0%	
G	0.0%	0.3%	0.0%	0.0%	
T	0.1%	33.8%	0.3%	0.0%	

도면91i

EMX1	C_3	C_6	C_{12}	indel %
비처리	A 0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C 99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G 0 ± 0	0 ± 0	0 ± 0	
	T 0 ± 0	0 ± 0	0 ± 0	

FANCF	C_6	C_7	C_8	C_{11}	indel %
비처리	A 0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	C 99.9 ± 0	99.9 ± 0	99.9 ± 0	99.9 ± 0	
	G 0 ± 0	0 ± 0	0 ± 0	0 ± 0	
	T 0 ± 0	0 ± 0	0 ± 0	0 ± 0	

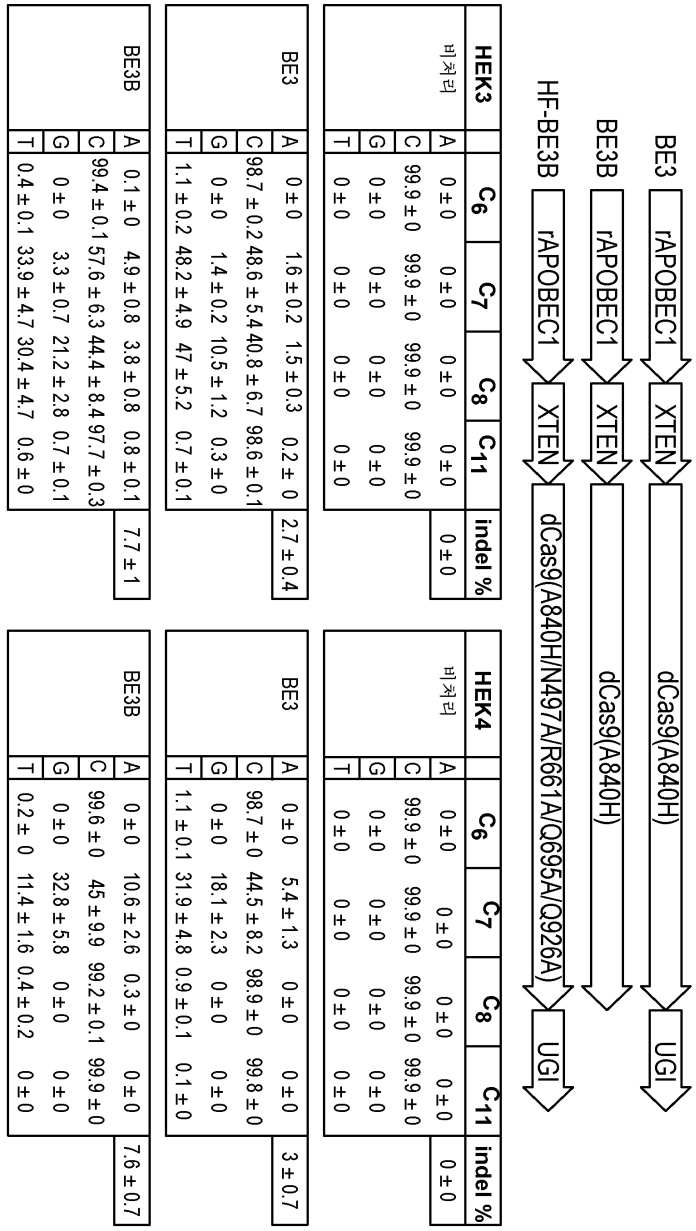
BE3					
	A	1.3 ± 0.1	0.7 ± 0.2	0 ± 0	2.6 ± 0.3
	C	51.7 ± 1	55.2 ± 0.1	99.9 ± 0	
	G	4 ± 0.7	1.9 ± 0.4	0 ± 0	
	T	42.8 ± 0.2	41.9 ± 0.4	0 ± 0	

BE3						
	A	0.8 ± 0.1	1.2 ± 0.3	1.2 ± 0.1	0.2 ± 0.1	5.8 ± 0.9
	C	64.7 ± 5.9	67.9 ± 5.1	68.9 ± 5.2	85.4 ± 2.7	
	G	0.6 ± 0	0.7 ± 0	0.7 ± 0.1	0.1 ± 0	
	T	33.8 ± 5.7	29.9 ± 4.8	29 ± 5	14.1 ± 2.6	

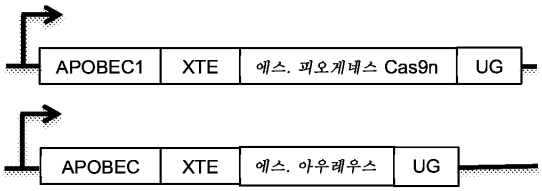
BE3B					
	A	4.9 ± 1.4	3.9 ± 1.1	0 ± 0	11.6 ± 2.6
	C	52.6 ± 12.1	61.9 ± 10.5	99.9 ± 0	
	G	12.3 ± 3.3	11.3 ± 3.1	0 ± 0	
	T	30.1 ± 7.3	22.7 ± 6.1	0 ± 0	

BE3B					
	A	2.9 ± 0.9	4.1 ± 1.4	2.8 ± 0.9	1.1 ± 0	20.5 ± 3.8
	C	67.3 ± 8.8	79.3 ± 5.8	79.5 ± 5.2	92 ± 1.9	
	G	2 ± 0.5	0.7 ± 0.3	0.9 ± 0.2	0.5 ± 0.1	
	T	27.7 ± 7.3	15.6 ± 4	16.5 ± 4	6.1 ± 1.3	

도면91ii



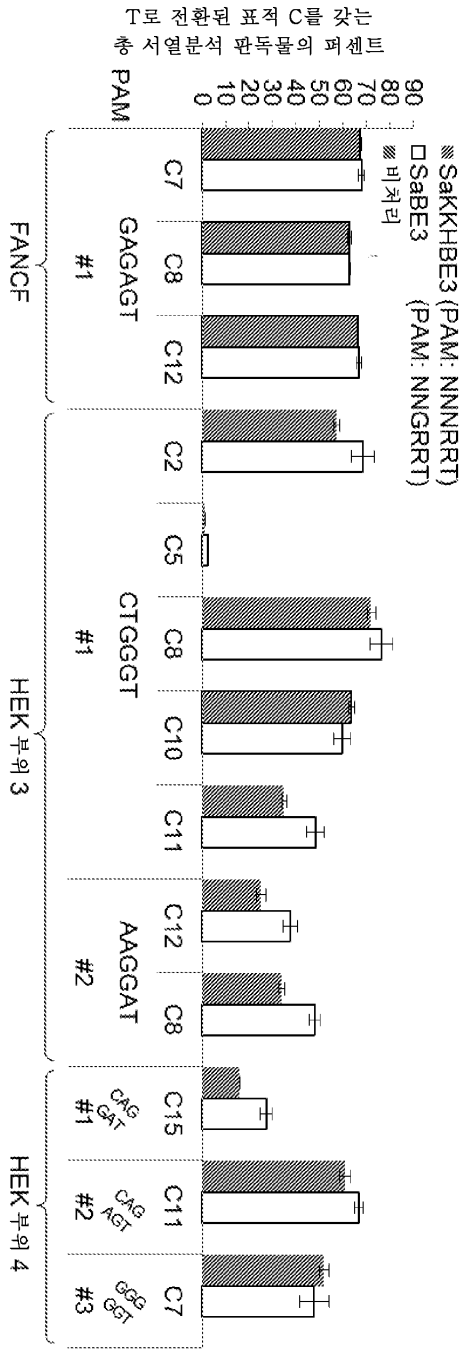
도면92a



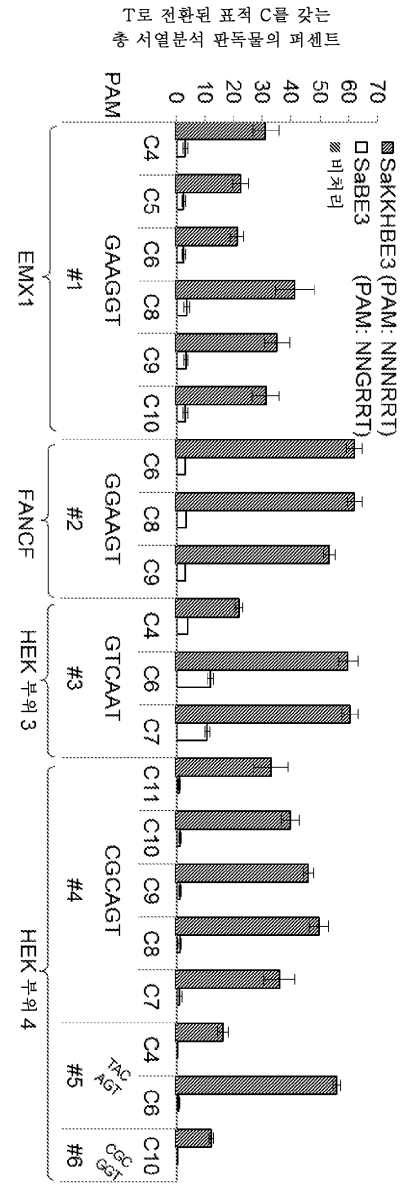
도면92b

종	PAM	염기 편집제	참조
에스. 피오게네스	NGG	BE3	야생형
	NGA	VQR, EQR BE3	Ref #7
	NGCG	VRER BE3	Ref #7
에스. 아우레우스	NNGRRT	SaBE3	야생형
	NNNRRT	SaKKHBE3	Ref #8

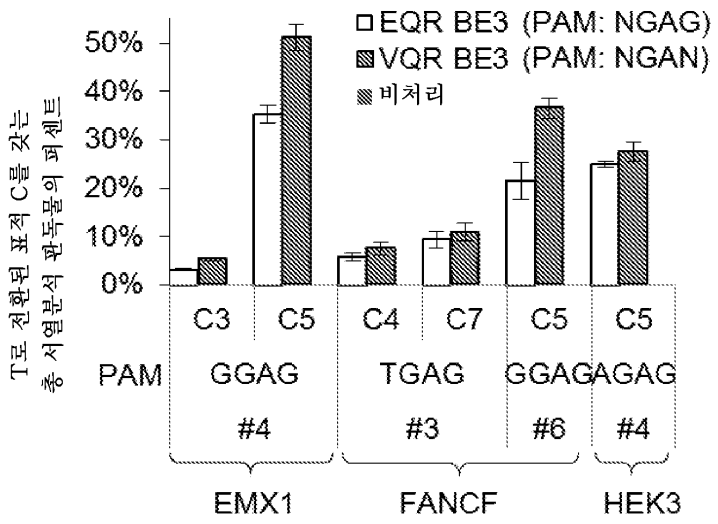
도면92c



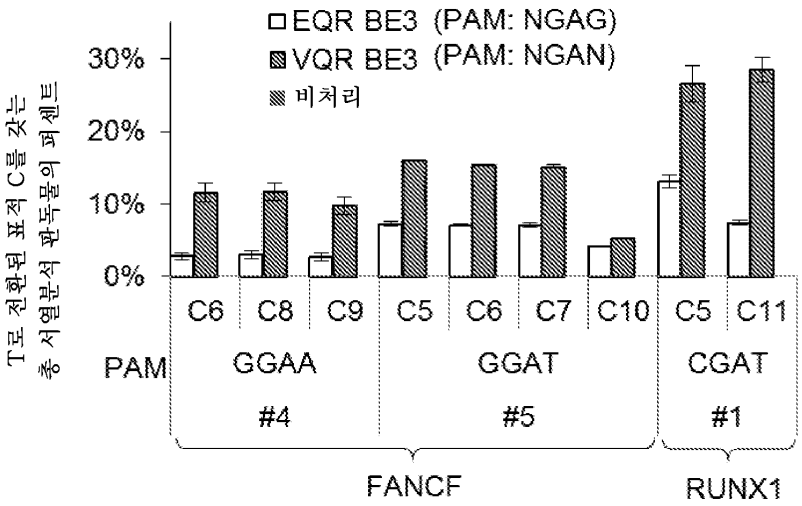
도면92d


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0
10
20
30
40
50
60
70
SaKKHBE3 (PAM: NNNRRT)
SaBE3 (PAM: NNGRRT)
비처리
PAM
EMX1
#1
GAAGGT
C4
C5
C6
C8
C9
C10
FANCF
#2
GGAAGT
C6
C8
C9
HEK 부위 3
#3
GTCAAT
C4
C6
C7
HEK 부위 4
#4
CGCAGT
C11
C10
C9
C8
C7
#5
TAC AGT
C4
C6
#6
CGC GGT
C10

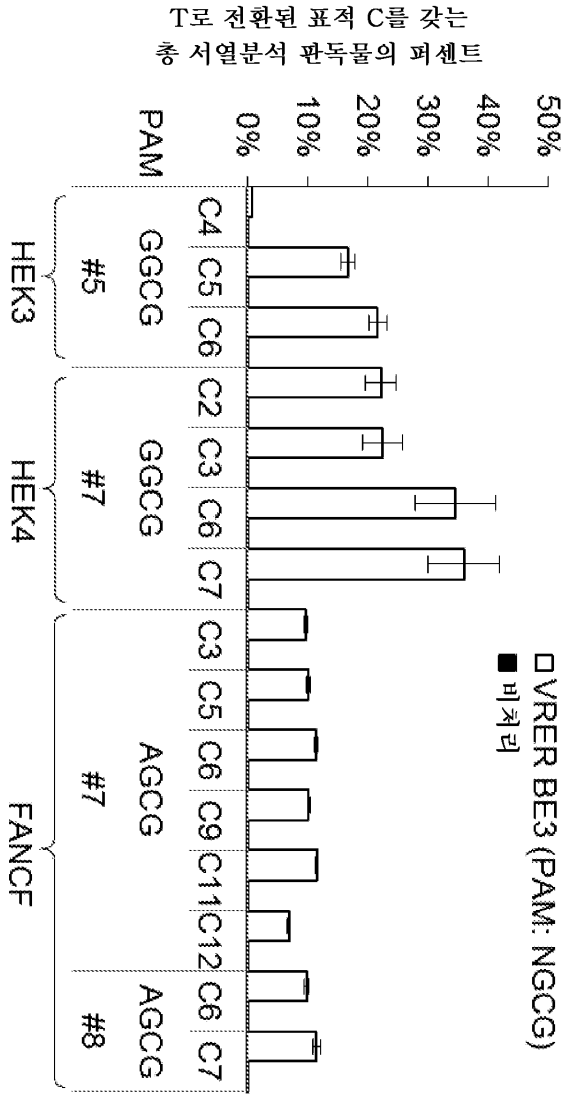
도면92e


EQR BE3 (PAM: NGAG)
VQR BE3 (PAM: NGAN)
비처리
T로 전환된 표적 C를 갖는 총 서열분석 판독물의 퍼센트
50%
40%
30%
20%
10%
0%
C3
C5
C4
C7
C5
C5
PAM
GGAG
TGAG
GGAG
AGAG
#4
#3
#6
#4
EMX1
FANCF
HEK3

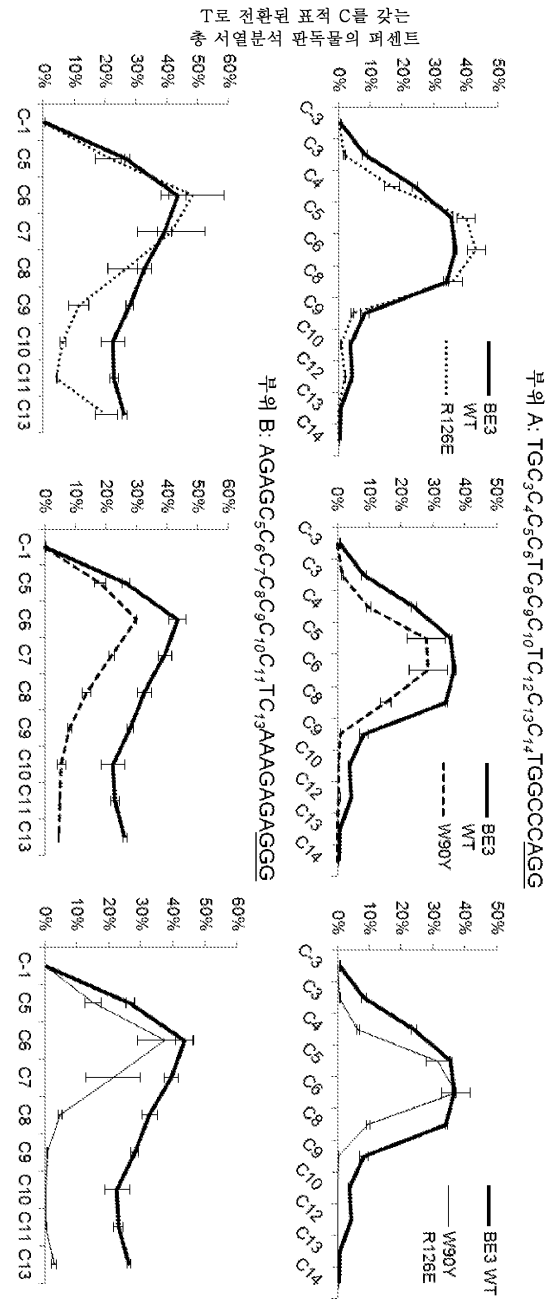
도면92f


EQR BE3 (PAM: NGAG)
VQR BE3 (PAM: NGAN)
비처리
T로 전환된 표적 C를 갖는 총 서열분석 판독물의 퍼센트
30%
20%
10%
0%
C6
C8
C9
C5
C6
C7
C10
C5
C11
PAM
GGAA
GGAT
CGAT
#4
#5
#1
FANCF
RUNX1

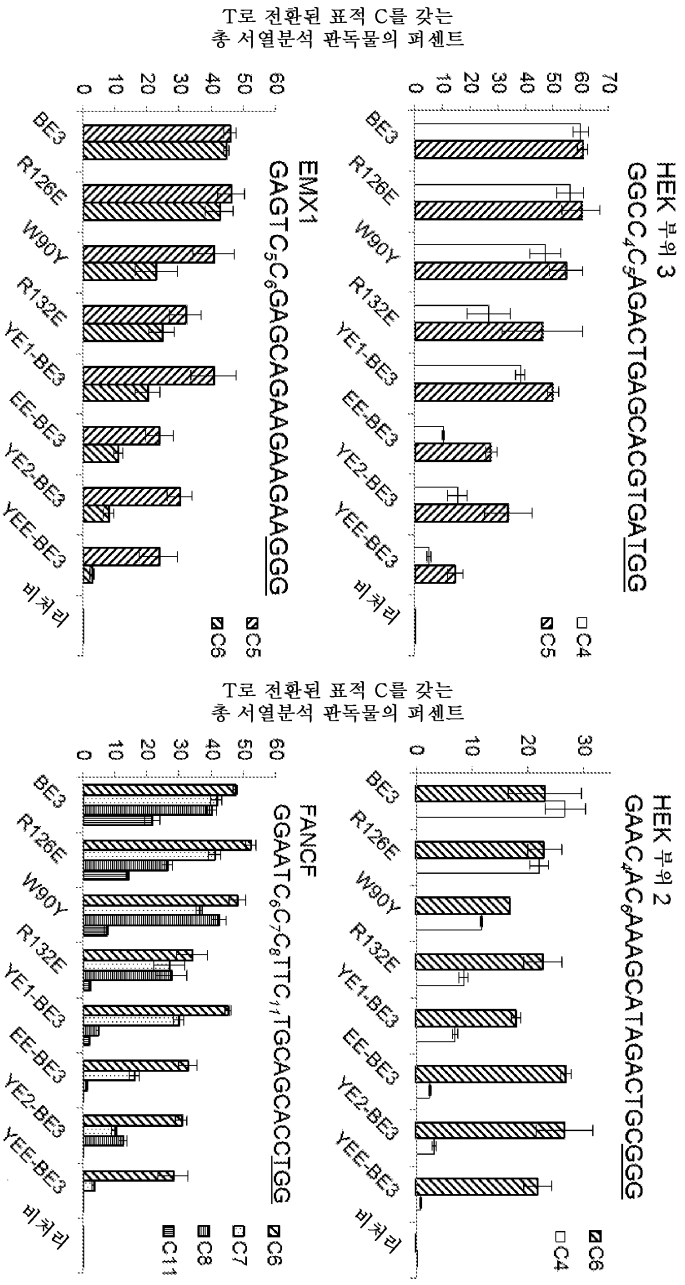
도면92g


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
50%
40%
30%
20%
10%
0%
□VRER BE3 (PAM: NGCG)
■ 비처리
HEK3
HEK4
FANCF
#5
#7
#7
#8
PAM
GGCG
GGCG
AGCG
AGCG
C4
C5
C6
C2
C3
C6
C7
C3
C5
C6
C9
C11
C12
C6
C7

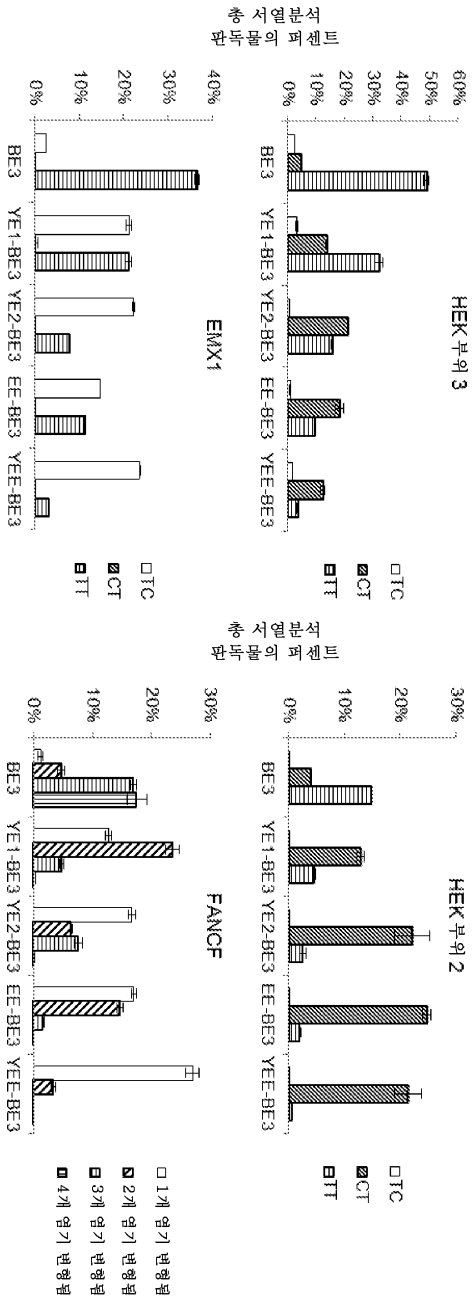
도면93a


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
부위 A: TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
부위 B: AGAGC5C6C7C8C9C10C11TC13AAAGAGAGGG
BE3
WT
R126E
W90Y
BE3 WT
W90Y
R126E
0%
10%
20%
30%
40%
50%
60%
C-1 C5 C6 C7 C8 C9 C10 C11 C13
C3 C4 C5 C6 C8 C9 C10 C12 C13 C14

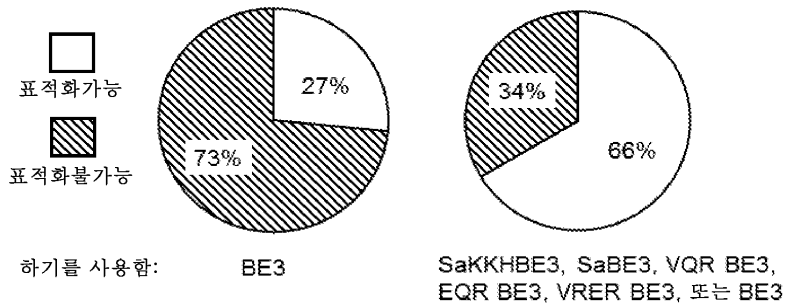
도면93b


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
HEK 부위 3
GGCC4C5AGACTGAGCACGTGATGG
C4
C5
EMX1
GAGTC5C6GAGCAGAAGAAGAAGGG
C5
C6
HEK 부위 2
GAAC4AC6AAAGCATAGACTGCGGG
C4
C6
FANCF
GGAATC6C7C8TTC11TGCAGCACCTGG
C7
C8
C11
C6
BE3
R126E
W90Y
R132E
YE1-BE3
EE-BE3
YE2-BE3
YEE-BE3
비처리

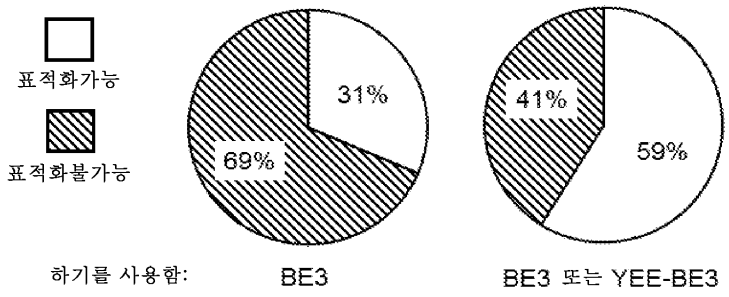
도면93c


총 서열분석
판독물의 퍼센트
HEK 부위 3
EMX1
HEK 부위 2
FANCF
BE3
YE1-BE3
YE2-BE3
EE-BE3
YEE-BE3
TT
CT
TC
1개 염기 변형됨
2개 염기 변형됨
3개 염기 변형됨
4개 염기 변형됨

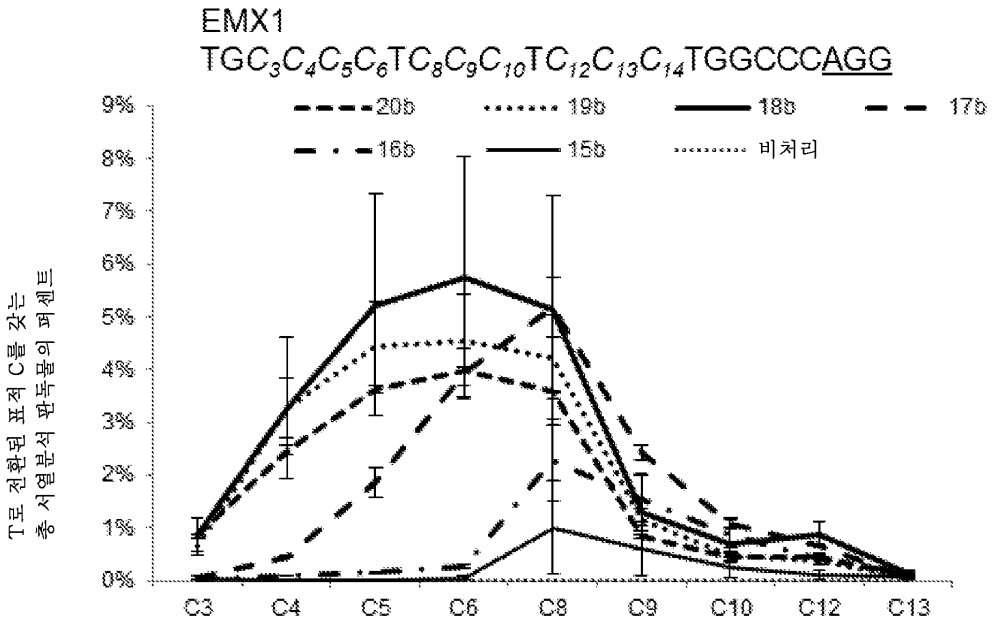
도면94a


표적화가능
표적화불가능
27%
73%
34%
66%
하기를 사용함:
BE3
SaKKHBE3, SaBE3, VQR BE3,
EQR BE3, VRER BE3, 또는 BE3

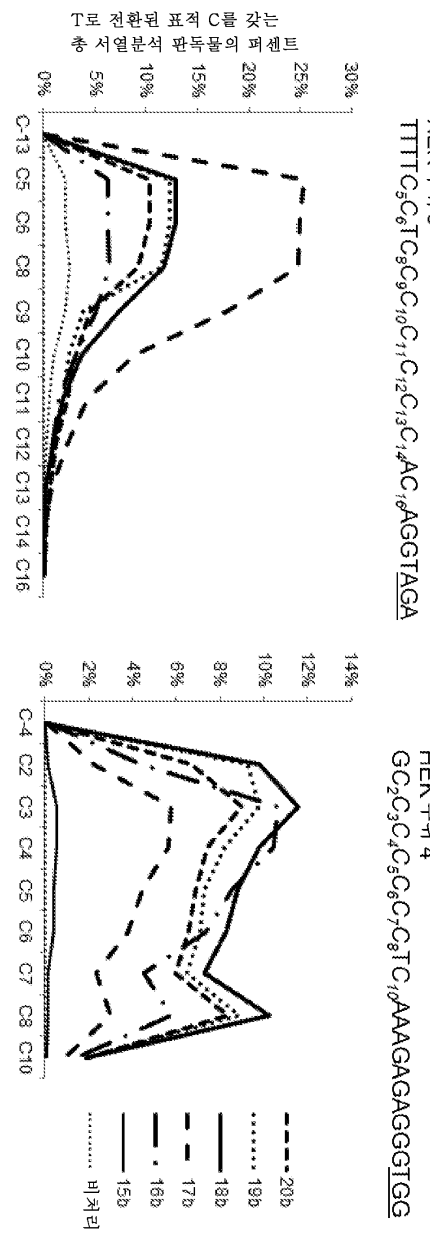
도면94b


표적화가능
표적화불가능
31%
69%
41%
59%
하기를 사용함:
BE3
BE3 또는 YEE-BE3

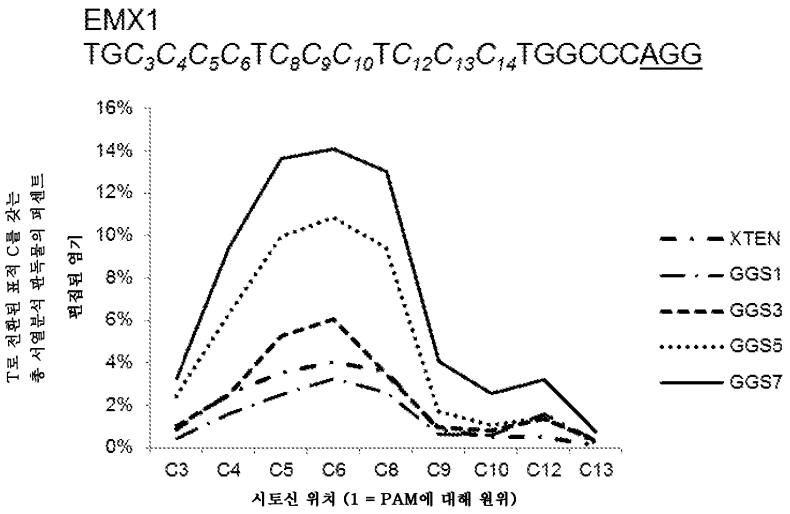
도면95a


EMX1
TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
20b
19b
18b
17b
16b
15b
비처리
T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
C3
C4
C5
C6
C8
C9
C10
C12
C13

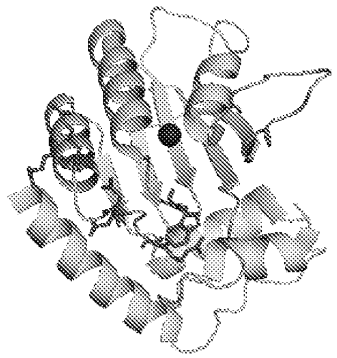
도면95b


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
HEK 부위 3
TTTTC5C6TC8C9C10C11C12C13C14AC16AGGTAGA
30%
25%
20%
15%
10%
5%
0%
C-13 C5 C6 C8 C9 C10 C11 C12 C13 C14 C16
HEK 부위 4
GC2C3C4C5C6C7C8TC10AAAGAGAGGGTGG
14%
12%
10%
8%
6%
4%
2%
0%
C-4 C2 C3 C4 C5 C6 C7 C8 C10
20b
19b
18b
17b
16b
15b
비처리

도면96



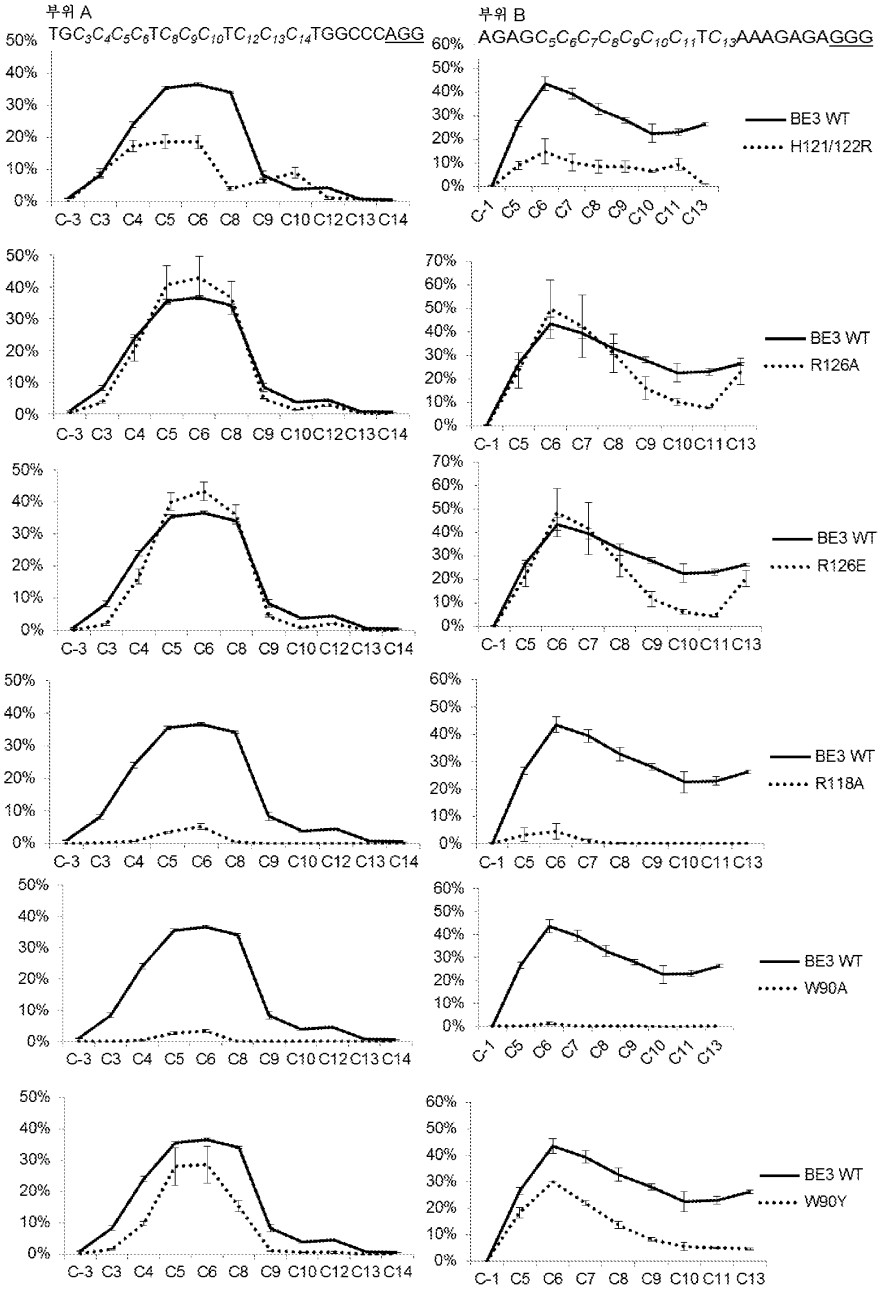
도면97a



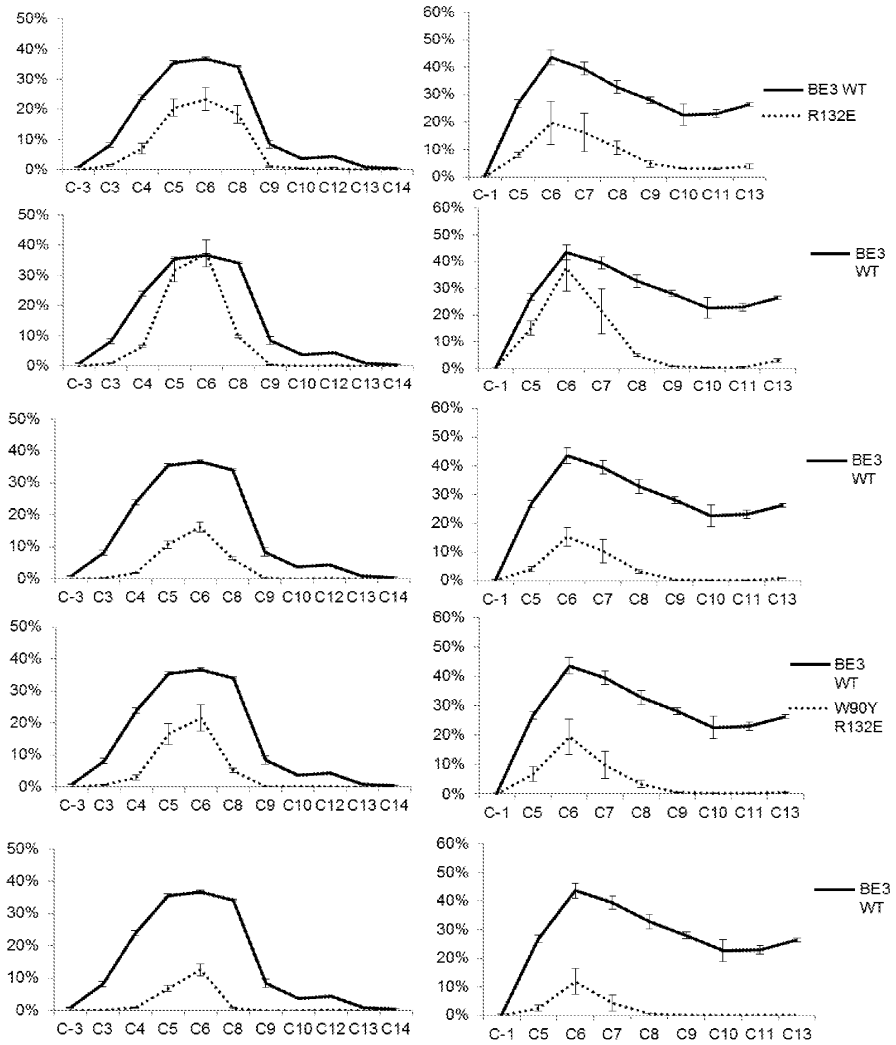
도면97b

APOBEC1 돌연변이	APOBEC3G 돌연변이	참조
R126A	R320A	#9,10
R126E	R320E	#9,10
W90A	W285A	#9,10
W90Y	W285Y	본 연구
R132E	R326E	본 연구

도면97ci


부위 A
TGC3C4C5C6TC8C9C10TC12C13C14TGGCCCAGG
부위 B
AGAGC5C6C7C8C9C10C11TC13AAAGAGAGGG
BE3 WT
H121/122R
R126A
R126E
R118A
W90A
W90Y
C-3 C3 C4 C5 C6 C8 C9 C10 C12 C13 C14
C-1 C5 C6 C7 C8 C9 C10 C11 C13

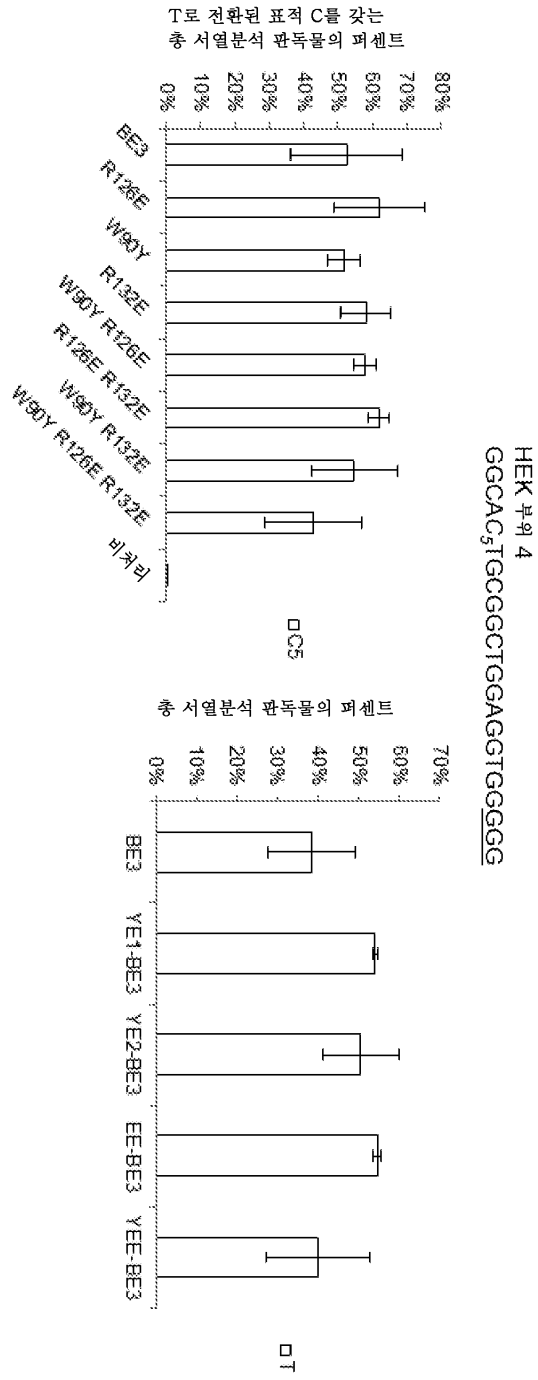
도면97cii



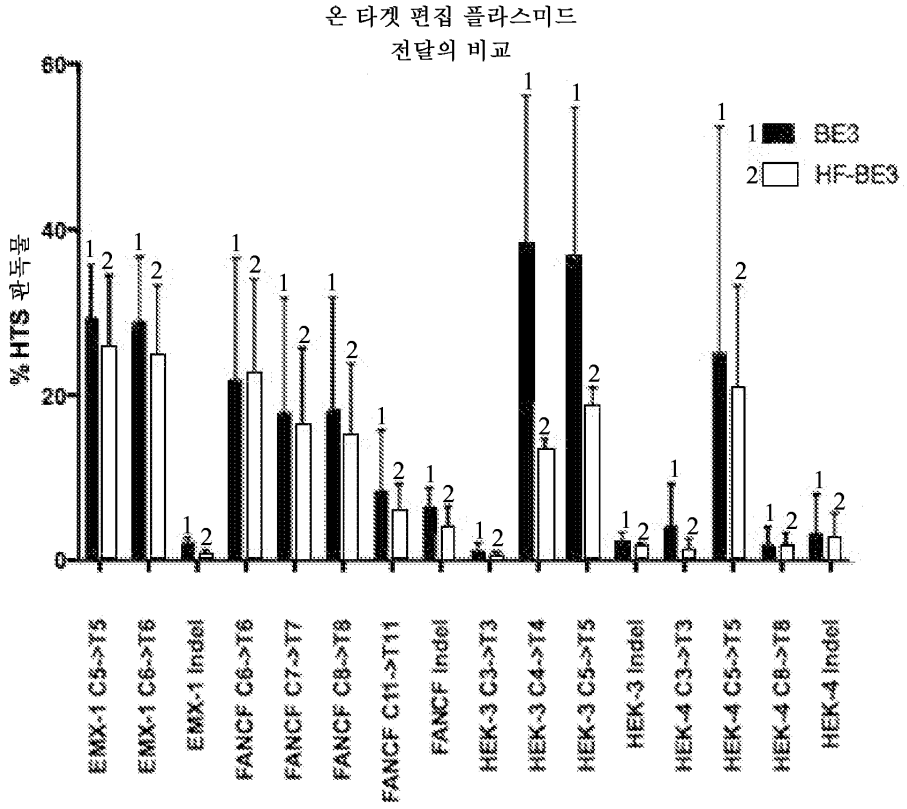
편집 윈도우 폭

	부위 A	부위 B		부위 A	부위 B
BE3 WT	5.0	6.1	W90Y	3.8	4.9
H121/122R	4.2	7.4	R132E	4.0	3.0
R126A	4.4	3.4	W90Y R126E	2.9	3.0
R126E	4.2	3.1	R126E R132E	2.9	3.0
R118A	2.4	3.6	W90Y R132E	2.7	2.8
W90A	2.5	1.1	W90Y R126E R132E	2.1	1.4

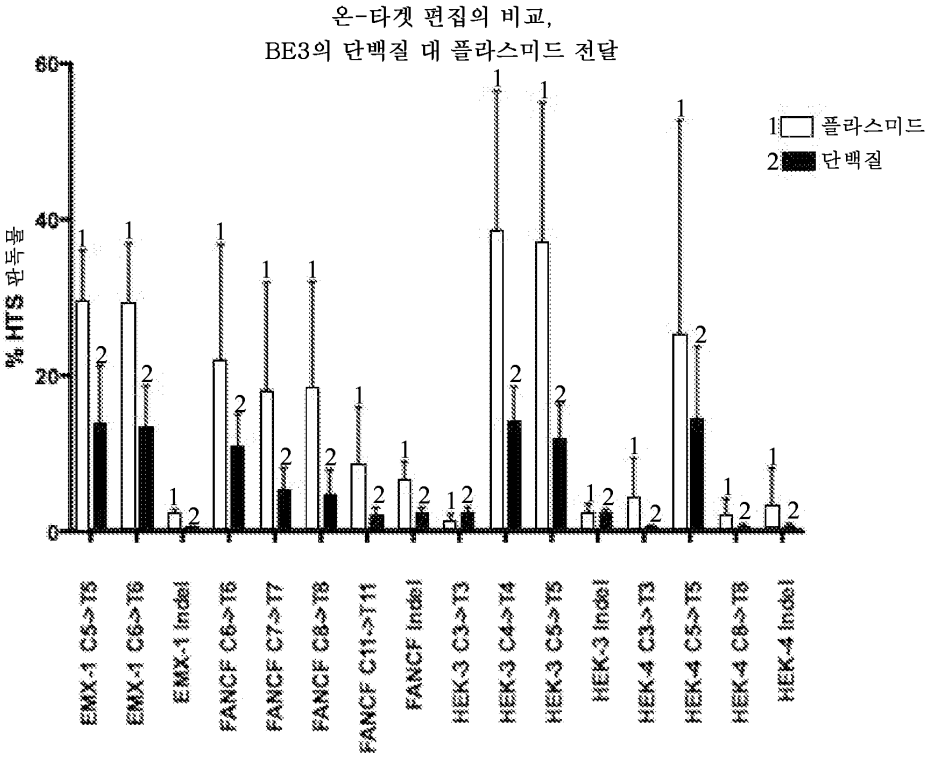
도면98


HEK 부위 4
GGCACGTGCGGCTGGAGGTGGGGG
T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0%
10%
20%
30%
40%
50%
60%
70%
80%
BE3
R126E
W90Y
R132E
W90Y R126E
R126E R132E
W90Y R132E
W90Y R126E R132E
비처리
□C5
총 서열분석 판독물의 퍼센트
0%
10%
20%
30%
40%
50%
60%
70%
BE3
YE1-BE3
YE2-BE3
EE-BE3
YEE-BE3
□T

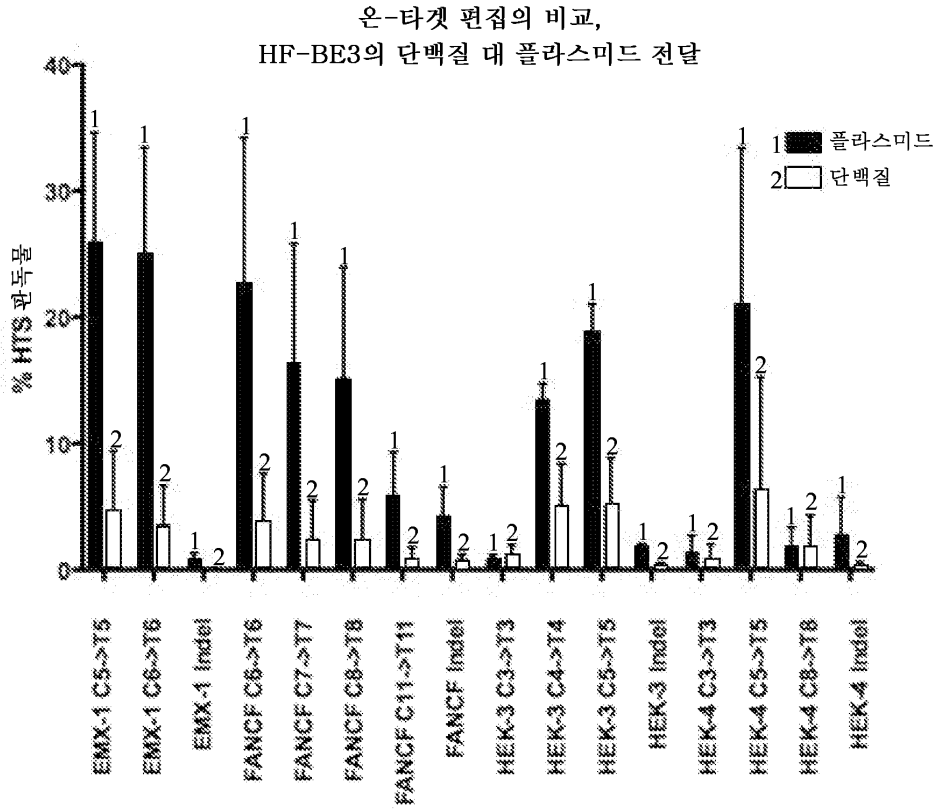
도면99


온 타겟 편집 플라스미드
전달의 비교
% HTS 판독물
60
40
20
0
1 BE3
2 HF-BE3
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 Indel
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF Indel
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 Indel
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 Indel

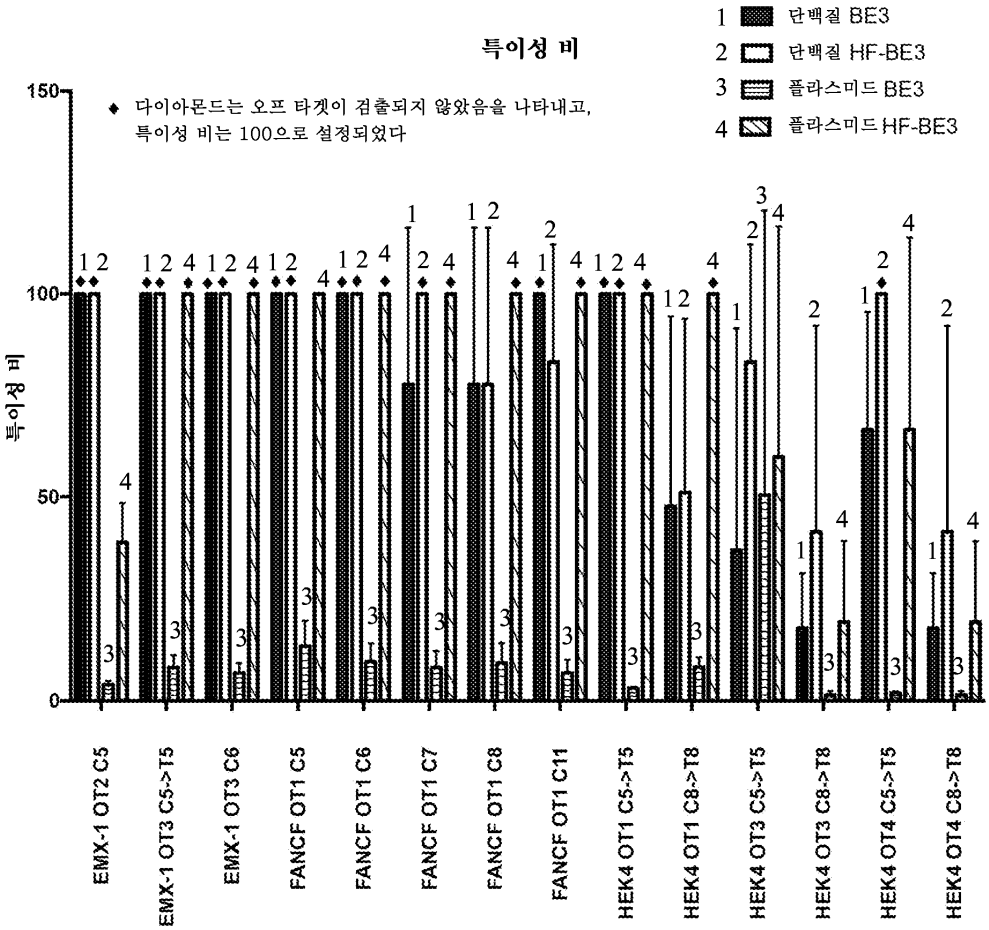
도면100


온-타겟 편집의 비교,
BE3의 단백질 대 플라스미드 전달
% HTS 판독물
60
40
20
0
1 플라스미드
2 단백질
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 Indel
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF Indel
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 Indel
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 Indel

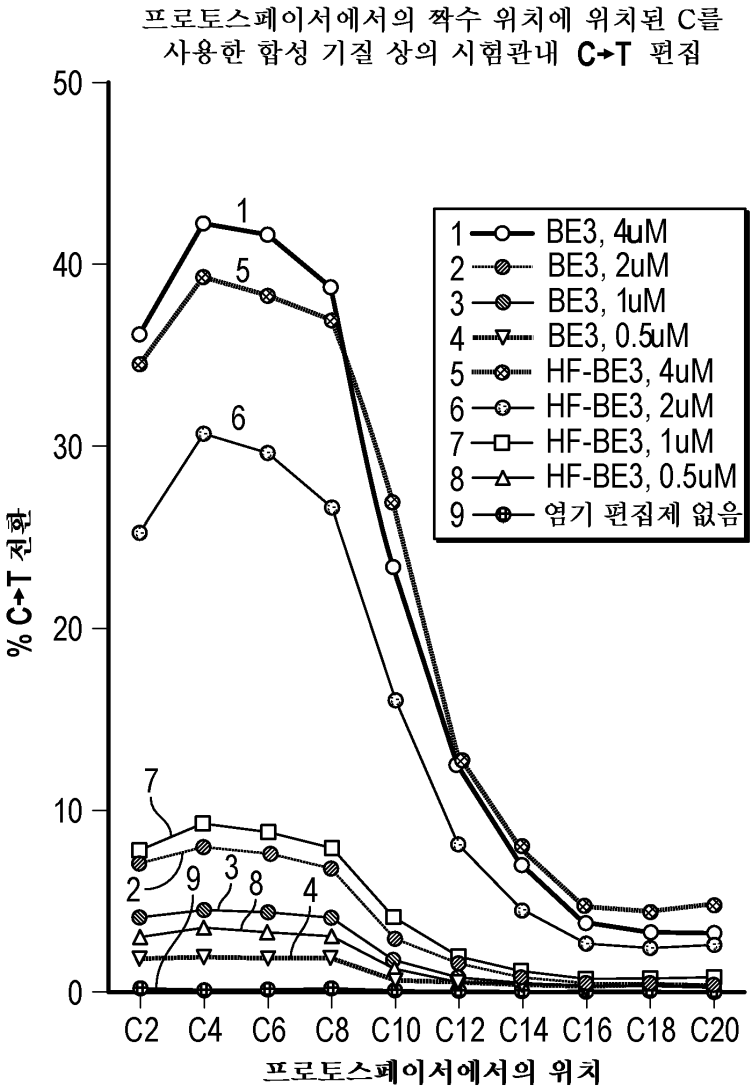
도면101


온-타겟 편집의 비교,
HF-BE3의 단백질 대 플라스미드 전달
% HTS 판독물
0
10
20
30
40
1 플라스미드
2 단백질
EMX-1 C5->T5
EMX-1 C6->T6
EMX-1 Indel
FANCF C6->T6
FANCF C7->T7
FANCF C8->T8
FANCF C11->T11
FANCF Indel
HEK-3 C3->T3
HEK-3 C4->T4
HEK-3 C5->T5
HEK-3 Indel
HEK-4 C3->T3
HEK-4 C5->T5
HEK-4 C8->T8
HEK-4 Indel

도면102

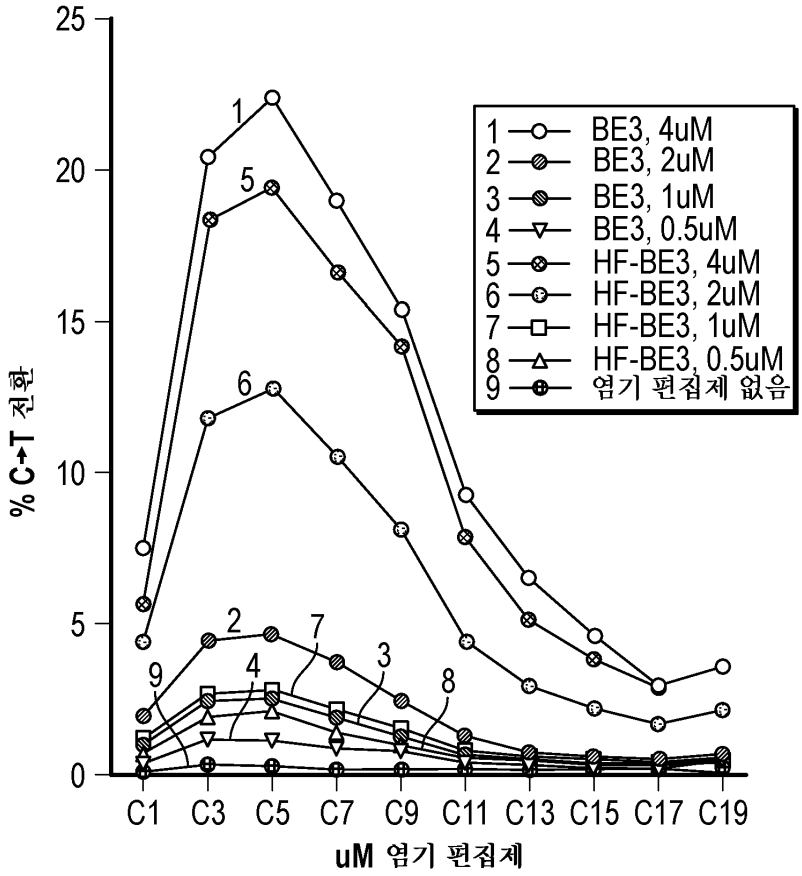

특이성 비
1 단백질 BE3
2 단백질 HF-BE3
3 플라스미드 BE3
4 플라스미드 HF-BE3
◆ 다이아몬드는 오프 타겟이 검출되지 않았음을 나타내고, 특이성 비는 100으로 설정되었다
특이성 비
150
100
50
0
EMX-1 OT2 C5
EMX-1 OT3 C5->T5
EMX-1 OT3 C6
FANCF OT1 C5
FANCF OT1 C6
FANCF OT1 C7
FANCF OT1 C8
FANCF OT1 C11
HEK4 OT1 C5->T5
HEK4 OT1 C8->T8
HEK4 OT3 C5->T5
HEK4 OT3 C8->T8
HEK4 OT4 C5->T5
HEK4 OT4 C8->T8

도면103

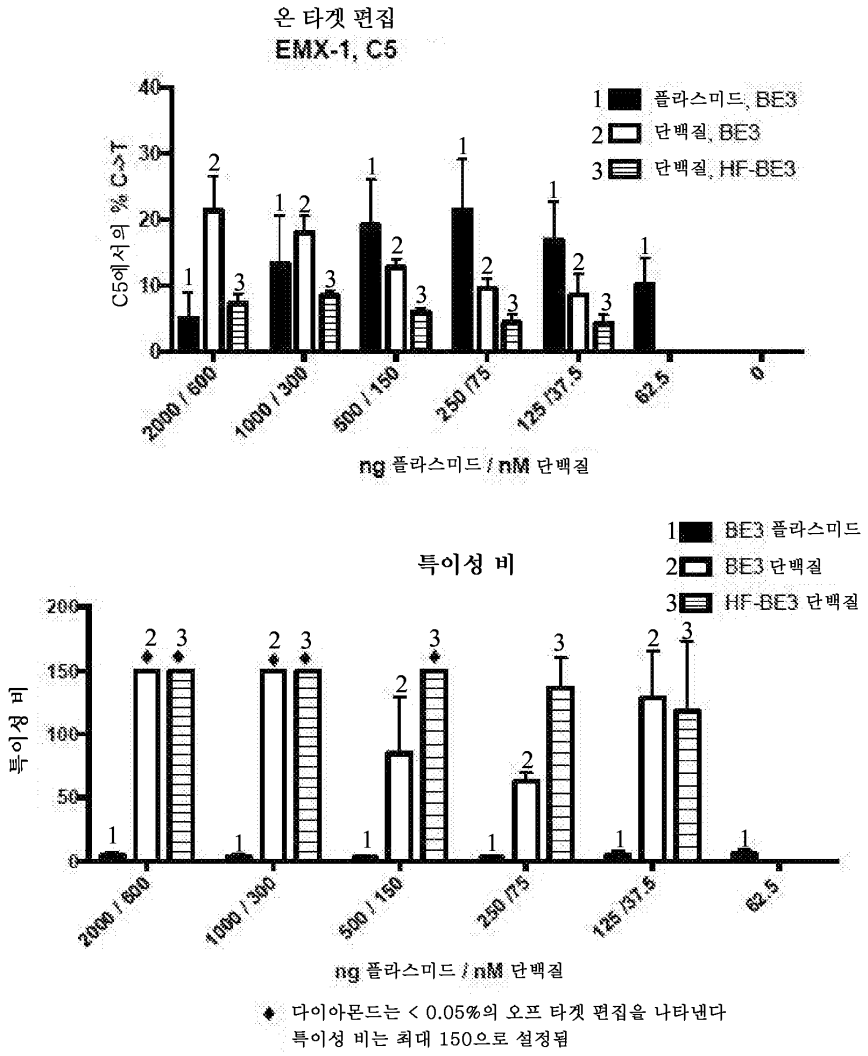

프로토스페이서에서의 짝수 위치에 위치된 C를
사용한 합성 기질 상의 시험관내 C→T 편집
% C→T 전환
50
40
30
20
10
0
C2
C4
C6
C8
C10
C12
C14
C16
C18
C20
프로토스페이서에서의 위치
1 BE3, 4uM
2 BE3, 2uM
3 BE3, 1uM
4 BE3, 0.5uM
5 HF-BE3, 4uM
6 HF-BE3, 2uM
7 HF-BE3, 1uM
8 HF-BE3, 0.5uM
9 염기 편집제 없음

도면104

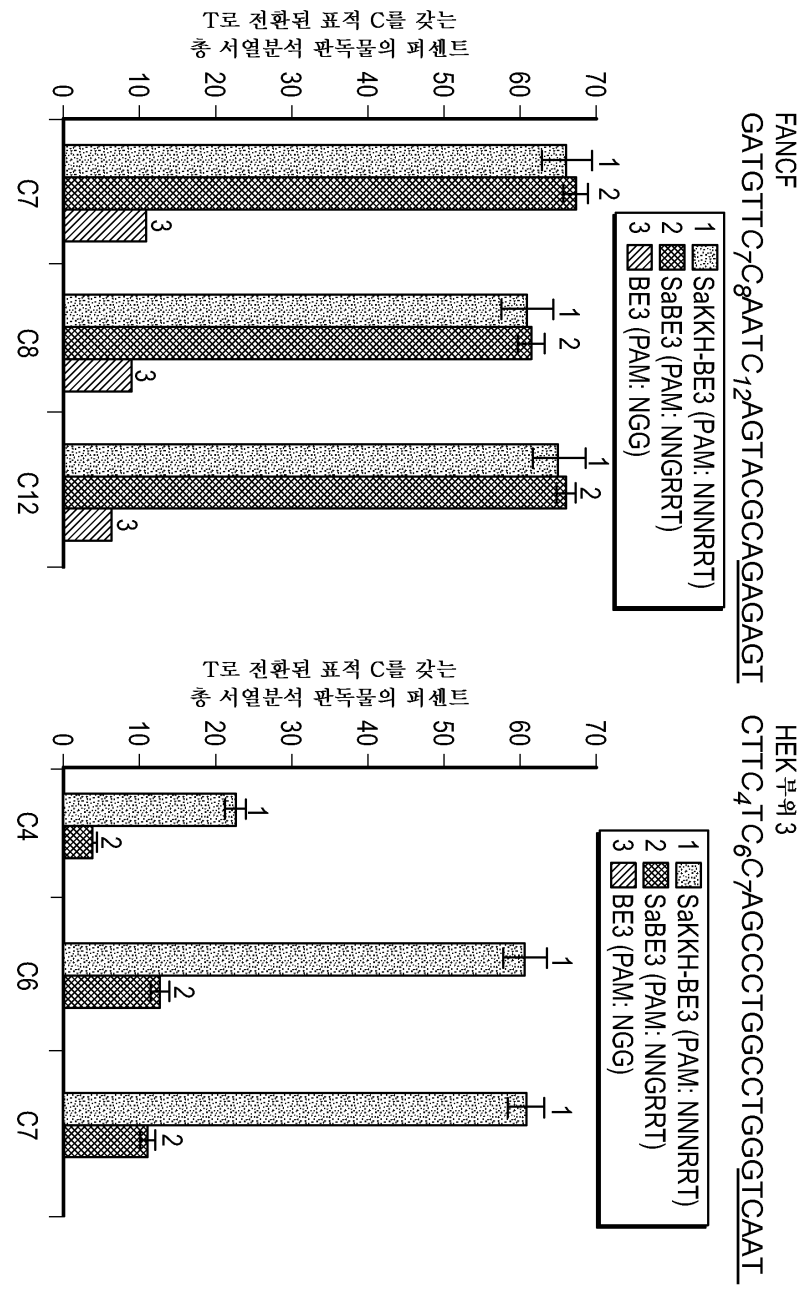
프로토스페이서에서의 홀수 위치에 위치된 C를
사용한 합성 기질 상의 시험관내 C→T 편집



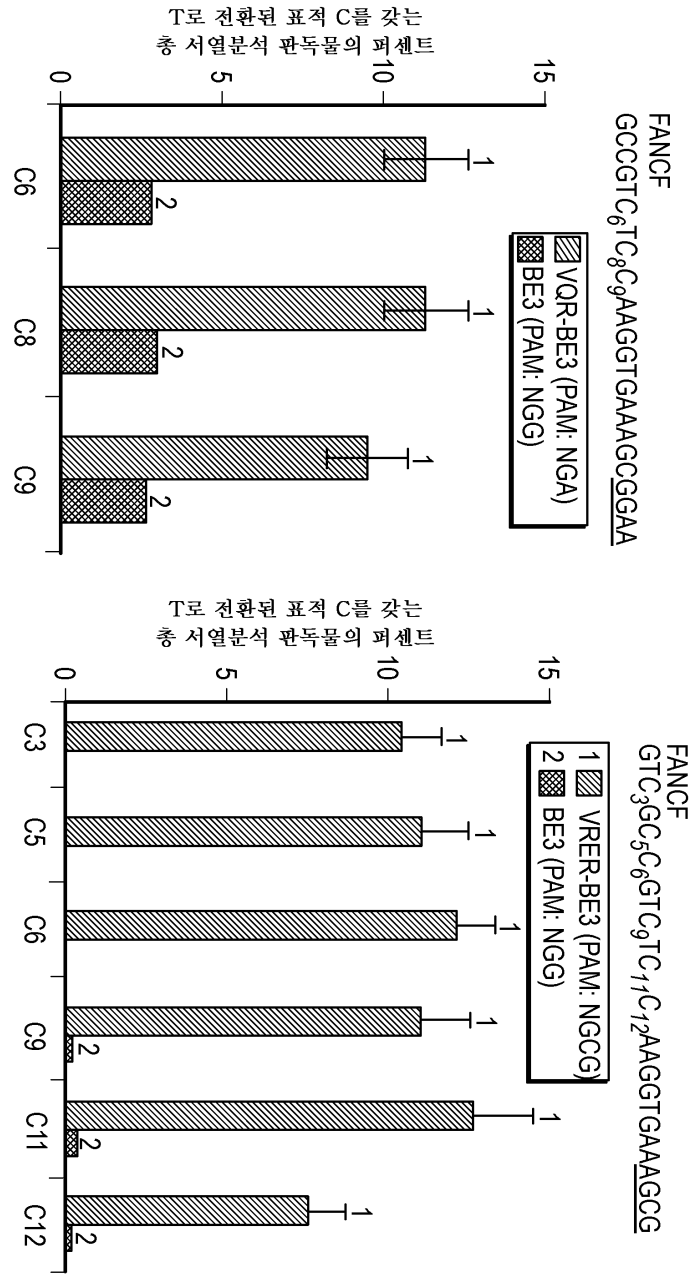
도면105


온 타겟 편집
EMX-1, C5
1 플라스미드, BE3
2 단백질, BE3
3 단백질, HF-BE3
C5에서의 % C>T
0
10
20
30
40
2000 / 600
1000 / 300
500 / 150
250 /75
125 /37.5
62.5
0
ng 플라스미드 / nM 단백질
특이성 비
1 BE3 플라스미드
2 BE3 단백질
3 HF-BE3 단백질
특이성 비
0
50
100
150
200
2000 / 600
1000 / 300
500 / 150
250 /75
125 /37.5
62.5
ng 플라스미드 / nM 단백질
◆ 다이아몬드는 < 0.05%의 오프 타겟 편집을 나타낸다
특이성 비는 최대 150으로 설정됨

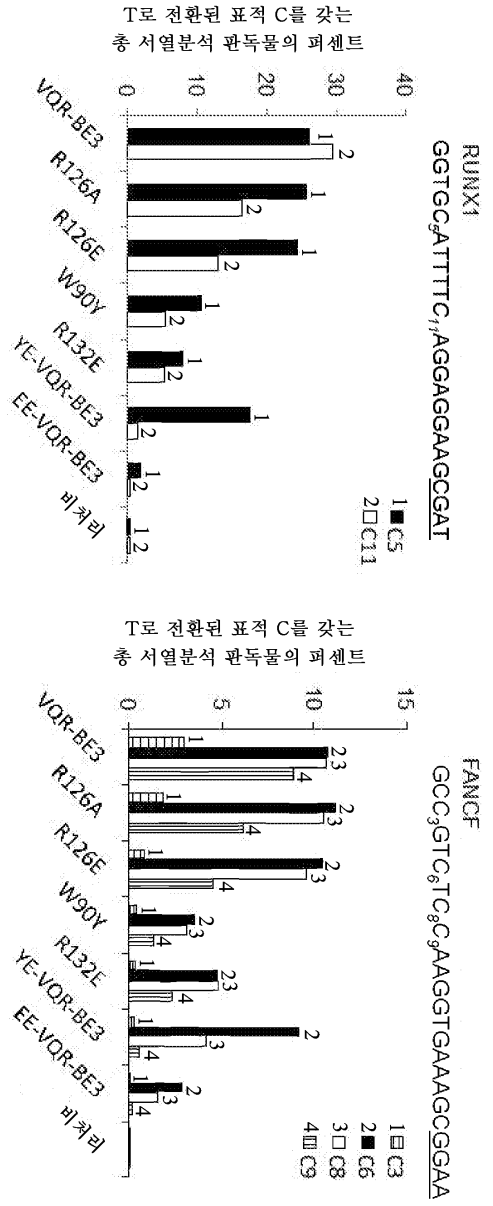
도면106a


FANCF
GATGTTC7C8AATC12AGTACGCAGAGAGT
T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0
10
20
30
40
50
60
70
C7
C8
C12
1 SaKKH-BE3 (PAM: NNNRRT)
2 SaBE3 (PAM: NNGRRT)
3 BE3 (PAM: NGG)
HEK 부위 3
CTTC4TC6C7AGCCCTGGCCTGGGTCAAT
T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0
10
20
30
40
50
60
70
C4
C6
C7
1 SaKKH-BE3 (PAM: NNNRRT)
2 SaBE3 (PAM: NNGRRT)
3 BE3 (PAM: NGG)

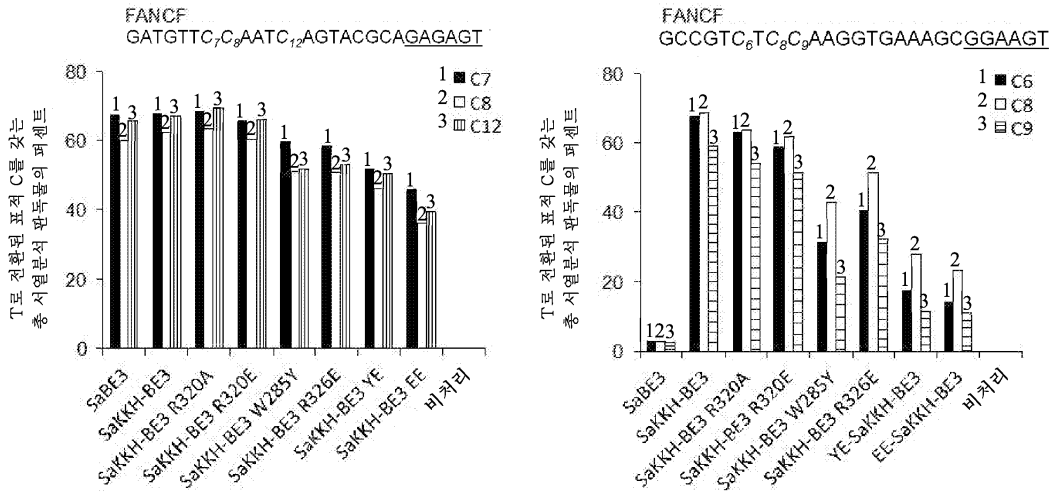
도면106b


T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0
5
10
15
C6
C8
C9
1
2
VQR-BE3 (PAM: NGA)
BE3 (PAM: NGG)
FANCF
GCCGTC6TC8C9AAGGTGAAAGCGGAA
T로 전환된 표적 C를 갖는
총 서열분석 판독물의 퍼센트
0
5
10
15
C3
C5
C6
C9
C11
C12
1
2
VRER-BE3 (PAM: NGCG)
BE3 (PAM: NGG)
FANCF
GTC3GC5C6GTC9TC11C12AAGGTGAAAGCG

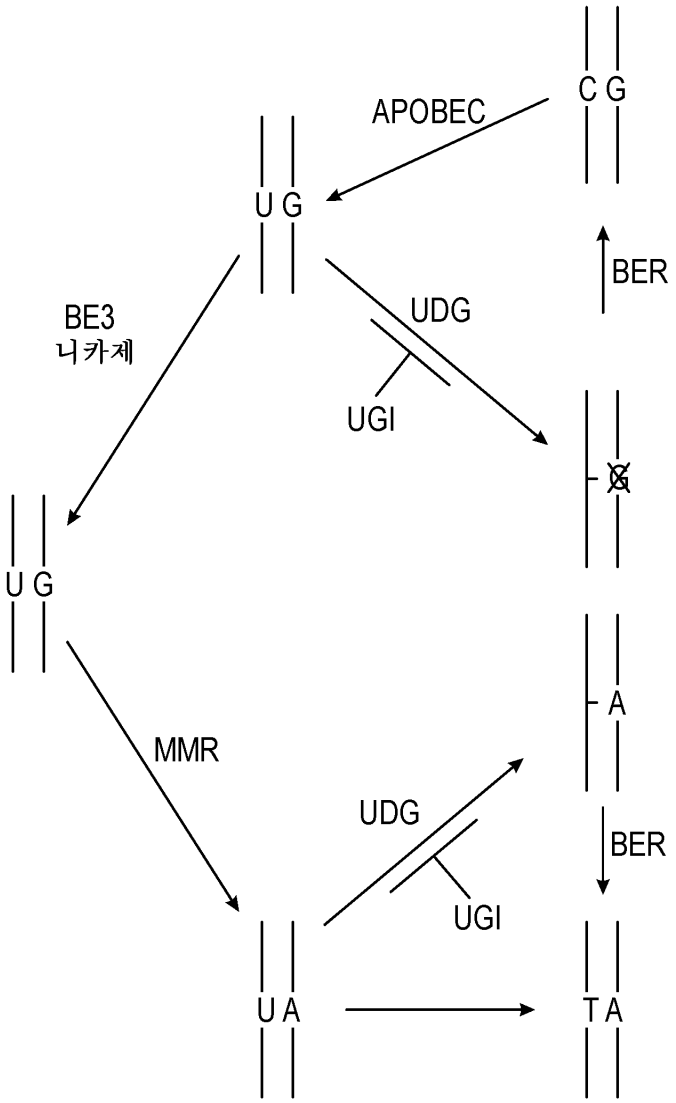
도면107a



도면107b



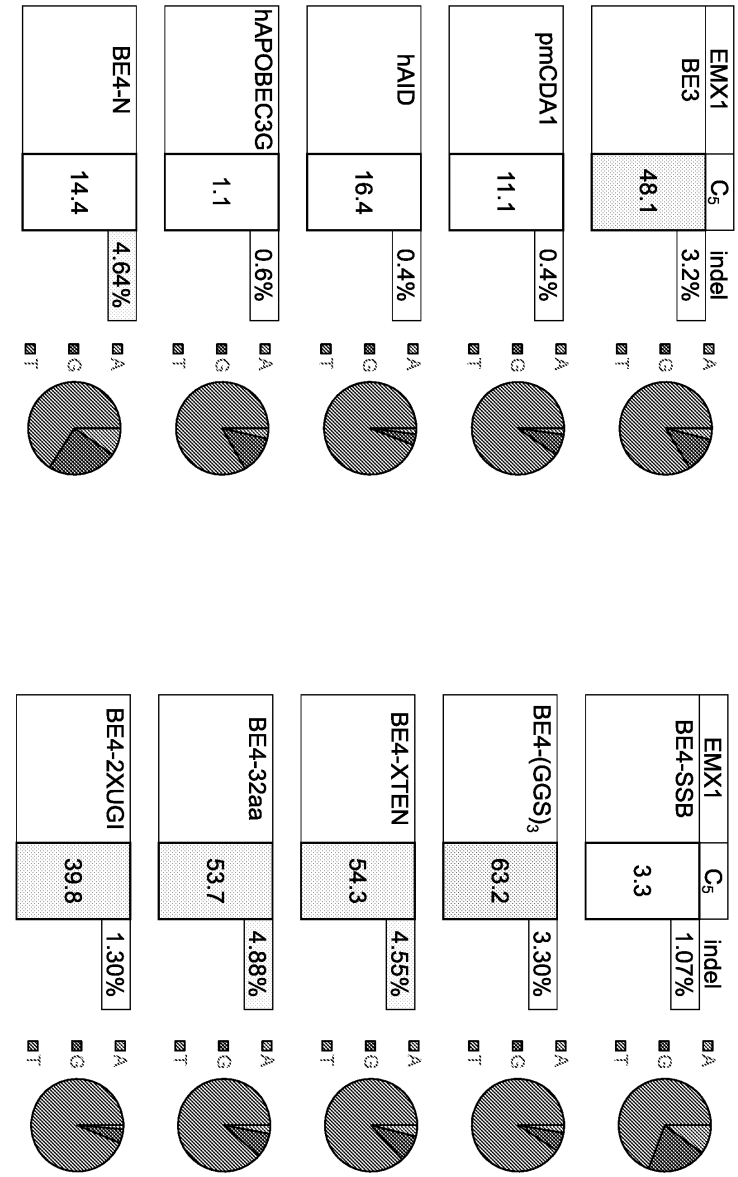
도면108



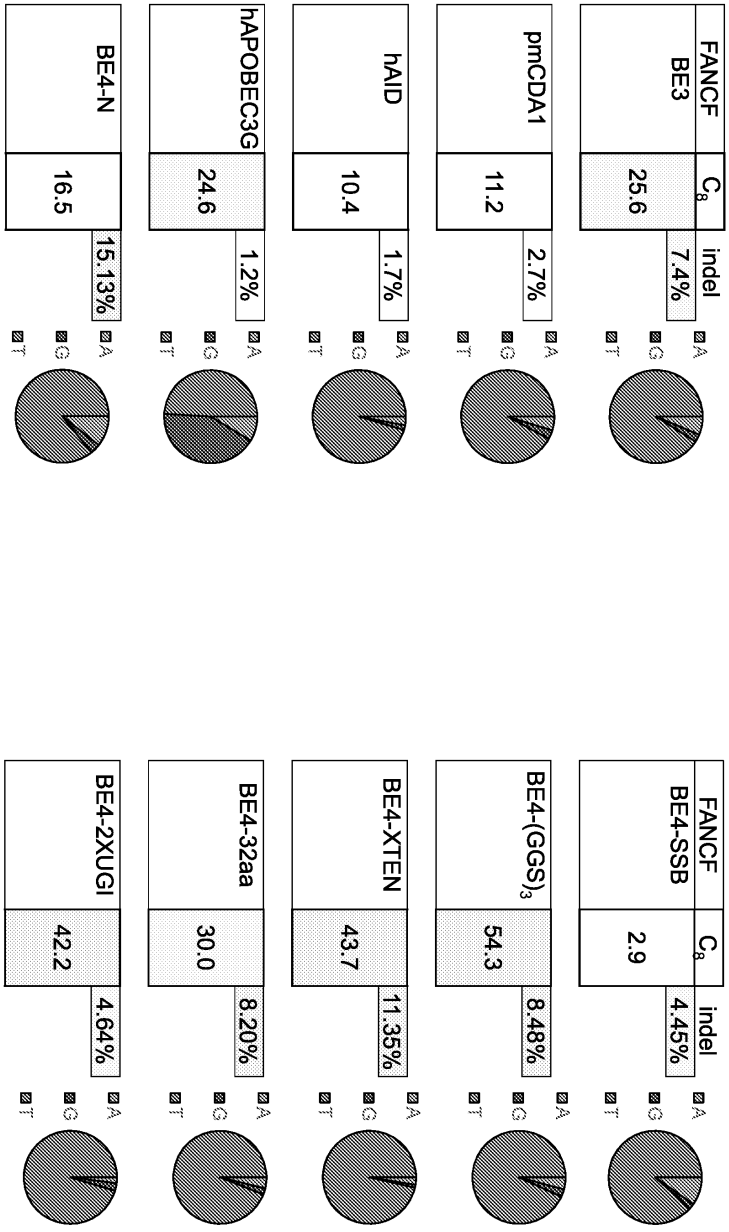
도면109

BE3:
rAPOBEC1
XTEN
dCas9(A840H)
SGGS
UGI
BE4-pmCDA1:
pmCDA1
XTEN
dCas9(A840H)
SGGS
UGI
BE4-hAID:
hAID
XTEN
dCas9(A840H)
SGGS
UGI
BE4-3G:
hAPOBEC3G
XTEN
dCas9(A840H)
SGGS
UGI
BE4-N:
UGI
(GGS)3
rAPOBEC1
XTEN
dCas9(A840H)
BE4-SSB:
rAPOBEC1
XTEN
dCas9(A840H)
(GGS)3
SSB
BE4-(GGS)3:
rAPOBEC1
XTEN
dCas9(A840H)
(GGS)3
UGI
BE4-XTEN:
rAPOBEC1
XTEN
dCas9(A840H)
XTEN
UGI
BE4-32aa:
rAPOBEC1
32aa
dCas9(A840H)
SGGS
UGI
BE4-2XUGI:
rAPOBEC1
XTEN
dCas9(A840H)
SGGS
UGI
SGGS
UGI
BE4:
rAPOBEC1
32aa
dCas9(A840H)
(GGS)3
UGI
(GGS)3
UGI
EMX1:
FANCF:
HEK2:
HEK3:
HEK4:
RNF2:
GAGTCCGAGCAGAAGAAGAAGGG
GGAATCCCTTCTGCAGCACCTGG
GAACACAAAGCATAGACTGCGGG
GGCCCAGACTGAGCACGTGATGG
GGCACTGCGGCTGGAGGTCCGGG
GTCATCTTAGTCATTACCTGAGG

도면110

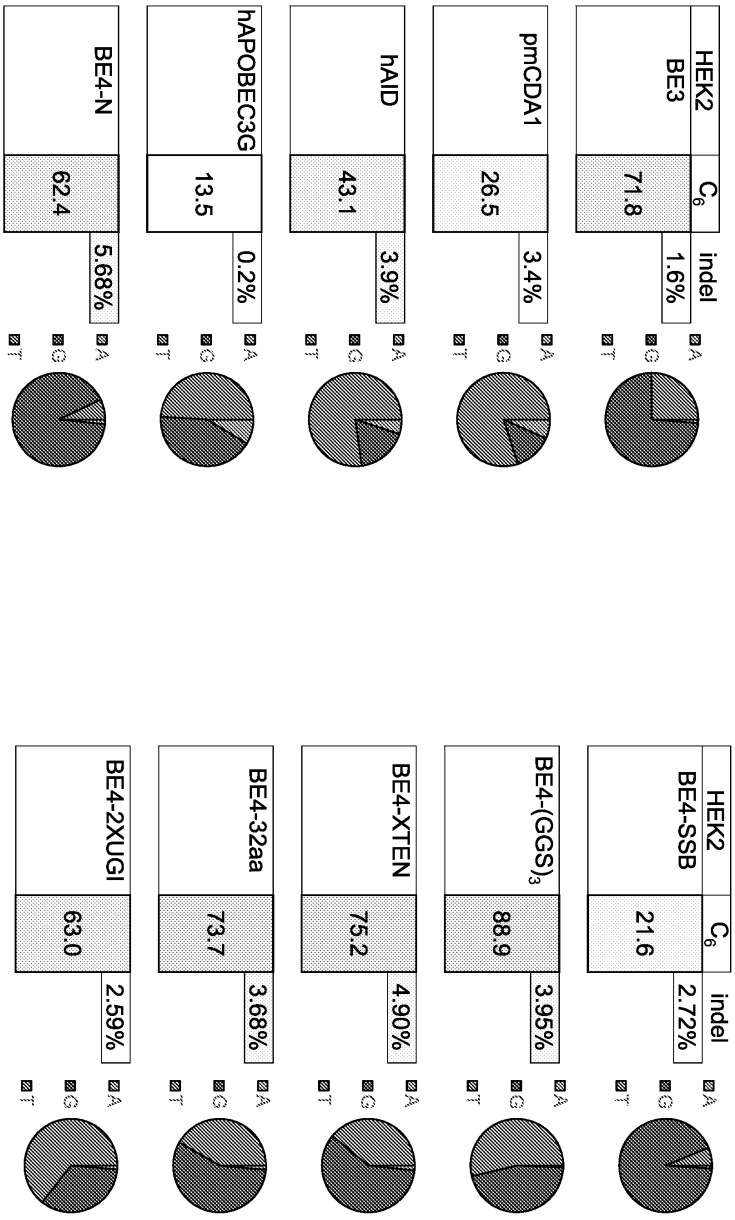

EMX1
BE3
C5
indel
48.1
3.2%
A
G
T
pmCDA1
11.1
0.4%
hAID
16.4
0.4%
hAPOBEC3G
1.1
0.6%
BE4-N
14.4
4.64%
EMX1
BE4-SSB
3.3
1.07%
BE4-(GGS)3
63.2
3.30%
BE4-XTEN
54.3
4.55%
BE4-32aa
53.7
4.88%
BE4-2XUGI
39.8
1.30%

도면111


FANCF
BE3
C8
25.6
indel
7.4%
A
G
T
pmCDA1
11.2
2.7%
hAID
10.4
1.7%
hAPOBEC3G
24.6
1.2%
BE4-N
16.5
15.13%

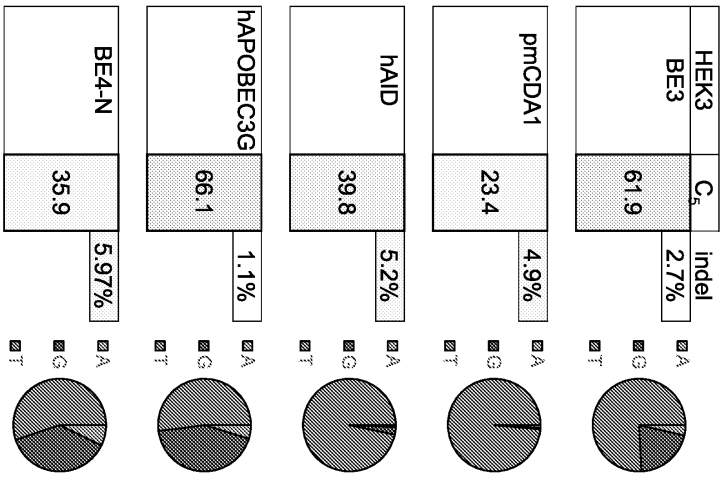
FANCF
BE4-SSB
C8
2.9
indel
4.45%
A
G
T
BE4-(GGS)3
54.3
8.48%
BE4-XTEN
43.7
11.35%
BE4-32aa
30.0
8.20%
BE4-2XUGI
42.2
4.64%

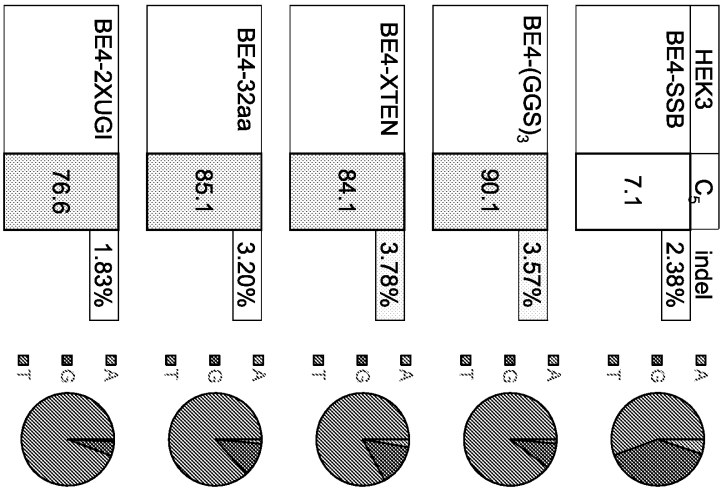
도면112


HEK2
BE3
C6
indel
71.8
1.6%
A
G
T
pmCDA1
26.5
3.4%
hAID
43.1
3.9%
hAPOBEC3G
13.5
0.2%
BE4-N
62.4
5.68%

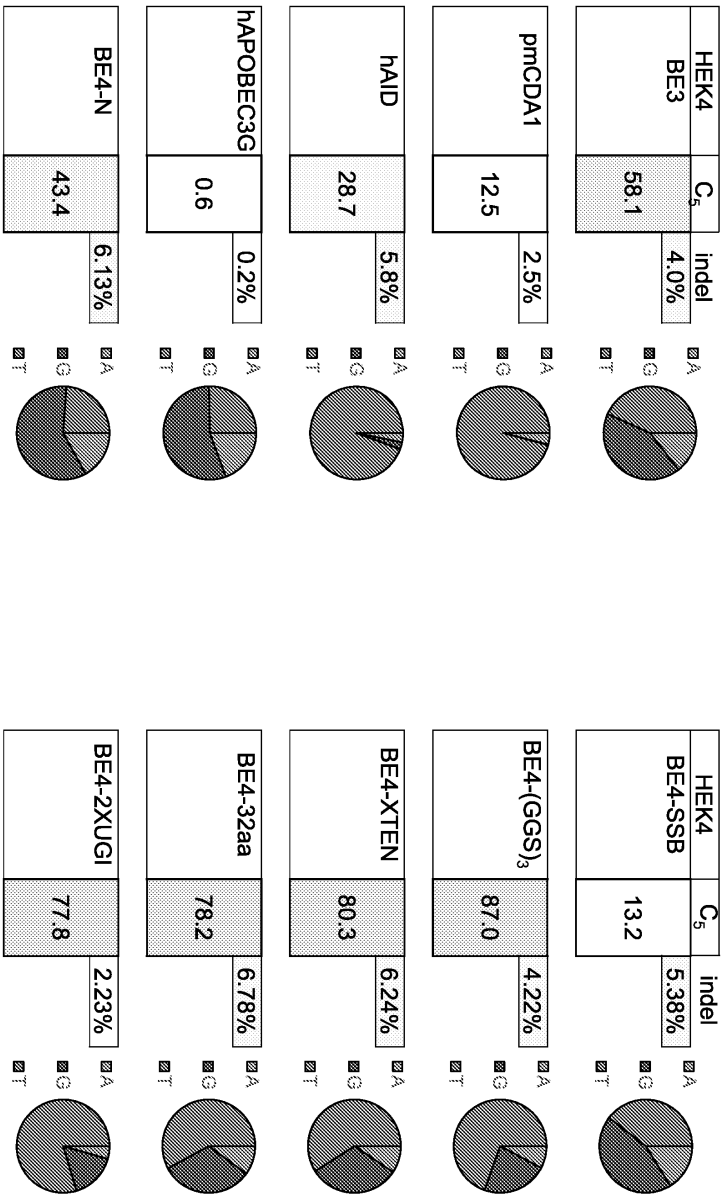
HEK2
BE4-SSB
C6
indel
21.6
2.72%
A
G
T
BE4-(GGS)3
88.9
3.95%
BE4-XTEN
75.2
4.90%
BE4-32aa
73.7
3.68%
BE4-2XUGI
63.0
2.59%

도면113


HEK3
BE3
C5
61.9
indel
2.7%
A
G
T
pmCDA1
23.4
4.9%
hAID
39.8
5.2%
hAPOBEC3G
66.1
1.1%
BE4-N
35.9
5.97%

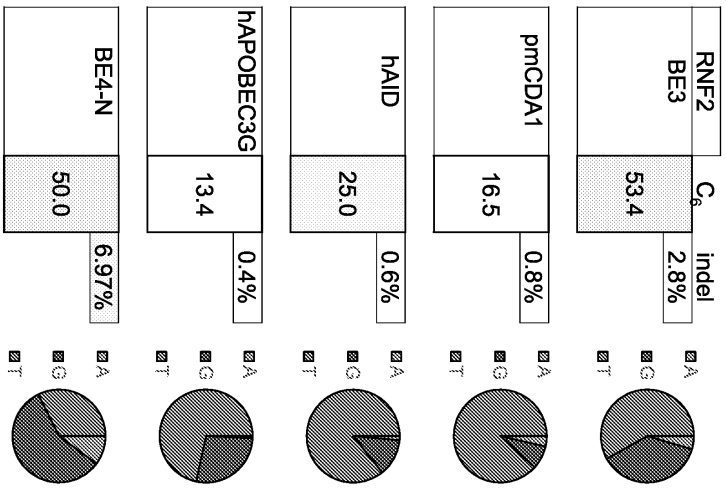

HEK3
BE4-SSB
C5
7.1
indel
2.38%
A
G
T
BE4-(GGS)3
90.1
3.57%
BE4-XTEN
84.1
3.78%
BE4-32aa
85.1
3.20%
BE4-2XUGI
76.6
1.83%

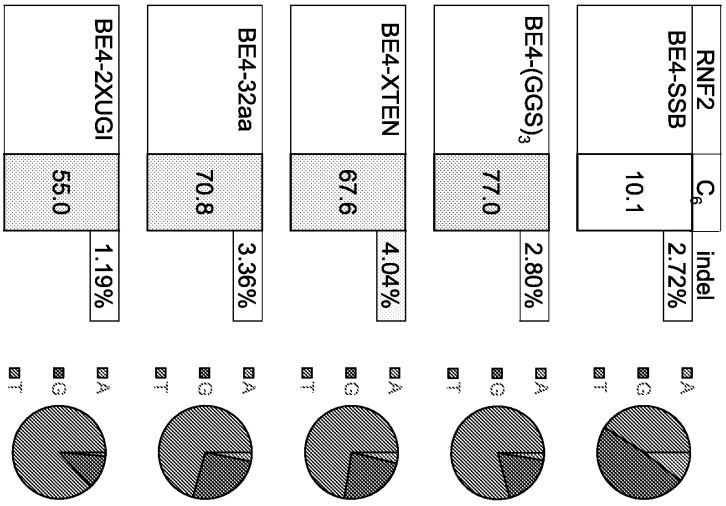
도면114


HEK4
BE3
C5
indel
58.1
4.0%
pmCDA1
12.5
2.5%
hAID
28.7
5.8%
hAPOBEC3G
0.6
0.2%
BE4-N
43.4
6.13%
A
G
T

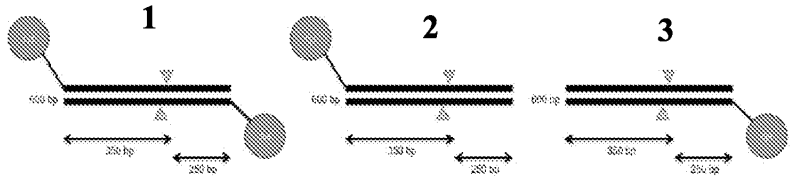
HEK4
BE4-SSB
C5
indel
13.2
5.38%
BE4-(GGS)3
87.0
4.22%
BE4-XTEN
80.3
6.24%
BE4-32aa
78.2
6.78%
BE4-2XUGI
77.8
2.23%
A
G
T

도면115



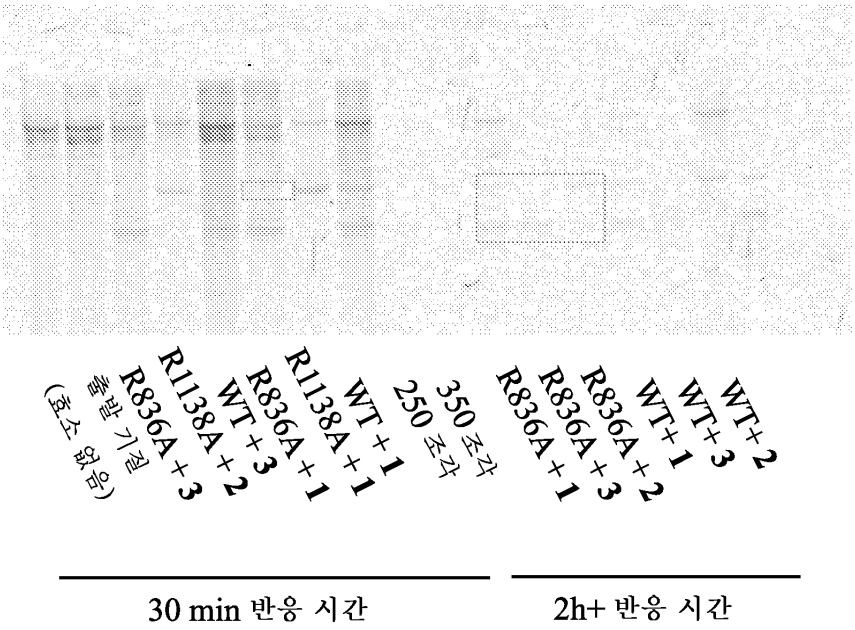


도면116



상부 가닥: 비-표적 가닥; 하부 가닥: 표적 가닥

도면117


출발 기질 (효소 없음)
R836A +3
R1138A +2
WT +3
R836A +1
R1138A +1
WT +1
250 초각
350 초각
R836A +1
R836A +3
R836A +2
WT+1
WT+3
WT+2
30 min 반응 시간
2h+ 반응 시간

도면118

BG –베경; R912A – B E 3 ; R1225A –자기–무력화 BE (비–표적 가닥을 니킹함)

PAM

FANCF1	TTTC	G1	C2	G3	G4	G7	C10	C11	C15	G17	C20	C22
BG		0.01005	0.03022	0.01383	0.00933	0.00502	0.01872	0.02014	0.03740	0.16504	0.01919	0.02099
AsCpf1		0.00385	0.02769	0.07439	0.00672	0.00495	0.03396	0.01795	0.02691	0.01570	0.02237	0.02406
APOBEC-AsCpf1(R912A)-UGI		0.00883	0.02606	0.01430	0.00539	0.00450	0.28158	0.04373	0.01425	0.02795	0.01631	
APOBEC-AsCpf1(R1225A)-UGI		0.00796	0.03754	0.01131	0.00749	0.01124	0.02769	0.02908	0.03346	0.02016	0.01732	0.01651

FANCF2	TTTC	C1	G2	G3	C4	C7	G8	G10	C11	G12	G13	C15	C16	G17	C18
BG		0.02628	0.01963	0.02346	0.00318	0.00798	0.03624	0.00850	0.08758	0.05168	0.00637	0.00599	0.07694	0.01382	0.01805
AsCpf1		0.03489	0.03325	0.05248	0.00392	0.00722	0.03404	0.00731	0.06139	0.18605	0.01851	0.02139	0.12988	0.00327	0.00992
APOBEC-AsCpf1(R912A)-UGI		0.02448	0.02677	0.00699	0.00481	0.01115	0.00534	0.11463	0.01660	0.08427	0.00722	0.10175	0.01119	0.01065	
APOBEC-AsCpf1(R1225A)-UGI		0.02855	0.04216	0.00322	0.00479	0.00587	0.01094	0.00585	0.06534	0.00586	0.00638	0.00444	0.07256	0.00797	0.01061

HEK3-3	TTTC	C3	C6	C8	C9	C12	C13	C14	C18	C19
BG		0.00307	0.00726	0.01440	0.01539	0.01084	0.01181	0.00729	0.01540	0.00634
AsCpf1		0.01118	0.00670	0.01994	0.00892	0.01334	0.01342	0.01707	0.01118	0.01351
APOBEC-AsCpf1(R912A)-UGI		0.00647	0.02264	0.44917	0.30191	0.03754	0.12192	0.01313	0.01291	0.00754
APOBEC-AsCpf1(R1225A)-UGI		0.00770	0.00768	0.10243	0.07236	0.01202	0.00770	0.00551	0.01365	0.00549

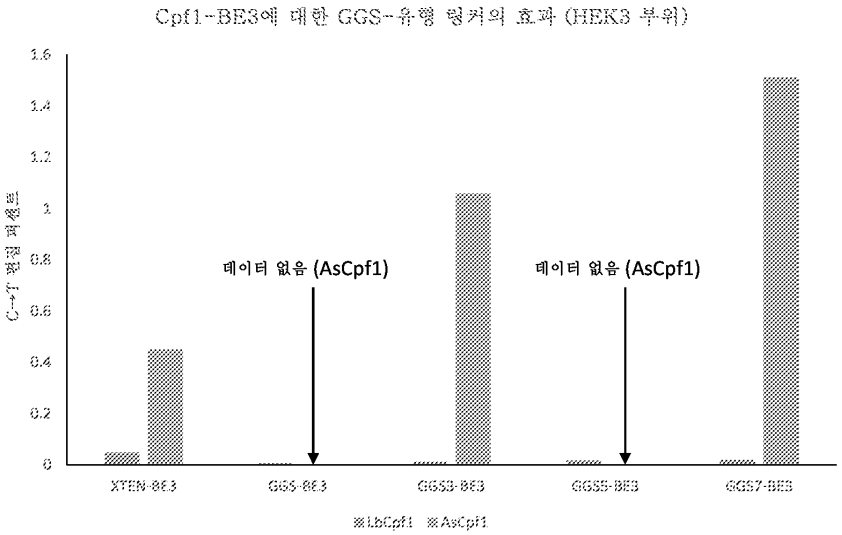
PAM

HEK3-4	GAAA	C23	C26	G28	G29	C30	C35	C33	G42	G50	G57	G58	C59	G1
BG		0.00638	0.01001	0.03459	0.01803	0.01008	0.01008	0.01285	0.01284	0.00548	0.01458	0.02537	0.01099	0.00636
AsCpf1		0.00986	0.00862	0.02466	0.01984	0.00343	0.02600	0.00865	0.00617	0.01718	0.02216	0.01232	0.00488	
APOBEC-AsCpf1(R912A)-UGI		0.00533	0.01185	0.03012	0.00759	0.01735	0.01734	0.04550	0.15978	0.02862	0.02266	0.00542	0.00644	
APOBEC-AsCpf1(R1225A)-UGI		0.01142	0.01045	0.03044	0.01435	0.01626	0.01104	0.01351	0.01238	0.01045	0.02428	0.01930	0.00473	

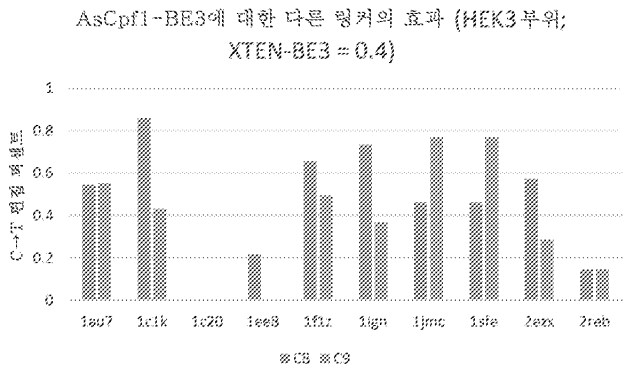
도면119



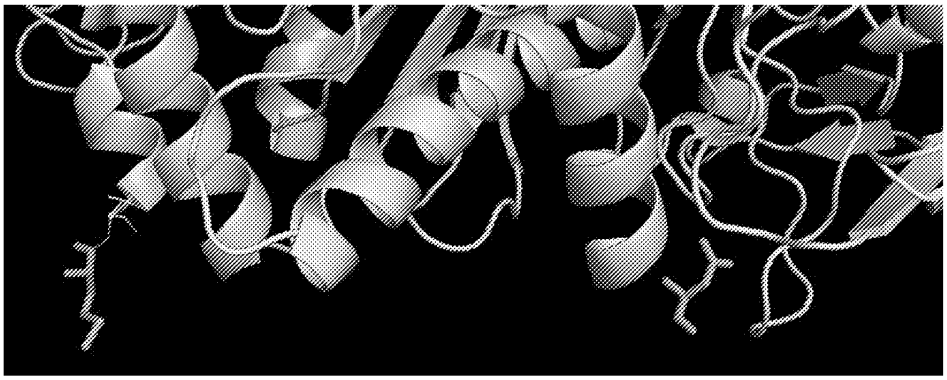
도면120



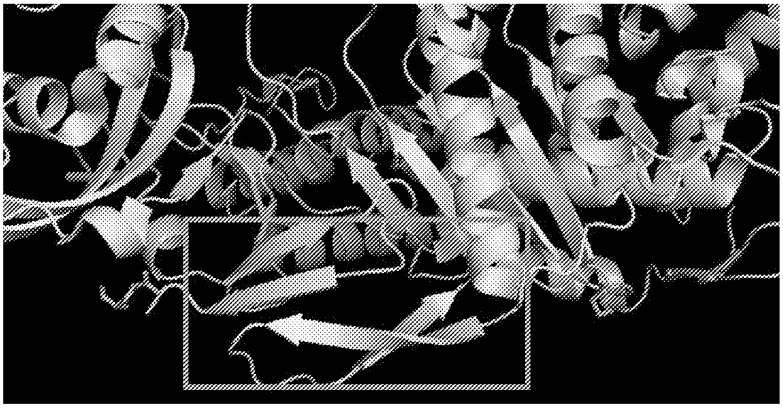
도면121



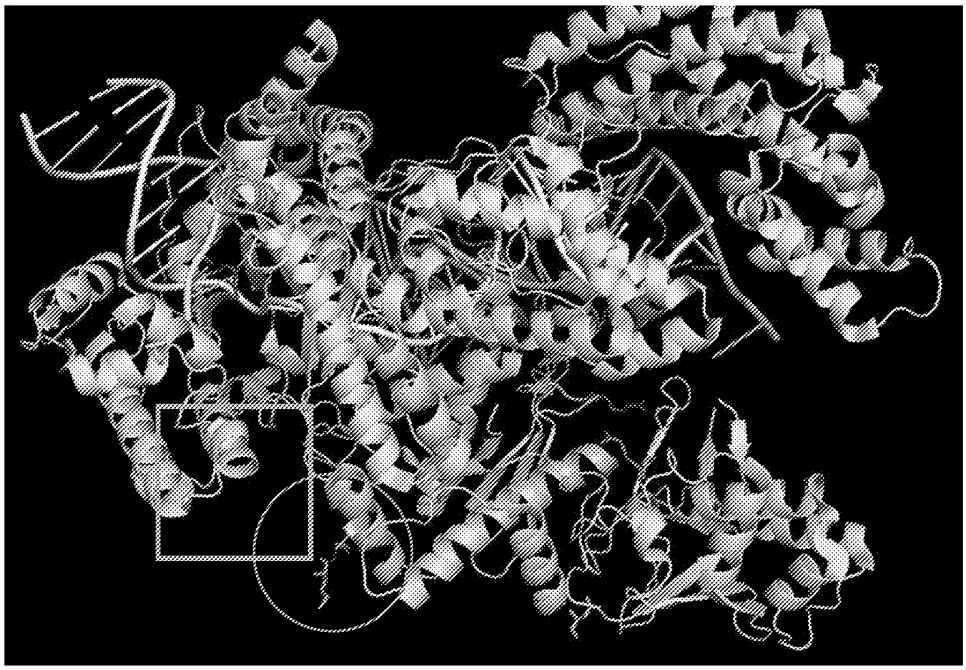
도면122



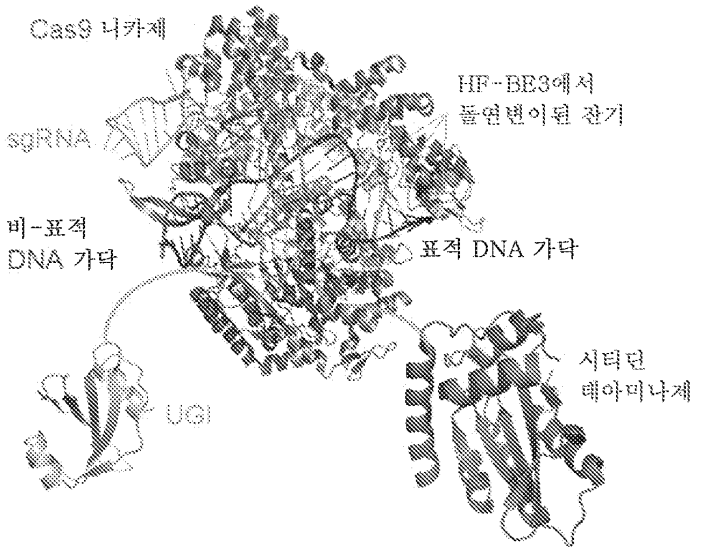
도면123



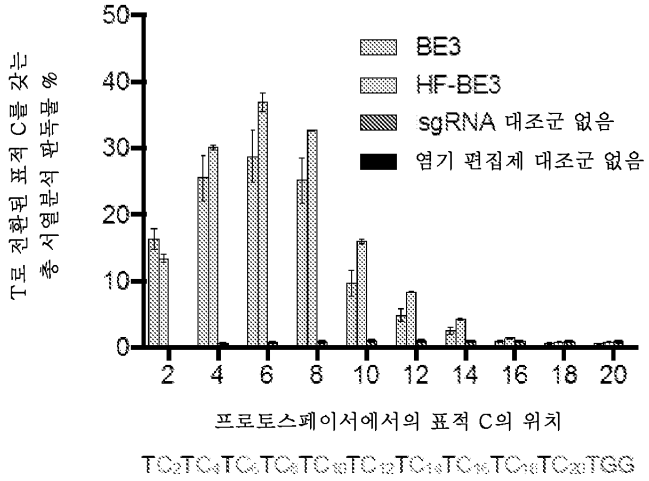
도면124



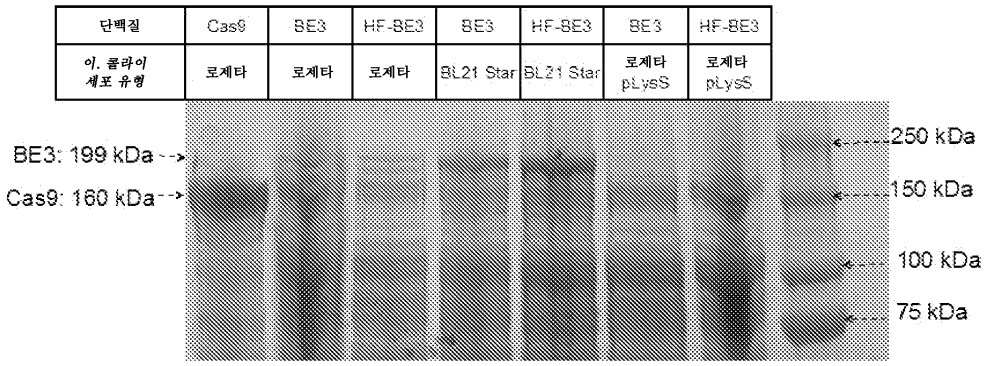
도면125a



도면125b

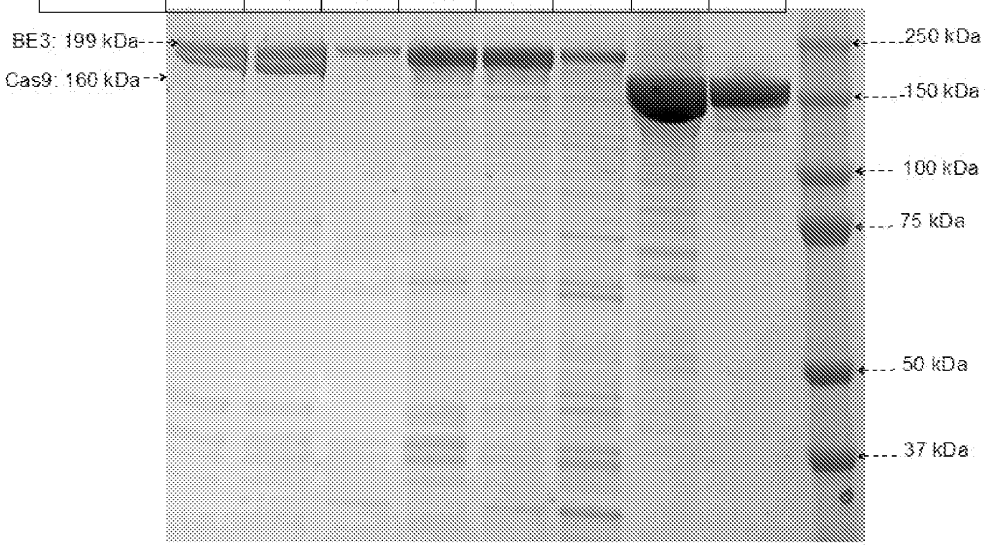


도면126a

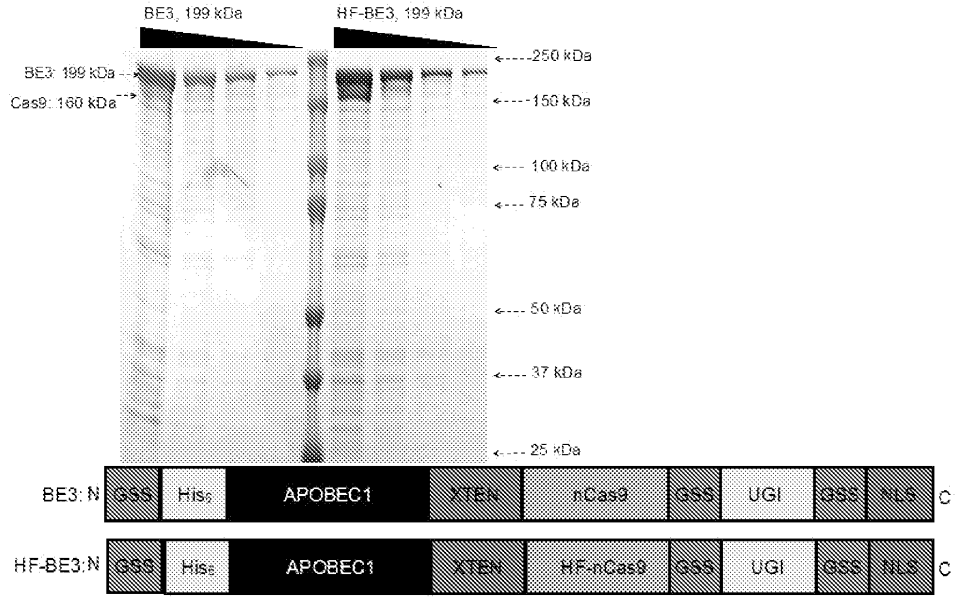


도면126b

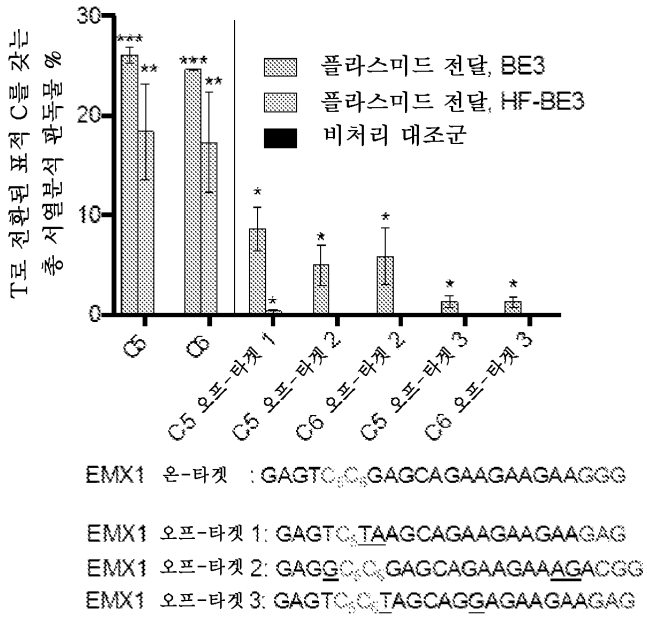
단백질	HF-BE3	BE3	HF-BE3	BE3	HF-BE3	BE3	Cas9	Cas9
His_6 태그	N-말단	N-말단	C-말단	C-말단	C-말단	C-말단	N-말단	N-말단
유도 OD_{600}	0.7	0.7	0.4	0.7	0.7	0.4	0.4	0.4
정제 방법	니켈 수지 + FPLC	니켈 수지 + FPLC	니켈 수지 + FPLC	니켈 수지 + FPLC	니켈 수지 + FPLC	니켈 수지 + FPLC	니켈 수지	니켈 수지 + FPLC



도면126c



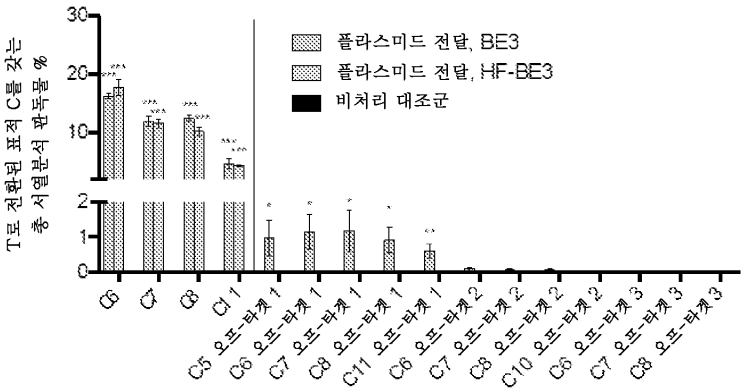
도면127a



EMX1 온-타겟 : GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

EMX1 오프-타겟 1: GAGTC$_5$TAAGCAGAAGAAGAAGAG
EMX1 오프-타겟 2: GAGGC$_5$C$_6$GAGCAGAAGAAAGACGG
EMX1 오프-타겟 3: GAGTC$_5$C$_6$TAGCAGGAGAAGAAGAG

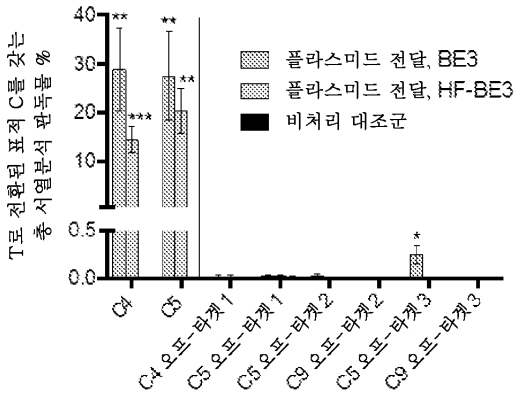
도면127b



FANCF 온-타겟: GGAATC$_5$C$_6$C$_7$TTC$_{11}$TGCAGCACCTGG

FANCF 오프-타겟 1: GGAAC$_5$C$_6$C$_7$GTC$_{11}$TGCAGCACCAGG
FANCF 오프-타겟 2: GGAGTC$_5$C$_6$C$_7$TC$_{10}$C$_{11}$TACAGCACCAGG
FANCF 오프-타겟 3: AGAGGC$_5$C$_6$C$_7$C$_8$TC$_{11}$TGCAGCACCAGG

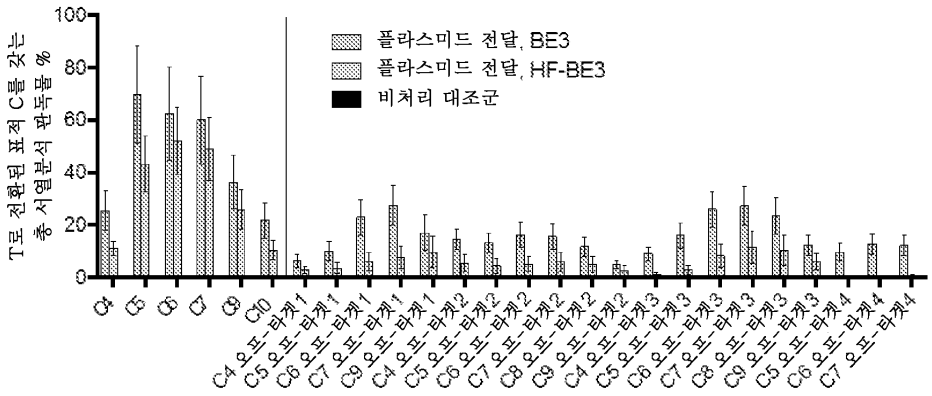
도면127c



HEK293 부위3 온-타겟 GCCCCCCAGACTGAGCACGTGATGG

HEK293 부위3 오프-타겟1: CACCCCCAGACTGAGCACGTGCTGG
HEK293 부위3 오프-타겟2: GACACCCAGACTGGGCACGTGAGGG
HEK293 부위3 오프-타겟3: AGCTCCCAGACTGAGCAAGTGAGGG

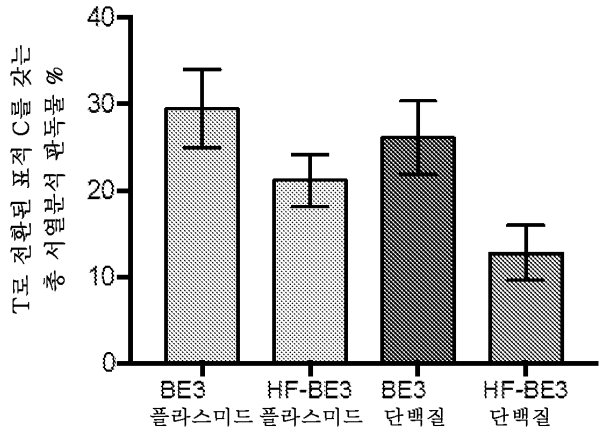
도면127d



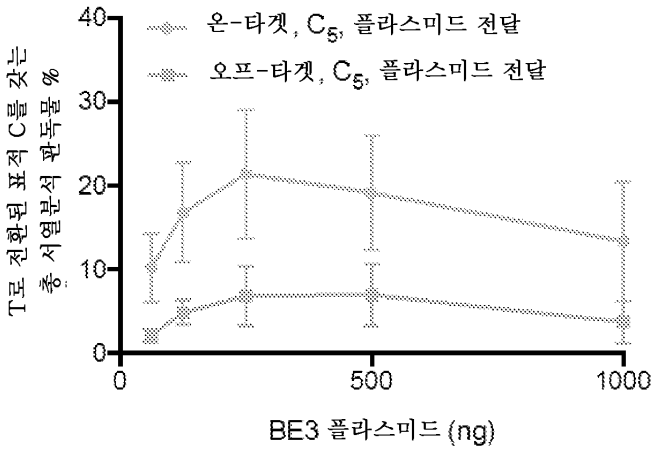
VEGFA 부위2 온-타겟 : GACCCCCCTCCCCGCCTCCGG

VEGFA 부위2 오프-타겟 1: CTACCCCCTCCCCGCCTCCGG
VEGFA 부위2 오프-타겟 2: ATTCCCCCTCCCCGCCTCAGG
VEGFA 부위2 오프-타겟 3: ACACCCCCTCCCCGCCTCAGG
VEGFA 부위2 오프-타겟 4: TGCCCCCCTCCCCACCTCTGG

도면128a



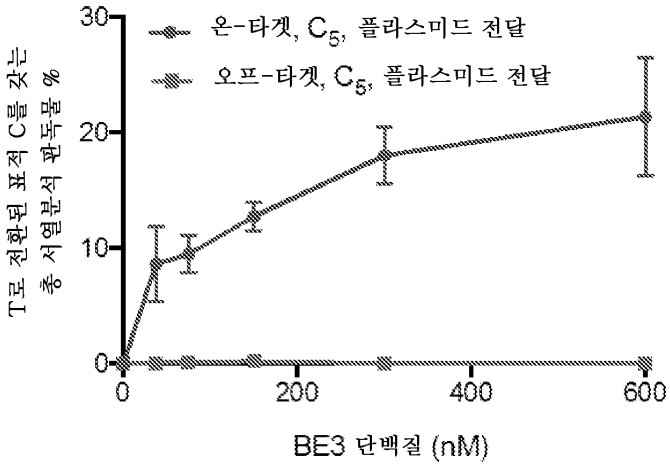
도면128b



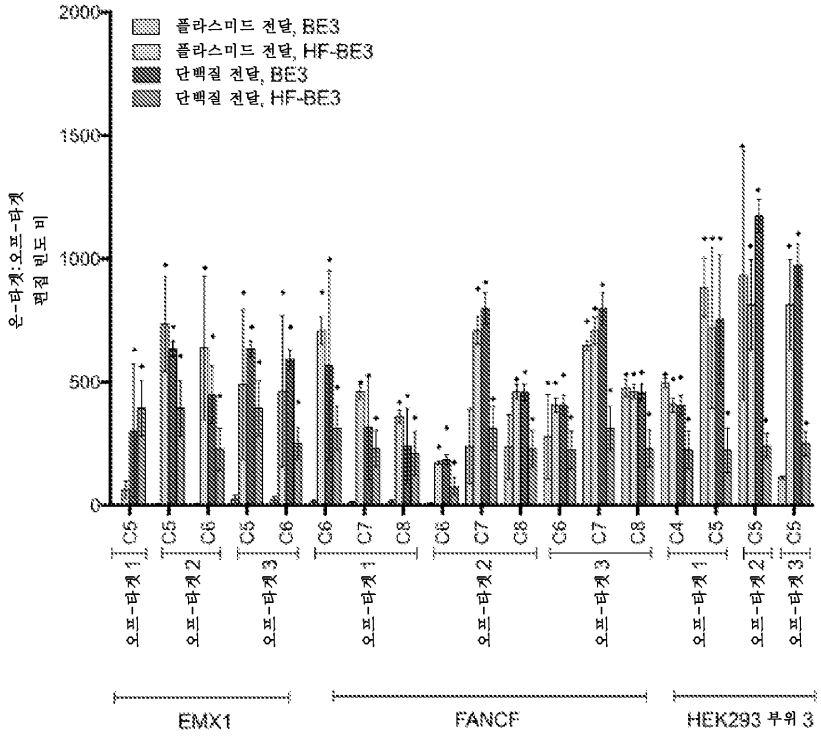
EMX1: GAGTC5C6GAGCAGAAGAAGAAGGG

EMX1 오프-타겟 1: GAGTC5TAAGCAGAAGAAGAAGAG

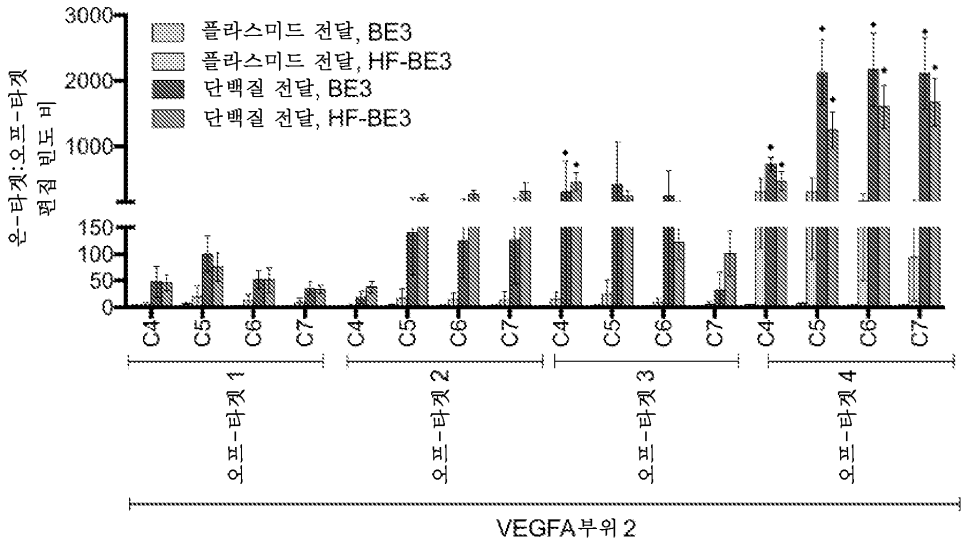
도면128c



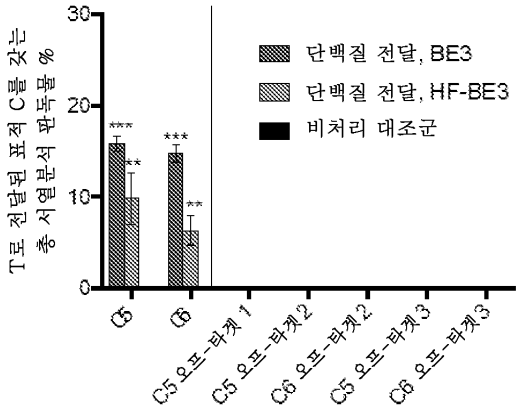
도면129a



도면129b



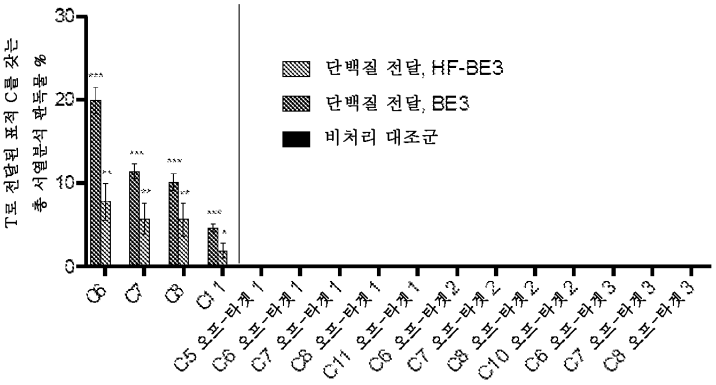
도면130a



EMX1 온-타겟 : GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

EMX1 오프-타겟 1: GAGTC$_5$TAAGCAGAAGAAGAAGAG
EMX1 오프-타겟 2: GAGGC$_5$C$_6$GAGCAGAAGAAAGACGG
EMX1 오프-타겟 3: GAGTC$_5$C$_6$TAGCAGGAGAAGAAGAG

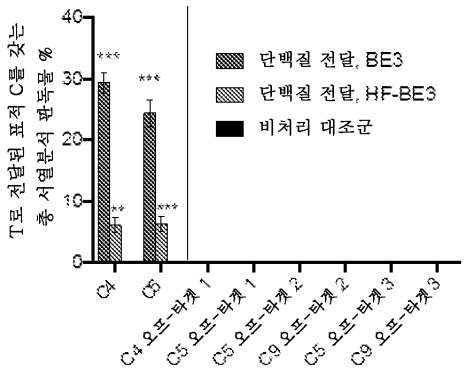
도면130b



FANCF 온-타겟: GGAATC$_5$C$_6$C$_7$C$_8$TTC$_{11}$TGCAGCACCTGG

FANCF 오프-타겟 1: GGAAC$_5$C$_6$C$_7$C$_8$GTC$_{11}$TGCAGCACCAGG
FANCF 오프-타겟 2: GGAGTC$_5$C$_6$C$_7$C$_8$TC$_{10}$C$_{11}$TACAGCACCAGG
FANCF 오프-타겟 3: AGAGGC$_5$C$_6$C$_7$C$_8$TC$_{11}$TGCAGCACCAGG

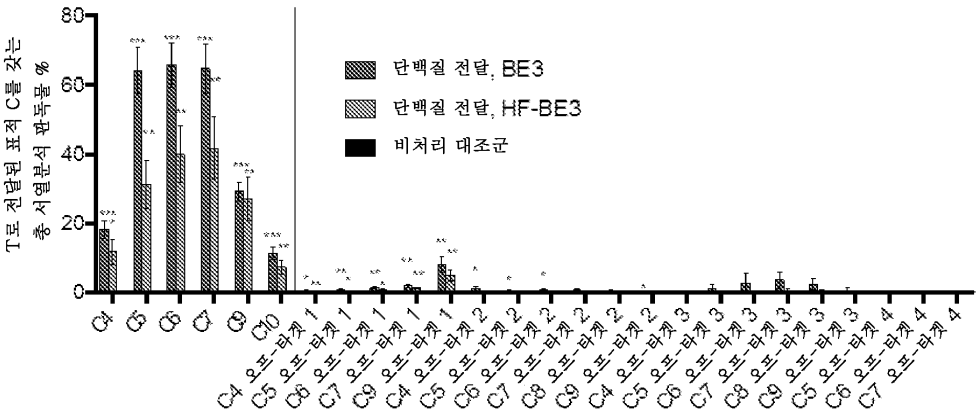
도면130c



HEK293 부위 3 온-타겟: GCCCCCCAGACTGAGCACGTGATGG

HEK293 부위 3 오프-타겟 1: CACCCCAGACTGAGCACGTGCTGG
HEK293 부위 3 오프-타겟 2: GACACCAGACTGGGCACGTGAGGG
HEK293 부위 3 오프-타겟 3: AGCTCCAGACTGAGCAAGTGAGGG

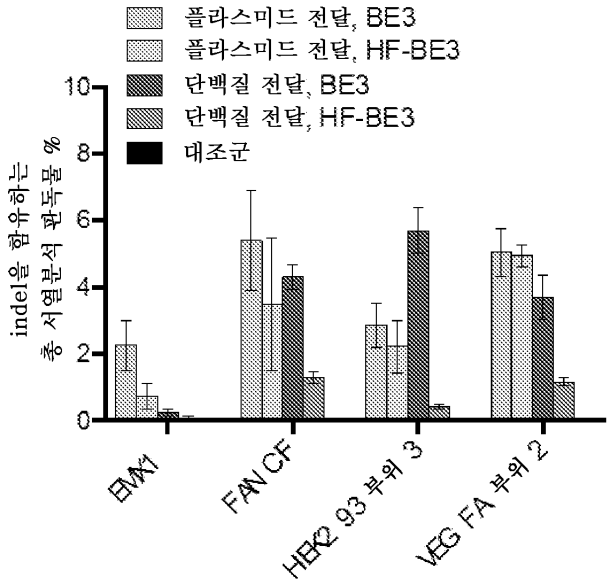
도면130d



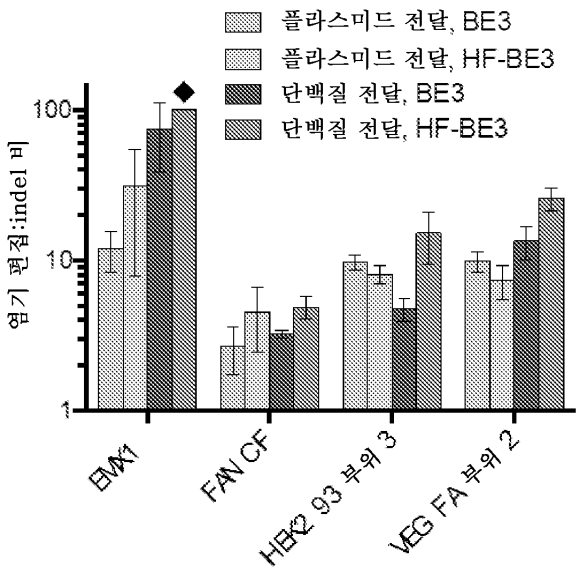
VEGFA 부위 2 온-타겟: GACCCCCCTCCACCCCGCCTCCGG

VEGFA 부위 2 오프-타겟 1: CTACCCCCTCTCCACCCCGCCTCCGG
VEGFA 부위 2 오프-타겟 2: ATTCCCCCCCCCCACCCCGCCTCAGG
VEGFA 부위 2 오프-타겟 3: ACACCCCCCCCCCACCCCGCCTCAGG
VEGFA 부위 2 오프-타겟 4: TGCCCCCCCCCCCCACCCCACCTCTGG

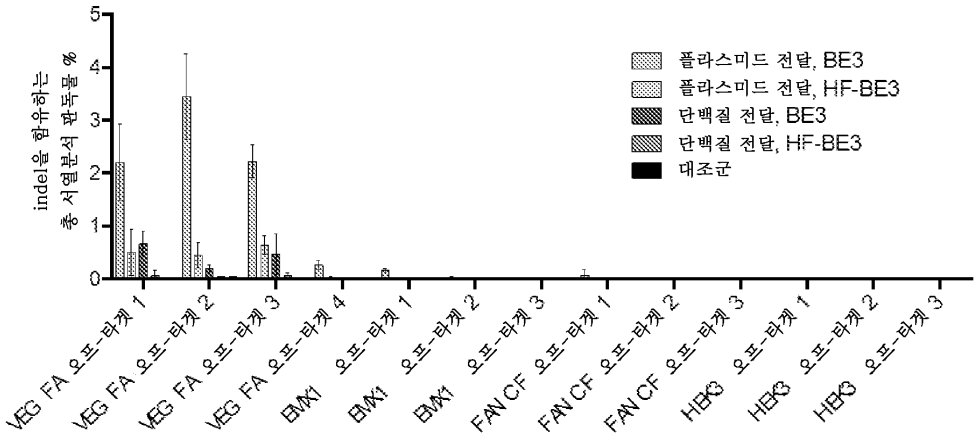
도면131a



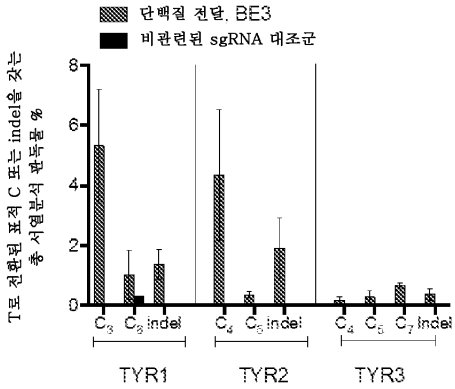
도면131b



도면131c

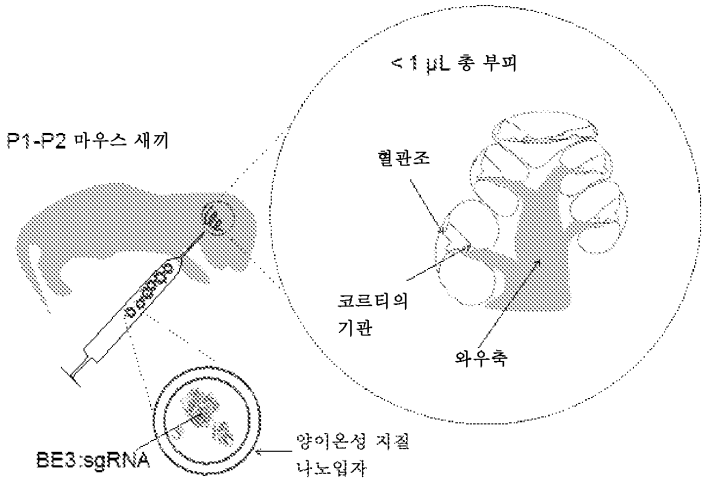


도면132a

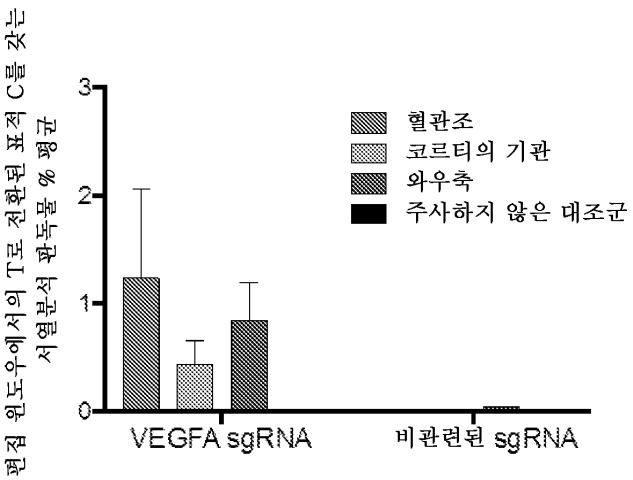


TYR1: GTC$_3$AGGTC$_6$GAGGGTTCTGTCAGG
TYR2: CTTC$_4$C$_5$AGGATGAGAACACAGAGG
TYR3: CAAC$_4$C$_5$AC$_7$TGCTCAAAGATGCTGG

도면132b

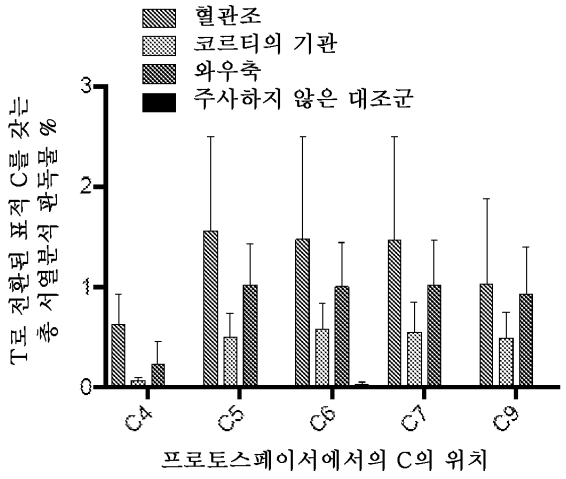


도면132c

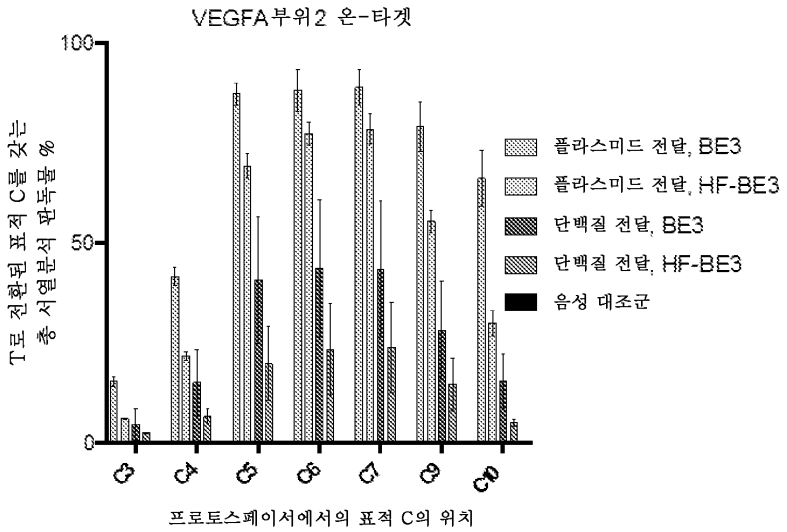


VEGFA 부위2: GAC$C_4C_5C_6C_7TC_9C_{10}$ACCCCGCCTCCGG

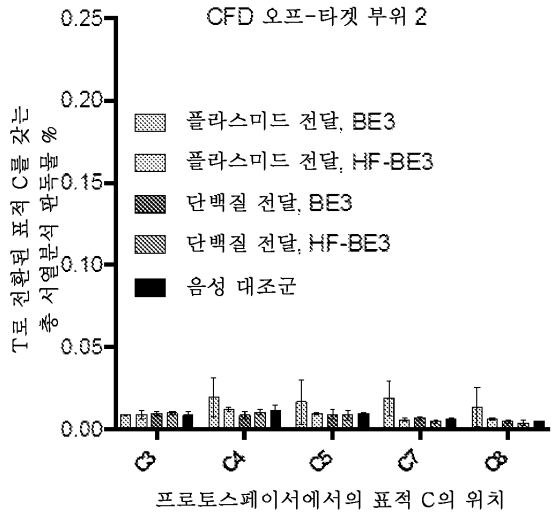
도면132d



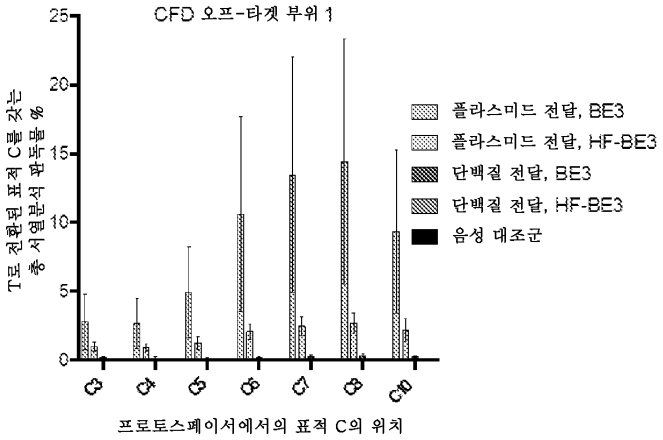
도면133a



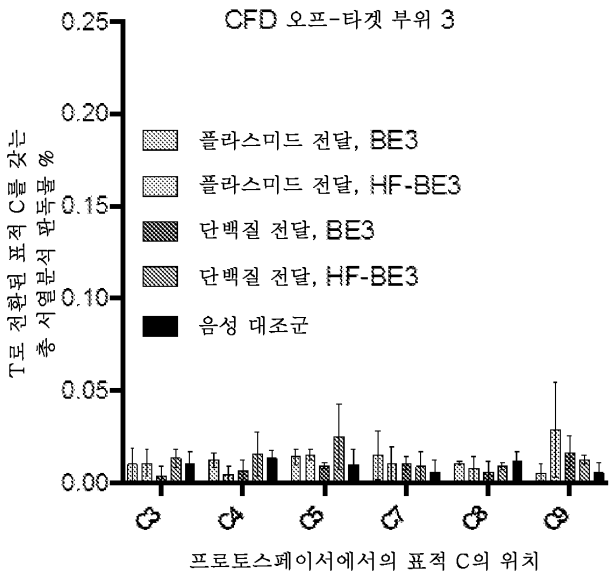
도면133b



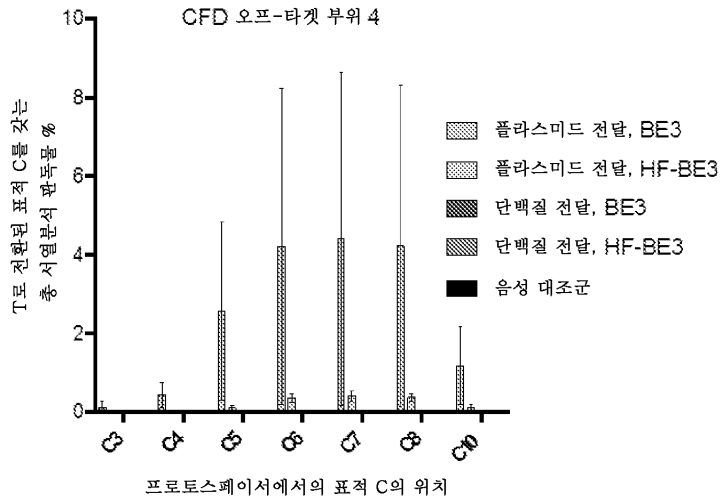
도면133c



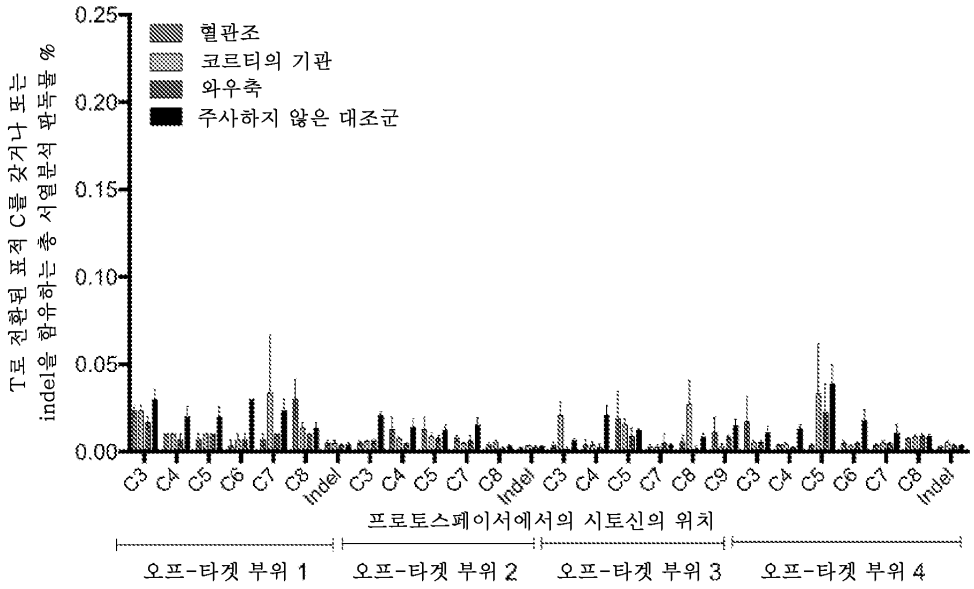
도면133d



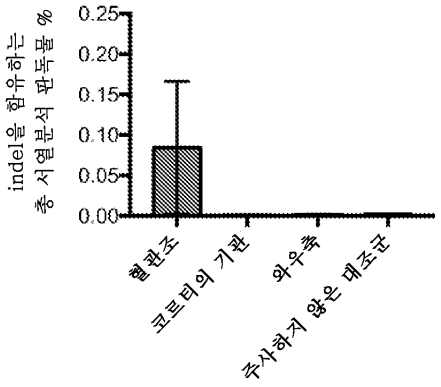
도면133e



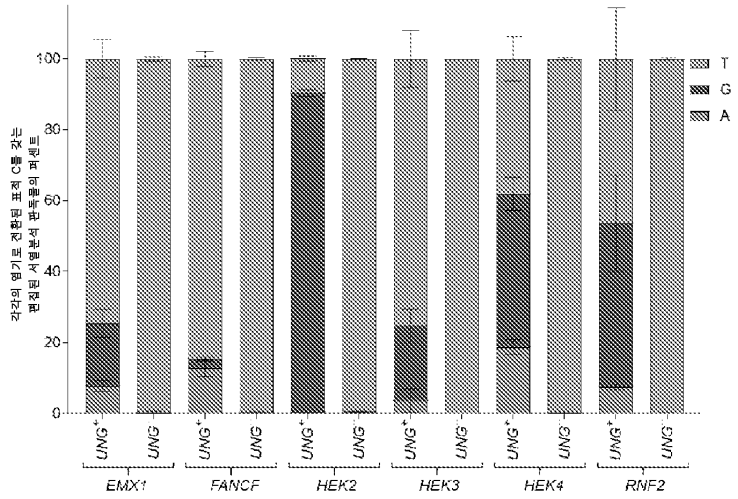
도면134a



도면134b



도면135a



도면135b

EMX1: GAGTC$_3$CGAGCAGAAGAAGAAGGG

FANCF: GGAATCC$_7$CTTCTGCAGCACCTGG

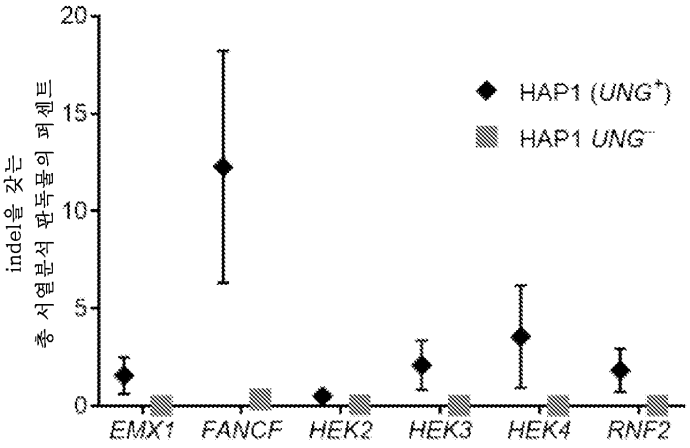
HEK2: GAACAC$_9$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

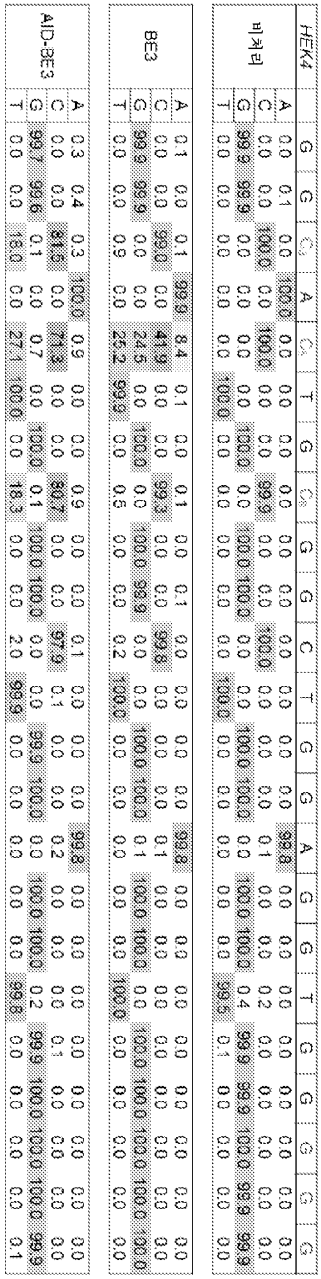
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_6$TTAGTCATTACCTGAGG

도면135c

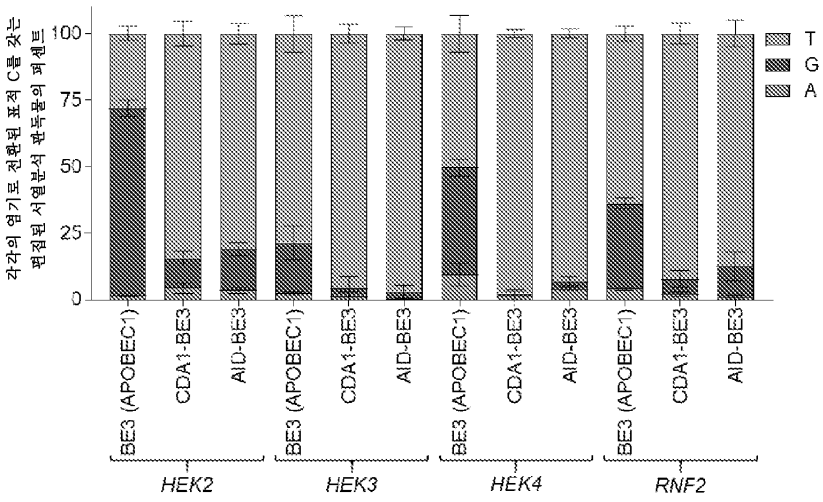


도면136a



HEK4		G	G	C	A	C	T	G	C	G	G	C	T	G	G	A	G	G	T	G	G	G	G	G	G
비처리	A	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.8	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	0.0	100.0	100.0	96.5	0.1	0.0	99.9	100.0	100.0	100.0
	T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.4	99.9	100.0	0.1	0.0	0.0	0.0
BE3	A	0.1	0.0	0.1	99.0	8.4	0.1	0.0	0.1	0.0	0.1	0.0	99.8	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	91.6	0.0	0.0	99.3	0.0	0.0	99.8	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.3	99.9	99.0	0.0	0.0	100.0	100.0	0.0	100.0	99.8	0.0	0.0	100.0	100.0	0.1	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T	0.0	0.0	0.9	0.9	0.0	0.0	0.0	0.5	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.9	0.0	0.0	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	81.8	0.0	71.8	0.0	0.0	80.7	0.0	0.0	97.9	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	100.0	100.0	0.1	100.0	100.0	0.0	99.9	100.0	100.0	0.2	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	0.0	0.0	2.0	99.9	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.1

도면136b



도면136c

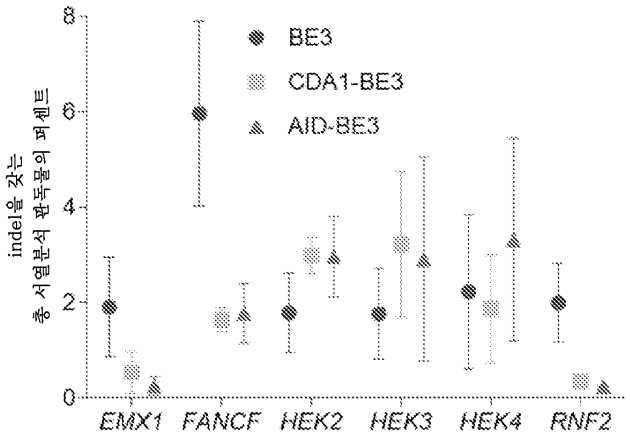
HEK2: GAACAC$_7$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_5$TTAGTCATTACCTGAGG

도면136d



도면137a

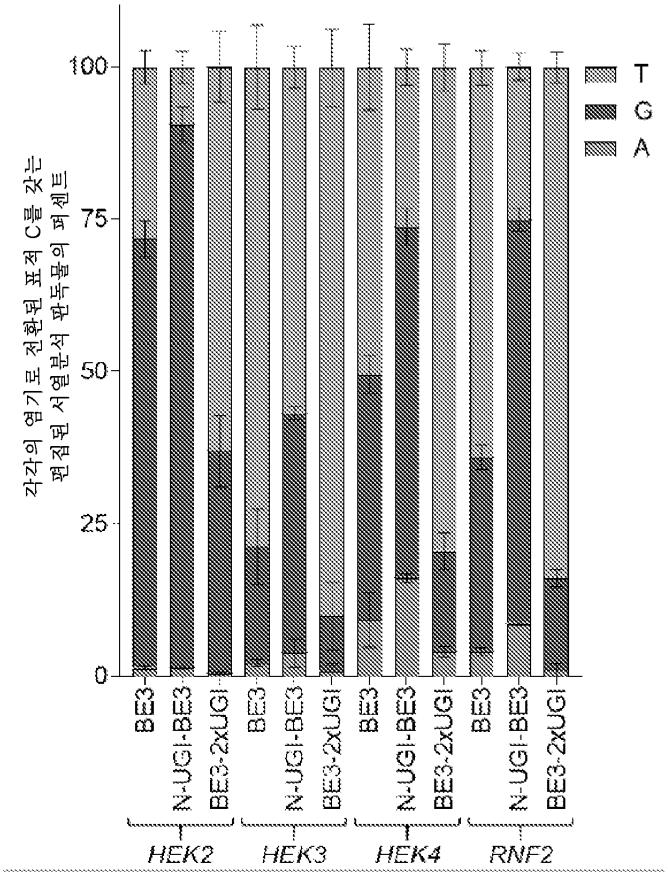
HEK2: GAAC$_4$AC$_3$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_2$AGACTGAGCACGTGATGG

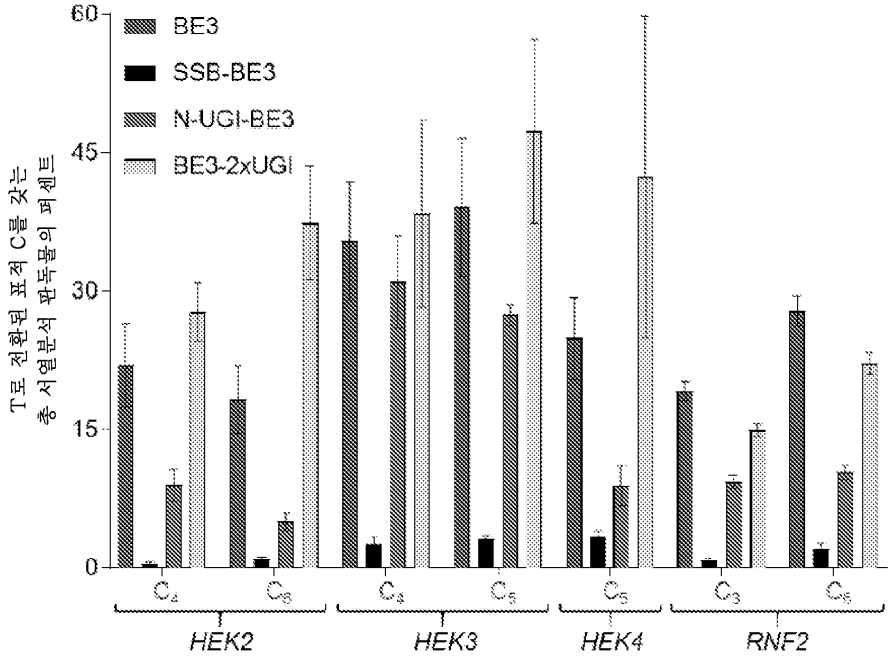
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_2$ATC$_6$TTAGTCATTACCTGAGG

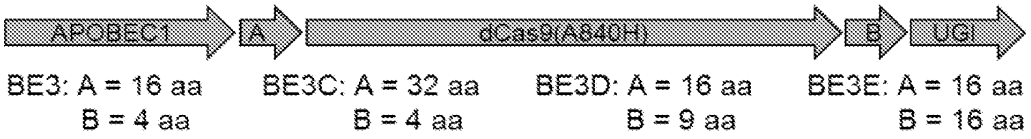
도면137b



도면137c



도면138a



도면138b

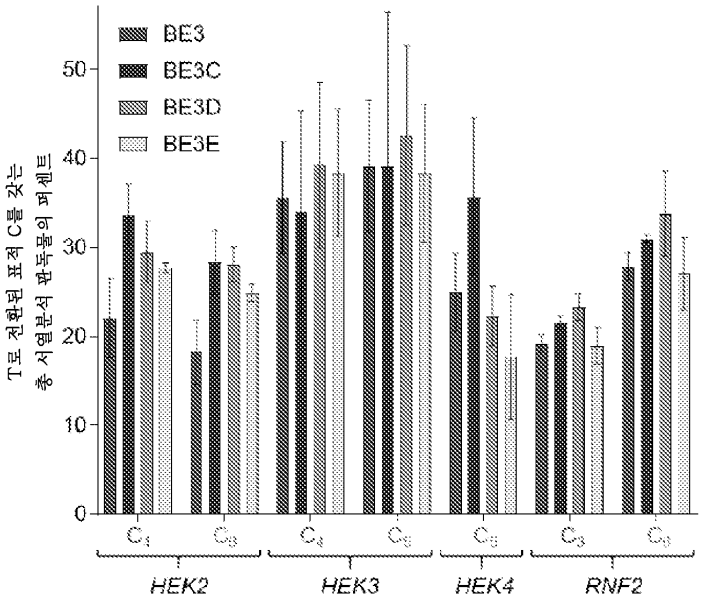
HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_6$AGACTGAGCACGTGATGG

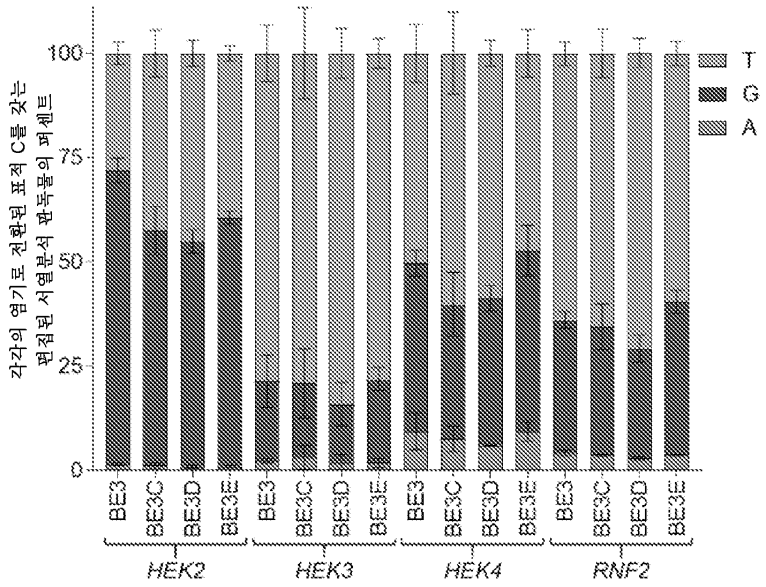
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_8$TTAGTCATTACCTGAGG

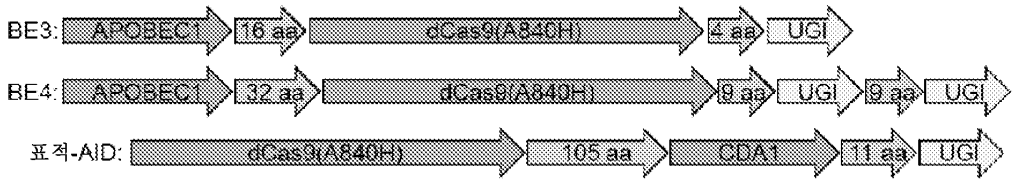
도면138c



도면138d



도면139a



도면139b

EMX1: GAGTC$_5$C$_6$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

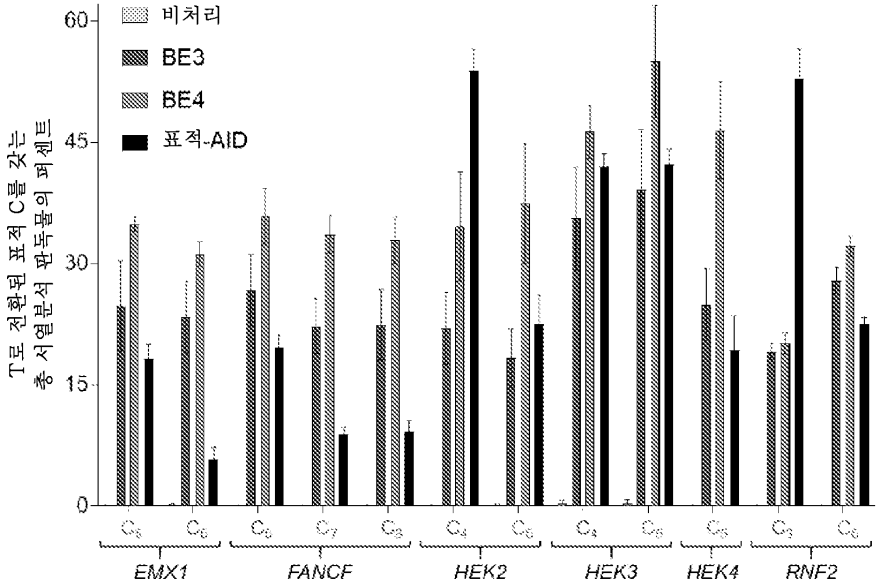
HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

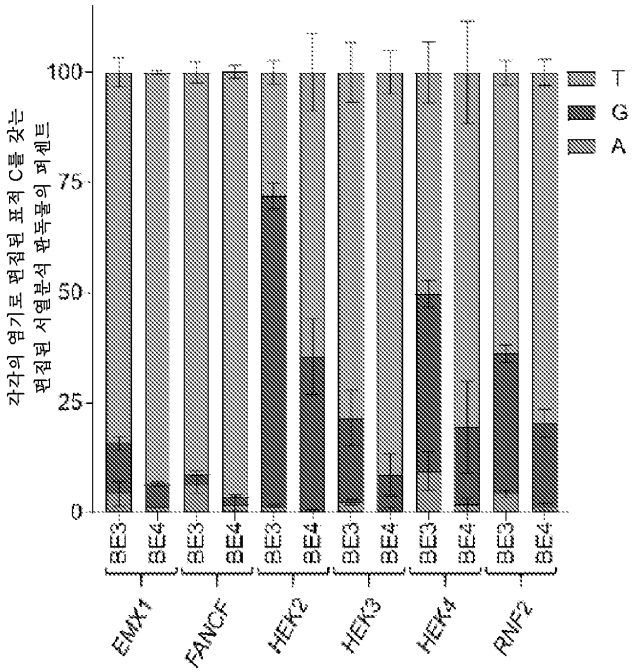
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

도면139c



도면139d



도면140a

EMX1: GAGTC$_3$C$_4$GAGCAGAAGAAGAAGGG

FANCF: GGAATC$_6$C$_7$C$_8$TTCTGCAGCACCTGG

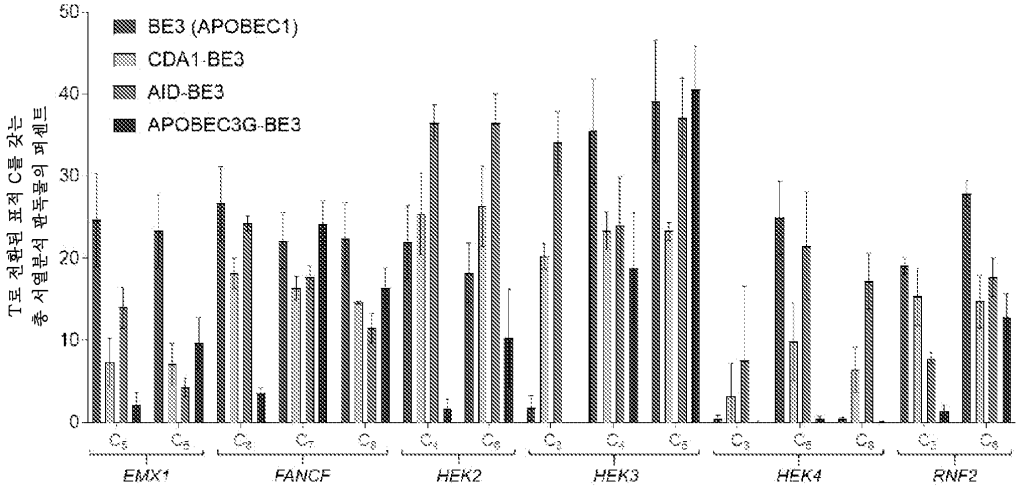
HEK2: GAAC$_3$AC$_5$AAAGCATAGACTGCGGG

HEK3: GGC$_3$C$_4$C$_5$AGACTGAGCACGTGATGG

HEK4: GGC$_3$AC$_5$TGC$_8$GGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_8$TTAGTCATTACCTGAGG

도면140b



도면140c

주어진 서열로 명확하게 편집된 판독물의 퍼센트

HEK2 프로토스페이서 서열	BE3	CDA1-BE3	AID-BE3
GAA$T_1$$T_2$$T_3$AAAGCATAGACTGCGGG	77 ± 1	88 ± 2	92 ± 2
GAA$C_1$$A_2$$T_3$AAAGCATAGACTGCGGG	23 ± 1	5 ± 1	4 ± 1
GAA$T_1$$A_2$$C_3$AAAGCATAGACTGCGGG	0 ± 0	7 ± 2	4 ± 2

도면141a

EMX1: GAGT$C_3$$C_4$GAGCAGAAGAAGAAGGG

FANCF: GGAAT$C_5$$C_6$$C_7$$C_8$TTCTGCAGCACCTGG

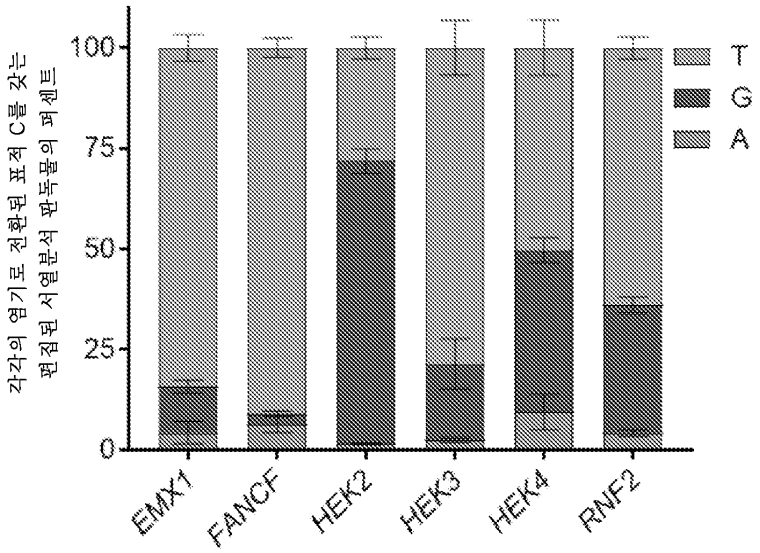
HEK2: GAAC_1AC_3AAAGCATAGACTGCGGG

HEK3: GGC$C_4$$C_5$AGACTGAGCACGTGATGG

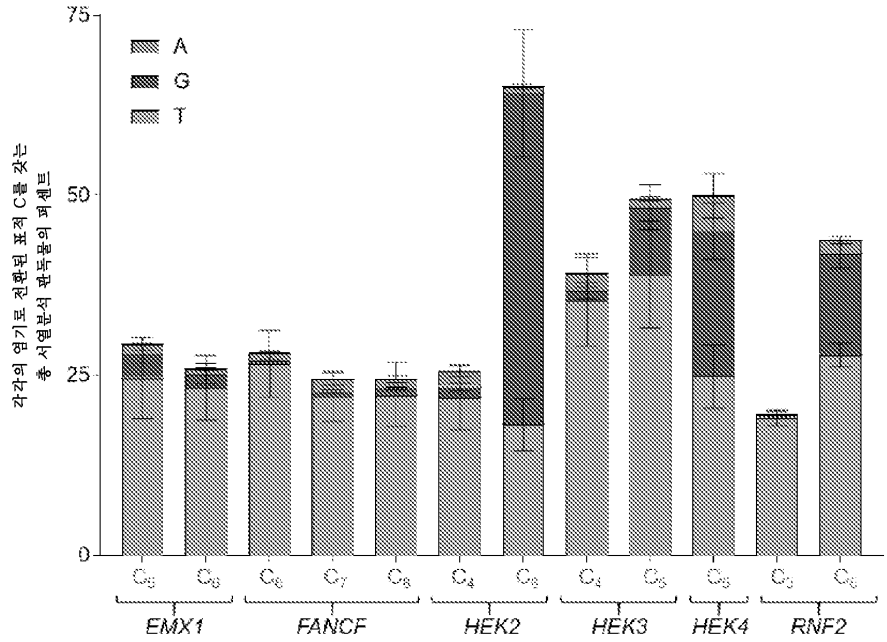
HEK4: GGCAC_5TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC_5TTAGTCATTACCTGAGG

도면141b



도면141c



도면142a

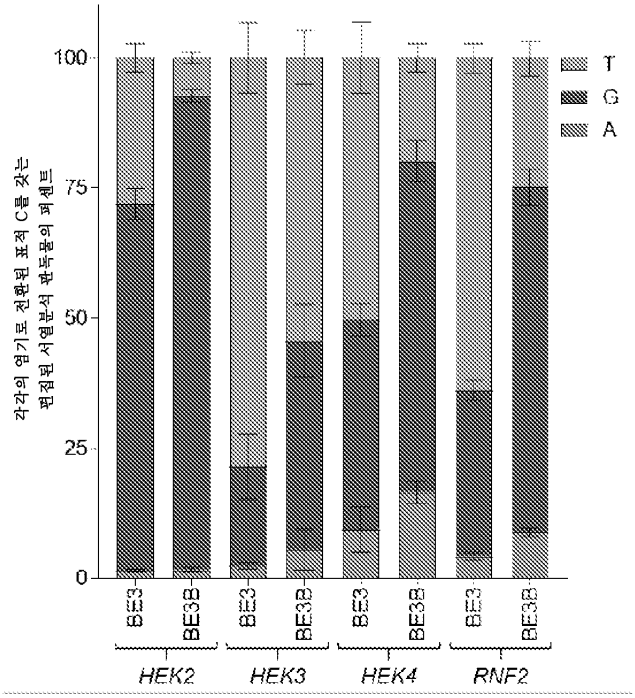
HEK2: GAAC$_4$AC$_6$AAAGCATAGACTGCGGG

HEK3: GGCC$_4$C$_5$AGACTGAGCACGTGATGG

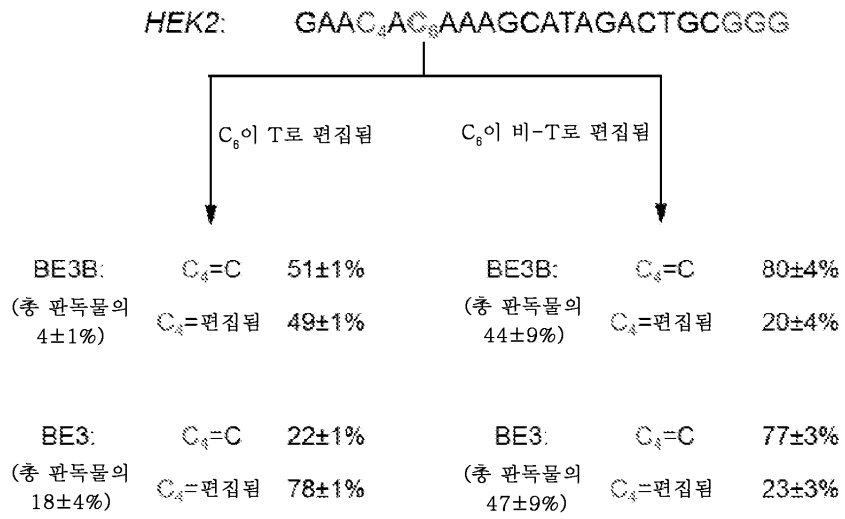
HEK4: GGCAC$_5$TGCGGCTGGAGGTCCGGG

RNF2: GTC$_3$ATC$_6$TTAGTCATTACCTGAGG

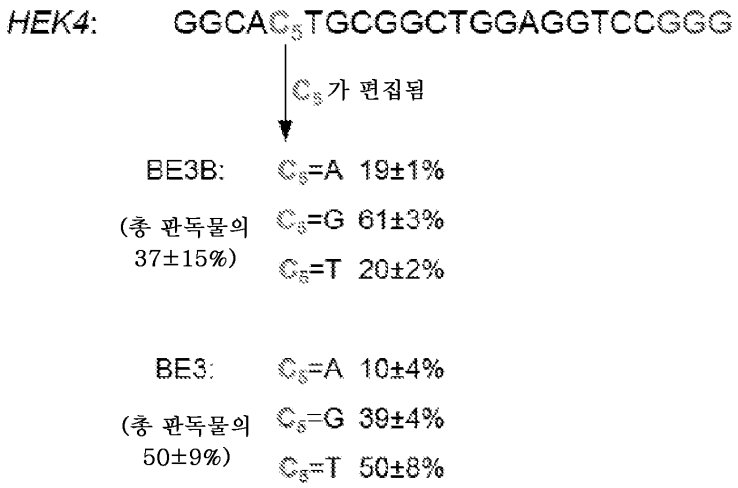
도면142b



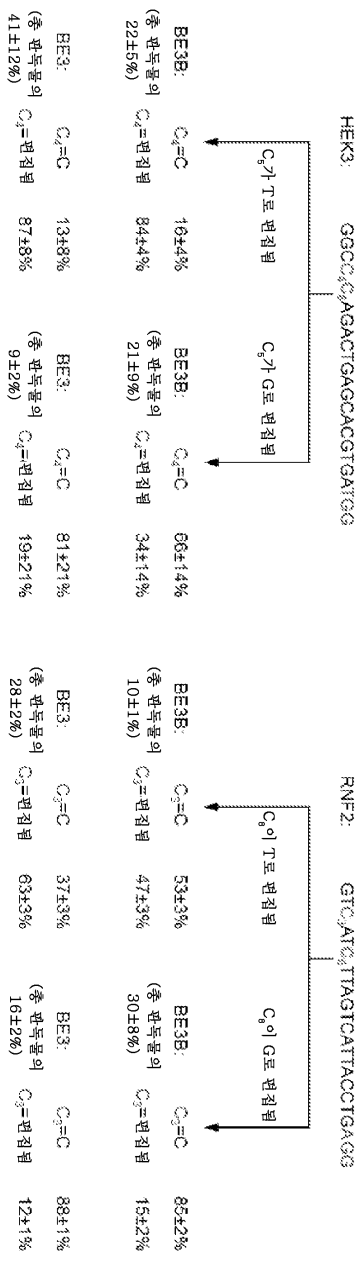
도면142c



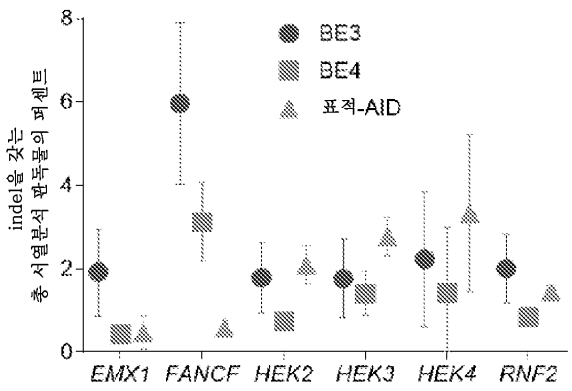
도면142d


HEK4: GGCAC8TGCGGCTGGAGGTCCGGG
C8 가 편집됨
BE3B:
(총 판독물의 37±15%)
C8=A 19±1%
C8=G 61±3%
C8=T 20±2%
BE3:
(총 판독물의 50±9%)
C8=A 10±4%
C8=G 39±4%
C8=T 50±8%

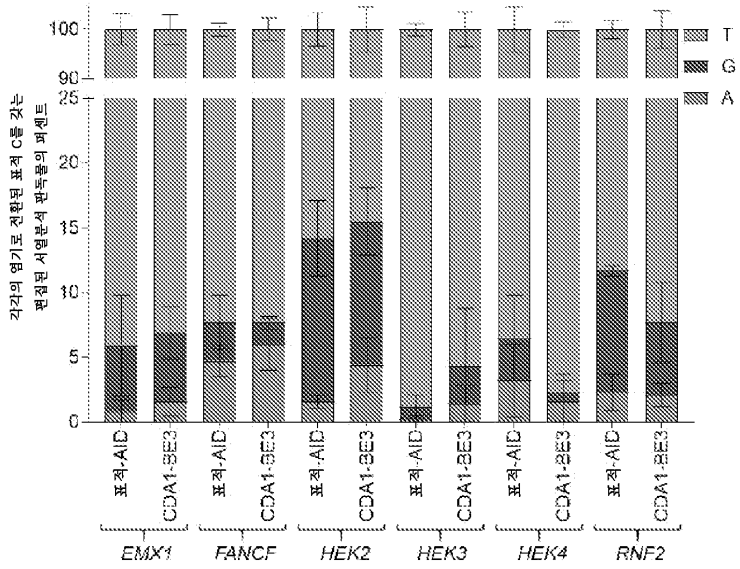
도면143



도면144a



도면144b



도면144c

EMX1: GAGTC$_5$CGAGCAGAAGAAGAAGGG

FANCF: GGAATCC$_7$CTTCTGCAGCACCTGG

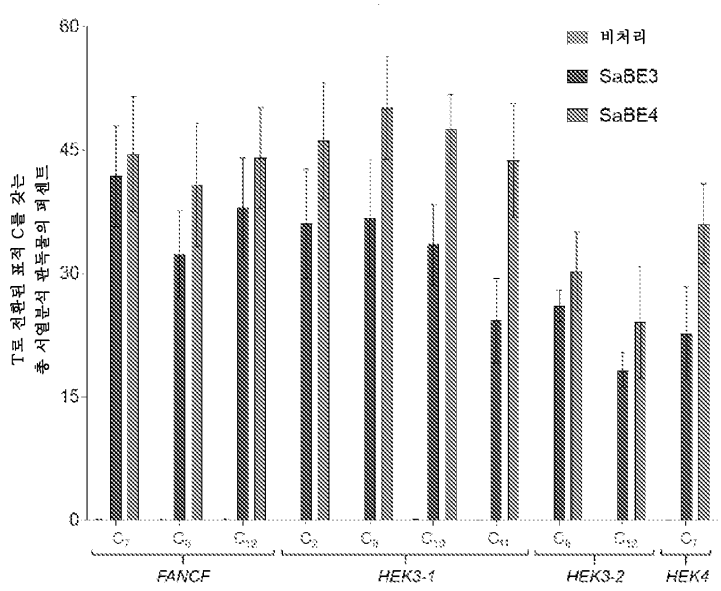
HEK2: GAACAC$_8$AAAGCATAGACTGCGGG

HEK3: GGCCC$_5$AGACTGAGCACGTGATGG

HEK4: GGCAC$_8$TGCGGCTGGAGGTCCGGG

RNF2: GTCATC$_8$TTAGTCATTACCTGAGG

도면145a



도면145b

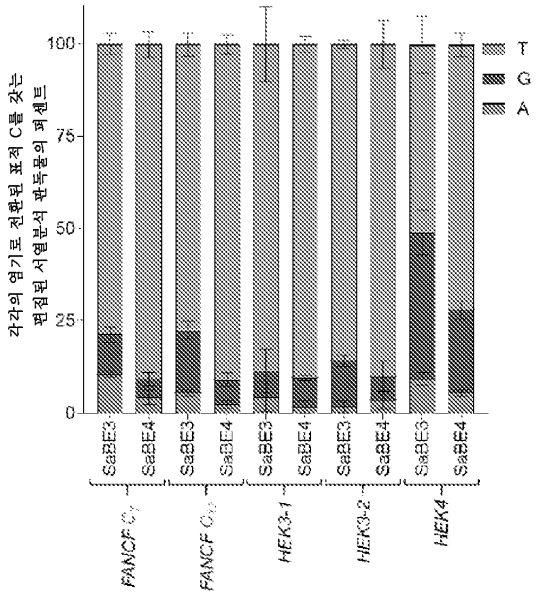
FANCF: GATGTTC$_7$C$_9$AATC$_{12}$AGTACGCAGAGAGT

HEK3-1: TC$_2$TGCTTC$_8$TC$_{10}$C$_{11}$AGCCCTGGCCTGGGT

HEK3-2: ACGTGCTC$_6$AGTC$_{12}$TGGGCCCCAAGGAT

HEK4: GTGGCAC$_7$TGCGGCTGGAGGTGGGGGT

도면145c

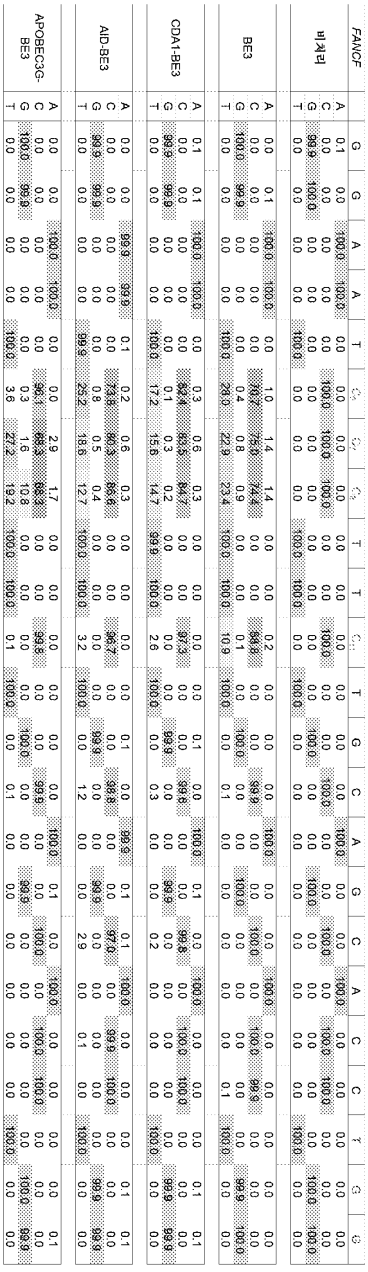


도면146

EMX1

		A	C	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T	G	A	G	T
비처리	A	0.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BE3	A	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	2.0	92.6	0.8	4.6	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	3.0	1.1	21.9	73.2	0.0	0.5	99.5	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	22.3	21.9	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	7.9	7.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0
CDA1-BE3	A	0.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	91.5	82.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	0.0	99.8	0.0	0.0	0.0	100.0	0.0	0.4	0.1	99.8	0.0	99.8	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	7.9	7.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0
AID-BE3	A	0.0	0.0	0.0	0.0	100.0	0.2	0.0	0.0	0.2	0.1	0.3	0.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	85.8	90.2	0.0	0.0	0.0	0.0	0.0	0.0	94.5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	99.8	0.0	0.0	0.0	100.0	0.0	0.7	0.0	99.7	0.0	99.7	0.0	0.0	0.0	99.6	0.0	0.0	0.0	0.3	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	13.4	4.6	0.0	0.0	0.0	0.0	0.0	100.0	5.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	88.9	88.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.2	2.3	99.9	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	1.0	8.9	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0

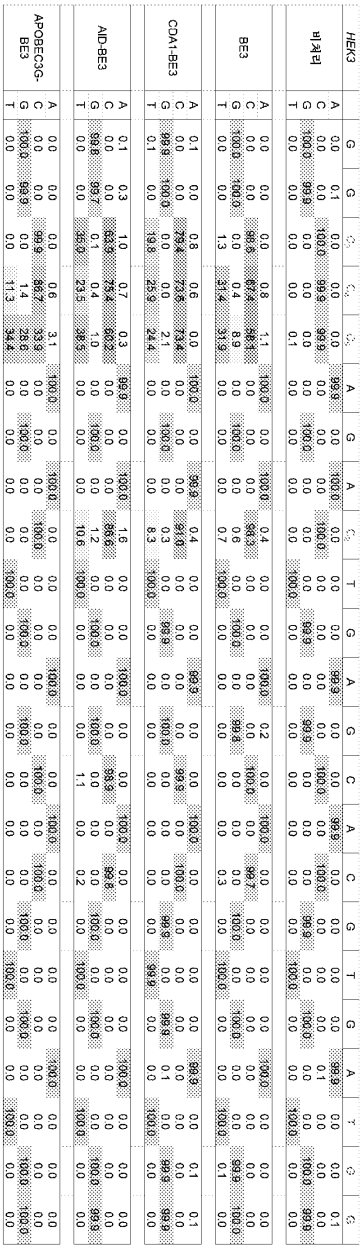
도면147



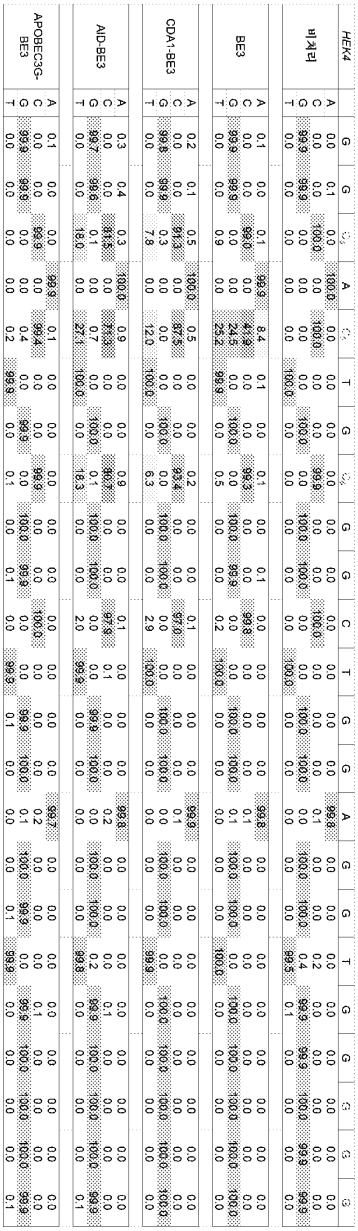
도면148

HEK2		G	A	A	C	A	C	A	A	A	G	C	A	T	A	G	A	C	T	G	C	G	G	G
비처리	A	0.0	100.0	100.0	0.0	100.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.1	99.9	0.0	0.0	100.0	0.0
	T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0
BE3	A	0.1	100.0	100.0	1.3	99.9	0.7	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	86.5	0.0	48.5	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	0.0	99.7	0.0	100.0	0.0	0.0	100.0	0.0	0.0
	G	99.9	0.0	0.0	0.9	0.1	34.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0
	T	0.0	0.0	0.0	12.2	0.0	14.8	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
CDA1-BE3	A	0.0	99.8	99.9	9.6	99.9	1.8	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	73.8	0.0	73.6	0.0	0.0	0.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.1	0.3	0.1	3.5	0.0	0.0	0.0	100.0	0.3	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.1	99.9	0.0	100.0	0.0	0.0	0.0
	T	0.0	0.0	0.0	19.8	0.0	21.2	0.0	0.0	0.0	0.0	8.8	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0
AID-BE3	A	0.0	99.8	99.9	6.9	99.9	1.2	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	54.5	0.0	52.4	0.0	0.0	0.0	0.0	83.8	0.0	0.0	0.0	0.0	0.0	99.7	0.0	100.0	0.0	0.0	99.8	0.0	0.0	0.0
	G	99.9	0.0	0.1	0.7	0.1	6.1	0.0	0.0	0.0	99.9	0.6	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.2	100.0	100.0	100.0
	T	0.1	0.1	0.0	37.9	0.0	40.3	0.0	0.0	0.0	0.1	15.5	0.0	100.0	0.0	0.0	0.0	0.2	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APOBEC3G-BE3	A	0.0	100.0	100.0	0.1	100.0	0.1	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	97.8	0.0	80.3	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.1	0.0	5.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0
	T	0.0	0.0	0.0	2.0	0.0	13.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0

도면149



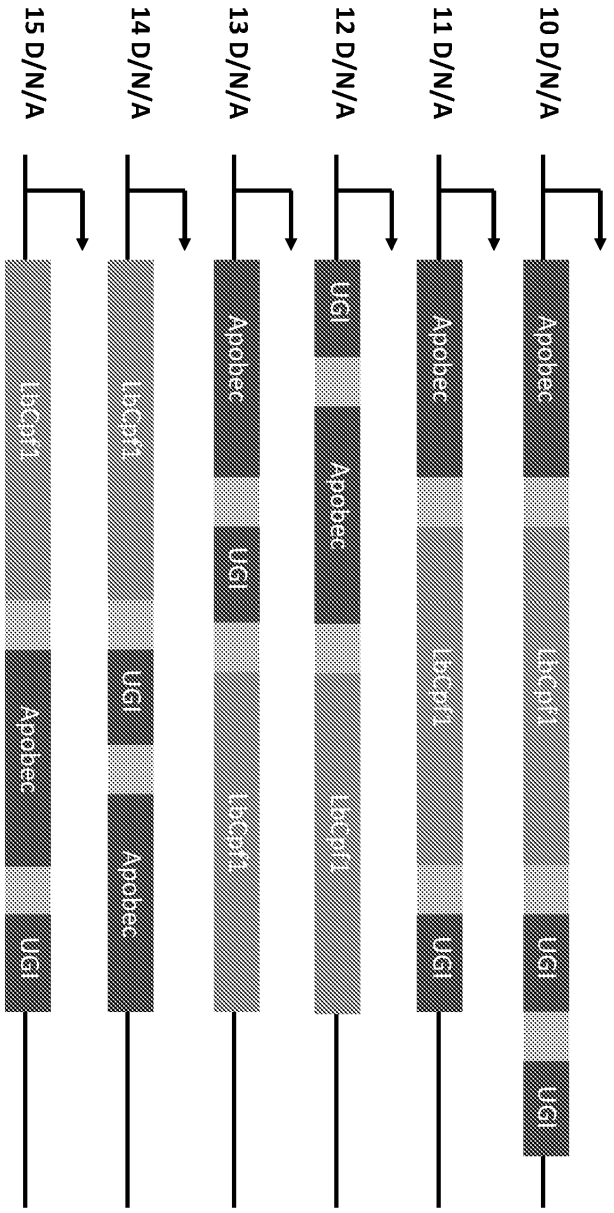
도면150



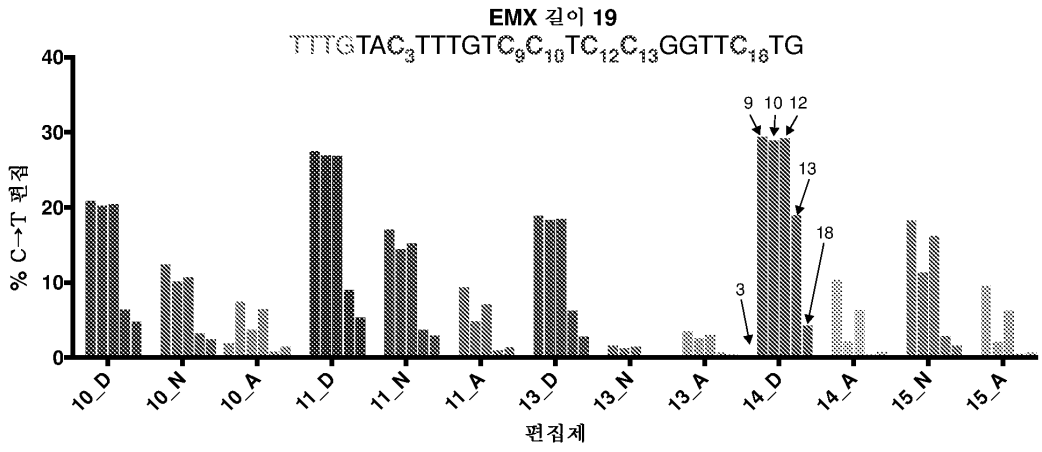
도면151

| RNF2 | | |

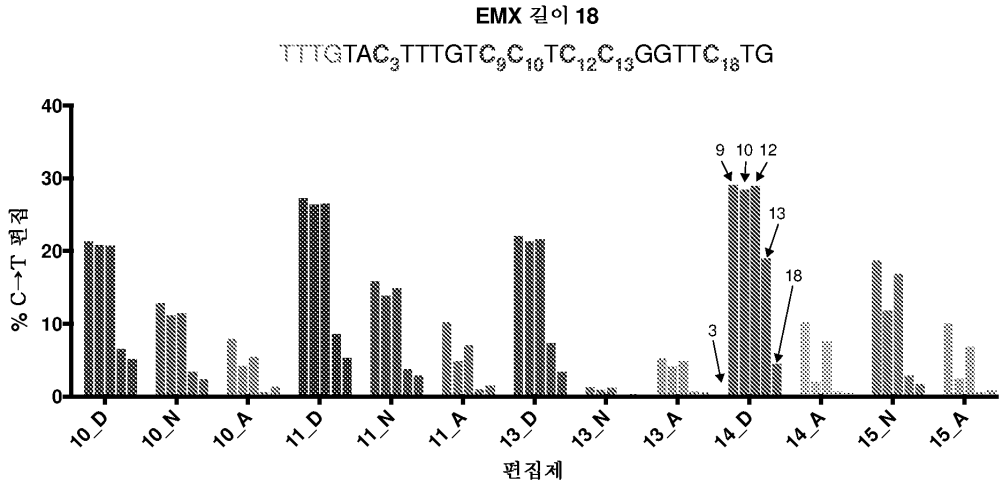
도면152


15 D/N/A
14 D/N/A
13 D/N/A
12 D/N/A
11 D/N/A
10 D/N/A
LbCpf1
Apobec
UGI

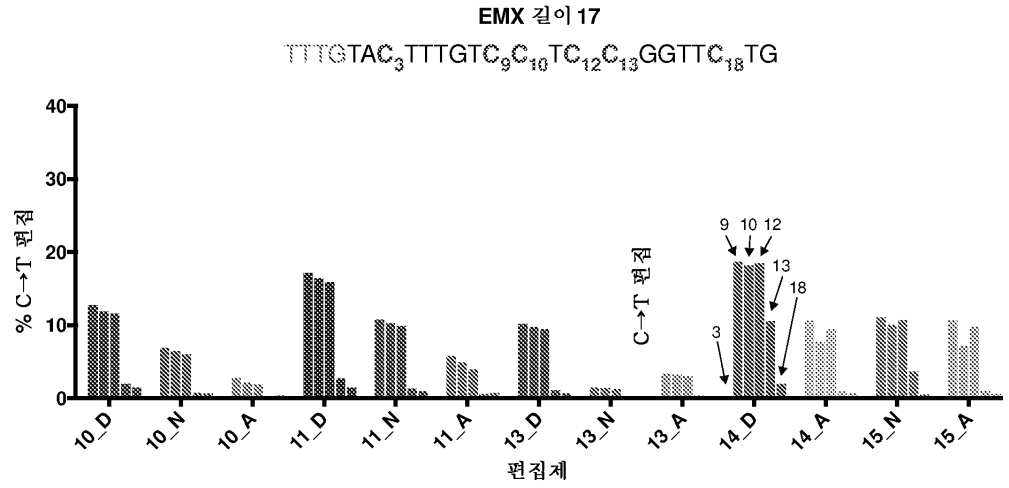
도면153


EMX 길이 19
TTTGTAC3TTTGTC9C10TC12C13GGTTC18TG
% C→T 편집
편집제
10_D
10_N
10_A
11_D
11_N
11_A
13_D
13_N
13_A
14_D
14_A
15_N
15_A
9 10 12
13
18
3

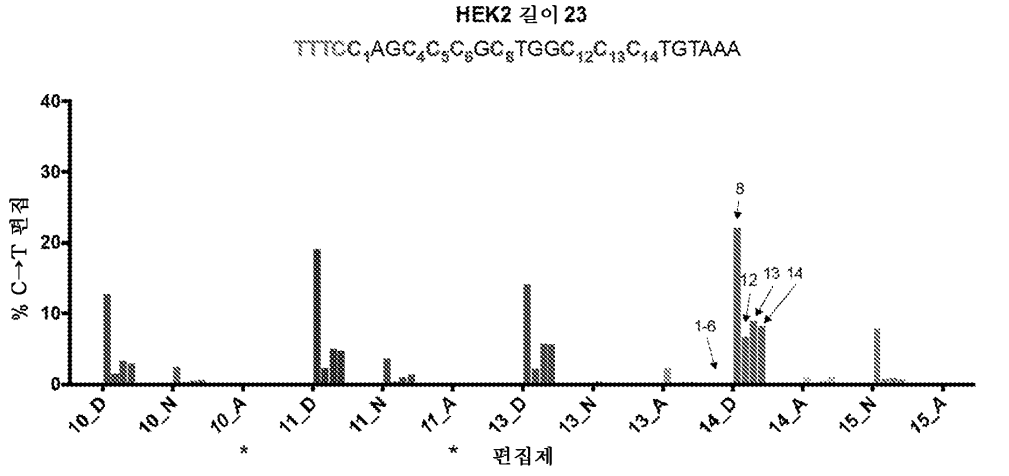
도면154



도면155

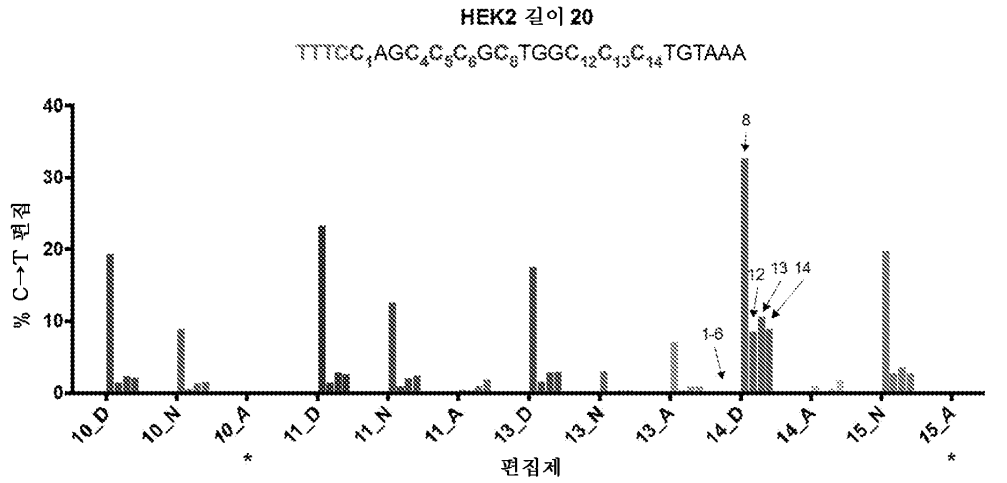


도면156



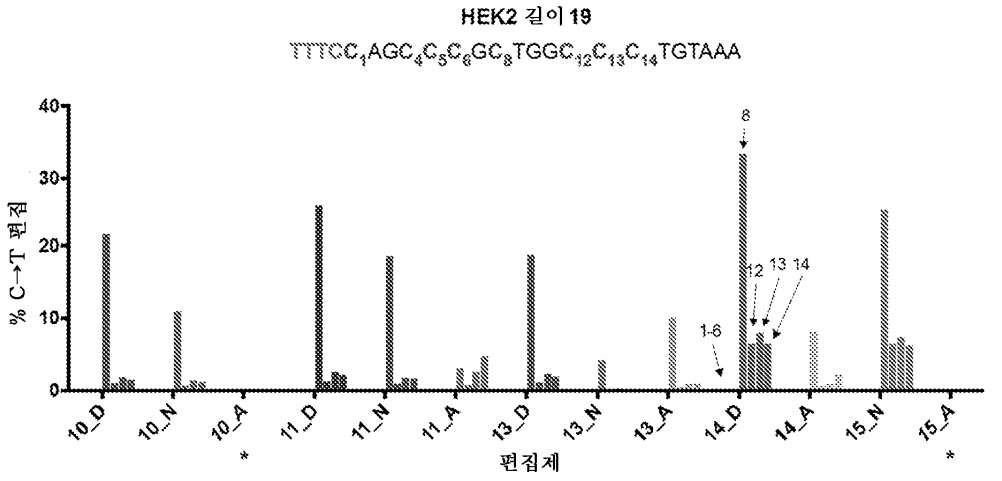
(*데이터는 수득되지 않음)

도면157



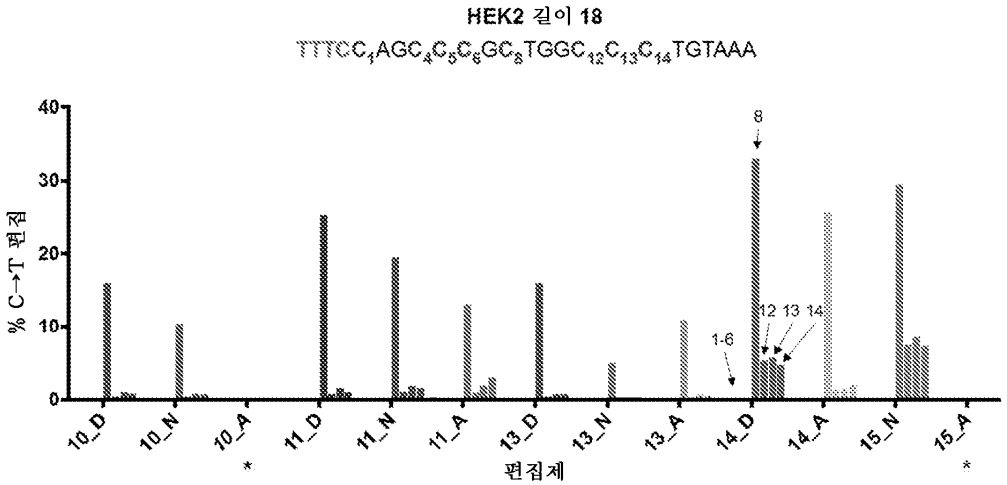
(*데이터는 수득되지 않음)

도면158



(*데이터는 수득되지 않음)

도면159



(*데이터는 수득되지 않음)

도면160

EMX 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.025	0.039	2.662	0.012	0.013	0.041	0.02	0.013	0.012	0.033	0.012	0.02	0.021
C9	20.899	13.15	10.456	27.535	18.921	13.072	19.043	1.591	3.638	29.39	14.271	20.95	13.924
C10	20.226	10.784	5.206	26.909	16.026	6.748	18.445	1.266	2.658	28.933	2.934	13.039	3.051
C12	20.453	11.385	9.095	26.849	16.893	9.983	18.589	1.437	3.127	29.165	8.695	18.597	9.07
C13	6.39	3.396	1.085	8.946	4.126	1.3	6.317	0.257	0.659	18.973	0.547	3.247	0.8
C18	4.757	2.603	2.013	5.304	3.233	1.914	2.805	0.235	0.367	4.302	0.992	1.848	0.994
indel (%)	0.028979	6.0862	28.9183	0.038623	9.8557	28.6068	0.94296	1.4573	3.78	0.024623	27.6872	13.0741	31.3295
C3	0.02499276	0.03662638	1.89219485	0.01199537	0.01171876	0.02927121	0.01981141	0.01281055	0.0115464	0.03299187	0.00867754	0.01738518	0.01442081
C9	20.8929437	12.3496647	7.43230255	27.5243652	17.056203	9.3325191	18.8634321	1.56781436	3.5004836	29.3827633	10.3197597	18.2109761	9.56168042
C10	20.2201387	10.1276642	3.7005133	26.8986069	14.4465255	4.81761314	18.271071	1.24755058	2.5575276	28.9258758	2.12165755	11.3342681	2.09513696
C12	20.4470729	10.6920861	6.46488062	26.8386301	15.2280766	7.12718316	18.4137132	1.4160586	3.0087994	29.1578187	6.28759796	16.1656096	6.22841435
C13	6.38814824	3.18931265	0.77123645	8.94254479	3.71935382	0.9281116	6.25743322	0.25325474	0.6340898	18.9683283	0.39555102	2.82248397	0.549364
C18	4.75562147	2.44457621	1.43087462	5.30195144	2.91436522	1.36646585	2.77854997	0.23157535	0.3531274	4.30094072	0.71734298	1.60639063	0.68258477

도면161

EMX 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.023	0.008	0.032	0.044	0.035	0.016	0.03	0.016	0.016	0.018	0.02	0.017	0.01
C9	21.333	13.606	10.333	27.294	17.216	14.646	22.315	1.297	5.485	29.094	14.566	21.847	13.66
C10	20.865	11.831	5.528	26.409	15.041	6.931	21.548	0.932	4.309	28.456	2.996	13.762	3.337
C12	20.773	12.165	7.147	26.572	16.178	10.157	21.839	1.202	5.059	28.937	10.88	19.625	9.359
C13	6.538	3.591	0.848	8.562	4.096	1.387	7.426	0.129	0.729	18.993	0.945	3.353	0.809
C18	5.174	2.54	1.788	5.313	3.146	2.196	3.443	0.243	0.565	4.517	0.647	2.03	1.154
indel (%)	0.039639	5.7886	23.1911	0.035862	7.9573	30.2671	1.1014	1.1121	4.6798	0.040521	29.8821	14.2217	26.5552
C3	0.02299088	0.00753691	0.02457885	0.04398422	0.03221495	0.01115726	0.02966958	0.01582206	0.01525123	0.01799271	0.01402358	0.01458231	0.00734448
C9	21.3245438	12.8184031	7.93666364	27.2842118	15.8460712	10.2130805	22.0692226	1.28257606	5.22831297	29.0822108	10.2133733	18.7399852	10.0325597
C10	20.8567293	11.1461507	4.24599599	26.3995292	13.8441425	4.8331873	21.3106703	0.92163523	4.10734742	28.4444693	2.10073228	11.8048096	2.45085298
C12	20.7647658	11.4608168	5.48953208	26.5624707	14.890668	7.08277065	21.5984653	1.18863256	4.82224892	28.9252744	7.62882752	16.8339914	6.87369883
C13	6.5354084	3.38313137	0.65133947	8.5589295	3.77006899	0.96719532	7.34421004	0.12756539	0.69488426	18.9853038	0.66261416	2.8761464	0.59416843
C18	5.17194908	2.39296956	1.37334313	5.31109465	2.89566334	1.53133448	3.4050788	0.2402976	0.53855913	4.51516967	0.45366281	1.74129949	0.84755299

도면162

EMX 17

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C3	0.016	0.014	0.044	0.012	0.018	0.012	0.013	0.016	0.03	0.014	0.011	0.016	0.005
C9	12.742	6.895	2.862	17.142	10.77	5.941	10.177	1.471	3.395	18.719	11.13	11.103	11.493
C10	11.901	6.427	2.199	16.385	10.277	5.064	9.741	1.393	3.216	18.206	8.1	10.086	7.747
C12	11.548	6.014	1.916	15.933	9.872	4.041	9.468	1.287	2.986	18.492	9.847	10.751	10.523
C13	1.993	0.768	0.252	2.696	1.357	0.599	1.155	0.083	0.285	10.532	0.971	3.674	1.082
C18	1.457	0.613	0.361	1.471	0.987	0.773	0.667	0.043	0.245	1.978	0.729	0.545	0.621
indel (%)	0.017245	0.052423	1.7787	0.025203	0.11499	2.7876	0.14938	0.019738	0.041079	0.04713	4.2923	0.37931	6.7995
C3	0.01599724	0.01399266	0.04321737	0.01199698	0.0179793	0.01166549	0.01298058	0.01599684	0.02998768	0.0139934	0.01052785	0.01593931	0.00466003
C9	12.7398026	6.89138543	2.81109361	17.1376797	10.7576156	5.77538868	10.1617976	1.47070965	3.39360537	18.7101777	10.652267	11.0608852	10.7115335
C10	11.8989477	6.42363077	2.15988639	16.3808705	10.2651825	4.92283594	9.72644889	1.39272505	3.2146789	18.1974195	7.7523237	10.0477428	7.22024274
C12	11.5460085	6.01084728	1.88192011	15.9289844	9.86064819	3.92835308	9.4538567	1.28674597	2.98477338	18.4832847	9.42433722	10.7102204	9.80748862
C13	1.99265631	0.76759739	0.24751768	2.69532053	1.35543959	0.58230228	1.15327466	0.08298362	0.28488292	10.5270363	0.92932177	3.66006415	1.00842941
C18	1.45674874	0.61267865	0.35457889	1.47062926	0.98586505	0.75145185	0.66600364	0.04299151	0.24489936	1.97706777	0.69770913	0.54293276	0.57877511

도면163

Hek2 23

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.042	0.021	0	0.039	0.084	0	0.045	0.07	0.04	0.045	0.056	0.057	0
C4	0.016	0.018	0	0.023	0.013	0	0.027	0.022	0.016	0.035	0.008	0.016	0
C5	0.031	0.018	0	0.037	0.052	0	0.048	0.021	0.044	0.047	0.157	0.017	0
C6	0.064	0.045	0	0.064	0.02	0	0.077	0.068	0.051	0.078	0.032	0.08	0
C8	12.758	2.444	0	19.198	3.66	0	14.134	0.423	2.22	22.158	0.977	7.763	0
C12	1.418	0.291	0	2.281	0.446	0	2.163	0.101	0.21	6.745	0.235	0.809	0
C13	3.282	0.603	0	5.021	1.121	0	5.713	0.042	0.308	8.933	0.599	0.921	0
C14	2.898	0.641	0	4.719	1.38	0	5.614	0.096	0.38	8.268	1.066	0.746	0
indel (%)	0.052343	11.0664		0.041898			2.2549	6.3138	13.9681	0.085777	59.2508	11.3052	
C1	0.04197802	0.01867606	0	0.03898366	0.084	0	0.0439853	0.06558034	0.03441276	0.0449614	0.0228195	0.05055604	0
C4	0.01599163	0.01600805	0	0.02299036	0.013	0	0.02639118	0.02061096	0.0137651	0.03496998	0.0032599	0.01419117	0
C5	0.03098377	0.01600805	0	0.0369845	0.052	0	0.04691765	0.0196741	0.03785404	0.04695968	0.0639762	0.01507812	0
C6	0.0639665	0.04002012	0	0.06397319	0.02	0	0.07526373	0.06370662	0.04387627	0.07793309	0.0130397	0.07095584	0
C8	12.7513221	2.17353718	0	19.1899564	3.66	0	13.8152924	0.3962963	1.90990818	22.1389935	0.3981196	6.88537732	0
C12	1.41725778	0.25879678	0	2.28004431	0.446	0	2.11422651	0.09462306	0.18066699	6.73921434	0.0957606	0.71754093	0
C13	3.2802821	0.53626961	0	5.0188963	1.121	0	5.58417756	0.0393482	0.26497825	8.92533754	0.2440877	0.81687911	0
C14	2.89648311	0.57006438	0	4.71702283	1.38	0	5.48740991	0.08993875	0.32692122	8.26090796	0.4343864	0.66166321	0

도면164

Hek2 20

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.024	0.051	0	0.036	0.116	0.126	0.035	0.022	0.042	0.032	0.02	0.133	0
C4	0.054	0.023	0	0.026	0.011	0.01	0.019	0.019	0.011	0.025	0.016	0.016	0
C5	0.033	0.017	0	0.051	0.054	0.048	0.046	0.02	0.029	0.035	0.075	0.043	0
C6	0.096	0.083	0	0.058	0.032	0.022	0.055	0.016	0.047	0.139	0.166	0.083	0
C8	19.339	9.265	0	23.342	13.199	1.173	17.974	3.225	7.353	32.768	2.835	20.891	0
C12	1.402	0.551	0	1.491	1.036	0.713	1.68	0.188	0.471	8.549	0.3	2.967	0
C13	2.377	1.409	0	2.974	2.174	2.814	3.043	0.356	0.895	10.654	1.81	3.81	0
C14	2.129	1.545	0	2.604	2.523	5.582	3.076	0.316	0.955	8.964	4.97	3.086	0
Indel (%)	0.10237	3.2815	24.3902	0.10375	4.3674	66.4779	1.8097	1.6132	3.0719	0.093678	63.4082	5.4543	
C1	0.02397543	0.04932644	0	0.03596265	0.11093382	0.04223785	0.03436661	0.0216451	0.0407098	0.03197002	0.00731836	0.12574578	0
C4	0.05394472	0.02224526	0	0.02597303	0.01051959	0.00335221	0.01865616	0.01869349	0.01066209	0.02497658	0.00585469	0.01512731	0
C5	0.03296622	0.01644215	0	0.05094709	0.0516416	0.01609061	0.04516754	0.01967736	0.02810915	0.03496721	0.02744385	0.04065465	0
C6	0.09590172	0.08027636	0	0.05793983	0.03060243	0.00737486	0.05400467	0.01574189	0.04555621	0.13886979	0.06074239	0.07847293	0
C8	19.3192027	8.96096903	0	23.3177827	12.6225469	0.39321423	17.6487245	3.1729743	7.12712319	32.7373036	1.03737753	19.7515422	0
C12	1.40056477	0.53291894	0	1.48945309	0.99075374	0.23901257	1.64959704	0.18496718	0.45653135	8.54099147	0.1097754	2.80517092	0
C13	2.37456667	1.36276367	0	2.97091448	2.07905272	0.94331189	2.98793083	0.35025701	0.8675065	10.6440195	0.66231158	3.60219117	0
C14	2.12682054	1.49430083	0	2.60129835	2.4128105	1.87120362	3.02033363	0.31090229	0.92566336	8.9556027	1.81861246	2.9176803	0

도면165

Hek2 19

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.072	0.031	0	0.039	0.062	0.406	0.023	0.019	0.025	0.06	0.224	0.04	0
C4	0.034	0.019	0	0.021	0.06	0.018	0.039	0.033	0.025	0.012	0.009	0.008	0
C5	0.03	0.037	0	0.044	0.045	0.027	0.018	0.031	0.033	0.024	0.021	0.012	0
C6	0.053	0.066	0	0.042	0.071	0.164	0.033	0.031	0.046	0.098	0.117	0.034	0
C8	21.611	11.001	0	25.556	18.67	6.38	18.944	4.272	10.143	33.441	15.227	25.017	0
C12	1.049	0.704	0	1.345	1.022	1.548	1.154	0.195	0.489	6.54	1.177	6.539	0
C13	1.914	1.39	0	2.629	1.819	5.23	2.301	0.253	0.881	8.065	1.823	7.452	0
C14	1.545	1.212	0	2.238	1.761	9.623	1.98	0.151	0.975	6.517	4.174	6.365	0
indel (%)	0.056704	0.17936		0.04541	0.43984	50.2032	0.61981	0.18691	0.22381	0.14694	45.816	0.52493	
C1	0.07195917	0.0309444	0	0.03898229	0.0617273	0.20217501	0.02285744	0.01896449	0.02494405	0.05991184	0.12137216	0.03979003	0
C4	0.03398072	0.01896592	0	0.02099046	0.0597361	0.00896342	0.03875827	0.03293832	0.02494405	0.01198237	0.00487656	0.00795801	0
C5	0.02998299	0.03693364	0	0.04398002	0.04480207	0.01344514	0.01788843	0.03094206	0.03292614	0.02396473	0.01137864	0.01193701	0
C6	0.05296995	0.06588162	0	0.04198093	0.07068771	0.08166675	0.03279546	0.03094206	0.04589705	0.097856	0.06339528	0.03382152	0
C8	21.5987457	10.9812686	0	25.544395	18.5878819	3.17703584	18.8265832	4.2640152	10.120299	33.3918618	8.25059768	24.8856783	0
C12	1.04840518	0.70273731	0	1.34438924	1.01750484	0.77085446	1.14684739	0.19463553	0.48790557	6.53039012	0.63774568	6.50467483	0
C13	1.91291469	1.3875069	0	2.62780617	1.81099931	2.60437264	2.28673817	0.25252712	0.87902823	8.05314929	0.98777432	7.41288222	0
C14	1.54412392	1.20982616	0	2.23698372	1.75325442	4.79194606	1.96772776	0.15071777	0.97281785	6.50742392	2.26164016	6.33158821	0

도면166

Hek2 18

	10_D	10_N	10_A	11_D	11_N	11_A	13_D	13_N	13_A	14_D	14_A	15_N	15_A
C1	0.037	0.027	0	0.035	0.041	0.315	0.039	0.034	0.034	0.028	0.125	0.074	0
C4	0.014	0.016	0	0.007	0.018	0.017	0.018	0.038	0.019	0.014	0.021	0.035	0
C5	0.052	0.018	0	0.051	0.034	0.045	0.033	0.039	0.022	0.014	0.088	0.035	0
C6	0.056	0.035	0	0.032	0.08	0.047	0.082	0.051	0.07	0.081	0.208	0.107	0
C8	16.035	10.427	0	25.298	19.509	14.808	15.991	5.043	10.926	33.064	28.458	29.702	0
C12	0.571	0.503	0	0.802	1.161	1.138	0.513	0.228	0.317	5.485	1.471	7.611	0
C13	1.12	0.903	0	1.607	1.869	2.299	0.803	0.315	0.709	5.902	1.741	8.752	0
C14	0.877	0.819	0	1.132	1.609	3.468	0.751	0.375	0.636	4.806	2.316	7.414	0
Indel (%)	0.076858	0.17892		0.048785	0.13947	11.5608	0.14868	0.1177	0.076621	0.23744	9.8678	0.50485	
C1	0.03697156	0.02695169	0	0.03498293	0.04094282	0.27858348	0.03894201	0.03395998	0.03397395	0.02793352	0.11266525	0.07362641	0
C4	0.01398924	0.01597137	0	0.00699659	0.0179749	0.01503466	0.01797324	0.03795527	0.01898544	0.01396676	0.01892776	0.0348233	0
C5	0.05196003	0.01796779	0	0.05097512	0.03395258	0.03979764	0.03295094	0.0389541	0.02198314	0.01396676	0.07931634	0.0348233	0
C6	0.05595696	0.03493738	0	0.03198439	0.07988842	0.04156642	0.08187808	0.05093997	0.06994637	0.08080767	0.18747498	0.10645981	0
C8	16.0226758	10.408344	0	25.2856584	19.4817908	13.0960767	15.9672246	5.03706439	10.9176284	32.9854928	25.6498215	29.5520495	0
C12	0.57056114	0.50210003	0	0.80160874	1.15938075	1.0064381	0.51223727	0.22773164	0.31675711	5.47197642	1.32584466	7.57257587	0
C13	1.11913919	0.90138435	0	1.60621603	1.86639331	2.03321721	0.80180611	0.31462925	0.70845676	5.88798629	1.5692016	8.70781553	0
C14	0.87632596	0.81753465	0	1.13144775	1.60675593	3.06707146	0.74988341	0.37455863	0.63551269	4.79458863	2.08746175	7.37657042	0

서 열 목 록

SEQUENCE LISTING

<110> PRESIDENT AND FELLOWS OF HARVARD COLLEGE

<120> NUCLEOBASE EDITORS COMPRISING NUCLEIC ACID PROGRAMMABLE DNA BINDING PROTEINS

<130> H0824.70257WO00

<140> Not Yet Assigned

<141> Concurrently Herewith

<150> 62/475,830

<151> 2017-03-23

<150> 62/490,587

<151> 2017-04-26

<150> 62/551,951

<151> 2017-08-30

<150> 62/511,934

<151> 2017-05-26

<160> 751

<170> PatentIn version 3.5

<210> 1

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 1

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> 2

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 2

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> 3
<211> 4212
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 3

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

caaaagggga 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> 4

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 4

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> 5
<211> 4107
<212> DNA
<213> Streptococcus pyogenes
<400> 5

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> 6
<211> 1368
<212> PRT
<213> Streptococcus pyogenes
<400> 6

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> 7

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 7

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> 8
<211> 1368
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 8

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> 9

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 9

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> 10

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 10

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 Ala 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> 11
<211> 1300
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 11
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 Ala 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> 12

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 12

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 Ala 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> 13

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 13

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 Ala 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 Ala 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> 14

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 14

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 Ala 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 Ala 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> 15

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 15

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 Ala 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 Ala 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> 16

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 16

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 Ala 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 Ala 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 Ala 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> 17
<211> 1307
<212> PRT
<213> Acidaminococcus species
<400> 17
Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr Leu
1 5 10 15
Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln Glu

20 25 30

Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys Glu

35 40 45

Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln Cys

50 55 60

Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile Asp

65 70 75 80

Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile Glu

85 90 95

Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly Arg

100 105 110

Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile Tyr

115 120 125

Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys Gln

130 135 140

Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg Ser

145 150 155 160

Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg Lys

165 170 175

Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg Ile

180 185 190

Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe Thr

195 200 205

Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn Val

210 215 220

Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val Phe

225 230 235 240

Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp Leu

245 250 255

Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu Lys

260 265 270

Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn Asp
275 280 285
Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro Leu
290 295 300

Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu Glu
305 310 315 320
Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr Lys
325 330 335
Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu Phe
340 345 350
Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His Lys
355 360 365

Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr Leu
370 375 380
Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys Ile
385 390 395 400
Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu Asp
405 410 415
Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser Glu
420 425 430

Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala Ala
435 440 445
Leu Asp Gln Pro Leu Pro Thr Thr Met 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> 18

<211> 1228

<212> PRT

<213> Lachnospiraceae species

<400> 18

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> 19

<211> 1307

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 19

Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr Leu

1 5 10 15

Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln Glu

20 25 30

Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys Glu

35 40 45

Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln Cys

50 55 60

Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile Asp

65 70 75 80

Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile Glu

85 90 95

Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly Arg

100 105 110

Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile Tyr

115 120 125

Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys Gln

130 135 140

Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg Ser

145 150 155 160

Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg Lys

165 170 175

Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg Ile

180 185 190

Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe Thr

195 200 205

Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn Val

210 215 220

Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val Phe

225 230 235 240

Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp Leu

245 250 255

Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu Lys

260 265 270

Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn Asp

275 280 285

Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro Leu

290 295 300

Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu Glu

305 310 315 320

Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr Lys

325 330 335

Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu Phe

340 345 350

Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His Lys
355 360 365
Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr Leu
370 375 380
Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys Ile
385 390 395 400

Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu Asp
405 410 415
Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser Glu
420 425 430
Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala Ala
435 440 445
Leu Asp Gln Pro Leu Pro Thr Thr Met 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 Ala
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> 20
<211> 1228

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 20

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 Ala 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> 21

<211> 1228

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 21

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 Ala 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> 22

<211> 1305

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 22

Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr Leu

1 5 10 15

Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln Glu

20 25 30

Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys Glu

35 40 45

Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln Cys

50 55 60

Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile Asp

65 70 75 80

Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile Glu

85 90 95

Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly Arg

100 105 110

Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile Tyr

115 120 125

Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys Gln

130 135 140

Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg Ser

145 150 155 160

Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg Lys

165 170 175

Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg Ile

180 185 190

Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe Thr

195 200 205

Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn Val

210 215 220

Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val Phe

225 230 235 240

Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp Leu

245 250 255

Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu Lys

260 265 270

Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn Asp

275 280 285

Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro Leu

290 295 300

Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu Glu

305 310 315 320

Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr Lys

325 330 335

Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu Phe

340 345 350

Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His Lys

355 360 365

Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr Leu
370 375 380
Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys Ile

385 390 395 400
Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu Asp
405 410 415
Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser Glu
420 425 430
Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala Ala
435 440 445
Leu Asp Gln Pro Leu Pro Thr Thr Met 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 Gly Gln Lys Gly Tyr Arg Glu Ala Leu Cys

660 665 670

Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr Lys Thr

675 680 685

Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr Lys Asp

690 695 700

Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His Ile Ser

705 710 715 720

Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu Thr Gly

725 730 735

Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys Gly His

740 745 750

His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu Phe Ser

755 760 765

Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln Ala Glu

770 775 780

Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His Arg Leu

785 790 795 800

Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr Pro Ile

805 810 815

Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His Arg Leu

820 825 830

Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn Val Ile

835 840 845

Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe Thr Ser

850 855 860

Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln Ala Ala
865 870 875 880
Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu Lys Glu
885 890 895
His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu

900 905 910
Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu Gln Arg
915 920 925
Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu Asp Asn
930 935 940
Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val Val Gly
945 950 955 960
Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile His Glu

965 970 975
Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu Glu Asn
980 985 990
Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu Lys Ala
995 1000 1005
Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Cys
1010 1015 1020
Leu Val Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly Val Leu

1025 1030 1035
Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala Lys Met
1040 1045 1050
Gly Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro Tyr Thr
1055 1060 1065
Ser Lys Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe Val Trp
1070 1075 1080
Lys Thr Ile Lys Asn His Glu Ser Arg Lys His Phe Leu Glu Gly

1085 1090 1095
Phe Asp Phe Leu His Tyr Asp Val Lys Thr Gly Asp Phe Ile Leu

1100 1105 1110

His Phe Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly Leu Pro

1115 1120 1125

Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn Glu Thr

1130 1135 1140

Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys Arg Ile

1145 1150 1155

Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr Arg Asp

1160 1165 1170

Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu Lys Gly

1175 1180 1185

Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu Leu Glu

1190 1195 1200

Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu Ile Arg

1205 1210 1215

Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly Glu Asp

1220 1225 1230

Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys Phe Asp

1235 1240 1245

Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp Ala Asn

1250 1255 1260

Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu Asn His

1265 1270 1275

Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile Ser Asn

1280 1285 1290

Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn

1295 1300 1305

<210> 23

<211> 1303

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 23

Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr Leu
1 5 10 15

Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln Glu
20 25 30

Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys Glu
35 40 45

Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln Cys
50 55 60

Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile Asp
65 70 75 80

Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile Glu
85 90 95

Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly Arg
100 105 110

Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile Tyr
115 120 125

Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys Gln
130 135 140

Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg Ser
145 150 155 160

Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg Lys
165 170 175

Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg Ile
180 185 190

Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe Thr
195 200 205

Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn Val
210 215 220

Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val Phe
225 230 235 240

Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp Leu
245 250 255
Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu Lys
260 265 270

Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn Asp
275 280 285
Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro Leu
290 295 300
Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu Glu
305 310 315 320
Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr Lys
325 330 335

Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu Phe
340 345 350
Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His Lys
355 360 365
Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr Leu
370 375 380
Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys Ile
385 390 395 400

Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu Asp
405 410 415
Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser Glu
420 425 430
Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala Ala
435 440 445
Leu Asp Gln Pro Leu Pro Thr Thr Met 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 Gly Lys Gly Tyr Arg Glu Ala Leu Cys Lys Trp

660 665 670

Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr Lys Thr Thr Ser

675 680 685

Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr Lys Asp Leu Gly

690 695 700

Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His Ile Ser Phe Gln

705 710 715 720

Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu Thr Gly Lys Leu

725 730 735

Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys Gly His His Gly
740 745 750

Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu Phe Ser Pro Glu
755 760 765

Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln Ala Glu Leu Phe
770 775 780

Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His Arg Leu Gly Glu
785 790 795 800

Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr Pro Ile Pro Asp
805 810 815

Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His Arg Leu Ser His
820 825 830

Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn Val Ile Thr Lys
835 840 845

Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe Thr Ser Asp Lys
850 855 860

Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln Ala Ala Asn Ser
865 870 875 880

Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu Lys Glu His Pro
885 890 895

Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr
900 905 910

Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu Gln Arg Ser Leu
915 920 925

Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu Asp Asn Arg Glu
930 935 940

Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val Val Gly Thr Ile
945 950 955 960

Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile His Glu Ile Val
965 970 975

Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu Glu Asn Leu Asn

980 985 990

Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu Lys Ala Val Tyr

995 1000 1005

Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Cys Leu Val

1010 1015 1020

Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly Val Leu Asn Pro

1025 1030 1035

Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala Lys Met Gly Thr

1040 1045 1050

Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro Tyr Thr Ser Lys

1055 1060 1065

Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe Val Trp Lys Thr

1070 1075 1080

Ile Lys Asn His Glu Ser Arg Lys His Phe Leu Glu Gly Phe Asp

1085 1090 1095

Phe Leu His Tyr Asp Val Lys Thr Gly Asp Phe Ile Leu His Phe

1100 1105 1110

Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly Leu Pro Gly Phe

1115 1120 1125

Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn Glu Thr Gln Phe

1130 1135 1140

Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys Arg Ile Val Pro

1145 1150 1155

Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr Arg Asp Leu Tyr

1160 1165 1170

Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu Lys Gly Ile Val

1175 1180 1185

Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu Leu Glu Asn Asp

1190 1195 1200

Asp Ser His Ala Ile Asp Thr Met Val Ala Leu Ile Arg Ser Val

1205 1210 1215

Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly Glu Asp Tyr Ile
1220 1225 1230
Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys Phe Asp Ser Arg
1235 1240 1245
Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp Ala Asn Gly Ala
1250 1255 1260
Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu Asn His Leu Lys
1265 1270 1275

Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile Ser Asn Gln Asp
1280 1285 1290
Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300

<210> 24
<211> 1301
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 24

Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr Leu
1 5 10 15
Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln Glu
20 25 30
Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys Glu
35 40 45
Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln Cys
50 55 60
Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile Asp
65 70 75 80
Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile Glu
85 90 95
Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly Arg
100 105 110

Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile Tyr
115 120 125
Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys Gln
130 135 140
Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg Ser

145 150 155 160
Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg Lys
165 170 175
Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg Ile
180 185 190
Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe Thr
195 200 205
Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn Val

210 215 220
Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val Phe
225 230 235 240
Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp Leu
245 250 255
Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu Lys
260 265 270
Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn Asp

275 280 285
Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro Leu
290 295 300
Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu Glu
305 310 315 320
Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr Lys
325 330 335
Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu Phe

340 345 350
Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His Lys

355 360 365

Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr Leu

370 375 380

Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys Ile

385 390 395 400

Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu Asp

405 410 415

Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser Glu

420 425 430

Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala Ala

435 440 445

Leu Asp Gln Pro Leu Pro Thr Thr Met 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 Gly Gly Tyr Arg Glu Ala Leu Cys Lys Trp Ile Asp

660 665 670
Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr Lys Thr Thr Ser Ile Asp
675 680 685
Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr Lys Asp Leu Gly Glu Tyr
690 695 700
Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His Ile Ser Phe Gln Arg Ile
705 710 715 720
Ala Glu Lys Glu Ile Met Asp Ala Val Glu Thr Gly Lys Leu Tyr Leu

725 730 735
Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys Gly His His Gly Lys Pro
740 745 750
Asn Leu His Thr Leu Tyr Trp Thr Gly Leu Phe Ser Pro Glu Asn Leu
755 760 765
Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln Ala Glu Leu Phe Tyr Arg
770 775 780
Pro Lys Ser Arg Met Lys Arg Met Ala His Arg Leu Gly Glu Lys Met

785 790 795 800
Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr Pro Ile Pro Asp Thr Leu
805 810 815
Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His Arg Leu Ser His Asp Leu
820 825 830
Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn Val Ile Thr Lys Glu Val
835 840 845
Ser His Glu Ile Ile Lys Asp Arg Arg Phe Thr Ser Asp Lys Phe Phe

850 855 860

Phe His Val Pro Ile Thr Leu Asn Tyr Gln Ala Ala Asn Ser Pro Ser

865 870 875 880

Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu Lys Glu His Pro Glu Thr

885 890 895

Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Ile Thr

900 905 910

Val Ile Asp Ser Thr Gly Lys Ile Leu Glu Gln Arg Ser Leu Asn Thr

915 920 925

Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu Asp Asn Arg Glu Lys Glu

930 935 940

Arg Val Ala Ala Arg Gln Ala Trp Ser Val Val Gly Thr Ile Lys Asp

945 950 955 960

Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile His Glu Ile Val Asp Leu

965 970 975

Met Ile His Tyr Gln Ala Val Val Val Leu Glu Asn Leu Asn Phe Gly

980 985 990

Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu Lys Ala Val Tyr Gln Gln

995 1000 1005

Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Cys Leu Val Leu Lys

1010 1015 1020

Asp Tyr Pro Ala Glu Lys Val Gly Gly Val Leu Asn Pro Tyr Gln

1025 1030 1035

Leu Thr Asp Gln Phe Thr Ser Phe Ala Lys Met Gly Thr Gln Ser

1040 1045 1050

Gly Phe Leu Phe Tyr Val Pro Ala Pro Tyr Thr Ser Lys Ile Asp

1055 1060 1065

Pro Leu Thr Gly Phe Val Asp Pro Phe Val Trp Lys Thr Ile Lys

1070 1075 1080

Asn His Glu Ser Arg Lys His Phe Leu Glu Gly Phe Asp Phe Leu

1085 1090 1095

His Tyr Asp Val Lys Thr Gly Asp Phe Ile Leu His Phe Lys Met

1100 1105 1110

Asn Arg Asn Leu Ser Phe Gln Arg Gly Leu Pro Gly Phe Met Pro

1115 1120 1125

Ala Trp Asp Ile Val Phe Glu Lys Asn Glu Thr Gln Phe Asp Ala

1130 1135 1140

Lys Gly Thr Pro Phe Ile Ala Gly Lys Arg Ile Val Pro Val Ile

1145 1150 1155

Glu Asn His Arg Phe Thr Gly Arg Tyr Arg Asp Leu Tyr Pro Ala

1160 1165 1170

Asn Glu Leu Ile Ala Leu Leu Glu Glu Lys Gly Ile Val Phe Arg

1175 1180 1185

Asp Gly Ser Asn Ile Leu Pro Lys Leu Leu Glu Asn Asp Asp Ser

1190 1195 1200

His Ala Ile Asp Thr Met Val Ala Leu Ile Arg Ser Val Leu Gln

1205 1210 1215

Met Arg Asn Ser Asn Ala Ala Thr Gly Glu Asp Tyr Ile Asn Ser

1220 1225 1230

Pro Val Arg Asp Leu Asn Gly Val Cys Phe Asp Ser Arg Phe Gln

1235 1240 1245

Asn Pro Glu Trp Pro Met Asp Ala Asp Ala Asn Gly Ala Tyr His

1250 1255 1260

Ile Ala Leu Lys Gly Gln Leu Leu Leu Asn His Leu Lys Glu Ser

1265 1270 1275

Lys Asp Leu Lys Leu Gln Asn Gly Ile Ser Asn Gln Asp Trp Leu

1280 1285 1290

Ala Tyr Ile Gln Glu Leu Arg Asn

1295 1300

<210> 25

<211> 887

<212> PRT

<213> Natronobacterium gregoryi

<400> 25

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> 26

<211> 1129

<212> PRT

<213> Alicyclobacillus acidoterrestris

<400> 26

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Asp Asp Met Pro

1 5 10 15

Glu Ile Arg Ala Gly Leu Trp Lys Leu His Lys Glu Val Asn Ala Gly

20 25 30

Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu

35 40 45

Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Asp Lys Thr

50 55 60

Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln

65 70 75 80

Val Glu Asn Gly His Arg Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu

85 90 95

Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly

100 105 110

Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu

115 120 125

Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn

130 135 140

Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu

145 150 155 160

Glu Glu Lys Glu Lys Ala Glu Thr Arg Lys Ser Ala Asp Arg Thr Ala

165 170 175

Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg

180 185 190

Val Tyr Thr Asp Ser Glu Met Ser Ser Val Glu Trp Lys Pro Leu Arg

195 200 205

Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala

210 215 220

Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Gln

225 230 235 240

Glu Tyr Ala Lys Leu Val Glu Gln Lys Asn Arg Phe Glu Gln Lys Asn

245 250 255

Phe Val Gly Gln Glu His Leu Val His Leu Val Asn Gln Leu Gln Gln

260 265 270

Asp Met Lys Glu Ala Ser Pro Gly Leu Glu Ser Lys Glu Gln Thr Ala

275 280 285

His Tyr Val Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu

290 295 300

Lys Trp Gly Lys Leu Ala Pro Asp Ala Pro Phe Asp Leu Tyr Asp Ala
305 310 315 320
Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His
325 330 335
Asp Leu Phe Ala Lys Leu Ala Glu Pro Glu Tyr Gln Ala Leu Trp Arg
340 345 350
Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Leu

355 360 365

Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp
370 375 380
Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn
385 390 395 400
Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Arg Arg His
405 410 415
Ala Ile Arg Phe His Lys Leu Leu Lys Val Glu Asn Gly Val Ala Arg

420 425 430

Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Glu Gln Leu Asp
435 440 445
Asn Leu Leu Pro Arg Asp Pro Asn Glu Pro Ile Ala Leu Tyr Phe Arg
450 455 460
Asp Tyr Gly Ala Glu Gln His Phe Thr Gly Glu Phe Gly Gly Ala Lys
465 470 475 480
Ile Gln Cys Arg Arg Asp Gln Leu Ala His Met His Arg Arg Arg Gly

485 490 495

Ala Arg Asp Val Tyr Leu Asn Val Ser Val Arg Val Gln Ser Gln Ser
500 505 510
Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu
515 520 525
Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp

530 535 540

Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu

545 550 555 560

Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser

565 570 575

Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro

580 585 590

Asn Ser Lys Gly Arg Val Pro Phe Phe Phe Pro Ile Lys Gly Asn Asp

595 600 605

Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly

610 615 620

Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg

625 630 635 640

Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val

645 650 655

Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys

660 665 670

Leu Ile Glu Gln Pro Val Asp Ala Ala Asn His Met Thr Pro Asp Trp

675 680 685

Arg Glu Ala Phe Glu Asn Glu Leu Gln Lys Leu Lys Ser Leu His Gly

690 695 700

Ile Cys Ser Asp Lys Glu Trp Met Asp Ala Val Tyr Glu Ser Val Arg

705 710 715 720

Arg Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp

725 730 735

Val Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Ala Lys Asp Val

740 745 750

Val Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr

755 760 765

Lys Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val

770 775 780

Ile Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His

785 790 795 800

Ile Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile

805 810 815

Ile Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Glu Arg Gly Lys

820 825 830

Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu

835 840 845

Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn

850 855 860

Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Ile

865 870 875 880

Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala

885 890 895

Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys

900 905 910

Arg Arg Val Pro Ala Arg Cys Thr Gln Glu His Asn Pro Glu Pro Phe

915 920 925

Pro Trp Trp Leu Asn Lys Phe Val Val Glu His Thr Leu Asp Ala Cys

930 935 940

Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Ile Phe

945 950 955 960

Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala

965 970 975

Asp Leu Asn Ala Ala Gln Asn Leu Gln Gln Arg Leu Trp Ser Asp Phe

980 985 990

Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp Gly

995 1000 1005

Glu Leu Val Leu Ile Pro Arg Leu Thr Gly Lys Arg Thr Ala Asp

1010 1015 1020

Ser Tyr Ser Asn Lys Val Phe Tyr Thr Asn Thr Gly Val Thr Tyr

1025 1030 1035

Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala Gln

1040 1045 1050

Glu Lys Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala Asp

1055 1060 1065

Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser Gly

1070 1075 1080

Ile Ile Asn Arg Gly Asn Trp Thr Arg Gln Lys Glu Phe Trp Ser

1085 1090 1095

Met Val Asn Gln Arg Ile Glu Gly Tyr Leu Val Lys Gln Ile Arg

1100 1105 1110

Ser Arg Val Pro Leu Gln Asp Ser Ala Cys Glu Asn Thr Gly Asp

1115 1120 1125

Ile

<210> 27

<211> 1389

<212> PRT

<213> Leptotrichia shahii

<400> 27

Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys

1 5 10 15

Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp

20 25 30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys

35 40 45

Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn

50 55 60

Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile

65 70 75 80

Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn

85 90 95

Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr
100 105 110

Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile
115 120 125
Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
130 135 140
Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145 150 155 160
Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
165 170 175

Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile
180 185 190
Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
195 200 205
Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
210 215 220
Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
225 230 235 240

Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu
245 250 255
Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser
260 265 270
Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
275 280 285
Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe
290 295 300

Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu
305 310 315 320
Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
325 330 335
Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp

340 345 350

Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys

355 360 365

Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile

370 375 380

Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile

385 390 395 400

Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys

405 410 415

Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr

420 425 430

Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys

435 440 445

Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn

450 455 460

Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr

465 470 475 480

Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp

485 490 495

Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu

500 505 510

Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu

515 520 525

Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile

530 535 540

Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys

545 550 555 560

Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp

565 570 575

Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile

580 585 590

Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
595 600 605
Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile
610 615 620

Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
625 630 635 640
Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
645 650 655
Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
660 665 670
Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
675 680 685

Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn
690 695 700
Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu
705 710 715 720
Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
725 730 735
Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
740 745 750

Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
755 760 765
Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
770 775 780
Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
785 790 795 800
Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
805 810 815

Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile
820 825 830
Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser

835 840 845

Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val

850 855 860

Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu

865 870 875 880

Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp

885 890 895

Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys

900 905 910

Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn

915 920 925

Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp

930 935 940

Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile

945 950 955 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu

965 970 975

Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys

980 985 990

Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser

995 1000 1005

Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys

1010 1015 1020

Tyr Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu

1025 1030 1035

Asn Glu Asn Lys Phe Gln Glu Ile Tyr Tyr Pro Lys Glu Arg Lys

1040 1045 1050

Asn Glu Leu Tyr Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly

1055 1060 1065

Asn Pro Asn Phe Asp Lys Ile Tyr Gly Leu Ile Ser Asn Asp Ile

1070 1075 1080

Lys Met Ala Asp Ala Lys Phe Leu Phe Asn Ile Asp Gly Lys Asn
1085 1090 1095
Ile Arg Lys Asn Lys Ile Ser Glu Ile Asp Ala Ile Leu Lys Asn
1100 1105 1110
Leu Asn Asp Lys Leu Asn Gly Tyr Ser Lys Glu Tyr Lys Glu Lys
1115 1120 1125

Tyr Ile Lys Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn
1130 1135 1140
Ile Gln Asn Lys Asn Tyr Lys Ser Phe Glu Lys Asp Tyr Asn Arg
1145 1150 1155
Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr
1160 1165 1170
Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu
1175 1180 1185

Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val
1190 1195 1200
Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn
1205 1210 1215
Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly
1220 1225 1230
Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser
1235 1240 1245

Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu
1250 1255 1260
Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg
1265 1270 1275
Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp
1280 1285 1290
Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser
1295 1300 1305

Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu

1310 1315 1320

Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys

1325 1330 1335

Lys Phe Lys Leu Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met

1340 1345 1350

Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser

1355 1360 1365

Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu

1370 1375 1380

Asn Thr Asn Asp Thr Leu

1385

<210> 28

<211> 1050

<212> PRT

<213> Marine metagenome

<400> 28

Met Arg Ser Asn Tyr His Gly Gly Arg Asn Ala Arg Gln Trp Arg Lys

1 5 10 15

Gln Ile Ser Gly Leu Ala Arg Arg Thr Lys Glu Thr Val Phe Thr Tyr

20 25 30

Lys Phe Pro Leu Glu Thr Asp Ala Ala Glu Ile Asp Phe Asp Lys Ala

35 40 45

Val Gln Thr Tyr Gly Ile Ala Glu Gly Val Gly His Gly Ser Leu Ile

50 55 60

Gly Leu Val Cys Ala Phe His Leu Ser Gly Phe Arg Leu Phe Ser Lys

65 70 75 80

Ala Gly Glu Ala Met Ala Phe Arg Asn Arg Ser Arg Tyr Pro Thr Asp

85 90 95

Ala Phe Ala Glu Lys Leu Ser Ala Ile Met Gly Ile Gln Leu Pro Thr

100 105 110

Leu Ser Pro Glu Gly Leu Asp Leu Ile Phe Gln Ser Pro Pro Arg Ser

115 120 125

```
Arg Asp Gly Ile Ala Pro Val Trp Ser Glu Asn Glu Val Arg Asn Arg
    130                 135                 140
Leu Tyr Thr Asn Trp Thr Gly Arg Gly Pro Ala Asn Lys Pro Asp Glu
145                 150                 155                 160

His Leu Leu Glu Ile Ala Gly Glu Ile Ala Lys Gln Val Phe Pro Lys
                165                 170                 175
Phe Gly Gly Trp Asp Asp Leu Ala Ser Asp Pro Asp Lys Ala Leu Ala
            180                 185                 190
Ala Ala Asp Lys Tyr Phe Gln Ser Gln Gly Asp Phe Pro Ser Ile Ala
        195                 200                 205
Ser Leu Pro Ala Ala Ile Met Leu Ser Pro Ala Asn Ser Thr Val Asp
    210                 215                 220

Phe Glu Gly Asp Tyr Ile Ala Ile Asp Pro Ala Ala Glu Thr Leu Leu
225                 230                 235                 240
His Gln Ala Val Ser Arg Cys Ala Ala Arg Leu Gly Arg Glu Arg Pro
                245                 250                 255
Asp Leu Asp Gln Asn Lys Gly Pro Phe Val Ser Ser Leu Gln Asp Ala
            260                 265                 270
Leu Val Ser Ser Gln Asn Asn Gly Leu Ser Trp Leu Phe Gly Val Gly
        275                 280                 285

Phe Gln His Trp Lys Glu Lys Ser Pro Lys Glu Leu Ile Asp Glu Tyr
    290                 295                 300
Lys Val Pro Ala Asp Gln His Gly Ala Val Thr Gln Val Lys Ser Phe
305                 310                 315                 320
Val Asp Ala Ile Pro Leu Asn Pro Leu Phe Asp Thr Thr His Tyr Gly
                325                 330                 335
Glu Phe Arg Ala Ser Val Ala Gly Lys Val Arg Ser Trp Val Ala Asn
            340                 345                 350

Tyr Trp Lys Arg Leu Leu Asp Leu Lys Ser Leu Leu Ala Thr Thr Glu
        355                 360                 365
Phe Thr Leu Pro Glu Ser Ile Ser Asp Pro Lys Ala Val Ser Leu Phe
```

370 375 380

Ser Gly Leu Leu Val Asp Pro Gln Gly Leu Lys Lys Val Ala Asp Ser

385 390 395 400

Leu Pro Ala Arg Leu Val Ser Ala Glu Glu Ala Ile Asp Arg Leu Met

405 410 415

Gly Val Gly Ile Pro Thr Ala Ala Asp Ile Ala Gln Val Glu Arg Val

420 425 430

Ala Asp Glu Ile Gly Ala Phe Ile Gly Gln Val Gln Gln Phe Asn Asn

435 440 445

Gln Val Lys Gln Lys Leu Glu Asn Leu Gln Asp Ala Asp Asp Glu Glu

450 455 460

Phe Leu Lys Gly Leu Lys Ile Glu Leu Pro Ser Gly Asp Lys Glu Pro

465 470 475 480

Pro Ala Ile Asn Arg Ile Ser Gly Gly Ala Pro Asp Ala Ala Ala Glu

485 490 495

Ile Ser Glu Leu Glu Glu Lys Leu Gln Arg Leu Leu Asp Ala Arg Ser

500 505 510

Glu His Phe Gln Thr Ile Ser Glu Trp Ala Glu Glu Asn Ala Val Thr

515 520 525

Leu Asp Pro Ile Ala Ala Met Val Glu Leu Glu Arg Leu Arg Leu Ala

530 535 540

Glu Arg Gly Ala Thr Gly Asp Pro Glu Glu Tyr Ala Leu Arg Leu Leu

545 550 555 560

Leu Gln Arg Ile Gly Arg Leu Ala Asn Arg Val Ser Pro Val Ser Ala

565 570 575

Gly Ser Ile Arg Glu Leu Leu Lys Pro Val Phe Met Glu Glu Arg Glu

580 585 590

Phe Asn Leu Phe Phe His Asn Arg Leu Gly Ser Leu Tyr Arg Ser Pro

595 600 605

Tyr Ser Thr Ser Arg His Gln Pro Phe Ser Ile Asp Val Gly Lys Ala

610 615 620

Lys Ala Ile Asp Trp Ile Ala Gly Leu Asp Gln Ile Ser Ser Asp Ile
625 630 635 640
Glu Lys Ala Leu Ser Gly Ala Gly Glu Ala Leu Gly Asp Gln Leu Arg
645 650 655
Asp Trp Ile Asn Leu Ala Gly Phe Ala Ile Ser Gln Arg Leu Arg Gly
660 665 670

Leu Pro Asp Thr Val Pro Asn Ala Leu Ala Gln Val Arg Cys Pro Asp
675 680 685
Asp Val Arg Ile Pro Pro Leu Leu Ala Met Leu Leu Glu Glu Asp Asp
690 695 700
Ile Ala Arg Asp Val Cys Leu Lys Ala Phe Asn Leu Tyr Val Ser Ala
705 710 715 720
Ile Asn Gly Cys Leu Phe Gly Ala Leu Arg Glu Gly Phe Ile Val Arg
725 730 735

Thr Arg Phe Gln Arg Ile Gly Thr Asp Gln Ile His Tyr Val Pro Lys
740 745 750
Asp Lys Ala Trp Glu Tyr Pro Asp Arg Leu Asn Thr Ala Lys Gly Pro
755 760 765
Ile Asn Ala Ala Val Ser Ser Asp Trp Ile Glu Lys Asp Gly Ala Val
770 775 780
Ile Lys Pro Val Glu Thr Val Arg Asn Leu Ser Ser Thr Gly Phe Ala
785 790 795 800

Gly Ala Gly Val Ser Glu Tyr Leu Val Gln Ala Pro His Asp Trp Tyr
805 810 815
Thr Pro Leu Asp Leu Arg Asp Val Ala His Leu Val Thr Gly Leu Pro
820 825 830
Val Glu Lys Asn Ile Thr Lys Leu Lys Arg Leu Thr Asn Arg Thr Ala
835 840 845
Phe Arg Met Val Gly Ala Ser Ser Phe Lys Thr His Leu Asp Ser Val
850 855 860

Leu Leu Ser Asp Lys Ile Lys Leu Gly Asp Phe Thr Ile Ile Ile Asp

865 870 875 880

Gln His Tyr Arg Gln Ser Val Thr Tyr Gly Gly Lys Val Lys Ile Ser

885 890 895

Tyr Glu Pro Glu Arg Leu Gln Val Glu Ala Ala Val Pro Val Val Asp

900 905 910

Thr Arg Asp Arg Thr Val Pro Glu Pro Asp Thr Leu Phe Asp His Ile

915 920 925

Val Ala Ile Asp Leu Gly Glu Arg Ser Val Gly Phe Ala Val Phe Asp

930 935 940

Ile Lys Ser Cys Leu Arg Thr Gly Glu Val Lys Pro Ile His Asp Asn

945 950 955 960

Asn Gly Asn Pro Val Val Gly Thr Val Ala Val Pro Ser Ile Arg Arg

965 970 975

Leu Met Lys Ala Val Arg Ser His Arg Arg Arg Arg Gln Pro Asn Gln

980 985 990

Lys Val Asn Gln Thr Tyr Ser Thr Ala Leu Gln Asn Tyr Arg Glu Asn

995 1000 1005

Val Ile Gly Asp Val Cys Asn Arg Ile Asp Thr Leu Met Glu Arg

1010 1015 1020

Tyr Asn Ala Phe Pro Val Leu Glu Phe Gln Ile Lys Asn Phe Gln

1025 1030 1035

Ala Gly Ala Lys Gln Leu Glu Ile Val Tyr Gly Ser

1040 1045 1050

<210> 29

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 29

Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln

1 5 10 15

Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu

20 25 30

Ile Asp Asp Glu Glu Leu Arg Asn Val Leu Asn Leu Ala Tyr Lys Ile
35 40 45
Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys

50 55 60
Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65 70 75 80
Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys
85 90 95
Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
100 105 110
Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys

115 120 125
Pro Glu Phe Tyr Glu Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
130 135 140
Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Ile Ala Ala
145 150 155 160
Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
165 170 175
Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr

180 185 190
Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
195 200 205
Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
210 215 220
Asn Pro Asn Val Ser Val Val Arg Ile Tyr Thr Ile Ser Asp Ala Val
225 230 235 240
Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr

245 250 255
Lys Leu Leu Glu Lys Arg Tyr Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270
Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile

275 280 285

Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu

290 295 300

Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn

305 310 315 320

Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val

325 330 335

Asn Gly Glu Leu Ile Arg Gly Glu Gly

340 345

<210> 30

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 30

Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln

1 5 10 15

Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu

20 25 30

Ile Asp Asp Glu Glu Leu Arg Asn Val Leu Asn Leu Ala Tyr Lys Ile

35 40 45

Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys

50 55 60

Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu

65 70 75 80

Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys

85 90 95

Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe

100 105 110

Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys

115 120 125

Pro Glu Phe Tyr Lys Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr

130 135 140

Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Met Ala Ala

145 150 155 160

Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly

165 170 175

Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr

180 185 190

Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu

195 200 205

Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr

210 215 220

Asn Pro Asn Val Ser Val Val Ser Ile Tyr Thr Ile Ser Asp Ala Val

225 230 235 240

Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr

245 250 255

Lys Leu Leu Glu Lys Arg Asp Leu Leu Ser Glu Arg Leu Glu Ala Ile

260 265 270

Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile

275 280 285

Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu

290 295 300

Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn

305 310 315 320

Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val

325 330 335

Asn Gly Glu Leu Ile Arg Gly Glu Gly

340 345

<210> 31

<211> 1210

<212> PRT

<213> uncultured Parcubacteria group bacterium

<400> 31

Met Ser Lys Arg His Pro Arg Ile Ser Gly Val Lys Gly Tyr Arg Leu
1 5 10 15
His Ala Gln Arg Leu Glu Tyr Thr Gly Lys Ser Gly Ala Met Arg Thr
20 25 30
Ile Lys Tyr Pro Leu Tyr Ser Ser Pro Ser Gly Gly Arg Thr Val Pro
35 40 45
Arg Glu Ile Val Ser Ala Ile Asn Asp Asp Tyr Val Gly Leu Tyr Gly

50 55 60
Leu Ser Asn Phe Asp Asp Leu Tyr Asn Ala Glu Lys Arg Asn Glu Glu
65 70 75 80
Lys Val Tyr Ser Val Leu Asp Phe Trp Tyr Asp Cys Val Gln Tyr Gly
85 90 95
Ala Val Phe Ser Tyr Thr Ala Pro Gly Leu Leu Lys Asn Val Ala Glu
100 105 110
Val Arg Gly Gly Ser Tyr Glu Leu Thr Lys Thr Leu Lys Gly Ser His

115 120 125
Leu Tyr Asp Glu Leu Gln Ile Asp Lys Val Ile Lys Phe Leu Asn Lys
130 135 140
Lys Glu Ile Ser Arg Ala Asn Gly Ser Leu Asp Lys Leu Lys Lys Asp
145 150 155 160
Ile Ile Asp Cys Phe Lys Ala Glu Tyr Arg Glu Arg His Lys Asp Gln
165 170 175
Cys Asn Lys Leu Ala Asp Asp Ile Lys Asn Ala Lys Lys Asp Ala Gly

180 185 190
Ala Ser Leu Gly Glu Arg Gln Lys Lys Leu Phe Arg Asp Phe Phe Gly
195 200 205
Ile Ser Glu Gln Ser Glu Asn Asp Lys Pro Ser Phe Thr Asn Pro Leu
210 215 220
Asn Leu Thr Cys Cys Leu Leu Pro Phe Asp Thr Val Asn Asn Asn Arg
225 230 235 240
Asn Arg Gly Glu Val Leu Phe Asn Lys Leu Lys Glu Tyr Ala Gln Lys

245 250 255

Leu Asp Lys Asn Glu Gly Ser Leu Glu Met Trp Glu Tyr Ile Gly Ile

260 265 270

Gly Asn Ser Gly Thr Ala Phe Ser Asn Phe Leu Gly Glu Gly Phe Leu

275 280 285

Gly Arg Leu Arg Glu Asn Lys Ile Thr Glu Leu Lys Lys Ala Met Met

290 295 300

Asp Ile Thr Asp Ala Trp Arg Gly Gln Glu Gln Glu Glu Glu Leu Glu

305 310 315 320

Lys Arg Leu Arg Ile Leu Ala Ala Leu Thr Ile Lys Leu Arg Glu Pro

325 330 335

Lys Phe Asp Asn His Trp Gly Gly Tyr Arg Ser Asp Ile Asn Gly Lys

340 345 350

Leu Ser Ser Trp Leu Gln Asn Tyr Ile Asn Gln Thr Val Lys Ile Lys

355 360 365

Glu Asp Leu Lys Gly His Lys Lys Asp Leu Lys Lys Ala Lys Glu Met

370 375 380

Ile Asn Arg Phe Gly Glu Ser Asp Thr Lys Glu Glu Ala Val Val Ser

385 390 395 400

Ser Leu Leu Glu Ser Ile Glu Lys Ile Val Pro Asp Asp Ser Ala Asp

405 410 415

Asp Glu Lys Pro Asp Ile Pro Ala Ile Ala Ile Tyr Arg Arg Phe Leu

420 425 430

Ser Asp Gly Arg Leu Thr Leu Asn Arg Phe Val Gln Arg Glu Asp Val

435 440 445

Gln Glu Ala Leu Ile Lys Glu Arg Leu Glu Ala Glu Lys Lys Lys Lys

450 455 460

Pro Lys Lys Arg Lys Lys Lys Ser Asp Ala Glu Asp Glu Lys Glu Thr

465 470 475 480

Ile Asp Phe Lys Glu Leu Phe Pro His Leu Ala Lys Pro Leu Lys Leu

485 490 495

Val Pro Asn Phe Tyr Gly Asp Ser Lys Arg Glu Leu Tyr Lys Lys Tyr

500 505 510

Lys Asn Ala Ala Ile Tyr Thr Asp Ala Leu Trp Lys Ala Val Glu Lys

515 520 525

Ile Tyr Lys Ser Ala Phe Ser Ser Ser Leu Lys Asn Ser Phe Phe Asp

530 535 540

Thr Asp Phe Asp Lys Asp Phe Phe Ile Lys Arg Leu Gln Lys Ile Phe

545 550 555 560

Ser Val Tyr Arg Arg Phe Asn Thr Asp Lys Trp Lys Pro Ile Val Lys

565 570 575

Asn Ser Phe Ala Pro Tyr Cys Asp Ile Val Ser Leu Ala Glu Asn Glu

580 585 590

Val Leu Tyr Lys Pro Lys Gln Ser Arg Ser Arg Lys Ser Ala Ala Ile

595 600 605

Asp Lys Asn Arg Val Arg Leu Pro Ser Thr Glu Asn Ile Ala Lys Ala

610 615 620

Gly Ile Ala Leu Ala Arg Glu Leu Ser Val Ala Gly Phe Asp Trp Lys

625 630 635 640

Asp Leu Leu Lys Lys Glu Glu His Glu Glu Tyr Ile Asp Leu Ile Glu

645 650 655

Leu His Lys Thr Ala Leu Ala Leu Leu Leu Ala Val Thr Glu Thr Gln

660 665 670

Leu Asp Ile Ser Ala Leu Asp Phe Val Glu Asn Gly Thr Val Lys Asp

675 680 685

Phe Met Lys Thr Arg Asp Gly Asn Leu Val Leu Glu Gly Arg Phe Leu

690 695 700

Glu Met Phe Ser Gln Ser Ile Val Phe Ser Glu Leu Arg Gly Leu Ala

705 710 715 720

Gly Leu Met Ser Arg Lys Glu Phe Ile Thr Arg Ser Ala Ile Gln Thr

725 730 735

Met Asn Gly Lys Gln Ala Glu Leu Leu Tyr Ile Pro His Glu Phe Gln

740 745 750

Ser Ala Lys Ile Thr Thr Pro Lys Glu Met Ser Arg Ala Phe Leu Asp

755 760 765

Leu Ala Pro Ala Glu Phe Ala Thr Ser Leu Glu Pro Glu Ser Leu Ser

770 775 780

Glu Lys Ser Leu Leu Lys Leu Lys Gln Met Arg Tyr Tyr Pro His Tyr

785 790 795 800

Phe Gly Tyr Glu Leu Thr Arg Thr Gly Gln Gly Ile Asp Gly Gly Val

805 810 815

Ala Glu Asn Ala Leu Arg Leu Glu Lys Ser Pro Val Lys Lys Arg Glu

820 825 830

Ile Lys Cys Lys Gln Tyr Lys Thr Leu Gly Arg Gly Gln Asn Lys Ile

835 840 845

Val Leu Tyr Val Arg Ser Ser Tyr Tyr Gln Thr Gln Phe Leu Glu Trp

850 855 860

Phe Leu His Arg Pro Lys Asn Val Gln Thr Asp Val Ala Val Ser Gly

865 870 875 880

Ser Phe Leu Ile Asp Glu Lys Lys Val Lys Thr Arg Trp Asn Tyr Asp

885 890 895

Ala Leu Thr Val Ala Leu Glu Pro Val Ser Gly Ser Glu Arg Val Phe

900 905 910

Val Ser Gln Pro Phe Thr Ile Phe Pro Glu Lys Ser Ala Glu Glu Glu

915 920 925

Gly Gln Arg Tyr Leu Gly Ile Asp Ile Gly Glu Tyr Gly Ile Ala Tyr

930 935 940

Thr Ala Leu Glu Ile Thr Gly Asp Ser Ala Lys Ile Leu Asp Gln Asn

945 950 955 960

Phe Ile Ser Asp Pro Gln Leu Lys Thr Leu Arg Glu Glu Val Lys Gly

965 970 975

Leu Lys Leu Asp Gln Arg Arg Gly Thr Phe Ala Met Pro Ser Thr Lys

980 985 990

Ile Ala Arg Ile Arg Glu Ser Leu Val His Ser Leu Arg Asn Arg Ile

995 1000 1005

His His Leu Ala Leu Lys His Lys Ala Lys Ile Val Tyr Glu Leu

1010 1015 1020

Glu Val Ser Arg Phe Glu Glu Gly Lys Gln Lys Ile Lys Lys Val

1025 1030 1035

Tyr Ala Thr Leu Lys Lys Ala Asp Val Tyr Ser Glu Ile Asp Ala

1040 1045 1050

Asp Lys Asn Leu Gln Thr Thr Val Trp Gly Lys Leu Ala Val Ala

1055 1060 1065

Ser Glu Ile Ser Ala Ser Tyr Thr Ser Gln Phe Cys Gly Ala Cys

1070 1075 1080

Lys Lys Leu Trp Arg Ala Glu Met Gln Val Asp Glu Thr Ile Thr

1085 1090 1095

Thr Gln Glu Leu Ile Gly Thr Val Arg Val Ile Lys Gly Gly Thr

1100 1105 1110

Leu Ile Asp Ala Ile Lys Asp Phe Met Arg Pro Pro Ile Phe Asp

1115 1120 1125

Glu Asn Asp Thr Pro Phe Pro Lys Tyr Arg Asp Phe Cys Asp Lys

1130 1135 1140

His His Ile Ser Lys Lys Met Arg Gly Asn Ser Cys Leu Phe Ile

1145 1150 1155

Cys Pro Phe Cys Arg Ala Asn Ala Asp Ala Asp Ile Gln Ala Ser

1160 1165 1170

Gln Thr Ile Ala Leu Leu Arg Tyr Val Lys Glu Glu Lys Lys Val

1175 1180 1185

Glu Asp Tyr Phe Glu Arg Phe Arg Lys Leu Lys Asn Ile Lys Val

1190 1195 1200

Leu Gly Gln Met Lys Lys Ile

1205 1210

<210> 32

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 32

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> 33

<211> 1052

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 33

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> 34

<211> 1052

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 34

Lys Arg Asn Tyr Ile Leu Gly Leu Ala 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> 35
<211> 1052
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 35
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 Ala 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> 36

<211> 1052

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 36

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 Ala 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 Lys 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 Lys 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 His 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> 37

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 37

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 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> 38
<211> 1367
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 38
Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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 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> 39

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 39

Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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 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 Glu 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 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> 40

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 40

Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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 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 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 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> 41
<211> 1367
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 41
Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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 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> 42
<211> 1395
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 42
Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240

Ala Thr Pro Glu Ser Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly
245 250 255
Ile Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val
260 265 270
Ile Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn
275 280 285
Glu Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg
290 295 300

Arg His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu
305 310 315 320
Leu Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg
325 330 335
Val Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala
340 345 350
Leu Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val
355 360 365

Glu Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg

370 375 380

Asn Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu

385 390 395 400

Arg Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys

405 410 415

Thr Ser Asp Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys

420 425 430

Ala Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu

435 440 445

Leu Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro

450 455 460

Phe Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His

465 470 475 480

Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn

485 490 495

Ala Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr

500 505 510

Arg Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile

515 520 525

Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala

530 535 540

Lys Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr

545 550 555 560

Ser Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile

565 570 575

Lys Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu

580 585 590

Asp Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile

595 600 605

Gln Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile

610 615 620

Glu Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser
625 630 635 640
Leu Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp
645 650 655
Asn Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val
660 665 670
Asp Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe
675 680 685

Ile Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val
690 695 700
Ile Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile
705 710 715 720
Glu Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn
725 730 735
Glu Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile
740 745 750

Ile Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile
755 760 765
Lys Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala
770 775 780
Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp
785 790 795 800
His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys
805 810 815

Val Leu Val Lys Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro
820 825 830
Phe Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe
835 840 845
Lys Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys
850 855 860
Thr Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser

865 870 875 880

Val Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala
885 890 895
Thr Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn
900 905 910
Leu Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu
915 920 925
Arg Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His
930 935 940

His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys
945 950 955 960
Glu Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met
965 970 975
Phe Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln
980 985 990
Glu Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys
995 1000 1005

Asp Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro
1010 1015 1020
Asn Arg Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp
1025 1030 1035
Asp Lys Gly Asn Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr
1040 1045 1050
Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro
1055 1060 1065

Glu Lys Leu Leu Met Tyr His His Asp Pro Gln Thr Tyr Gln Lys
1070 1075 1080
Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp Glu Lys Asn Pro Leu
1085 1090 1095
Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu Thr Lys Tyr Ser
1100 1105 1110

Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys Tyr Tyr Gly
1115 1120 1125

Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr Pro Asn
1130 1135 1140
Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg Phe
1145 1150 1155
Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
1160 1165 1170
Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
1175 1180 1185

Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln
1190 1195 1200
Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile
1205 1210 1215
Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu
1220 1225 1230
Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr
1235 1240 1245

Leu Glu Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr
1250 1255 1260
Ile Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile
1265 1270 1275
Leu Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile
1280 1285 1290
Ile Lys Lys Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile
1295 1300 1305

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu
1310 1315 1320
Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu
1325 1330 1335
Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1340 1345 1350

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1355 1360 1365

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1370 1375 1380

Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1385 1390 1395

<210> 43

<211> 1395

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 43

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240

Ala Thr Pro Glu Ser Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly

245 250 255

Ile Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val

260 265 270

Ile Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn

275 280 285

Glu Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg

290 295 300

Arg His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu

305 310 315 320

Leu Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg

325 330 335

Val Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala

340 345 350

Leu Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val

355 360 365

Glu Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg

370 375 380

Asn Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu

385 390 395 400

Arg Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys

405 410 415

Thr Ser Asp Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys

420 425 430

Ala Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu

435 440 445

Leu Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro

450 455 460

Phe Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His

465 470 475 480

Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn

485 490 495

Ala Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr

500 505 510

Arg Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile

515 520 525

Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala

530 535 540

Lys Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr

545 550 555 560

Ser Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile

565 570 575

Lys Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu

580 585 590

Asp Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile

595 600 605

Gln Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile

610 615 620

Glu Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser

625 630 635 640

Leu Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp
645 650 655
Asn Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val

660 665 670
Asp Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe
675 680 685
Ile Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val
690 695 700
Ile Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile
705 710 715 720
Glu Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn

725 730 735
Glu Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile
740 745 750
Ile Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile
755 760 765
Lys Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala
770 775 780
Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp

785 790 795 800
His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys
805 810 815
Val Leu Val Lys Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro
820 825 830
Phe Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe
835 840 845
Lys Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys

850 855 860
Thr Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser
865 870 875 880
Val Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala

885 890 895

Thr Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn

900 905 910

Leu Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu

915 920 925

Arg Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His

930 935 940

His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys

945 950 955 960

Glu Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met

965 970 975

Phe Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln

980 985 990

Glu Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys

995 1000 1005

Asp Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro

1010 1015 1020

Asn Arg Lys Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp

1025 1030 1035

Asp Lys Gly Asn Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr

1040 1045 1050

Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro

1055 1060 1065

Glu Lys Leu Leu Met Tyr His His Asp Pro Gln Thr Tyr Gln Lys

1070 1075 1080

Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp Glu Lys Asn Pro Leu

1085 1090 1095

Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu Thr Lys Tyr Ser

1100 1105 1110

Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys Tyr Tyr Gly

1115 1120 1125

Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr Pro Asn

1130 1135 1140

Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg Phe

1145 1150 1155

Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys

1160 1165 1170

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser

1175 1180 1185

Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln

1190 1195 1200

Ala Glu Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile

1205 1210 1215

Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu

1220 1225 1230

Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr

1235 1240 1245

Leu Glu Asn Met Asn Asp Lys Arg Pro Pro His Ile Ile Lys Thr

1250 1255 1260

Ile Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile

1265 1270 1275

Leu Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile

1280 1285 1290

Ile Lys Lys Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1295 1300 1305

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

1310 1315 1320

Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu

1325 1330 1335

Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1340 1345 1350

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1355 1360 1365

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1370 1375 1380

Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1385 1390 1395

<210> 44

<211> 1710

<212> PRT

<213> Artificial Sequence

<220><223>

Synthetic Polypeptide

<400> 44

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560
Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Glu Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1385 1390 1395
Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1400 1405 1410
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1415 1420 1425
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1445 1450 1455
Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1460 1465 1470
Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545
Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575
Gln Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605
Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1610 1615 1620

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu
1625 1630 1635
Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu
1640 1645 1650
Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu
1655 1660 1665
Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp
1670 1675 1680

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met
1685 1690 1695
Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val
1700 1705 1710

<210> 45
<211> 1710
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 45
Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala

995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu

1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

```
His Ile  Val Pro Gln Ser Phe  Leu Lys Asp Asp Ser  Ile Asp Asn
    1085                1090                1095
Lys Val  Leu Thr Arg Ser Asp  Lys Asn Arg Gly Lys  Ser Asp Asn

    1100                1105                1110
Val Pro  Ser Glu Glu Val Val  Lys Lys Met Lys Asn  Tyr Trp Arg
    1115                1120                1125
Gln Leu  Leu Asn Ala Lys Leu  Ile Thr Gln Arg Lys  Phe Asp Asn
    1130                1135                1140
Leu Thr  Lys Ala Glu Arg Gly  Gly Leu Ser Glu Leu  Asp Lys Ala
    1145                1150                1155
Gly Phe  Ile Lys Arg Gln Leu  Val Glu Thr Arg Gln  Ile Thr Lys

    1160                1165                1170
His Val  Ala Gln Ile Leu Asp  Ser Arg Met Asn Thr  Lys Tyr Asp
    1175                1180                1185
Glu Asn  Asp Lys Leu Ile Arg  Glu Val Lys Val Ile  Thr Leu Lys
    1190                1195                1200
Ser Lys  Leu Val Ser Asp Phe  Arg Lys Asp Phe Gln  Phe Tyr Lys
    1205                1210                1215
Val Arg  Glu Ile Asn Asn Tyr  His His Ala His Asp  Ala Tyr Leu

    1220                1225                1230
Asn Ala  Val Val Gly Thr Ala  Leu Ile Lys Lys Tyr  Pro Lys Leu
    1235                1240                1245
Glu Ser  Glu Phe Val Tyr Gly  Asp Tyr Lys Val Tyr  Asp Val Arg
    1250                1255                1260
Lys Met  Ile Ala Lys Ser Glu  Gln Glu Ile Gly Lys  Ala Thr Ala
    1265                1270                1275
Lys Tyr  Phe Phe Tyr Ser Asn  Ile Met Asn Phe Phe  Lys Thr Glu

    1280                1285                1290
Ile Thr  Leu Ala Asn Gly Glu  Ile Arg Lys Arg Pro  Leu Ile Glu
    1295                1300                1305
Thr Asn  Gly Glu Thr Gly Glu  Ile Val Trp Asp Lys  Gly Arg Asp
```

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Gln Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1610 1615 1620

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

1625 1630 1635

Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu

1640 1645 1650

Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1655 1660 1665

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1670 1675 1680

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1685 1690 1695

Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1700 1705 1710

<210> 46

<211> 1710

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 46

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala

995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110
Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1145 1150 1155
Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
1160 1165 1170
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
1190 1195 1200
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
1205 1210 1215
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1220 1225 1230
Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575
Glu Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile
1610 1615 1620
Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu
1625 1630 1635
Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu
1640 1645 1650
Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu
1655 1660 1665

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp
1670 1675 1680
Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met
1685 1690 1695
Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val
1700 1705 1710
<210> 47
<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 47

Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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 Ala
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 Ala 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 Ala 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 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 Ala 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> 48

<211> 1612

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 48

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735

Glu Arg Met Thr Ala Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

```
        755                 760                 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
    770                 775                 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785                 790                 795                 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
                805                 810                 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
            820                 825                 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
        835                 840                 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
    850                 855                 860
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865                 870                 875                 880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
                885                 890                 895
Arg Arg Arg Tyr Thr Gly Trp Gly Ala Leu Ser Arg Lys Leu Ile Asn
            900                 905                 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
        915                 920                 925
Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Ala Leu Ile His Asp Asp
    930                 935                 940
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945                 950                 955                 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
                965                 970                 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
            980                 985                 990
Lys Val Met Gly Arg His Lys Pro  Glu Asn Ile Val Ile  Glu Met Ala
        995                 1000                1005
```

```
Arg Glu  Asn Gln Thr Thr Gln  Lys Gly Gln Lys Asn  Ser Arg Glu
    1010                1015                1020
Arg Met  Lys Arg Ile Glu Glu  Gly Ile Lys Glu Leu  Gly Ser Gln
    1025                1030                1035

Ile Leu  Lys Glu His Pro Val  Glu Asn Thr Gln Leu  Gln Asn Glu
    1040                1045                1050
Lys Leu  Tyr Leu Tyr Tyr Leu  Gln Asn Gly Arg Asp  Met Tyr Val
    1055                1060                1065
Asp Gln  Glu Leu Asp Ile Asn  Arg Leu Ser Asp Tyr  Asp Val Asp
    1070                1075                1080
His Ile  Val Pro Gln Ser Phe  Leu Lys Asp Asp Ser  Ile Asp Asn
    1085                1090                1095

Lys Val  Leu Thr Arg Ser Asp  Lys Asn Arg Gly Lys  Ser Asp Asn
    1100                1105                1110
Val Pro  Ser Glu Glu Val Val  Lys Lys Met Lys Asn  Tyr Trp Arg
    1115                1120                1125
Gln Leu  Leu Asn Ala Lys Leu  Ile Thr Gln Arg Lys  Phe Asp Asn
    1130                1135                1140
Leu Thr  Lys Ala Glu Arg Gly  Gly Leu Ser Glu Leu  Asp Lys Ala
    1145                1150                1155

Gly Phe  Ile Lys Arg Gln Leu  Val Glu Thr Arg Ala  Ile Thr Lys
    1160                1165                1170
His Val  Ala Gln Ile Leu Asp  Ser Arg Met Asn Thr  Lys Tyr Asp
    1175                1180                1185
Glu Asn  Asp Lys Leu Ile Arg  Glu Val Lys Val Ile  Thr Leu Lys
    1190                1195                1200
Ser Lys  Leu Val Ser Asp Phe  Arg Lys Asp Phe Gln  Phe Tyr Lys
    1205                1210                1215

Val Arg  Glu Ile Asn Asn Tyr  His His Ala His Asp  Ala Tyr Leu
    1220                1225                1230
Asn Ala  Val Val Gly Thr Ala  Leu Ile Lys Lys Tyr  Pro Lys Leu
```

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545
Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605
Leu Gly Gly Asp
1610

<210> 49
<211> 198
<212> PRT
<213> Homo sapiens
<400> 49

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 Tyr Val

20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45
Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu

195

<210> 50

<211> 181

<212> PRT

<213> Homo sapiens

<400> 50

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 Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr

35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu

180

<210> 51

<211> 198

<212> PRT

<213> Mus musculus

<400> 51

Met Asp Ser Leu Leu Met Lys Gln Lys Lys Phe Leu Tyr His Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Cys Ser Leu Asp Phe Gly His

35 40 45

Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Glu
85 90 95
Phe Leu Arg Trp Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110
Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125
Leu His Arg Ala Gly Val Gln Ile Gly Ile Met Thr Phe Lys Asp Tyr

130 135 140
Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Arg Thr Phe Lys
145 150 155 160
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Thr Arg Gln Leu
165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190
Phe Arg Met Leu Gly Phe
195

<210> 52
<211> 198
<212> PRT
<213> Canis lupus familiaris
<400> 52
Met Asp Ser Leu Leu Met Lys Gln Arg Lys Phe Leu Tyr His Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly His
35 40 45
Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Ala Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Lys Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu

195

<210> 53

<211> 199

<212> PRT

<213> Bos taurus

<400> 53

Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His

35 40 45

Leu Arg Asn Lys Ala Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Asp Lys Glu Arg Lys Ala Glu Pro Glu Gly Leu Arg

115 120 125

Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp

130 135 140

Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe

145 150 155 160

Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln

165 170 175

Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp

180 185 190

Ala Phe Arg Thr Leu Gly Leu

195

<210> 54

<211> 239

<212> PRT

<213> Rattus norvegicus

<400> 54

Met Ala Val Gly Ser Lys Pro Lys Ala Ala Leu Val Gly Pro His Trp

1 5 10 15

Glu Arg Glu Arg Ile Trp Cys Phe Leu Cys Ser Thr Gly Leu Gly Thr

20 25 30

Gln Gln Thr Gly Gln Thr Ser Arg Trp Leu Arg Pro Ala Ala Thr Gln

35 40 45

Asp Pro Val Ser Pro Pro Arg Ser Leu Leu Met Lys Gln Arg Lys Phe

50 55 60

Leu Tyr His Phe Lys Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr

65 70 75 80

Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser

85 90 95

Leu Asp Phe Gly Tyr Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu
100 105 110
Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys
115 120 125
Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala
130 135 140
Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg

145 150 155 160
Ile Phe Thr Ala Arg Leu Thr Gly Trp Gly Ala Leu Pro Ala Gly Leu
165 170 175
Met Ser Pro Ala Arg Pro Ser Asp Tyr Phe Tyr Cys Trp Asn Thr Phe
180 185 190
Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu
195 200 205
Asn Ser Val Arg Leu Ser Arg Arg Leu Arg Arg Ile Leu Leu Pro Leu

210 215 220
Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly Leu
225 230 235
<210> 55
<211> 429
<212> PRT
<213> Mus musculus
<400> 55
Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro
1 5 10 15
Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn
20 25 30

Leu Gly Tyr Ala Lys Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val
35 40 45
Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe
50 55 60
Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe

65 70 75 80

His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile

85 90 95

Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile

100 105 110

Val Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser

115 120 125

Ser Arg Leu Tyr Asn Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys

130 135 140

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu

145 150 155 160

Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe

165 170 175

Arg Pro Trp Lys Arg Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys

180 185 190

Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser

195 200 205

Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr

210 215 220

Arg Phe Cys Val Glu Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu

225 230 235 240

Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr

245 250 255

His Arg Met Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly

260 265 270

Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His

275 280 285

Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln

290 295 300

Val Thr Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala

305 310 315 320

Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His
325 330 335
Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys
340 345 350

Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp
355 360 365
Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg
370 375 380
Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln
385 390 395 400
Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val
405 410 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser
420 425

<210> 56
<211> 429
<212> PRT
<213> Rattus norvegicus
<400> 56
Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro
1 5 10 15
Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn
20 25 30
Leu Arg Tyr Ala Ile Asp Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val

35 40 45
Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe
50 55 60
Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe
65 70 75 80
His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile
85 90 95

Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Val

100 105 110

Leu Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser

115 120 125

Ser Arg Leu Tyr Asn Ile Arg Asp Pro Glu Asn Gln Gln Asn Leu Cys

130 135 140

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu

145 150 155 160

Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe

165 170 175

Arg Pro Trp Lys Lys Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys

180 185 190

Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser

195 200 205

Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr

210 215 220

Arg Phe Cys Val Glu Arg Arg Arg Val His Leu Leu Ser Glu Glu Glu

225 230 235 240

Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr

245 250 255

His Gly Val Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly

260 265 270

Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His

275 280 285

Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln

290 295 300

Val Ile Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala

305 310 315 320

Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His

325 330 335

Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys

340 345 350

Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp

355 360 365

Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg

370 375 380

Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln

385 390 395 400

Arg Arg Leu His Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val

405 410 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser

420 425

<210> 57

<211> 370

<212> PRT

<213> Macaca mulatta

<400> 57

Met Val Glu Pro Met Asp Pro Arg Thr Phe Val Ser Asn Phe Asn Asn

1 5 10 15

Arg Pro Ile Leu Ser Gly Leu Asn Thr Val Trp Leu Cys Cys Glu Val

20 25 30

Lys Thr Lys Asp Pro Ser Gly Pro Pro Leu Asp Ala Lys Ile Phe Gln

35 40 45

Gly Lys Val Tyr Ser Lys Ala Lys Tyr His Pro Glu Met Arg Phe Leu

50 55 60

Arg Trp Phe His Lys Trp Arg Gln Leu His His Asp Gln Glu Tyr Lys

65 70 75 80

Val Thr Trp Tyr Val Ser Trp Ser Pro Cys Thr Arg Cys Ala Asn Ser

85 90 95

Val Ala Thr Phe Leu Ala Lys Asp Pro Lys Val Thr Leu Thr Ile Phe

100 105 110

Val Ala Arg Leu Tyr Tyr Phe Trp Lys Pro Asp Tyr Gln Gln Ala Leu

115 120 125

Arg Ile Leu Cys Gln Lys Arg Gly Gly Pro His Ala Thr Met Lys Ile

130 135 140

Met Asn Tyr Asn Glu Phe Gln Asp Cys Trp Asn Lys Phe Val Asp Gly

145 150 155 160

Arg Gly Lys Pro Phe Lys Pro Arg Asn Asn Leu Pro Lys His Tyr Thr

165 170 175

Leu Leu Gln Ala Thr Leu Gly Glu Leu Leu Arg His Leu Met Asp Pro

180 185 190

Gly Thr Phe Thr Ser Asn Phe Asn Asn Lys Pro Trp Val Ser Gly Gln

195 200 205

His Glu Thr Tyr Leu Cys Tyr Lys Val Glu Arg Leu His Asn Asp Thr

210 215 220

Trp Val Pro Leu Asn Gln His Arg Gly Phe Leu Arg Asn Gln Ala Pro

225 230 235 240

Asn Ile His Gly Phe Pro Lys Gly Arg His Ala Glu Leu Cys Phe Leu

245 250 255

Asp Leu Ile Pro Phe Trp Lys Leu Asp Gly Gln Gln Tyr Arg Val Thr

260 265 270

Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala

275 280 285

Lys Phe Ile Ser Asn Asn Glu His Val Ser Leu Cys Ile Phe Ala Ala

290 295 300

Arg Ile Tyr Asp Asp Gln Gly Arg Tyr Gln Glu Gly Leu Arg Ala Leu

305 310 315 320

His Arg Asp Gly Ala Lys Ile Ala Met Met Asn Tyr Ser Glu Phe Glu

325 330 335

Tyr Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe Gln Pro

340 345 350

Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg

355 360 365

Ala Ile

370

<210> 58

<211> 384

<212> PRT

<213> Pan troglodytes

<400> 58

Met Lys Pro His Phe Arg Asn Pro Val Glu Arg Met Tyr Gln Asp Thr
1 5 10 15
Phe Ser Asp Asn Phe Tyr Asn Arg Pro Ile Leu Ser His Arg Asn Thr
20 25 30
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
35 40 45
Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Lys Leu Lys Tyr

50 55 60
His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65 70 75 80
His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95
Cys Thr Lys Cys Thr Arg Asp Val Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110
Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp

115 120 125
Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140
Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145 150 155 160
Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175
Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile

180 185 190
Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Ser Asn Phe Asn Asn

195 200 205

Glu Leu Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val

210 215 220

Glu Arg Leu His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly

225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg

245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp

260 265 270

Leu His Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys

275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Asn Asn Lys His

290 295 300

Val Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg

305 310 315 320

Cys Gln Glu Gly Leu Arg Thr Leu Ala Lys Ala Gly Ala Lys Ile Ser

325 330 335

Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp

340 345 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser

355 360 365

Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn

370 375 380

<210> 59

<211> 377

<212> PRT

<213> Chlorocebus sabaeus

<400> 59

Met Asn Pro Gln Ile Arg Asn Met Val Glu Gln Met Glu Pro Asp Ile

1 5 10 15

Phe Val Tyr Tyr Phe Asn Asn Arg Pro Ile Leu Ser Gly Arg Asn Thr

20 25 30

```
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Asp Pro Ser Gly Pro Pro
        35                  40                  45

Leu Asp Ala Asn Ile Phe Gln Gly Lys Leu Tyr Pro Glu Ala Lys Asp
    50                  55                  60
His Pro Glu Met Lys Phe Leu His Trp Phe Arg Lys Trp Arg Gln Leu
65                  70                  75                  80
His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Val Ser Trp Ser Pro
                85                  90                  95
Cys Thr Arg Cys Ala Asn Ser Val Ala Thr Phe Leu Ala Glu Asp Pro
            100                 105                 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Lys
        115                 120                 125
Pro Asp Tyr Gln Gln Ala Leu Arg Ile Leu Cys Gln Glu Arg Gly Gly
    130                 135                 140
Pro His Ala Thr Met Lys Ile Met Asn Tyr Asn Glu Phe Gln His Cys
145                 150                 155                 160
Trp Asn Glu Phe Val Asp Gly Gln Gly Lys Pro Phe Lys Pro Arg Lys
                165                 170                 175

Asn Leu Pro Lys His Tyr Thr Leu Leu His Ala Thr Leu Gly Glu Leu
            180                 185                 190
Leu Arg His Val Met Asp Pro Gly Thr Phe Thr Ser Asn Phe Asn Asn
        195                 200                 205
Lys Pro Trp Val Ser Gly Gln Arg Glu Thr Tyr Leu Cys Tyr Lys Val
    210                 215                 220
Glu Arg Ser His Asn Asp Thr Trp Val Leu Leu Asn Gln His Arg Gly
225                 230                 235                 240

Phe Leu Arg Asn Gln Ala Pro Asp Arg His Gly Phe Pro Lys Gly Arg
                245                 250                 255
His Ala Glu Leu Cys Phe Leu Asp Leu Ile Pro Phe Trp Lys Leu Asp
            260                 265                 270
Asp Gln Gln Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe
```

275 280 285

Ser Cys Ala Gln Lys Met Ala Lys Phe Ile Ser Asn Asn Lys His Val

290 295 300

Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys

305 310 315 320

Gln Glu Gly Leu Arg Thr Leu His Arg Asp Gly Ala Lys Ile Ala Val

325 330 335

Met Asn Tyr Ser Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Asp Arg

340 345 350

Gln Gly Arg Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln

355 360 365

Ala Leu Ser Gly Arg Leu Arg Ala Ile

370 375

<210> 60

<211> 384

<212> PRT

<213> Homo sapiens

<400> 60

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr

20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro

35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr

50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu

65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro

85 90 95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro

100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
115 120 125
Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140
Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145 150 155 160
Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
180 185 190
Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
195 200 205
Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
210 215 220
Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
245 250 255
His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
260 265 270
Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
275 280 285
Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His
290 295 300

Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg
305 310 315 320
Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser
325 330 335
Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
340 345 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser
355 360 365

Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
370 375 380

<210> 61

<211> 373

<212> PRT

<213> Homo sapiens

<400> 61

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Arg
35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Gln Pro Glu His
50 55 60

His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro
65 70 75 80

Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys
85 90 95

Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ala Glu His Pro Asn
100 105 110

Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg
115 120 125

Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val
130 135 140

Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val
145 150 155 160

Tyr Ser Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn
165 170 175

Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met
180 185 190
Glu Ala Met Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg
195 200 205
Lys Ala Tyr Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val
210 215 220
Val Lys His His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn

225 230 235 240
Gln Val Asp Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser
245 250 255
Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr
260 265 270
Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala
275 280 285
Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala

290 295 300
Arg Leu Tyr Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser
305 310 315 320
Leu Ser Gln Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe
325 330 335
Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys
340 345 350
Pro Trp Lys Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu

355 360 365
Gln Glu Ile Leu Glu
370
<210> 62
<211> 382
<212> PRT
<213> Homo sapiens
<400> 62
Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr

20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu

35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln

50 55 60

Tyr His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu

65 70 75 80

Pro Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro

85 90 95

Cys Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ser Glu His Pro

100 105 110

Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu

115 120 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg

130 135 140

Val Thr Ile Met Asp Tyr Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Glu Gly Gln Gln Phe Met Pro Trp Tyr Lys Phe Asp Glu

165 170 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Tyr Leu

180 185 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val

195 200 205

Leu Arg Arg Arg Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp

210 215 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn

225 230 235 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu

245 250 255

Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile
260 265 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly
275 280 285

Cys Ala Gly Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg

290 295 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys
305 310 315 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met
325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala

355 360 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
370 375 380

<210> 63

<211> 396

<212> PRT

<213> Rattus norvegicus

<400> 63

Met Gln Pro Gln Gly Leu Gly Pro Asn Ala Gly Met Gly Pro Val Cys
1 5 10 15

Leu Gly Cys Ser His Arg Arg Pro Tyr Ser Pro Ile Arg Asn Pro Leu
20 25 30

Lys Lys Leu Tyr Gln Gln Thr Phe Tyr Phe His Phe Lys Asn Val Arg
35 40 45

Tyr Ala Trp Gly Arg Lys Asn Asn Phe Leu Cys Tyr Glu Val Asn Gly
50 55 60

Met Asp Cys Ala Leu Pro Val Pro Leu Arg Gln Gly Val Phe Arg Lys
65 70 75 80

Gln Gly His Ile His Ala Glu Leu Cys Phe Ile Tyr Trp Phe His Asp

85 90 95

Lys Val Leu Arg Val Leu Ser Pro Met Glu Glu Phe Lys Val Thr Trp
100 105 110
Tyr Met Ser Trp Ser Pro Cys Ser Lys Cys Ala Glu Gln Val Ala Arg
115 120 125
Phe Leu Ala Ala His Arg Asn Leu Ser Leu Ala Ile Phe Ser Ser Arg
130 135 140
Leu Tyr Tyr Tyr Leu Arg Asn Pro Asn Tyr Gln Gln Lys Leu Cys Arg
145 150 155 160

Leu Ile Gln Glu Gly Val His Val Ala Ala Met Asp Leu Pro Glu Phe
165 170 175
Lys Lys Cys Trp Asn Lys Phe Val Asp Asn Asp Gly Gln Pro Phe Arg
180 185 190
Pro Trp Met Arg Leu Arg Ile Asn Phe Ser Phe Tyr Asp Cys Lys Leu
195 200 205
Gln Glu Ile Phe Ser Arg Met Asn Leu Leu Arg Glu Asp Val Phe Tyr
210 215 220

Leu Gln Phe Asn Asn Ser His Arg Val Lys Pro Val Gln Asn Arg Tyr
225 230 235 240
Tyr Arg Arg Lys Ser Tyr Leu Cys Tyr Gln Leu Glu Arg Ala Asn Gly
245 250 255
Gln Glu Pro Leu Lys Gly Tyr Leu Leu Tyr Lys Lys Gly Glu Gln His
260 265 270
Val Glu Ile Leu Phe Leu Glu Lys Met Arg Ser Met Glu Leu Ser Gln
275 280 285

Val Arg Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
290 295 300
Arg Gln Leu Ala Ala Phe Lys Lys Asp His Pro Asp Leu Ile Leu Arg
305 310 315 320
Ile Tyr Thr Ser Arg Leu Tyr Phe Tyr Trp Arg Lys Lys Phe Gln Lys
325 330 335

Gly Leu Cys Thr Leu Trp Arg Ser Gly Ile His Val Asp Val Met Asp
340 345 350

Leu Pro Gln Phe Ala Asp Cys Trp Thr Asn Phe Val Asn Pro Gln Arg
355 360 365

Pro Phe Arg Pro Trp Asn Glu Leu Glu Lys Asn Ser Trp Arg Ile Gln
370 375 380

Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu
385 390 395

<210> 64

<211> 227

<212> PRT

<213> Bos taurus

<400> 64

Asp Gly Trp Glu Val Ala Phe Arg Ser Gly Thr Val Leu Lys Ala Gly
1 5 10 15

Val Leu Gly Val Ser Met Thr Glu Gly Trp Ala Gly Ser Gly His Pro
20 25 30

Gly Gln Gly Ala Cys Val Trp Thr Pro Gly Thr Arg Asn Thr Met Asn
35 40 45

Leu Leu Arg Glu Val Leu Phe Lys Gln Gln Phe Gly Asn Gln Pro Arg
50 55 60

Val Pro Ala Pro Tyr Tyr Arg Arg Lys Thr Tyr Leu Cys Tyr Gln Leu
65 70 75 80

Lys Gln Arg Asn Asp Leu Thr Leu Asp Arg Gly Cys Phe Arg Asn Lys
85 90 95

Lys Gln Arg His Ala Glu Ile Arg Phe Ile Asp Lys Ile Asn Ser Leu
100 105 110

Asp Leu Asn Pro Ser Gln Ser Tyr Lys Ile Ile Cys Tyr Ile Thr Trp
115 120 125

Ser Pro Cys Pro Asn Cys Ala Asn Glu Leu Val Asn Phe Ile Thr Arg
130 135 140

Asn Asn His Leu Lys Leu Glu Ile Phe Ala Ser Arg Leu Tyr Phe His
145 150 155 160
Trp Ile Lys Ser Phe Lys Met Gly Leu Gln Asp Leu Gln Asn Ala Gly
165 170 175
Ile Ser Val Ala Val Met Thr His Thr Glu Phe Glu Asp Cys Trp Glu
180 185 190
Gln Phe Val Asp Asn Gln Ser Arg Pro Phe Gln Pro Trp Asp Lys Leu
195 200 205

Glu Gln Tyr Ser Ala Ser Ile Arg Arg Arg Leu Gln Arg Ile Leu Thr
210 215 220
Ala Pro Ile
225

<210> 65
<211> 490
<212> PRT
<213> Pan troglodytes
<400> 65
Met Asn Pro Gln Ile Arg Asn Pro Met Glu Trp Met Tyr Gln Arg Thr
1 5 10 15
Phe Tyr Tyr Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30
Thr Trp Leu Cys Tyr Glu Val Lys Ile Arg Arg Gly His Ser Asn Leu

35 40 45
Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Met Tyr Ser Gln Pro Glu
50 55 60
His His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65 70 75 80
Ser Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
85 90 95
Cys Pro Asp Cys Val Ala Lys Leu Ala Lys Phe Leu Ala Glu His Pro

100 105 110
Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu

115 120 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg

130 135 140

Val Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp

165 170 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Ile Arg His Leu

180 185 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val

195 200 205

Leu Arg Arg His Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp

210 215 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn

225 230 235 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu

245 250 255

Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile

260 265 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly

275 280 285

Cys Ala Gly Gln Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg

290 295 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys

305 310 315 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met

325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln

340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala

355 360 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Val Arg Ala Ser Ser Leu
370 375 380
Cys Met Val Pro His Arg Pro Pro Pro Pro Pro Gln Ser Pro Gly Pro
385 390 395 400
Cys Leu Pro Leu Cys Ser Glu Pro Pro Leu Gly Ser Leu Leu Pro Thr
405 410 415
Gly Arg Pro Ala Pro Ser Leu Pro Phe Leu Leu Thr Ala Ser Phe Ser

420 425 430
Phe Pro Pro Pro Ala Ser Leu Pro Pro Leu Pro Ser Leu Ser Leu Ser
435 440 445
Pro Gly His Leu Pro Val Pro Ser Phe His Ser Leu Thr Ser Cys Ser
450 455 460
Ile Gln Pro Pro Cys Ser Ser Arg Ile Arg Glu Thr Glu Gly Trp Ala
465 470 475 480
Ser Val Ser Lys Glu Gly Arg Asp Leu Gly
485 490

<210> 66
<211> 190
<212> PRT
<213> Homo sapiens
<400> 66
Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
1 5 10 15
Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
20 25 30
Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
35 40 45
Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His

50 55 60
Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu
65 70 75 80
Ser Pro Asn Thr Lys Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro

85 90 95

Cys Pro Asp Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser

100 105 110

Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Tyr

115 120 125

Pro Cys Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala

130 135 140

Val Glu Ile Met Asp Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Asp Asn Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Thr

165 170 175

Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ser Leu Gln

180 185 190

<210> 67

<211> 190

<212> PRT

<213> Gorilla gorilla

<400> 67

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr

1 5 10 15

Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu

20 25 30

Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val

35 40 45

Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His

50 55 60

Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu

65 70 75 80

Ser Pro Asn Thr Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro

85 90 95

Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser

100 105 110

Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Asp
115 120 125
Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
130 135 140
Val Lys Ile Met Asp Tyr Lys Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145 150 155 160
Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
165 170 175

Asn Phe Arg Phe Leu Lys Arg Arg Leu Gln Glu Ile Leu Glu
180 185 190

<210> 68
<211> 199
<212> PRT
<213> Homo sapiens
<400> 68

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His
1 5 10 15
Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr
20 25 30
Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met

35 40 45
Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys
50 55 60
Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro
65 70 75 80
Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile
85 90 95
Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala

100 105 110
Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg

115 120 125

Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg

130 135 140

Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His

145 150 155 160

Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp

165 170 175

Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala

180 185 190

Ile Leu Gln Asn Gln Gly Asn

195

<210> 69

<211> 202

<212> PRT

<213> Macaca mulatta

<400> 69

Met Asp Gly Ser Pro Ala Ser Arg Pro Arg His Leu Met Asp Pro Asn

1 5 10 15

Thr Phe Thr Phe Asn Phe Asn Asn Asp Leu Ser Val Arg Gly Arg His

20 25 30

Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Trp

35 40 45

Val Pro Met Asp Glu Arg Arg Gly Phe Leu Cys Asn Lys Ala Lys Asn

50 55 60

Val Pro Cys Gly Asp Tyr Gly Cys His Val Glu Leu Arg Phe Leu Cys

65 70 75 80

Glu Val Pro Ser Trp Gln Leu Asp Pro Ala Gln Thr Tyr Arg Val Thr

85 90 95

Trp Phe Ile Ser Trp Ser Pro Cys Phe Arg Arg Gly Cys Ala Gly Gln

100 105 110

Val Arg Val Phe Leu Gln Glu Asn Lys His Val Arg Leu Arg Ile Phe

115 120 125

Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Gln Glu Ala Leu Arg

130 135 140

Thr Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Glu Glu

145 150 155 160

Phe Lys His Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe

165 170 175

Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg

180 185 190

Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn

195 200

<210> 70

<211> 185

<212> PRT

<213> Bos taurus

<400> 70

Met Asp Glu Tyr Thr Phe Thr Glu Asn Phe Asn Asn Gln Gly Trp Pro

1 5 10 15

Ser Lys Thr Tyr Leu Cys Tyr Glu Met Glu Arg Leu Asp Gly Asp Ala

20 25 30

Thr Ile Pro Leu Asp Glu Tyr Lys Gly Phe Val Arg Asn Lys Gly Leu

35 40 45

Asp Gln Pro Glu Lys Pro Cys His Ala Glu Leu Tyr Phe Leu Gly Lys

50 55 60

Ile His Ser Trp Asn Leu Asp Arg Asn Gln His Tyr Arg Leu Thr Cys

65 70 75 80

Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Gln Lys Leu Thr Thr

85 90 95

Phe Leu Lys Glu Asn His His Ile Ser Leu His Ile Leu Ala Ser Arg

100 105 110

Ile Tyr Thr His Asn Arg Phe Gly Cys His Gln Ser Gly Leu Cys Glu

115 120 125

Leu Gln Ala Ala Gly Ala Arg Ile Thr Ile Met Thr Phe Glu Asp Phe

130 135 140

Lys His Cys Trp Glu Thr Phe Val Asp His Lys Gly Lys Pro Phe Gln

145 150 155 160

Pro Trp Glu Gly Leu Asn Val Lys Ser Gln Ala Leu Cys Thr Glu Leu

165 170 175

Gln Ala Ile Leu Lys Thr Gln Gln Asn

180 185

<210> 71

<211> 200

<212> PRT

<213> Homo sapiens

<400> 71

Met Ala Leu Leu Thr Ala Glu Thr Phe Arg Leu Gln Phe Asn Asn Lys

1 5 10 15

Arg Arg Leu Arg Arg Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr

20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Glu

35 40 45

Asn Lys Lys Lys Cys His Ala Glu Ile Cys Phe Ile Asn Glu Ile Lys

50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu

65 70 75 80

Thr Trp Ser Pro Cys Ser Ser Cys Ala Trp Glu Leu Val Asp Phe Ile

85 90 95

Lys Ala His Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr

100 105 110

Tyr His Trp Cys Lys Pro Gln Gln Lys Gly Leu Arg Leu Leu Cys Gly

115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Phe Pro Lys Phe Ala Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Glu Lys Pro Leu Ser Phe Asn Pro Tyr
145 150 155 160
Lys Met Leu Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg
165 170 175
Leu Glu Arg Ile Lys Ile Pro Gly Val Arg Ala Gln Gly Arg Tyr Met
180 185 190
Asp Ile Leu Cys Asp Ala Glu Val
195 200

<210> 72
<211> 210
<212> PRT
<213> Macaca mulatta
<400> 72
Met Ala Leu Leu Thr Ala Lys Thr Phe Ser Leu Gln Phe Asn Asn Lys
1 5 10 15
Arg Arg Val Asn Lys Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr
20 25 30
Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly His Leu Lys
35 40 45
Asn Lys Lys Lys Asp His Ala Glu Ile Arg Phe Ile Asn Lys Ile Lys

50 55 60
Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu
65 70 75 80
Thr Trp Ser Pro Cys Pro Ser Cys Ala Gly Glu Leu Val Asp Phe Ile
85 90 95
Lys Ala His Arg His Leu Asn Leu Arg Ile Phe Ala Ser Arg Leu Tyr
100 105 110
Tyr His Trp Arg Pro Asn Tyr Gln Glu Gly Leu Leu Leu Leu Cys Gly

115 120 125
Ser Gln Val Pro Val Glu Val Met Gly Leu Pro Glu Phe Thr Asp Cys
130 135 140
Trp Glu Asn Phe Val Asp His Lys Glu Pro Pro Ser Phe Asn Pro Ser

145 150 155 160

Glu Lys Leu Glu Glu Leu Asp Lys Asn Ser Gln Ala Ile Lys Arg Arg

165 170 175

Leu Glu Arg Ile Lys Ser Arg Ser Val Asp Val Leu Glu Asn Gly Leu

180 185 190

Arg Ser Leu Gln Leu Gly Pro Val Thr Pro Ser Ser Ser Ile Arg Asn

195 200 205

Ser Arg

210

<210> 73

<211> 386

<212> PRT

<213> Homo sapiens

<400> 73

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr

20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu

35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Pro Val Leu Pro Lys Arg Gln

50 55 60

Ser Asn His Arg Gln Glu Val Tyr Phe Arg Phe Glu Asn His Ala Glu

65 70 75 80

Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Arg Leu Pro Ala Asn Arg

85 90 95

Arg Phe Gln Ile Thr Trp Phe Val Ser Trp Asn Pro Cys Leu Pro Cys

100 105 110

Val Val Lys Val Thr Lys Phe Leu Ala Glu His Pro Asn Val Thr Leu

115 120 125

Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Arg Asp Arg Asp Trp Arg

130 135 140

Trp Val Leu Leu Arg Leu His Lys Ala Gly Ala Arg Val Lys Ile Met
145 150 155 160

Asp Tyr Glu Asp Phe Ala Tyr Cys Trp Glu Asn Phe Val Cys Asn Glu
165 170 175
Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn Tyr Ala Ser
180 185 190
Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met Glu Ala Met
195 200 205
Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Leu Lys Ala Cys
210 215 220

Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val Thr Lys His
225 230 235 240
His Ser Ala Val Phe Arg Lys Arg Gly Val Phe Arg Asn Gln Val Asp
245 250 255
Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys
260 265 270
Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr
275 280 285

Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu
290 295 300
Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Cys
305 310 315 320
Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Cys Ser Leu Ser Gln
325 330 335
Glu Gly Ala Ser Val Lys Ile Met Gly Tyr Lys Asp Phe Val Ser Cys
340 345 350

Trp Lys Asn Phe Val Tyr Ser Asp Asp Glu Pro Phe Lys Pro Trp Lys
355 360 365
Gly Leu Gln Thr Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ile
370 375 380
Leu Gln

385

<210> 74

<211> 236

<212> PRT

<213> Homo sapiens

<400> 74

Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu

20 25 30
Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
35 40 45
Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
50 55 60
Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met
65 70 75 80
Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys

85 90 95
Ser Gln Ala Ile Arg Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu
100 105 110
Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160
Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
165 170 175
Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
180 185 190
Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu
195 200 205

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu

210 215 220

Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg

225 230 235

<210> 75

<211> 229

<212> PRT

<213> Mus musculus

<400> 75

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Val Trp Arg His Thr Ser Gln Asn Thr Ser Asn His Val Glu Val

50 55 60

Asn Phe Leu Glu Lys Phe Thr Thr Glu Arg Tyr Phe Arg Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg His Pro Tyr Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160

Pro Ser Asn Glu Ala Tyr Trp Pro Arg Tyr Pro His Leu Trp Val Lys

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Lys Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Thr Leu Gln Thr Cys His Tyr Gln Arg Ile Pro Pro His Leu Leu Trp
210 215 220
Ala Thr Gly Leu Lys
225

<210
> 76

<211> 229

<212> PRT

<213> Rattus norvegicus

<400> 76

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys

225

<210> 77

<211> 60

<212> PRT

<213> Homo sapiens

<400> 77

Met Ala Gln Lys Glu Glu Ala Ala Val Ala Thr Glu Ala Ala Ser Gln

1 5 10 15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro

35 40 45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu

50 55 60

<210> 78

<211> 164

<212> PRT

<213> Mus musculus

<400> 78

Tyr Ser Ser Gly Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala

1 5 10 15

Gln Gly Lys Gly Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp

20 25 30

Glu His Ala Ala Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu

35 40 45

Pro Ala Phe Asp Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser

50 55 60

Ser Ser Pro Cys Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser

65 70 75 80

Lys Thr Lys Asn Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met

85 90 95

Trp Glu Glu Pro Glu Ile Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala

100 105 110

Gly Cys Lys Leu Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp

115 120 125

Gln Asn Phe Val Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro

130 135 140

Trp Glu Asp Ile Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala

145 150 155 160

Asp Ile Leu Lys

<210> 79

<211> 224

<212> PRT

<213> Rattus norvegicus

<400> 79

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln

1 5 10 15

Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro

35 40 45

Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly

50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly
65 70 75 80
Gly Gln Val Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly
85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110
Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125
Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140
Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175
Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Leu Trp Gln Asn Phe Val
180 185 190
Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205
Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 80
<211> 60
<212> PRT
<213> Bos taurus
<400> 80
Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Ala Glu Pro Ala Ser Gln
1 5 10 15
Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
20 25 30
Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45

Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu

50 55 60

<210> 81

<211> 164

<212> PRT

<213> Petromyzon marinus

<400> 81

Tyr Ser Ser Gly Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala
1 5 10 15

Gln Ser Lys Gly Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp
20 25 30

Glu His Ala Thr Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met
35 40 45

Pro Thr Phe Asp Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser
50 55 60

Ser Ser Pro Cys Ala Ala Cys Ala Asp Arg Ile Val Lys Thr Leu Asn
65 70 75 80

Lys Thr Lys Asn Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met
85 90 95

Trp Glu Glu Pro Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala
100 105 110

Gly Cys Arg Leu Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp
115 120 125

Gln Asn Phe Val Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro
130 135 140

Trp Glu Asp Ile Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala
145 150 155 160

Asp Ile Leu Lys

<210> 82

<211> 384

<212> PRT

<213> Homo sapiens

<400> 82

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr

20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro

35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr

50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu

65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro

85 90 95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro

100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp

115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly

130 135 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys

145 150 155 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn

165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile

180 185 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn

195 200 205

Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val

210 215 220

Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly

225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg

245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp

260 265 270

Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys

275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His

290 295 300

Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg

305 310 315 320

Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser

325 330 335

Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp

340 345 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser

355 360 365

Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn

370 375 380

<210> 83

<211> 184

<212> PRT

<213> Homo sapiens

<400> 83

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val

1 5 10 15

Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His

20 25 30

Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn

35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu

50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp
65 70 75 80
Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala
85 90 95
Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys
100 105 110
Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly
115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr
130 135 140
Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145 150 155 160
Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
165 170 175
Gly Arg Leu Arg Ala Ile Leu Gln
180

<210> 84
<211> 184
<212> PRT
<213> Homo sapiens

<400> 84
Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val
1 5 10 15
Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His
20 25 30
Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn
35 40 45
Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu
50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp

65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala

85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys

100 105 110

Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln Glu Gly

115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr

130 135 140

Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys

145 150 155 160

Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser

165 170 175

Gly Arg Leu Arg Ala Ile Leu Gln

180

<210> 85

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 85

gtaggtagtt aggatgaatg gaaggttggt atagtcccgc ggatttattt atttaatgga 60

tgacctctgg atccatggac 80

<210> 86

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 86

gtaggtagtt aggatgaatg gaaggttggt atagcttccg cggatttatt tatttatgga 60

tgacctctgg atccatggac 80

<210> 87

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 87

gtaggtagtt aggatgaatg gaaggttggt atagccttcc gcggatttat ttattatgga 60

tgacctctgg atccatggac 80

<210> 88

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 88

gtaggtagtt aggatgaatg gaaggttggt atagccattc cgcggattta tttatttgga 60

tgacctctgg atccatggac 80

<210> 89

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 89

gtaggtagtt aggatgaatg gaaggttggt atagcctatt ccgcggattt atttattgga 60

tgacctctgg atccatggac 80

<210> 90

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 90

gtaggtagtt aggatgaatg gaaggttggt atagccttat tccgcggatt tatttatgga 60

tgacctctgg atccatggac 80

<210> 91

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 91

gtaggtagtt aggatgaatg gaaggttggt atagccatta ttccgcggat ttattttgga 60

tgacctctgg atccatggac 80

<210> 92

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 92

gtaggtagtt aggatgaatg gaaggttggt atagcctatt attccgcgga tttatttgga 60

tgacctctgg atccatggac 80

<210> 93

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 93

gtaggtagtt aggatgaatg gaaggttggt atagccatta tattccgcgg atttattgga 60

tgacctctgg atccatggac 80

<210> 94

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 94

gtaggtagtt aggatgaatg gaaggttggt atagcctatt atattccgcg gatttatgga 60

tgacctctgg atccatggac 80

<210> 95

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 95

gtaggtagtt aggatgaatg gaaggttggt atagccttat tatattccgc ggattttgga 60

tgacctctgg atccatggac 80

<210> 96

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 96

gtaggtagtt aggatgaatg gaaggttggt atagccatta ttatattccg cggatttgga 60

tgacctctgg atccatggac 80

<210> 97

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 97

gtaggtagtt aggatgaatg gaaggttggt atagcctatt attatattcc gcggattgga 60

tgacctctgg atccatggac 80

<210> 98

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 98

gtaggtagtt aggatgaatg gaaggttggt atagccatta ttattattac cgcggatgga 60

tgacctctgg atccatggac 80

<210> 99

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 99

gtaggtagtt aggatgaatg gaaggttggt atagccatta ttattattat taccgctgga 60

tgacctctgg atccatggac 80

<210> 100

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 100

gtaggtagtt aggatgaatg gaaggttggt atagtaatat taatttattt atttaatgga 60

tgacctctgg atccatggac 80

<210> 101

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 101

gtaggtagtt aggatgaatg gaaggttggt gtagattatt atctgcggat ttattggatg 60

acctctggat ccatggacat 80

<210> 102

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 102

gtaggtagtt aggatgaatg gaaggttggt atagccatta ttccgcggat ttattttgga 60

tgacctctgg atccatggac 80

<210> 103

<211> 33

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 103

auuauuccgc ggauuuauuu guuuuagagc uag 33

<210> 104

<211> 33

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 104

cguaggccag gguggucacg guuuuagagc uag 33

<210> 105

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 105

gtaggtagtt aggatgaatg gaaggttggt aactatgctg ccgcccagtg ggactttgga 60

aatacaatgt gtcaactctt 80

<210> 106

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 106

gtaggtagtt aggatgaatg gaaggttggt aaaatacaat gtgtcaactc ttgacagggc 60

tctattttat aggcttcttc 80

<210> 107

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 107

gtaggtagtt aggatgaatg gaaggttggt attttccata cagtcagtat caattctgga 60

agaatttcca gacattaaag 80

<210> 108

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 108

gtaggtagtt aggatgaatg gaaggttggt agcttcggtg tcgaaatgag aagaagaggc 60

acagggctgt gaggcttatc 80

<210> 109

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 109

gtaggtagtt aggatgaatg gaaggttggt agtgagccca gaaggggaca gtaagaagga 60

aaaacaggtc agagatggcc 80

<210> 110

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 110

gtaggtagtt aggatgaatg gaaggttggt atcctgaaca ccttccagga attctttggc 60

ctgaataatt gcagtagctc 80

<210> 111

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 111

gtaggtagtt aggatgaatg gaaggttggt agacatggag gacgtgcgcg gccgcctggt 60

gcagtaccgc ggcgaggtgc 80

<210> 112

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 112

gtaggtagtt aggatgaatg gaaggttggt actgcagaag cgcctggcag tgtaccaggc 60

cggggcccgc gagggcgccg 80

<210> 113

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 113

gtaggtagtt aggatgaatg gaaggttggt agccgacaag cagaagaacg gcatcaaggt 60

gaacttcaag atccgccaca 80

<210> 114

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 114

gtaggtagtt aggatgaatg gaaggttggt aaccactacc agcagaacac ccccatcggc 60

gacggccccg tgctgctgcc 80

<210> 115

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 115

gtaggtagtt aggatgaatg gaaggttggt acctcgccct tgctcaccat ctcgagtcgg 60

ccgccagtgt gatggatatc t 81

<210> 116

<211> 79

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 116

gtaggtagtt aggatgaatg gaaggttggt acacgcgtag gccagggtgg tcacgagggt 60

gggccagggc acgggcagc 79

<210> 117

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 117

gtaggtagtt aggatgaatg gaaggttggt aaagcactgc actccgcagg tcagggtggt 60

cacgagggtt ggccagggca 80

<210> 118

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 118

gtaggtagtt aggatgaatg gaaggttggt acactccgca ggtcagggtg gtcacgaggg 60

ttggccaggg cacgggcagg 80

<210> 119

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 119

gtaggtagtt aggatgaatg gaaggttggt aggatctctt cttcacggtt gcctactggt 60

tcaattactt ttaaaaatgg 80

<210> 120

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 120

gtaggtagtt aggatgaatg gaaggttggt attctcgatt gaggatctct tcttcacggt 60

tgcctactgg ttcaattact 80

<210> 121

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 121

gtaggtagtt aggatgaatg gaaggttggt aaggagctgt ggcagtggca ccagaatgga 60

ttccagagtc caggtaagac 80

<210> 122

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 122

gtaggtagtt aggatgaatg gaaggttggt agtactcctc ccggccggcg gtatccagga 60

tgtccaacag gcacgtctcc 80

<210> 123

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 123

gtaggtagtt aggatgaatg gaaggttggt atgactgctt gcagatggcc atggcgcgga 60

cgcgggtgcc gggcgggggt 80

<210> 124

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 124

gtaggtagtt aggatgaatg gaaggttggt actgttacac atgcagttgt agtggatggt 60

ggtacagtca gagccaacct 80

<210> 125

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 125

gtaggtagtt aggatgaatg gaaggttggt aggaactgtc acacatgtag ttgtagtgga 60

tggtggtaca gtcagagcca 80

<210> 126

<211> 1623

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 126

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala

995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu

1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1610 1615 1620

<210> 127

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 127

gagtccgagc agaagaagaa ggg 23

<210> 128

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 128

ggaatccctt ctgcagcacc tgg 23

<210> 129

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 129

gaacacaaag catagactgc ggg 23

<210> 130

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 130

ggcccagact gagcacgtga tgg 23

<210> 131

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 131

ggcactgcgg ctggaggtcc ggg 23

<210> 132

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 132

gtcatcttag tcattacctg agg 23

<210> 133

<211> 1710

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 133

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala

995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu

1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1610 1615 1620

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

1625 1630 1635

Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu

1640 1645 1650

Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1655 1660 1665

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1670 1675 1680

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1685 1690 1695

Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1700 1705 1710

<210> 134

<211> 84

<212> PRT

<213> Bacillus phage PBS2

<400> 134

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> 135

<211> 181

<212> PRT

<213> Erwinia tasmaniensis

<400> 135

Met Ala Ser Arg Gly Val Asn Lys Val Ile Leu Val Gly Asn Leu Gly

1 5 10 15

Gln Asp Pro Glu Val Arg Tyr Met Pro Asn Gly Gly Ala Val Ala Asn

20 25 30

Ile Thr Leu Ala Thr Ser Glu Ser Trp Arg Asp Lys Gln Thr Gly Glu

35 40 45

Thr Lys Glu Lys Thr Glu Trp His Arg Val Val Leu Phe Gly Lys Leu

50 55 60

Ala Glu Val Ala Gly Glu Tyr Leu Arg Lys Gly Ser Gln Val Tyr Ile

65 70 75 80

Glu Gly Ala Leu Gln Thr Arg Lys Trp Thr Asp Gln Ala Gly Val Glu

85 90 95

Lys Tyr Thr Thr Glu Val Val Val Asn Val Gly Gly Thr Met Gln Met

100 105 110

Leu Gly Gly Arg Ser Gln Gly Gly Gly Ala Ser Ala Gly Gly Gln Asn

115 120 125

Gly Gly Ser Asn Asn Gly Trp Gly Gln Pro Gln Gln Pro Gln Gly Gly

130 135 140

Asn Gln Phe Ser Gly Gly Ala Gln Gln Gln Ala Arg Pro Gln Gln Gln
145 150 155 160
Pro Gln Gln Asn Asn Ala Pro Ala Asn Asn Glu Pro Pro Ile Asp Phe
165 170 175
Asp Asp Asp Ile Pro
180

<210> 136
<211> 209
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 136

Met Ala Gly Ala Gln Asp Phe Val Pro His Thr Ala Asp Leu Ala Glu
1 5 10 15

Leu Ala Ala Ala Ala Gly Glu Cys Arg Gly Cys Gly Leu Tyr Arg Asp
20 25 30
Ala Thr Gln Ala Val Phe Gly Ala Gly Gly Arg Ser Ala Arg Ile Met
35 40 45
Met Ile Gly Glu Gln Pro Gly Asp Lys Glu Asp Leu Ala Gly Leu Pro
50 55 60
Phe Val Gly Pro Ala Gly Arg Leu Leu Asp Arg Ala Leu Glu Ala Ala
65 70 75 80

Asp Ile Asp Arg Asp Ala Leu Tyr Val Thr Asn Ala Val Lys His Phe
85 90 95
Lys Phe Thr Arg Ala Ala Gly Gly Lys Arg Arg Ile His Lys Thr Pro
100 105 110
Ser Arg Thr Glu Val Val Ala Cys Arg Pro Trp Leu Ile Ala Glu Met
115 120 125
Thr Ser Val Glu Pro Asp Val Val Val Leu Leu Gly Ala Thr Ala Ala
130 135 140

Lys Ala Leu Leu Gly Asn Asp Phe Arg Val Thr Gln His Arg Gly Glu
145 150 155 160
Val Leu His Val Asp Asp Val Pro Gly Asp Pro Ala Leu Val Ala Thr
165 170 175
Val His Pro Ser Ser Leu Leu Arg Gly Pro Lys Glu Glu Arg Glu Ser
180 185 190
Ala Phe Ala Gly Leu Val Asp Asp Leu Arg Val Ala Ala Asp Val Arg
195 200 205

Pro

<210> 137
<211> 313
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 137
Met Ile Gly Gln Lys Thr Leu Tyr Ser Phe Phe Ser Pro Ser Pro Ala
1 5 10 15
Arg Lys Arg His Ala Pro Ser Pro Glu Pro Ala Val Gln Gly Thr Gly
20 25 30
Val Ala Gly Val Pro Glu Glu Ser Gly Asp Ala Ala Ala Ile Pro Ala
35 40 45

Lys Lys Ala Pro Ala Gly Gln Glu Glu Pro Gly Thr Pro Pro Ser Ser
50 55 60
Pro Leu Ser Ala Glu Gln Leu Asp Arg Ile Gln Arg Asn Lys Ala Ala
65 70 75 80
Ala Leu Leu Arg Leu Ala Ala Arg Asn Val Pro Val Gly Phe Gly Glu
85 90 95
Ser Trp Lys Lys His Leu Ser Gly Glu Phe Gly Lys Pro Tyr Phe Ile
100 105 110

Lys Leu Met Gly Phe Val Ala Glu Glu Arg Lys His Tyr Thr Val Tyr
115 120 125

Pro Pro Pro His Gln Val Phe Thr Trp Thr Gln Met Cys Asp Ile Lys
130 135 140
Asp Val Lys Val Val Ile Leu Gly Gln Glu Pro Tyr His Gly Pro Asn
145 150 155 160
Gln Ala His Gly Leu Cys Phe Ser Val Gln Arg Pro Val Pro Pro Pro
165 170 175

Pro Ser Leu Glu Asn Ile Tyr Lys Glu Leu Ser Thr Asp Ile Glu Asp
180 185 190
Phe Val His Pro Gly His Gly Asp Leu Ser Gly Trp Ala Lys Gln Gly
195 200 205
Val Leu Leu Leu Asn Ala Val Leu Thr Val Arg Ala His Gln Ala Asn
210 215 220
Ser His Lys Glu Arg Gly Trp Glu Gln Phe Thr Asp Ala Val Val Ser
225 230 235 240

Trp Leu Asn Gln Asn Ser Asn Gly Leu Val Phe Leu Leu Trp Gly Ser
245 250 255
Tyr Ala Gln Lys Lys Gly Ser Ala Ile Asp Arg Lys Arg His His Val
260 265 270
Leu Gln Thr Ala His Pro Ser Pro Leu Ser Val Tyr Arg Gly Phe Phe
275 280 285
Gly Cys Arg His Phe Ser Lys Thr Asn Glu Leu Leu Gln Lys Ser Gly
290 295 300

Lys Lys Pro Ile Asp Trp Lys Glu Leu
305 310
<210> 138
<211> 148
<212> PRT
<213> Homo sapiens
<400> 138
Met Phe Arg Arg Pro Val Leu Gln Val Leu Arg Gln Phe Val Arg His
1 5 10 15
Glu Ser Glu Thr Thr Thr Ser Leu Val Leu Glu Arg Ser Leu Asn Arg

20 25 30

Val His Leu Leu Gly Arg Val Gly Gln Asp Pro Val Leu Arg Gln Val

35 40 45

Glu Gly Lys Asn Pro Val Thr Ile Phe Ser Leu Ala Thr Asn Glu Met

50 55 60

Trp Arg Ser Gly Asp Ser Glu Val Tyr Gln Leu Gly Asp Val Ser Gln

65 70 75 80

Lys Thr Thr Trp His Arg Ile Ser Val Phe Arg Pro Gly Leu Arg Asp

85 90 95

Val Ala Tyr Gln Tyr Val Lys Lys Gly Ser Arg Ile Tyr Leu Glu Gly

100 105 110

Lys Ile Asp Tyr Gly Glu Tyr Met Asp Lys Asn Asn Val Arg Arg Gln

115 120 125

Ala Thr Thr Ile Ile Ala Asp Asn Ile Ile Phe Leu Ser Asp Gln Thr

130 135 140

Lys Glu Lys Glu

145

<210> 139

<211> 388

<212> PRT

<213> Mus musculus

<400> 139

Met Phe Ala Lys Gly Lys Gly Ser Ala Val Pro Ser Asp Gly Gln Ala

1 5 10 15

Arg Glu Lys Leu Ala Leu Tyr Val Tyr Glu Tyr Leu Leu His Val Gly

20 25 30

Ala Gln Lys Ser Ala Gln Thr Phe Leu Ser Glu Ile Arg Trp Glu Lys

35 40 45

Asn Ile Thr Leu Gly Glu Pro Pro Gly Phe Leu His Ser Trp Trp Cys

50 55 60

Val Phe Trp Asp Leu Tyr Cys Ala Ala Pro Glu Arg Arg Asp Thr Cys

65 70 75 80

Glu His Ser Ser Glu Ala Lys Ala Phe His Asp Tyr Ser Ala Ala Ala

85 90 95

Ala Pro Ser Pro Val Leu Gly Asn Ile Pro Pro Asn Asp Gly Met Pro

100 105 110

Gly Gly Pro Ile Pro Pro Gly Phe Phe Gln Gly Pro Pro Gly Ser Gln

115 120 125

Pro Ser Pro His Ala Gln Pro Pro Pro His Asn Pro Ser Ser Met Met

130 135 140

Gly Pro His Ser Gln Pro Phe Met Ser Pro Arg Tyr Ala Gly Gly Pro

145 150 155 160

Arg Pro Pro Ile Arg Met Gly Asn Gln Pro Pro Gly Gly Val Pro Gly

165 170 175

Thr Gln Pro Leu Leu Pro Asn Ser Met Asp Pro Thr Arg Gln Gln Gly

180 185 190

His Pro Asn Met Gly Gly Ser Met Gln Arg Met Asn Pro Pro Arg Gly

195 200 205

Met Gly Pro Met Gly Pro Gly Pro Gln Asn Tyr Gly Ser Gly Met Arg

210 215 220

Pro Pro Pro Asn Ser Leu Gly Pro Ala Met Pro Gly Ile Asn Met Gly

225 230 235 240

Pro Gly Ala Gly Arg Pro Trp Pro Asn Pro Asn Ser Ala Asn Ser Ile

245 250 255

Pro Tyr Ser Ser Ser Ser Pro Gly Thr Tyr Val Gly Pro Pro Gly Gly

260 265 270

Gly Gly Pro Pro Gly Thr Pro Ile Met Pro Ser Pro Ala Asp Ser Thr

275 280 285

Asn Ser Ser Asp Asn Ile Tyr Thr Met Ile Asn Pro Val Pro Pro Gly

290 295 300

Gly Ser Arg Ser Asn Phe Pro Met Gly Pro Gly Ser Asp Gly Pro Met

305 310 315 320

Gly Gly Met Gly Gly Met Glu Pro His His Met Asn Gly Ser Leu Gly

325 330 335

Ser Gly Asp Ile Asp Gly Leu Pro Lys Asn Ser Pro Asn Asn Ile Ser

340 345 350

Gly Ile Ser Asn Pro Pro Gly Thr Pro Arg Asp Asp Gly Glu Leu Gly

355 360 365

Gly Asn Phe Leu His Ser Phe Gln Asn Asp Asn Tyr Ser Pro Ser Met

370 375 380

Thr Met Ser Val

385

<210> 140

<211> 616

<212> PRT

<213> Homo sapiens

<400> 140

Met Val Gly Gln Leu Ser Glu Gly Ala Ile Ala Ala Ile Met Gln Lys

1 5 10 15

Gly Asp Thr Asn Ile Lys Pro Ile Leu Gln Val Ile Asn Ile Arg Pro

20 25 30

Ile Thr Thr Gly Asn Ser Pro Pro Arg Tyr Arg Leu Leu Met Ser Asp

35 40 45

Gly Leu Asn Thr Leu Ser Ser Phe Met Leu Ala Thr Gln Leu Asn Pro

50 55 60

Leu Val Glu Glu Glu Gln Leu Ser Ser Asn Cys Val Cys Gln Ile His

65 70 75 80

Arg Phe Ile Val Asn Thr Leu Lys Asp Gly Arg Arg Val Val Ile Leu

85 90 95

Met Glu Leu Glu Val Leu Lys Ser Ala Glu Ala Val Gly Val Lys Ile

100 105 110

Gly Asn Pro Val Pro Tyr Asn Glu Gly Leu Gly Gln Pro Gln Val Ala

115 120 125

Pro Pro Ala Pro Ala Ala Ser Pro Ala Ala Ser Ser Arg Pro Gln Pro

130 135 140

Gln Asn Gly Ser Ser Gly Met Gly Ser Thr Val Ser Lys Ala Tyr Gly
145 150 155 160
Ala Ser Lys Thr Phe Gly Lys Ala Ala Gly Pro Ser Leu Ser His Thr
165 170 175
Ser Gly Gly Thr Gln Ser Lys Val Val Pro Ile Ala Ser Leu Thr Pro
180 185 190
Tyr Gln Ser Lys Trp Thr Ile Cys Ala Arg Val Thr Asn Lys Ser Gln
195 200 205

Ile Arg Thr Trp Ser Asn Ser Arg Gly Glu Gly Lys Leu Phe Ser Leu
210 215 220
Glu Leu Val Asp Glu Ser Gly Glu Ile Arg Ala Thr Ala Phe Asn Glu
225 230 235 240
Gln Val Asp Lys Phe Phe Pro Leu Ile Glu Val Asn Lys Val Tyr Tyr
245 250 255
Phe Ser Lys Gly Thr Leu Lys Ile Ala Asn Lys Gln Phe Thr Ala Val
260 265 270

Lys Asn Asp Tyr Glu Met Thr Phe Asn Asn Glu Thr Ser Val Met Pro
275 280 285
Cys Glu Asp Asp His His Leu Pro Thr Val Gln Phe Asp Phe Thr Gly
290 295 300
Ile Asp Asp Leu Glu Asn Lys Ser Lys Asp Ser Leu Val Asp Ile Ile
305 310 315 320
Gly Ile Cys Lys Ser Tyr Glu Asp Ala Thr Lys Ile Thr Val Arg Ser
325 330 335

Asn Asn Arg Glu Val Ala Lys Arg Asn Ile Tyr Leu Met Asp Thr Ser
340 345 350
Gly Lys Val Val Thr Ala Thr Leu Trp Gly Glu Asp Ala Asp Lys Phe
355 360 365
Asp Gly Ser Arg Gln Pro Val Leu Ala Ile Lys Gly Ala Arg Val Ser
370 375 380

Asp Phe Gly Gly Arg Ser Leu Ser Val Leu Ser Ser Ser Thr Ile Ile
385 390 395 400

Ala Asn Pro Asp Ile Pro Glu Ala Tyr Lys Leu Arg Gly Trp Phe Asp
405 410 415
Ala Glu Gly Gln Ala Leu Asp Gly Val Ser Ile Ser Asp Leu Lys Ser
420 425 430
Gly Gly Val Gly Gly Ser Asn Thr Asn Trp Lys Thr Leu Tyr Glu Val
435 440 445
Lys Ser Glu Asn Leu Gly Gln Gly Asp Lys Pro Asp Tyr Phe Ser Ser
450 455 460

Val Ala Thr Val Val Tyr Leu Arg Lys Glu Asn Cys Met Tyr Gln Ala
465 470 475 480
Cys Pro Thr Gln Asp Cys Asn Lys Lys Val Ile Asp Gln Gln Asn Gly
485 490 495
Leu Tyr Arg Cys Glu Lys Cys Asp Thr Glu Phe Pro Asn Phe Lys Tyr
500 505 510
Arg Met Ile Leu Ser Val Asn Ile Ala Asp Phe Gln Glu Asn Gln Trp
515 520 525

Val Thr Cys Phe Gln Glu Ser Ala Glu Ala Ile Leu Gly Gln Asn Ala
530 535 540
Ala Tyr Leu Gly Glu Leu Lys Asp Lys Asn Glu Gln Ala Phe Glu Glu
545 550 555 560
Val Phe Gln Asn Ala Asn Phe Arg Ser Phe Ile Phe Arg Val Arg Val
565 570 575
Lys Val Glu Thr Tyr Asn Asp Glu Ser Arg Ile Lys Ala Thr Val Met
580 585 590

Asp Val Lys Pro Val Asp Tyr Arg Glu Tyr Gly Arg Arg Leu Val Met
595 600 605
Ser Ile Arg Arg Ser Ala Leu Met
610 615
<210> 141

<211> 270

<212> PRT

<213> Homo sapiens

<400> 141

Met Trp Asn Ser Gly Phe Glu Ser Tyr Gly Ser Ser Ser Tyr Gly Gly
1 5 10 15

Ala Gly Gly Tyr Thr Gln Ser Pro Gly Gly Phe Gly Ser Pro Ala Pro
20 25 30

Ser Gln Ala Glu Lys Lys Ser Arg Ala Arg Ala Gln His Ile Val Pro
35 40 45

Cys Thr Ile Ser Gln Leu Leu Ser Ala Thr Leu Val Asp Glu Val Phe
50 55 60

Arg Ile Gly Asn Val Glu Ile Ser Gln Val Thr Ile Val Gly Ile Ile
65 70 75 80

Arg His Ala Glu Lys Ala Pro Thr Asn Ile Val Tyr Lys Ile Asp Asp
85 90 95

Met Thr Ala Ala Pro Met Asp Val Arg Gln Trp Val Asp Thr Asp Asp
100 105 110

Thr Ser Ser Glu Asn Thr Val Val Pro Pro Glu Thr Tyr Val Lys Val
115 120 125

Ala Gly His Leu Arg Ser Phe Gln Asn Lys Lys Ser Leu Val Ala Phe
130 135 140

Lys Ile Met Pro Leu Glu Asp Met Asn Glu Phe Thr Thr His Ile Leu
145 150 155 160

Glu Val Ile Asn Ala His Met Val Leu Ser Lys Ala Asn Ser Gln Pro
165 170 175

Ser Ala Gly Arg Ala Pro Ile Ser Asn Pro Gly Met Ser Glu Ala Gly
180 185 190

Asn Phe Gly Gly Asn Ser Phe Met Pro Ala Asn Gly Leu Thr Val Ala
195 200 205

Gln Asn Gln Val Leu Asn Leu Ile Lys Ala Cys Pro Arg Pro Glu Gly

210 215 220

Leu Asn Phe Gln Asp Leu Lys Asn Gln Leu Lys His Met Ser Val Ser
225 230 235 240
Ser Ile Lys Gln Ala Val Asp Phe Leu Ser Asn Glu Gly His Ile Tyr
245 250 255
Ser Thr Val Asp Asp Asp His Phe Lys Ser Thr Asp Ala Glu
260 265 270

<210> 142
<211> 121
<212> PRT
<213> Homo sapiens
<400> 142

Met Val Asp Met Met Asp Leu Pro Arg Ser Arg Ile Asn Ala Gly Met
1 5 10 15
Leu Ala Gln Phe Ile Asp Lys Pro Val Cys Phe Val Gly Arg Leu Glu
20 25 30
Lys Ile His Pro Thr Gly Lys Met Phe Ile Leu Ser Asp Gly Glu Gly
35 40 45
Lys Asn Gly Thr Ile Glu Leu Met Glu Pro Leu Asp Glu Glu Ile Ser
50 55 60
Gly Ile Val Glu Val Val Gly Arg Val Thr Ala Lys Ala Thr Ile Leu
65 70 75 80
Cys Thr Ser Tyr Val Gln Phe Lys Glu Asp Ser His Pro Phe Asp Leu
85 90 95
Gly Leu Tyr Asn Glu Ala Val Lys Ile Ile His Asp Phe Pro Gln Phe
100 105 110
Tyr Pro Leu Gly Ile Val Gln His Asp
115 120

<210> 143
<211> 171
<212> PRT
<213> Bacillus subtilis

<400> 143

Met Leu Asn Arg Val Val Leu Val Gly Arg Leu Thr Lys Asp Pro Glu

1 5 10 15

Leu Arg Tyr Thr Pro Asn Gly Ala Ala Val Ala Thr Phe Thr Leu Ala

20 25 30

Val Asn Arg Thr Phe Thr Asn Gln Gly Glu Arg Glu Ala Asp Phe Ile

35 40 45

Asn Cys Val Thr Trp Arg Arg Gln Ala Glu Asn Val Ala Asn Phe Leu

50 55 60

Lys Lys Gly Ser Leu Ala Gly Val Asp Gly Arg Leu Gln Thr Arg Asn

65 70 75 80

Tyr Glu Asn Gln Gln Gly Gln Arg Val Phe Val Thr Glu Val Gln Ala

85 90 95

Glu Ser Val Gln Phe Leu Glu Pro Lys Asn Gly Gly Gly Ser Gly Ser

100 105 110

Gly Gly Tyr Asn Glu Gly Asn Ser Gly Gly Gly Gln Tyr Phe Gly Gly

115 120 125

Gly Gln Asn Asp Asn Pro Phe Gly Gly Asn Gln Asn Asn Gln Arg Arg

130 135 140

Asn Gln Gly Asn Ser Phe Asn Asp Asp Pro Phe Ala Asn Asp Gly Lys

145 150 155 160

Pro Ile Asp Ile Ser Asp Asp Asp Leu Pro Phe

165 170

<210> 144

<211> 199

<212> PRT

<213> Streptomyces coelicolor

<400> 144

Met Ala Gly Glu Thr Val Ile Thr Val Val Gly Asn Leu Val Asp Asp

1 5 10 15

Pro Glu Leu Arg Phe Thr Pro Ser Gly Ala Ala Val Ala Lys Phe Arg

20 25 30

Val Ala Ser Thr Pro Arg Thr Phe Asp Arg Gln Thr Asn Glu Trp Lys

35 40 45

Asp Gly Glu Ser Leu Phe Leu Thr Cys Ser Val Trp Arg Gln Ala Ala

50 55 60

Glu Asn Val Ala Glu Ser Leu Gln Arg Gly Met Arg Val Ile Val Gln

65 70 75 80

Gly Arg Leu Lys Gln Arg Ser Tyr Glu Asp Arg Glu Gly Val Lys Arg

85 90 95

Thr Val Tyr Glu Leu Asp Val Asp Glu Val Gly Ala Ser Leu Arg Ser

100 105 110

Ala Thr Ala Lys Val Thr Lys Thr Ser Gly Gln Gly Arg Gly Gly Gln

115 120 125

Gly Gly Tyr Gly Gly Gly Gly Gly Gly Gln Gly Gly Gly Gly Trp Gly

130 135 140

Gly Gly Pro Gly Gly Gly Gln Gln Gly Gly Gly Ala Pro Ala Asp Asp

145 150 155 160

Pro Trp Ala Thr Gly Gly Ala Pro Ala Gly Gly Gln Gln Gly Gly Gly

165 170 175

Gly Gln Gly Gly Gly Gly Trp Gly Gly Gly Ser Gly Gly Gly Gly Gly

180 185 190

Tyr Ser Asp Glu Pro Pro Phe

195

<210> 145

<211> 156

<212> PRT

<213> Streptococcus pneumoniae

<400> 145

Met Ile Asn Asn Val Val Leu Val Gly Arg Met Thr Arg Asp Ala Glu

1 5 10 15

Leu Arg Tyr Thr Pro Ser Asn Val Ala Val Ala Thr Phe Thr Leu Ala

20 25 30

Val Asn Arg Thr Phe Lys Ser Gln Asn Gly Glu Arg Glu Ala Asp Phe
35 40 45
Ile Asn Val Val Met Trp Arg Gln Gln Ala Glu Asn Leu Ala Asn Trp
50 55 60
Ala Lys Lys Gly Ser Leu Ile Gly Val Thr Gly Arg Ile Gln Thr Arg

65 70 75 80
Ser Tyr Asp Asn Gln Gln Gly Gln Arg Val Tyr Val Thr Glu Val Val
85 90 95
Ala Glu Asn Phe Gln Met Leu Glu Ser Arg Ser Val Arg Glu Gly His
100 105 110
Thr Gly Gly Ala Tyr Ser Ala Pro Thr Ala Asn Tyr Ser Ala Pro Thr
115 120 125
Asn Ser Val Pro Asp Phe Ser Arg Asn Glu Asn Pro Phe Gly Ala Thr

130 135 140
Asn Pro Leu Asp Ile Ser Asp Asp Asp Leu Pro Phe
145 150 155
<210> 146
<211> 1196
<212> PRT
<213> Human alphaherpesvirus
<400> 146
Met Glu Thr Lys Pro Lys Thr Ala Thr Thr Ile Lys Val Pro Pro Gly
1 5 10 15
Pro Leu Gly Tyr Val Tyr Ala Arg Ala Cys Pro Ser Glu Gly Ile Glu
20 25 30

Leu Leu Ala Leu Leu Ser Ala Arg Ser Gly Asp Ser Asp Val Ala Val
35 40 45
Ala Pro Leu Val Val Gly Leu Thr Val Glu Ser Gly Phe Glu Ala Asn
50 55 60
Val Ala Val Val Val Gly Ser Arg Thr Thr Gly Leu Gly Gly Thr Ala
65 70 75 80
Val Ser Leu Lys Leu Thr Pro Ser His Tyr Ser Ser Ser Val Tyr Val

85 90 95

Phe His Gly Gly Arg His Leu Asp Pro Ser Thr Gln Ala Pro Asn Leu
100 105 110
Thr Arg Leu Cys Glu Arg Ala Arg Arg His Phe Gly Phe Ser Asp Tyr
115 120 125
Thr Pro Arg Pro Gly Asp Leu Lys His Glu Thr Thr Gly Glu Ala Leu
130 135 140
Cys Glu Arg Leu Gly Leu Asp Pro Asp Arg Ala Leu Leu Tyr Leu Val
145 150 155 160

Val Thr Glu Gly Phe Lys Glu Ala Val Cys Ile Asn Asn Thr Phe Leu
165 170 175
His Leu Gly Gly Ser Asp Lys Val Thr Ile Gly Gly Ala Glu Val His
180 185 190
Arg Ile Pro Val Tyr Pro Leu Gln Leu Phe Met Pro Asp Phe Ser Arg
195 200 205
Val Ile Ala Glu Pro Phe Asn Ala Asn His Arg Ser Ile Gly Glu Asn
210 215 220

Phe Thr Tyr Pro Leu Pro Phe Phe Asn Arg Pro Leu Asn Arg Leu Leu
225 230 235 240
Phe Glu Ala Val Val Gly Pro Ala Ala Val Ala Leu Arg Cys Arg Asn
245 250 255
Val Asp Ala Val Ala Arg Ala Ala Ala His Leu Ala Phe Asp Glu Asn
260 265 270
His Glu Gly Ala Ala Leu Pro Ala Asp Ile Thr Phe Thr Ala Phe Glu
275 280 285

Ala Ser Gln Gly Lys Thr Pro Arg Gly Gly Arg Asp Gly Gly Gly Lys
290 295 300
Gly Pro Ala Gly Gly Phe Glu Gln Arg Leu Ala Ser Val Met Ala Gly
305 310 315 320
Asp Ala Ala Leu Ala Leu Glu Ser Ile Val Ser Met Ala Val Phe Asp
325 330 335

Glu Pro Pro Thr Asp Ile Ser Ala Trp Pro Leu Phe Glu Gly Gln Asp
340 345 350

Thr Ala Ala Ala Arg Ala Asn Ala Val Gly Ala Tyr Leu Ala Arg Ala
355 360 365
Ala Gly Leu Val Gly Ala Met Val Phe Ser Thr Asn Ser Ala Leu His
370 375 380
Leu Thr Glu Val Asp Asp Ala Gly Pro Ala Asp Pro Lys Asp His Ser
385 390 395 400
Lys Pro Ser Phe Tyr Arg Phe Phe Leu Val Pro Gly Thr His Val Ala
405 410 415

Ala Asn Pro Gln Val Asp Arg Glu Gly His Val Val Pro Gly Phe Glu
420 425 430
Gly Arg Pro Thr Ala Pro Leu Val Gly Gly Thr Gln Glu Phe Ala Gly
435 440 445
Glu His Leu Ala Met Leu Cys Gly Phe Ser Pro Ala Leu Leu Ala Lys
450 455 460
Met Leu Phe Tyr Leu Glu Arg Cys Asp Gly Gly Val Ile Val Gly Arg
465 470 475 480

Gln Glu Met Asp Val Phe Arg Tyr Val Ala Asp Ser Asn Gln Thr Asp
485 490 495
Val Pro Cys Asn Leu Cys Thr Phe Asp Thr Arg His Ala Cys Val His
500 505 510
Thr Thr Leu Met Arg Leu Arg Ala Arg His Pro Lys Phe Ala Ser Ala
515 520 525
Ala Arg Gly Ala Ile Gly Val Phe Gly Thr Met Asn Ser Met Tyr Ser
530 535 540

Asp Cys Asp Val Leu Gly Asn Tyr Ala Ala Phe Ser Ala Leu Lys Arg
545 550 555 560
Ala Asp Gly Ser Glu Thr Ala Arg Thr Ile Met Gln Glu Thr Tyr Arg
565 570 575
Ala Ala Thr Glu Arg Val Met Ala Glu Leu Glu Thr Leu Gln Tyr Val

580 585 590

Asp Gln Ala Val Pro Thr Ala Met Gly Arg Leu Glu Thr Ile Ile Thr

595 600 605

Asn Arg Glu Ala Leu His Thr Val Val Asn Asn Val Arg Gln Val Val

610 615 620

Asp Arg Glu Val Glu Gln Leu Met Arg Asn Leu Val Glu Gly Arg Asn

625 630 635 640

Phe Lys Phe Arg Asp Gly Leu Gly Glu Ala Asn His Ala Met Ser Leu

645 650 655

Thr Leu Asp Pro Tyr Ala Cys Gly Pro Cys Pro Leu Leu Gln Leu Leu

660 665 670

Gly Arg Arg Ser Asn Leu Ala Val Tyr Gln Asp Leu Ala Leu Ser Gln

675 680 685

Cys His Gly Val Phe Ala Gly Gln Ser Val Glu Gly Arg Asn Phe Arg

690 695 700

Asn Gln Phe Gln Pro Val Leu Arg Arg Arg Val Met Asp Met Phe Asn

705 710 715 720

Asn Gly Phe Leu Ser Ala Lys Thr Leu Thr Val Ala Leu Ser Glu Gly

725 730 735

Ala Ala Ile Cys Ala Pro Ser Leu Thr Ala Gly Gln Thr Ala Pro Ala

740 745 750

Glu Ser Ser Phe Glu Gly Asp Val Ala Arg Val Thr Leu Gly Phe Pro

755 760 765

Lys Glu Leu Arg Val Lys Ser Arg Val Leu Phe Ala Gly Ala Ser Ala

770 775 780

Asn Ala Ser Glu Ala Ala Lys Ala Arg Val Ala Ser Leu Gln Ser Ala

785 790 795 800

Tyr Gln Lys Pro Asp Lys Arg Val Asp Ile Leu Leu Gly Pro Leu Gly

805 810 815

Phe Leu Leu Lys Gln Phe His Ala Ala Ile Phe Pro Asn Gly Lys Pro

820 825 830

Pro Gly Ser Asn Gln Pro Asn Pro Gln Trp Phe Trp Thr Ala Leu Gln
835 840 845
Arg Asn Gln Leu Pro Ala Arg Leu Leu Ser Arg Glu Asp Ile Glu Thr
850 855 860

Ile Ala Phe Ile Lys Lys Phe Ser Leu Asp Tyr Gly Ala Ile Asn Phe
865 870 875 880
Ile Asn Leu Ala Pro Asn Asn Val Ser Glu Leu Ala Met Tyr Tyr Met
885 890 895
Ala Asn Gln Ile Leu Arg Tyr Cys Asp His Ser Thr Tyr Phe Ile Asn
900 905 910
Thr Leu Thr Ala Ile Ile Ala Gly Ser Arg Arg Pro Pro Ser Val Gln
915 920 925

Ala Ala Ala Ala Trp Ser Ala Gln Gly Gly Ala Gly Leu Glu Ala Gly
930 935 940
Ala Arg Ala Leu Met Asp Ala Val Asp Ala His Pro Gly Ala Trp Thr
945 950 955 960
Ser Met Phe Ala Ser Cys Asn Leu Leu Arg Pro Val Met Ala Ala Arg
965 970 975
Pro Met Val Val Leu Gly Leu Ser Ile Ser Lys Tyr Tyr Gly Met Ala
980 985 990

Gly Asn Asp Arg Val Phe Gln Ala Gly Asn Trp Ala Ser Leu Met Gly
995 1000 1005
Gly Lys Asn Ala Cys Pro Leu Leu Ile Phe Asp Arg Thr Arg Lys
1010 1015 1020
Phe Val Leu Ala Cys Pro Arg Ala Gly Phe Val Cys Ala Ala Ser
1025 1030 1035
Ser Leu Gly Gly Gly Ala His Glu Ser Ser Leu Cys Glu Gln Leu
1040 1045 1050

Arg Gly Ile Ile Ser Glu Gly Gly Ala Ala Val Ala Ser Ser Val
1055 1060 1065
Phe Val Ala Thr Val Lys Ser Leu Gly Pro Arg Thr Gln Gln Leu

1070 1075 1080

Gln Ile Glu Asp Trp Leu Ala Leu Leu Glu Asp Glu Tyr Leu Ser

1085 1090 1095

Glu Glu Met Met Glu Leu Thr Ala Arg Ala Leu Glu Arg Gly Asn

1100 1105 1110

Gly Glu Trp Ser Thr Asp Ala Ala Leu Glu Val Ala His Glu Ala

1115 1120 1125

Glu Ala Leu Val Ser Gln Leu Gly Asn Ala Gly Glu Val Phe Asn

1130 1135 1140

Phe Gly Asp Phe Gly Cys Glu Asp Asp Asn Ala Thr Pro Phe Gly

1145 1150 1155

Gly Pro Gly Ala Pro Gly Pro Ala Phe Ala Gly Arg Lys Arg Ala

1160 1165 1170

Phe His Gly Asp Asp Pro Phe Gly Glu Gly Pro Pro Asp Lys Lys

1175 1180 1185

Gly Asp Leu Thr Leu Asp Met Leu

1190 1195

<210> 147

<211> 124

<212> PRT

<213> Bacillus virus phi29

<400> 147

Met Glu Asn Thr Asn Ile Val Lys Ala Thr Phe Asp Thr Glu Thr Leu

1 5 10 15

Glu Gly Gln Ile Lys Ile Phe Asn Ala Gln Thr Gly Gly Gly Gln Ser

20 25 30

Phe Lys Asn Leu Pro Asp Gly Thr Ile Ile Glu Ala Asn Ala Ile Ala

35 40 45

Gln Tyr Lys Gln Val Ser Asp Thr Tyr Gly Asp Ala Lys Glu Glu Thr

50 55 60

Val Thr Thr Ile Phe Ala Ala Asp Gly Ser Leu Tyr Ser Ala Ile Ser

65 70 75 80

Lys Thr Val Ala Glu Ala Ala Ser Asp Leu Ile Asp Leu Val Thr Arg
85 90 95

His Lys Leu Glu Thr Phe Lys Val Lys Val Val Gln Gly Thr Ser Ser
100 105 110
Lys Gly Asn Val Phe Phe Ser Leu Gln Leu Ser Leu
115 120

<210> 148
<211> 162
<212> PRT
<213> Burkholderia virus
<400> 148

Met Ala Ser Val Asn Lys Val Ile Leu Val Gly Asn Leu Gly Ala Asp
1 5 10 15
Pro Glu Thr Arg Tyr Leu Pro Ser Gly Asp Ala Ile Ser Asn Ile Arg

20 25 30
Leu Ala Thr Thr Asp Arg Tyr Lys Asp Lys Ala Ser Gly Glu Met Lys
35 40 45
Glu Ser Thr Glu Trp His Arg Val Ser Phe Phe Gly Arg Leu Ala Glu
50 55 60
Ile Val Asp Glu Tyr Leu Arg Lys Gly Ala Pro Val Tyr Ile Glu Gly
65 70 75 80
Arg Ile Arg Thr Arg Lys Trp Gln Asp Asn Ala Gly Gln Asp Arg Tyr

85 90 95
Thr Thr Glu Ile Val Ala Glu Lys Met Gln Met Leu Gly Asp Arg Arg
100 105 110
Asp Gly Gly Glu Arg Gln Gln Arg Ala Pro Gln Gln Gln Gln Gln Arg
115 120 125
Thr Gln Arg Asn Gly Tyr Ala Asp Ala Thr Gly Arg Ala Gln Pro Ser
130 135 140
Gln Arg Pro Ala Ala Gly Gly Gly Phe Asp Glu Met Asp Asp Asp Ile

145 150 155 160

Pro Phe

<210> 149

<211> 1599

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 149

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu

225 230 235 240
Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
245 250 255
Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
260 265 270
Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
275 280 285
Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr

290 295 300
Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
305 310 315 320
Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
325 330 335
Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
340 345 350
Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His

355 360 365
Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
370 375 380
Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
385 390 395 400
Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
405 410 415
Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile

420 425 430
Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser

435 440 445

Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys

450 455 460

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr

465 470 475 480

Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln

485 490 495

Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln

500 505 510

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser

515 520 525

Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr

530 535 540

Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His

545 550 555 560

Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu

565 570 575

Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly

580 585 590

Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys

595 600 605

Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu

610 615 620

Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser

625 630 635 640

Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg

645 650 655

Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu

660 665 670

Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg

675 680 685

Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
690 695 700
Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln
705 710 715 720
Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
725 730 735
Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr

740 745 750
Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
755 760 765
Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe
770 775 780
Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe
785 790 795 800
Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp

805 810 815
Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile
820 825 830
Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
835 840 845
Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
850 855 860
Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys

865 870 875 880
Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
885 890 895
Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
900 905 910
Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
915 920 925
His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val

930 935 940

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly

945 950 955 960

Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp

965 970 975

Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile

980 985 990

Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser

995 1000 1005

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly

1010 1015 1020

Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln

1025 1030 1035

Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met

1040 1045 1050

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1055 1060 1065

Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

1070 1075 1080

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser

1085 1090 1095

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr

1100 1105 1110

Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

1115 1120 1125

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

1130 1135 1140

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile

1145 1150 1155

Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

1160 1165 1170

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr

1175 1180 1185

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe

1190 1195 1200

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala

1205 1210 1215

Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro

1220 1225 1230

Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1235 1240 1245

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1250 1255 1260

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1265 1270 1275

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1280 1285 1290

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1295 1300 1305

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1310 1315 1320

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1325 1330 1335

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1340 1345 1350

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1355 1360 1365

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1370 1375 1380

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1385 1390 1395

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1400 1405 1410

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1415 1420 1425

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1430 1435 1440

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1445 1450 1455

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1460 1465 1470

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1475 1480 1485

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1490 1495 1500

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1505 1510 1515

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1520 1525 1530

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1535 1540 1545

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1550 1555 1560

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1565 1570 1575

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1580 1585 1590

Ser Gln Leu Gly Gly Asp

1595

<210> 150

<211> 1606

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 150

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220

Ala Thr Gly Leu Lys Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Asp

225 230 235 240

Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp
245 250 255
Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val
260 265 270
Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala
275 280 285
Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg
290 295 300

Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu
305 310 315 320
Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe
325 330 335
His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu
340 345 350
Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu
355 360 365

Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr
370 375 380
Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile
385 390 395 400
Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn
405 410 415
Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln
420 425 430

Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala
435 440 445
Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile
450 455 460
Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile
465 470 475 480

Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu
485 490 495

Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp
500 505 510
Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe
515 520 525
Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu
530 535 540
Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile
545 550 555 560

Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu
565 570 575
Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln
580 585 590
Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu
595 600 605
Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr
610 615 620

Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln
625 630 635 640
Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu
645 650 655
Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys
660 665 670
Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr
675 680 685

Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr
690 695 700
Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val
705 710 715 720
Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe

725 730 735

Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu

740 745 750

Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val

755 760 765

Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys

770 775 780

Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys

785 790 795 800

Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val

805 810 815

Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr

820 825 830

His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu

835 840 845

Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe

850 855 860

Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu

865 870 875 880

Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly

885 890 895

Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln

900 905 910

Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn

915 920 925

Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu

930 935 940

Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu

945 950 955 960

His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu

965 970 975

Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His
980 985 990
Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr
995 1000 1005

Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
1010 1015 1020
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
1025 1030 1035
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr
1040 1045 1050
Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
1055 1060 1065

Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser
1070 1075 1080
Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser
1085 1090 1095
Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val
1100 1105 1110
Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys
1115 1120 1125

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
1130 1135 1140
Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln
1145 1150 1155
Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu
1160 1165 1170
Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile
1175 1180 1185

Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp
1190 1195 1200
Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn

1205 1210 1215

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1220 1225 1230

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1235 1240 1245

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1250 1255 1260

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1265 1270 1275

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly

1280 1285 1290

Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly

1295 1300 1305

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys

1310 1315 1320

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val

1325 1330 1335

Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn

1340 1345 1350

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys

1355 1360 1365

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val

1370 1375 1380

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val

1385 1390 1395

Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu

1400 1405 1410

Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

1415 1420 1425

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu

1430 1435 1440

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu
1445 1450 1455
Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
1460 1465 1470
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu
1475 1480 1485

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1490 1495 1500
Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1505 1510 1515
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1520 1525 1530
Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1535 1540 1545

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1550 1555 1560
Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1565 1570 1575
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1580 1585 1590
Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1595 1600 1605

<210>
151
<211> 1599
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 151
Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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 Gly

1355 1360 1365

Gly Ser Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr

1370 1375 1380

Leu Arg Arg Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp

1385 1390 1395

Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn

1400 1405 1410

Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser Gln Asn Thr

1415 1420 1425

Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe Thr Thr Glu

1430 1435 1440

Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp Phe Leu

1445 1450 1455

Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu Phe

1460 1465 1470

Leu Ser Arg Tyr Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg

1475 1480 1485

Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln Gly Leu Arg Asp

1490 1495 1500

Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr Glu Gln Glu

1505 1510 1515

Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser Pro Ser Asn

1520 1525 1530

Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg Leu Tyr

1535 1540 1545

Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu

1550 1555 1560

Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

1565 1570 1575

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu

1580 1585 1590

Trp Ala Thr Gly Leu Lys

1595

<210> 152

<211> 1605

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 152

Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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 Gly
1355 1360 1365
Gly Ser Gly Gly Ser Gly Gly Ser Met Ser Ser Glu Thr Gly Pro
1370 1375 1380
Val Ala Val Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu

1385 1390 1395
Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys
1400 1405 1410
Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg
1415 1420 1425
His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val Asn Phe Ile
1430 1435 1440
Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys

1445 1450 1455
Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser
1460 1465 1470
Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
1475 1480 1485
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn
1490 1495 1500
Arg Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln

1505 1510 1515
Ile Met Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val

1520 1525 1530

Asn Tyr Ser Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His

1535 1540 1545

Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu

1550 1555 1560

Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln

1565 1570 1575

Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg

1580 1585 1590

Leu Pro Pro His Ile Leu Trp Ala Thr Gly Leu Lys

1595 1600 1605

<210> 153

<211> 1612

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 153

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510
Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560
Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895
Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
915 920 925
Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990
Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005
Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080
Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1340 1345 1350
Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1355 1360 1365
Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1370 1375 1380
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1400 1405 1410
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1415 1420 1425
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1430 1435 1440
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1460 1465 1470
Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545
Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605
Leu Gly Gly Asp
1610

<210> 154
<211> 772
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 154

Met Gly Ser Ser His His His His His His Met Ser Ser Glu Thr Gly
1 5 10 15

Pro Val Ala Val Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu
20 25 30
Phe Glu Val Glu Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu
35 40 45
Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr
50 55 60
Ser Gln Asn Thr Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe
65 70 75 80

Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp
85 90 95
Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu
100 105 110
Phe Leu Ser Arg Tyr Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg
115 120 125
Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln Gly Leu Val Lys Val
130 135 140

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu
145 150 155 160
Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
165 170 175
Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu
180 185 190
His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr
195 200 205

Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
210 215 220
Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe
225 230 235 240
Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys
245 250 255
Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys
260 265 270

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln
275 280 285
Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
290 295 300
Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln
305 310 315 320
Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
325 330 335

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
340 345 350
Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
355 360 365
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
370 375 380
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser

385 390 395 400

Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
405 410 415
Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
420 425 430
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Leu Ala Asn Gly
435 440 445
Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu
450 455 460

Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu
465 470 475 480
Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly
485 490 495
Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu
500 505 510
Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp
515 520 525

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys
530 535 540
Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
545 550 555 560
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu
565 570 575
Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
580 585 590

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
595 600 605
Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys
610 615 620
Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly
625 630 635 640

Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
645 650 655

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
660 665 670

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
675 680 685

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His
690 695 700

Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe
705 710 715 720

Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu
725 730 735

Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
740 745 750

Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly Gly Ser Pro Lys Lys
755 760 765

Lys Arg Lys Val
770

<210> 155

<211> 1622

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic Polypeptide

<400> 155

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400
Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

690 695 700

Arg Phe Ala Trp Met Phe Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Lys Leu Ile Asn Gly
900 905 910
Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser
915 920 925
Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser
930 935 940
Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
945 950 955 960

Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile
965 970 975
Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
980 985 990
Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
995 1000 1005
Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg
1010 1015 1020

Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile
1025 1030 1035
Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

1040 1045 1050

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1055 1060 1065

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala

1070 1075 1080

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1085 1090 1095

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1100 1105 1110

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1115 1120 1125

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1130 1135 1140

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1145 1150 1155

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1160 1165 1170

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1175 1180 1185

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1190 1195 1200

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1205 1210 1215

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1220 1225 1230

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1235 1240 1245

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1250 1255 1260

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1265 1270 1275

```
Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1280                1285                1290
Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1295                1300                1305
Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1310                1315                1320

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1325                1330                1335
Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1340                1345                1350
Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1355                1360                1365
Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp Ser Pro  Thr Val Ala
    1370                1375                1380

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1385                1390                1395
Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1400                1405                1410
Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1415                1420                1425
Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1430                1435                1440

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1445                1450                1455
Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1460                1465                1470
Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1475                1480                1485
Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1490                1495                1500

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
```

1505 1510 1515

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1520 1525 1530

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1535 1540 1545

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1550 1555 1560

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1565 1570 1575

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1580 1585 1590

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1595 1600 1605

Gly Gly Asp Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1610 1615 1620

<210> 156

<211> 1709

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 156

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400
Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510
Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Glu

580 585 590

Ile Leu Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

755 760 765

Tyr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Tyr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1310 1315 1320
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
1325 1330 1335
Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1340 1345 1350
Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1370 1375 1380
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1385 1390 1395
Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1400 1405 1410
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1430 1435 1440
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1445 1450 1455
Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1460 1465 1470
Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575
Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile
1610 1615 1620
Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu
1625 1630 1635
Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu
1640 1645 1650
Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu
1655 1660 1665

Asn Val Met Leu Leu Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala
1670 1675 1680
Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu
1685 1690 1695
Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val
1700 1705

<210> 157
<211> 1708
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 157
Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala

290 295 300

Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn

305 310 315 320

Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val

325 330 335

Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu

340 345 350

Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu

355 360 365

Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys

370 375 380

Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala

385 390 395 400

Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp

405 410 415

Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val

420 425 430

Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly

435 440 445

Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg

450 455 460

Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu

465 470 475 480

Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys

485 490 495

Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp

500 505 510

Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln

515 520 525

Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu

530 535 540

Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu

545 550 555 560

Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr

565 570 575

Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu

580 585 590

Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly

595 600 605

Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu

610 615 620

Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp

625 630 635 640

Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln

645 650 655

Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe

660 665 670

Tyr Pro Phe Leu Lys Asp Asn Arg Glu Asp Ile Glu Lys Ile Leu Thr

675 680 685

Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg

690 695 700

Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn

705 710 715 720

Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu

725 730 735

Arg Met Thr Asn Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys

740 745 750

His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys

```
        755                 760                 765
Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Ser

    770                 775                 780
Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg
785                 790                 795                 800
Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu
                805                 810                 815
Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala
            820                 825                 830
Ser Leu Gly Thr Tyr Gly Asp Leu Leu Lys Ile Ile Lys Asp Ile Asp

        835                 840                 845
Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu
    850                 855                 860
Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys
865                 870                 875                 880
Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg
                885                 890                 895
Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly

            900                 905                 910
Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser
        915                 920                 925
Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser
    930                 935                 940
Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
945                 950                 955                 960
Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile

                965                 970                 975
Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
            980                 985                 990
Val Met Gly Arg His Lys Pro Glu  Asn Ile Val Ile Glu  Met Ala Arg
        995                 1000                1005
```

Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg

1010 1015 1020

Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile

1025 1030 1035

Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

1040 1045 1050

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1055 1060 1065

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1070 1075 1080

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1085 1090 1095

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1100 1105 1110

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1115 1120 1125

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Leu Thr

1130 1135 1140

Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe

1145 1150 1155

Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val

1160 1165 1170

Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn

1175 1180 1185

Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

1190 1195 1200

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

1205 1210 1215

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala

1220 1225 1230

Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser

1235 1240 1245

Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met

1250 1255 1260

Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr

1265 1270 1275

Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr

1280 1285 1290

Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn

1295 1300 1305

Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala

1310 1315 1320

Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys

1325 1330 1335

Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu

1340 1345 1350

Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp

1355 1360 1365

Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr

1370 1375 1380

Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys

1385 1390 1395

Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg

1400 1405 1410

Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly

1415 1420 1425

Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr

1430 1435 1440

Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser

1445 1450 1455

Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys

1460 1465 1470

Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys

1475 1480 1485

Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln

1490 1495 1500

His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe

1505 1510 1515

Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu

1520 1525 1530

Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala

1535 1540 1545

Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro

1550 1555 1560

Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr

1565 1570 1575

Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser

1580 1585 1590

Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly

1595 1600 1605

Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys

1610 1615 1620

Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu

1625 1630 1635

Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp

1640 1645 1650

Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val

1655 1660 1665

Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu

1670 1675 1680

Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser

1685 1690 1695

Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1700 1705

<210> 158

<211> 1629

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 158

Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
35 40 45
Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
50 55 60
Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met
65 70 75 80
Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95

Ser Gln Ala Ile Arg Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu
100 105 110
Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160

Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
165 170 175
Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
180 185 190
Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu

195 200 205

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu

210 215 220

Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg Gly Ser Glu Thr

225 230 235 240

Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser

245 250 255

Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr

260 265 270

Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr

275 280 285

Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp

290 295 300

Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg

305 310 315 320

Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe

325 330 335

Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu

340 345 350

Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile

355 360 365

Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr

370 375 380

Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp

385 390 395 400

Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly

405 410 415

His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp

420 425 430

Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu

435 440 445

Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala
450 455 460
Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro
465 470 475 480

Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu
485 490 495
Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala
500 505 510
Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu
515 520 525
Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys
530 535 540

Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr
545 550 555 560
Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp
565 570 575
Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln
580 585 590
Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly
595 600 605

Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys
610 615 620
Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu
625 630 635 640
Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp
645 650 655
Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile
660 665 670

Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu
675 680 685
Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro

690 695 700

Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu

705 710 715 720

Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala

725 730 735

Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu

740 745 750

Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe

755 760 765

Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met

770 775 780

Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp

785 790 795 800

Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu

805 810 815

Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly

820 825 830

Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu

835 840 845

Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp

850 855 860

Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu

865 870 875 880

Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys

885 890 895

Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu

900 905 910

Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr

915 920 925

Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met

930 935 940

Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys
945 950 955 960
Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn
965 970 975
Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys
980 985 990

Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn
995 1000 1005
Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly
1010 1015 1020
Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
1025 1030 1035
Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn
1040 1045 1050

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn
1055 1060 1065
Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
1070 1075 1080
Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
1085 1090 1095
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn
1100 1105 1110

Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys
1115 1120 1125
Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
1130 1135 1140
Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu
1145 1150 1155
Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu
1160 1165 1170

Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg

1175 1180 1185

Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val

1190 1195 1200

Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys

1205 1210 1215

Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His

1220 1225 1230

Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile

1235 1240 1245

Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr

1250 1255 1260

Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu

1265 1270 1275

Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met

1280 1285 1290

Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg

1295 1300 1305

Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val

1310 1315 1320

Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser

1325 1330 1335

Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly

1340 1345 1350

Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys

1355 1360 1365

Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly

1370 1375 1380

Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys

1385 1390 1395

Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu

1400 1405 1410

Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro
1415 1420 1425
Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp
1430 1435 1440
Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn
1445 1450 1455
Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly
1460 1465 1470

Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu
1475 1480 1485
Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu
1490 1495 1500
Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu
1505 1510 1515
Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala
1520 1525 1530

Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg
1535 1540 1545
Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe
1550 1555 1560
Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
1565 1570 1575
Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu
1580 1585 1590

Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr
1595 1600 1605
Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly Gly Ser Pro
1610 1615 1620
Lys Lys Lys Arg Lys Val
1625
<210> 159
<211> 1679

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 159

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr
1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
20 25 30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65 70 75 80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
85 90 95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
130 135 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145 150 155 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
180 185 190

Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

210 215 220

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly

225 230 235 240

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

245 250 255

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

260 265 270

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

275 280 285

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

290 295 300

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

305 310 315 320

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

325 330 335

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

340 345 350

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

355 360 365

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

370 375 380

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

385 390 395 400

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

405 410 415

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

420 425 430

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

435 440 445

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

450 455 460

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
465 470 475 480
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
485 490 495
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
500 505 510
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
515 520 525

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
530 535 540
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
545 550 555 560
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
565 570 575
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
580 585 590

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
595 600 605
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
610 615 620
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
625 630 635 640
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
645 650 655

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
660 665 670
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
675 680 685
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
690 695 700
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

705 710 715 720

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
725 730 735
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
740 745 750
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
755 760 765
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
770 775 780

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
785 790 795 800
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
805 810 815
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
820 825 830
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
835 840 845

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
850 855 860
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
865 870 875 880
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
885 890 895
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
900 905 910

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
915 920 925
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
930 935 940
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
945 950 955 960

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
965 970 975

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
980 985 990
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
995 1000 1005
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
1010 1015 1020
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr
1025 1030 1035

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile
1040 1045 1050
Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser
1055 1060 1065
Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser
1070 1075 1080
Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val
1085 1090 1095

Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys
1100 1105 1110
Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
1115 1120 1125
Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln
1130 1135 1140
Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu
1145 1150 1155

Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile
1160 1165 1170
Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp
1175 1180 1185
Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn

1190 1195 1200

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1205 1210 1215

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1220 1225 1230

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1235 1240 1245

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1250 1255 1260

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly

1265 1270 1275

Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly

1280 1285 1290

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys

1295 1300 1305

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val

1310 1315 1320

Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn

1325 1330 1335

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys

1340 1345 1350

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val

1355 1360 1365

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val

1370 1375 1380

Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu

1385 1390 1395

Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

1400 1405 1410

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu

1415 1420 1425

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu
1430 1435 1440
Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
1445 1450 1455

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu
1460 1465 1470
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1475 1480 1485
Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1490 1495 1500
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1505 1510 1515

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1520 1525 1530
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1535 1540 1545
Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1550 1555 1560
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1565 1570 1575

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly
1580 1585 1590
Gly Ser Met Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly
1595 1600 1605
Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu
1610 1615 1620
Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val
1625 1630 1635

His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu
1640 1645 1650
Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln

1655 1660 1665

Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu

1670 1675

<210> 160

<211> 1689

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 160

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr

1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg

20 25 30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys

35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly

50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg

65 70 75 80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro

85 90 95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu

100 105 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr

115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn

130 135 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg

145 150 155 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp

165 170 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser

180 185 190

Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val

195 200 205

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

210 215 220

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly

225 230 235 240

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

245 250 255

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

260 265 270

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

275 280 285

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

290 295 300

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

305 310 315 320

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

325 330 335

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

340 345 350

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

355 360 365

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

370 375 380

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

385 390 395 400

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

405 410 415

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

420 425 430

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
435 440 445
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

450 455 460
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
465 470 475 480
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
485 490 495
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
500 505 510
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

515 520 525
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
530 535 540
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
545 550 555 560
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
565 570 575
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

580 585 590
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
595 600 605
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
610 615 620
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
625 630 635 640
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

645 650 655
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
660 665 670
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

675 680 685

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

690 695 700

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

705 710 715 720

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

725 730 735

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

740 745 750

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

755 760 765

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

770 775 780

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

785 790 795 800

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

805 810 815

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

820 825 830

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

835 840 845

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

850 855 860

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

865 870 875 880

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

885 890 895

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

900 905 910

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

915 920 925

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

930 935 940

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

945 950 955 960

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

965 970 975

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

980 985 990

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

995 1000 1005

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

1010 1015 1020

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr

1025 1030 1035

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

1040 1045 1050

Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser

1055 1060 1065

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser

1070 1075 1080

Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val

1085 1090 1095

Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys

1100 1105 1110

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg

1115 1120 1125

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln

1130 1135 1140

Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu

1145 1150 1155

Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile

1160 1165 1170

Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp

1175 1180 1185

Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn

1190 1195 1200

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1205 1210 1215

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1220 1225 1230

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1235 1240 1245

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1250 1255 1260

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly

1265 1270 1275

Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly

1280 1285 1290

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys

1295 1300 1305

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val

1310 1315 1320

Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn

1325 1330 1335

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys

1340 1345 1350

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val

1355 1360 1365

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val

1370 1375 1380

Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu

1385 1390 1395

Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

1400 1405 1410

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu

1415 1420 1425

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu

1430 1435 1440

Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe

1445 1450 1455

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu

1460 1465 1470

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1475 1480 1485

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1490 1495 1500

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1505 1510 1515

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1520 1525 1530

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1535 1540 1545

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1550 1555 1560

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1565 1570 1575

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly

1580 1585 1590

Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys

1595 1600 1605

Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val

1610 1615 1620

Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His

1625 1630 1635

Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr

1640 1645 1650

Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp

1655 1660 1665

Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Pro

1670 1675 1680

Lys Lys Lys Arg Lys Val

1685

<210> 161

<211> 1655

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 161

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val

1 5 10 15

Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His

20 25 30

Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn

35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu

50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp

65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala

85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys

100 105 110

Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly

115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr

130 135 140

Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145 150 155 160

Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
165 170 175

Gly Arg Leu Arg Ala Ile Leu Gln Ser Gly Ser Glu Thr Pro Gly Thr
180 185 190

Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu
195 200 205

Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
210 215 220

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
225 230 235 240

Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
245 250 255

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
260 265 270

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
275 280 285

Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
290 295 300

Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
305 310 315 320

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
325 330 335

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
340 345 350

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
355 360 365

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
370 375 380

Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
385 390 395 400
Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
405 410 415
Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
420 425 430

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
435 440 445
Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
450 455 460
Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
465 470 475 480
Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
485 490 495

Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr
500 505 510
Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
515 520 525
Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu
530 535 540
Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly
545 550 555 560

Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys
565 570 575
Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu
580 585 590
Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser
595 600 605
Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg
610 615 620

Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu

625 630 635 640

Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg

645 650 655

Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile

660 665 670

Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln

675 680 685

Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu

690 695 700

Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr

705 710 715 720

Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro

725 730 735

Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe

740 745 750

Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe

755 760 765

Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp

770 775 780

Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile

785 790 795 800

Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu

805 810 815

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu

820 825 830

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys

835 840 845

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys

850 855 860

Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp

865 870 875 880

Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
885 890 895
His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
900 905 910
Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
915 920 925
Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
930 935 940

Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
945 950 955 960
Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
965 970 975
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
980 985 990
Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
995 1000 1005

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1010 1015 1020
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1025 1030 1035
Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1040 1045 1050
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1055 1060 1065

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1070 1075 1080
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1085 1090 1095
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1100 1105 1110
Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1115 1120 1125

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1130 1135 1140
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
1145 1150 1155
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
1160 1165 1170
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1175 1180 1185

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1190 1195 1200
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1205 1210 1215
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1220 1225 1230
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1235 1240 1245

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1250 1255 1260
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1265 1270 1275
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
1280 1285 1290
Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1295 1300 1305

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1310 1315 1320
Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1325 1330 1335
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1340 1345 1350

```
Lys Lys  Leu Lys Ser Val Lys  Glu Leu Leu Gly Ile  Thr Ile Met
    1355                 1360                 1365

Glu Arg  Ser Ser Phe Glu Lys  Asn Pro Ile Asp Phe  Leu Glu Ala
    1370                 1375                 1380
Lys Gly  Tyr Lys Glu Val Lys  Lys Asp Leu Ile Ile  Lys Leu Pro
    1385                 1390                 1395
Lys Tyr  Ser Leu Phe Glu Leu  Glu Asn Gly Arg Lys  Arg Met Leu
    1400                 1405                 1410
Ala Ser  Ala Gly Glu Leu Gln  Lys Gly Asn Glu Leu  Ala Leu Pro
    1415                 1420                 1425

Ser Lys  Tyr Val Asn Phe Leu  Tyr Leu Ala Ser His  Tyr Glu Lys
    1430                 1435                 1440
Leu Lys  Gly Ser Pro Glu Asp  Asn Glu Gln Lys Gln  Leu Phe Val
    1445                 1450                 1455
Glu Gln  His Lys His Tyr Leu  Asp Glu Ile Ile Glu  Gln Ile Ser
    1460                 1465                 1470
Glu Phe  Ser Lys Arg Val Ile  Leu Ala Asp Ala Asn  Leu Asp Lys
    1475                 1480                 1485

Val Leu  Ser Ala Tyr Asn Lys  His Arg Asp Lys Pro  Ile Arg Glu
    1490                 1495                 1500
Gln Ala  Glu Asn Ile Ile His  Leu Phe Thr Leu Thr  Asn Leu Gly
    1505                 1510                 1515
Ala Pro  Ala Ala Phe Lys Tyr  Phe Asp Thr Thr Ile  Asp Arg Lys
    1520                 1525                 1530
Arg Tyr  Thr Ser Thr Lys Glu  Val Leu Asp Ala Thr  Leu Ile His
    1535                 1540                 1545

Gln Ser  Ile Thr Gly Leu Tyr  Glu Thr Arg Ile Asp  Leu Ser Gln
    1550                 1555                 1560
Leu Gly  Gly Asp Ser Gly Gly  Ser Met Thr Asn Leu  Ser Asp Ile
    1565                 1570                 1575
Ile Glu  Lys Glu Thr Gly Lys  Gln Leu Val Ile Gln  Glu Ser Ile
```

1580 1585 1590

Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro

1595 1600 1605

Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp

1610 1615 1620

Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro

1625 1630 1635

Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys

1640 1645 1650

Met Leu

1655

<210> 162

<211> 1665

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 162

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val

1 5 10 15

Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His

20 25 30

Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn

35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu

50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp

65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala

85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys

100 105 110

Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly

115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr
130 135 140
Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145 150 155 160
Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
165 170 175
Gly Arg Leu Arg Ala Ile Leu Gln Ser Gly Ser Glu Thr Pro Gly Thr
180 185 190

Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu
195 200 205
Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
210 215 220
Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
225 230 235 240
Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
245 250 255

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
260 265 270
Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
275 280 285
Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
290 295 300
Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
305 310 315 320

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
325 330 335
Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
340 345 350
Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
355 360 365

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
370 375 380

Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
385 390 395 400
Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
405 410 415
Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
420 425 430
Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
435 440 445

Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
450 455 460
Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
465 470 475 480
Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
485 490 495
Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr
500 505 510

Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
515 520 525
Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu
530 535 540
Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly
545 550 555 560
Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys
565 570 575

Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu
580 585 590
Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser
595 600 605
Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg

610 615 620

Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu
625 630 635 640

Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
645 650 655

Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
660 665 670

Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln
675 680 685

Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
690 695 700

Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
705 710 715 720

Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
725 730 735

Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe
740 745 750

Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe
755 760 765

Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp
770 775 780

Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile
785 790 795 800

Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
805 810 815

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
820 825 830

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
835 840 845

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
850 855 860

Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
865 870 875 880
Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
885 890 895

His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
900 905 910
Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
915 920 925
Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
930 935 940
Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
945 950 955 960

Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
965 970 975
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
980 985 990
Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
995 1000 1005
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1010 1015 1020

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1025 1030 1035
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1040 1045 1050
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1055 1060 1065
Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1070 1075 1080

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1085 1090 1095
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1100 1105 1110

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1115 1120 1125

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1130 1135 1140

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1145 1150 1155

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1160 1165 1170

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1175 1180 1185

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1190 1195 1200

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1205 1210 1215

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1220 1225 1230

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1235 1240 1245

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1250 1255 1260

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1265 1270 1275

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1280 1285 1290

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1295 1300 1305

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1310 1315 1320

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1325 1330 1335

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1340 1345 1350
Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1355 1360 1365
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1370 1375 1380

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1385 1390 1395
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1400 1405 1410
Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1415 1420 1425
Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1430 1435 1440

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1445 1450 1455
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1460 1465 1470
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1475 1480 1485
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1490 1495 1500

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1505 1510 1515
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1520 1525 1530
Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1535 1540 1545
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1550 1555 1560

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1565 1570 1575

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

1580 1585 1590

Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu

1595 1600 1605

Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1610 1615 1620

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1625 1630 1635

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1640 1645 1650

Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1655 1660 1665

<210> 163

<211> 1665

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 163

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val

1 5 10 15

Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His

20 25 30

Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn

35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu

50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp

65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala

85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys

100 105 110

Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln Glu Gly

115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr

130 135 140

Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys

145 150 155 160

Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser

165 170 175

Gly Arg Leu Arg Ala Ile Leu Gln Ser Gly Ser Glu Thr Pro Gly Thr

180 185 190

Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu

195 200 205

Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr

210 215 220

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His

225 230 235 240

Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu

245 250 255

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr

260 265 270

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu

275 280 285

Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe

290 295 300

Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn

305 310 315 320

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His

325 330 335

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu

340 345 350

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
355 360 365
Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
370 375 380
Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile

385 390 395 400
Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
405 410 415
Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
420 425 430
Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
435 440 445
Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln

450 455 460
Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
465 470 475 480
Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
485 490 495
Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr
500 505 510
Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His

515 520 525
Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu
530 535 540
Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly
545 550 555 560
Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys
565 570 575
Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu

580 585 590
Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser

595 600 605

Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg

610 615 620

Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu

625 630 635 640

Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg

645 650 655

Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile

660 665 670

Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln

675 680 685

Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu

690 695 700

Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr

705 710 715 720

Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro

725 730 735

Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe

740 745 750

Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe

755 760 765

Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp

770 775 780

Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile

785 790 795 800

Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu

805 810 815

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu

820 825 830

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys

835 840 845

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
850 855 860
Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
865 870 875 880
Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
885 890 895
His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val

900 905 910
Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
915 920 925
Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
930 935 940
Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
945 950 955 960
Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser

965 970 975
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
980 985 990
Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
995 1000 1005
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1010 1015 1020
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1025 1030 1035
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1040 1045 1050
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1055 1060 1065
Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1070 1075 1080
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1085 1090 1095

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1100 1105 1110

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1115 1120 1125

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1130 1135 1140

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1145 1150 1155

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1160 1165 1170

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1175 1180 1185

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1190 1195 1200

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1205 1210 1215

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1220 1225 1230

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1235 1240 1245

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1250 1255 1260

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1265 1270 1275

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1280 1285 1290

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1295 1300 1305

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1310 1315 1320

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1325 1330 1335

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1340 1345 1350

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1355 1360 1365

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1370 1375 1380

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1385 1390 1395

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1400 1405 1410

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1415 1420 1425

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1430 1435 1440

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1445 1450 1455

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1460 1465 1470

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1475 1480 1485

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1490 1495 1500

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1505 1510 1515

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1520 1525 1530

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1535 1540 1545

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1550 1555 1560

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1565 1570 1575

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

1580 1585 1590

Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu

1595 1600 1605

Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1610 1615 1620

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1625 1630 1635

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1640 1645 1650

Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1655 1660 1665

<210> 164

<211> 246

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 164

gcggacatgg aggacgtgcg cggccgcctg gtgcagtacc gcggcgaggt gcaggccatg 60

ctcggccaga gcaccgagga gctgcgggtg cgcctcgcct cccacctgcg caagctgcgt 120

aagcggctcc tccgcgatgc cgatgacctg cagaagcgcc tggcagtgta ccaggccggg 180

gcccgcgagg gcgccgagcg cggcctcagc gccatccgcg agcgcctggg gcccctggtg 240

gaacag 246

<210> 165

<211> 1689

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 165

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr
1 5 10 15
Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg

20 25 30
Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45
Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
50 55 60
Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65 70 75 80
Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro

85 90 95
Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110
Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
115 120 125
Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
130 135 140
Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg

145 150 155 160
Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175
Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
180 185 190
Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205
Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

210 215 220
Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
225 230 235 240
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

245 250 255

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

260 265 270

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

275 280 285

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

290 295 300

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

305 310 315 320

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

325 330 335

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

340 345 350

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

355 360 365

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

370 375 380

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

385 390 395 400

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

405 410 415

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

420 425 430

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

435 440 445

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

450 455 460

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

465 470 475 480

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

485 490 495

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
500 505 510
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
515 520 525
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

530 535 540
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
545 550 555 560
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
565 570 575
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
580 585 590
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

595 600 605
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
610 615 620
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
625 630 635 640
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
645 650 655
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

660 665 670
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
675 680 685
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
690 695 700
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
705 710 715 720
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

725 730 735
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

740 745 750

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

755 760 765

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

770 775 780

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

785 790 795 800

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

805 810 815

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

820 825 830

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

835 840 845

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

850 855 860

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

865 870 875 880

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

885 890 895

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

900 905 910

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

915 920 925

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

930 935 940

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

945 950 955 960

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

965 970 975

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

980 985 990

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

995 1000 1005

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

1010 1015 1020

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr

1025 1030 1035

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

1040 1045 1050

Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser

1055 1060 1065

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser

1070 1075 1080

Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val

1085 1090 1095

Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys

1100 1105 1110

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg

1115 1120 1125

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln

1130 1135 1140

Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu

1145 1150 1155

Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile

1160 1165 1170

Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp

1175 1180 1185

Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn

1190 1195 1200

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1205 1210 1215

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1220 1225 1230

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1235 1240 1245

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1250 1255 1260

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly

1265 1270 1275

Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly

1280 1285 1290

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys

1295 1300 1305

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val

1310 1315 1320

Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn

1325 1330 1335

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys

1340 1345 1350

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val

1355 1360 1365

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val

1370 1375 1380

Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu

1385 1390 1395

Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

1400 1405 1410

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu

1415 1420 1425

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu

1430 1435 1440

Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe

1445 1450 1455

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu

1460 1465 1470

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1475 1480 1485

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1490 1495 1500

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1505 1510 1515

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1520 1525 1530

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1535 1540 1545

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1550 1555 1560

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1565 1570 1575

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly

1580 1585 1590

Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys

1595 1600 1605

Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val

1610 1615 1620

Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His

1625 1630 1635

Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr

1640 1645 1650

Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp

1655 1660 1665

Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Pro

1670 1675 1680

Lys Lys Lys Arg Lys Val

1685

<210> 166

<211> 1679

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 166

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 Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr

35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu

195 200 205

Ser Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile

210 215 220

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val

225 230 235 240

Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile

245 250 255

Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala

260 265 270

Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg

275 280 285

Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala

290 295 300

Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val

305 310 315 320

Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val

325 330 335

Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg

340 345 350

Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr

355 360 365

Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu

370 375 380

Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln

385 390 395 400

Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala

405 410 415

Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser

420 425 430

Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn

435 440 445

Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn

450 455 460

Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser

465 470 475 480

Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly

485 490 495

Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala

500 505 510

Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala

515 520 525

Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp

530 535 540

Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr

545 550 555 560

Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile

565 570 575

Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile

580 585 590

Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg

595 600 605

Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro

610 615 620

His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu

625 630 635 640

Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile

645 650 655

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn

660 665 670

Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro

675 680 685

```
Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe
    690                 695                 700
Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val

705                 710                 715                 720
Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu
                725                 730                 735
Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe
            740                 745                 750
Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr
        755                 760                 765
Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys

    770                 775                 780
Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe
785                 790                 795                 800
Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp
                805                 810                 815
Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile
            820                 825                 830
Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg

        835                 840                 845
Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu
    850                 855                 860
Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile
865                 870                 875                 880
Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu
                885                 890                 895
Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp

            900                 905                 910
Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly
        915                 920                 925
Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro
```

930 935 940

Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu

945 950 955 960

Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met

965 970 975

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

980 985 990

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile

995 1000 1005

Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

1010 1015 1020

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1025 1030 1035

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1040 1045 1050

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1055 1060 1065

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1070 1075 1080

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1085 1090 1095

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1100 1105 1110

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1115 1120 1125

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1130 1135 1140

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1145 1150 1155

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1160 1165 1170

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1175 1180 1185
Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1190 1195 1200
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1205 1210 1215
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1220 1225 1230
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1235 1240 1245
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1250 1255 1260
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1265 1270 1275
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1280 1285 1290
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1295 1300 1305
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1310 1315 1320
Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1325 1330 1335
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala
1340 1345 1350
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1355 1360 1365
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1370 1375 1380
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1385 1390 1395
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1400 1405 1410

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1415 1420 1425

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1430 1435 1440

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1445 1450 1455

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1460 1465 1470

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1475 1480 1485

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1490 1495 1500

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1505 1510 1515

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1520 1525 1530

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1535 1540 1545

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1550 1555 1560

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1565 1570 1575

Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu

1580 1585 1590

Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met

1595 1600 1605

Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser

1610 1615 1620

Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn

1625 1630 1635

Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala
1640 1645 1650
Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu
1655 1660 1665
Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val
1670 1675

<210> 167

<211> 1824

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 167

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser
225 230 235 240
Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly
245 250 255

Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
260 265 270
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
275 280 285
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
290 295 300
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
305 310 315 320

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
325 330 335
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
340 345 350
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
355 360 365
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
370 375 380

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
385 390 395 400
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

405 410 415

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

420 425 430

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

435 440 445

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

450 455 460

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

465 470 475 480

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

485 490 495

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

500 505 510

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

515 520 525

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

530 535 540

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

545 550 555 560

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

565 570 575

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

580 585 590

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

595 600 605

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

610 615 620

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

625 630 635 640

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

645 650 655

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
660 665 670

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
675 680 685

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
690 695 700

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
705 710 715 720

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
725 730 735

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
740 745 750

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
755 760 765

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
770 775 780

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
785 790 795 800

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
805 810 815

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
820 825 830

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
835 840 845

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
850 855 860

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
865 870 875 880

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
885 890 895

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

900 905 910

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

915 920 925

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

930 935 940

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

945 950 955 960

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

965 970 975

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

980 985 990

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

995 1000 1005

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met

1010 1015 1020

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

1025 1030 1035

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser

1040 1045 1050

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn

1055 1060 1065

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr

1070 1075 1080

Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val

1085 1090 1095

Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp

1100 1105 1110

Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp

1115 1120 1125

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp

1130 1135 1140

Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp
1145 1150 1155
Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
1160 1165 1170
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
1175 1180 1185
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr
1190 1195 1200

Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu
1205 1210 1215
Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr
1220 1225 1230
Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr
1235 1240 1245
Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys
1250 1255 1260

Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val
1265 1270 1275
Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr
1280 1285 1290
Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr
1295 1300 1305
Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile
1310 1315 1320

Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg
1325 1330 1335
Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn
1340 1345 1350
Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu
1355 1360 1365
Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys

1370 1375 1380

Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr
1385 1390 1395
Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys
1400 1405 1410
Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
1415 1420 1425
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu
1430 1435 1440

Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu
1445 1450 1455
Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
1460 1465 1470
Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu
1475 1480 1485
Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu
1490 1495 1500

Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
1505 1510 1515
Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile
1520 1525 1530
Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp
1535 1540 1545
Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg
1550 1555 1560

Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu
1565 1570 1575
Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg
1580 1585 1590
Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile
1595 1600 1605

His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
1610 1615 1620

Gln Leu Gly Gly Asp Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
1625 1630 1635
Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu
1640 1645 1650
Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu
1655 1660 1665
Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala
1670 1675 1680

Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp
1685 1690 1695
Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn
1700 1705 1710
Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Gly Gly Ser
1715 1720 1725
Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly
1730 1735 1740

Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu
1745 1750 1755
Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val
1760 1765 1770
His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu
1775 1780 1785
Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln
1790 1795 1800

Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser
1805 1810 1815
Pro Lys Lys Lys Arg Lys
1820
<210> 168

<211> 1544

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 168

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser
225 230 235 240
Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly
245 250 255
Ser Ser Gly Gly Ser Gly Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile
260 265 270
Gly Ile Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp
275 280 285

Val Ile Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn
290 295 300
Asn Glu Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg
305 310 315 320
Arg Arg His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn
325 330 335
Leu Leu Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala
340 345 350

Arg Val Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala
355 360 365
Ala Leu Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu
370 375 380
Val Glu Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser
385 390 395 400
Arg Asn Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu
405 410 415

Glu Arg Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe
420 425 430
Lys Thr Ser Asp Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln
435 440 445
Lys Ala Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp
450 455 460

Leu Leu Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser
465 470 475 480

Pro Phe Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly
485 490 495

His Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr
500 505 510

Asn Ala Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile
515 520 525

Thr Arg Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile
530 535 540

Ile Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile
545 550 555 560

Ala Lys Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val
565 570 575

Thr Ser Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp
580 585 590

Ile Lys Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu
595 600 605

Leu Asp Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp
610 615 620

Ile Gln Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu
625 630 635 640

Ile Glu Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu
645 650 655

Ser Leu Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn
660 665 670

Asp Asn Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys
675 680 685

Val Asp Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp
690 695 700

Phe Ile Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys

705 710 715 720

Val Ile Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile

725 730 735

Ile Glu Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile

740 745 750

Asn Glu Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu

755 760 765

Ile Ile Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys

770 775 780

Ile Lys Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu

785 790 795 800

Ala Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val

805 810 815

Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn

820 825 830

Lys Val Leu Val Lys Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr

835 840 845

Pro Phe Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr

850 855 860

Phe Lys Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser

865 870 875 880

Lys Thr Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe

885 890 895

Ser Val Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr

900 905 910

Ala Thr Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn

915 920 925

Asn Leu Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe

930 935 940

Leu Arg Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys

945 950 955 960

His His Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe
965 970 975
Lys Glu Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln
980 985 990

Met Phe Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu
995 1000 1005
Gln Glu Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His
1010 1015 1020
Ile Lys Asp Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys
1025 1030 1035
Lys Pro Asn Arg Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg
1040 1045 1050

Lys Asp Asp Lys Gly Asn Thr Leu Ile Val Asn Asn Leu Asn Gly
1055 1060 1065
Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu Ile Asn Lys
1070 1075 1080
Ser Pro Glu Lys Leu Leu Met Tyr His His Asp Pro Gln Thr Tyr
1085 1090 1095
Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp Glu Lys Asn
1100 1105 1110

Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu Thr Lys
1115 1120 1125
Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys Tyr
1130 1135 1140
Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
1145 1150 1155
Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
1160 1165 1170

Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr
1175 1180 1185
Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val

1190 1195 1200

Asn Ser Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser

1205 1210 1215

Asn Gln Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile

1220 1225 1230

Lys Ile Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp

1235 1240 1245

Leu Leu Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg

1250 1255 1260

Glu Tyr Leu Glu Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile

1265 1270 1275

Lys Thr Ile Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr

1280 1285 1290

Asp Ile Leu Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro

1295 1300 1305

Gln Ile Ile Lys Lys Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1310 1315 1320

Ser Ser Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp

1325 1330 1335

Ile Asp Tyr Lys Asp Asp Asp Asp Lys Ser Gly Gly Ser Gly Gly

1340 1345 1350

Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr

1355 1360 1365

Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu

1370 1375 1380

Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu

1385 1390 1395

Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu

1400 1405 1410

Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile

1415 1420 1425

Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly
1430 1435 1440

Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu
1445 1450 1455

Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met
1460 1465 1470

Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser
1475 1480 1485

Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn
1490 1495 1500

Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala
1505 1510 1515

Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu
1520 1525 1530

Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val
1535 1540

<210> 169

<211> 661

<212> DNA

<213> Homo sapiens

<400> 169

catccttggt accgagctcg gatccagcca ccatggatag cctcttgatg aatagacgca 60

agttcctgta tcagtttaaa aacgtgagat gggcaaaagg ccgacgagag acatatctgt 120

gctatgtcgt taagcgcaga gattcagcca ccagtttctc tctcgacttc ggctacctgc 180

ggaacaagaa tggttgccat gttgagctcc tgttcctgag gtatatcagc gactgggatt 240

tggacccagg gcggtgctat agggtgacat ggtttacctc ctggtcacct tgttatgact 300

gcgcgcggca tgttgccgat tttctgagag ggaaccctaa cctgtctctg aggatcttca 360

ccgcgcgact gtacttctgt gaggaccgga aagccgaacc cgagggactg agacgcctcc 420

acagagcggg tgtgcagatt gccataatga cctttaagga ctacttctac tgctggaaca 480

ccttcgtcga aaatcacgag cggactttca aggcttggga aggattgcat gaaaacagcg 540

tcaggctttc caggcagctt cgccgcattc ttctcccgtt gtacgaggtt gatgacctca 600

gagatgcctt tagaacactg ggactgtagg cggccgctcg attggtttgg tgtggctcta 660

a 661

<210> 170

<211> 755

<212> DNA

<213> Homo sapiens

<400> 170

catccttggt accgagctcg gatccagcca ccatgagctc agagactggc ccagtggctg 60

tggaccccac attgagacgg cggatcgagc cccatgagtt tgaggtattc ttcgatccga 120

gagagctccg caaggagacc tgcctgcttt acgaaattaa ttgggggggc cggcactcca 180

tttggcgaca tacatcacag aacactaaca agcacgtcga agtcaacttc atcgagaagt 240

tcacgacaga aagatatttc tgtccgaaca caaggtgcag cattacctgg tttctcagct 300

ggagcccatg cgggcgaatg tagtagggcc atcactgaat tcctgtcaag gtatccccac 360

gtcactctgt ttatttacat cgcaaggctg taccaccacg ctgacccccg caatcgacaa 420

ggcctgcggg atttgatctc ttcaggtgtg actatccaaa ttatgactga gcaggagtca 480

ggatactgct ggagaaactt tctgaattat agcccgagta atgaagccca ctggcctagg 540

tatccccatc tgtgggtacg actgtacgtt cttgaactgt actgcatcat actgggcctg 600

cctccttgtc tcaacattct gagaaggaag cagccacagc tgacattctt taccatcgct 660

cttcagtctt gtcattacca gcgactgccc ccacacattc tctgggccac cgggttgaaa 720

tgagcggccg ctcgattggt ttggtgtggc tctaa 755

<210> 171

<211> 574

<212> DNA

<213> Petromyzon marinus

<400> 171

ctcacagatg ctatgtcctg ttcgaactca agagaagagg agagaaaggc gggcctgttt 60

ctgggggtac gcggttaata aaccccagtc cgggaccgag agggggggatt cacgccgaga 120

tcttttcaat taggaaggtt gaagagtatc ttcgcgacaa tcccggtcag ttcacaatta 180

actggtacag ctcctggagc ccttgcgctg attgcgccga gaaaatactc gaatggtaca 240

accaggactt gagaggcaat ggccacactc tcaagatttg ggcttgcaag ctttactacg 300

agaagaacgc gagaaatcag attggcttgt ggaacctcag ggacaacggg gtcgggttga 360

atgttatggt gtccgaacat taccagtgct gtagaaagat cttcattcag tccagtcaca 420

atcagctgaa cgagaacaga tggctggaga aaacactgaa acgggcagag aaaaggcgct 480

cagagctgag tatcatgatc caggtcaaaa tcctgcatac aaccaaaagc ccggctgtat 540

aagcggccgc tcgattggtt tggtgtggct ctaa 574

<210> 172

<211> 1045

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 172

catccttggt accgagctcg gatccagcca ccatggagct gaagtatcac cctgagatgc 60

ggtttttcca ctggtttagt aagtggcgca aacttcatcg ggatcaggag tatgaagtga 120

cctggtatat ctcttggtct ccctgcacaa aatgtacacg cgavatggaa acatttctgg 180

ccgaggatcc aaaggtgacg ctcacaaatc tttgtggccc gcctgtatta tttctgggac 240

ccggattatc aggaggcact taggtcattg tgccaaaagc gcgacggacc acgggcgact 300

atgaaaatca tgaattatga cgaattccag cattgctgga gtaagtttgt gtacagccag 360

cgggagctgt tcgagccctg gaacaatctt cccaagtact acatactgct tcacattatg 420

ttgggggaga tccttcggca ctctatggat cctcctaccc tttacgttta aactttaata 480

atgagccttg ggttcgcggg cgccatgaaa cctatttgtg ctacgaggtc gagcggatcc 540

ataatgatac gtgggtcctg ctgaatcaga ggaggggttt tctgtgtaac caggctccac 600

ataaagatgg atttctcgag gggcggcacg ccgaactgtg tttccttgat gtgatacctt 660

tctggaagct cgaccttgat caagattaca gggtgacgtg tttcacctcc tggtcaccct 720

gcttcagttg cgcccaagag atggctaaat ttagcagtaa gaacaagcat gtgtccctct 780

gtatttttac agccagaatt tatgatgacc agggccggtg ccaggagggg ctgcggacac 840

tcgctgaggc gggcgcgaag atcagcataa tgacatactc cgaattcaaa cactgttggg 900

acacttttgt ggaccaccag ggctgcccat ttcagccgtg ggatgggctc gacgaacata 960

gtcaggatct ctcaggccgg ctgcgagcca tattgcagaa ccaggagaat taggcggccg 1020

ctcgattggt ttggtgtggc tctaa 1045

<210> 173

<211> 785

<212> DNA

<213> Escherichia coli

<400> 173

ggccggggat tctagaaata attttgttta actttaagaa ggagatatac catggatgtc 60

ttctgaaacc ggtccggttg cggttgaccc gaccctgcgt cgtcgtatcg aaccgcacga 120

attcgaagtt ttcttcgacc ccgcgtgaac tgcgtaaaga aacctgcctg ctgtacgaaa 180

tcaactgggg tggtcgtcac tctatctggc gtcacacctc tcagaacacc aacaaacacg 240

ttgaagttaa cttcatcgaa aaattcacca ccgaacgtta cttctgcccg aacacccgtt 300

gctctatcac ctggttcctg tcttggtctc cgtgcggtga atgctcctcg tgcgatcacc 360

gaattcctgt ctcgttaccc gcacgttacc ctgttcatct acatcgcgcg tctgtaccac 420

cacgcggacc cgcgtaaccg tcagggtctg cgtgacctga tctcttctgg tgttaccatc 480

cagatcatga ccgaacagga atctggttac tgctggcgta acttcgttaa ctactctccg 540

tctaacgaag cgcactggcc gcgttacccg cacctgtggg ttcgtctgta cgttctggaa 600

ctgtactgca tcatcctggg tctgccgccg tgcctgaaca tcctgcgtcg taaacagccg 660

cagctgacct cttcaccatc gcgctgcagt cttgccacta ccagcgtctg ccgccgcaca 720

tcctgtgggc gaccggtctg aaaggtggta gtggagggag cggcggttca atggataaga 780

aatac 785

<210> 174

<211> 1710

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 174

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala

995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu

1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1610 1615 1620

Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

1625 1630 1635

Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu

1640 1645 1650

Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1655 1660 1665

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1670 1675 1680

Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1685 1690 1695

Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1700 1705 1710

<210> 175

<211> 1689

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 175

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr

1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg

```
            20                  25                  30
Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
        35                  40                  45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
    50                  55                  60
Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65                  70                  75                  80
Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
                85                  90                  95
Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
            100                 105                 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
        115                 120                 125
Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
    130                 135                 140
Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145                 150                 155                 160
Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
                165                 170                 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
            180                 185                 190
Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
        195                 200                 205
Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
    210                 215                 220
Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
225                 230                 235                 240

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
                245                 250                 255
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
            260                 265                 270
```

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
275 280 285
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
290 295 300

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
305 310 315 320
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
325 330 335
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
340 345 350
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
355 360 365

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
370 375 380
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
385 390 395 400
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
405 410 415
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
420 425 430

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
435 440 445
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
450 455 460
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
465 470 475 480
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
485 490 495

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
500 505 510
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

515 520 525

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

530 535 540

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

545 550 555 560

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

565 570 575

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

580 585 590

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

595 600 605

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

610 615 620

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

625 630 635 640

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

645 650 655

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

660 665 670

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

675 680 685

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

690 695 700

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

705 710 715 720

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

725 730 735

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

740 745 750

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

755 760 765

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
770 775 780
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
785 790 795 800
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
805 810 815

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
820 825 830
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
835 840 845
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
850 855 860
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
865 870 875 880

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
885 890 895
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
900 905 910
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
915 920 925
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
930 935 940

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
945 950 955 960
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
965 970 975
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
980 985 990
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
995 1000 1005

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

1010 1015 1020

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr

1025 1030 1035

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

1040 1045 1050

Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser

1055 1060 1065

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser

1070 1075 1080

Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val

1085 1090 1095

Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys

1100 1105 1110

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg

1115 1120 1125

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln

1130 1135 1140

Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu

1145 1150 1155

Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile

1160 1165 1170

Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp

1175 1180 1185

Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn

1190 1195 1200

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1205 1210 1215

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1220 1225 1230

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1235 1240 1245

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1250 1255 1260
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly
1265 1270 1275
Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly
1280 1285 1290
Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys
1295 1300 1305

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val
1310 1315 1320
Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn
1325 1330 1335
Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys
1340 1345 1350
Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val
1355 1360 1365

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val
1370 1375 1380
Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu
1385 1390 1395
Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val
1400 1405 1410
Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu
1415 1420 1425

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu
1430 1435 1440
Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
1445 1450 1455
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu
1460 1465 1470
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1475 1480 1485

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1490 1495 1500
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1505 1510 1515
Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1520 1525 1530
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1535 1540 1545

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1550 1555 1560
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1565 1570 1575
Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly
1580 1585 1590
Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys
1595 1600 1605

Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val
1610 1615 1620
Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His
1625 1630 1635
Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr
1640 1645 1650
Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp
1655 1660 1665

Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Pro
1670 1675 1680
Lys Lys Lys Arg Lys Val
1685
<210> 176
<211> 1679

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 176

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 Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr
130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys
145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190

Phe Arg Thr Leu Gly Leu Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
195 200 205

Ser Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile

210 215 220

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val
225 230 235 240
Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile
245 250 255
Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala
260 265 270
Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
275 280 285

Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala
290 295 300
Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val
305 310 315 320
Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val
325 330 335
Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg
340 345 350

Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr
355 360 365
Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu
370 375 380
Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln
385 390 395 400
Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala
405 410 415

Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser
420 425 430
Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn
435 440 445
Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn
450 455 460

Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser
465 470 475 480

Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly
485 490 495

Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala
500 505 510

Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala
515 520 525

Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp
530 535 540

Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr
545 550 555 560

Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile
565 570 575

Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile
580 585 590

Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg
595 600 605

Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro
610 615 620

His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu
625 630 635 640

Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile
645 650 655

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn
660 665 670

Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro
675 680 685

Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe
690 695 700

Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val

705 710 715 720

Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu

725 730 735

Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe

740 745 750

Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr

755 760 765

Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys

770 775 780

Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe

785 790 795 800

Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp

805 810 815

Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile

820 825 830

Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg

835 840 845

Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu

850 855 860

Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile

865 870 875 880

Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu

885 890 895

Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp

900 905 910

Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly

915 920 925

Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro

930 935 940

Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu

945 950 955 960

Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met
965 970 975
Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
980 985 990

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile
995 1000 1005
Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
1010 1015 1020
Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1025 1030 1035
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His
1040 1045 1050

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1055 1060 1065
Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1070 1075 1080
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1085 1090 1095
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1100 1105 1110

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1115 1120 1125
Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1130 1135 1140
Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1145 1150 1155
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1160 1165 1170

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1175 1180 1185
Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1190 1195 1200

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1205 1210 1215

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1220 1225 1230

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1235 1240 1245

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1250 1255 1260

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1265 1270 1275

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1280 1285 1290

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1295 1300 1305

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1310 1315 1320

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1325 1330 1335

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala

1340 1345 1350

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1355 1360 1365

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1370 1375 1380

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1385 1390 1395

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1400 1405 1410

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1415 1420 1425

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1430 1435 1440
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1445 1450 1455
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1460 1465 1470

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1475 1480 1485
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1490 1495 1500
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1505 1510 1515
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1520 1525 1530

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1535 1540 1545
Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1550 1555 1560
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1565 1570 1575
Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu
1580 1585 1590

Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met
1595 1600 1605
Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser
1610 1615 1620
Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn
1625 1630 1635
Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala
1640 1645 1650

Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu

1655 1660 1665

Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1670 1675

<210> 177

<211> 1806

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 177

Met Glu Leu Lys Tyr His Pro Glu Met Arg Phe Phe His Trp Phe Ser

1 5 10 15

Lys Trp Arg Lys Leu His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr

20 25 30

Ile Ser Trp Ser Pro Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe

35 40 45

Leu Ala Glu Asp Pro Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu

50 55 60

Tyr Tyr Phe Trp Asp Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys

65 70 75 80

Gln Lys Arg Asp Gly Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp

85 90 95

Glu Phe Gln His Cys Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu

100 105 110

Phe Glu Pro Trp Asn Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile

115 120 125

Met Leu Gly Glu Ile Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr

130 135 140

Phe Asn Phe Asn Asn Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr

145 150 155 160

Leu Cys Tyr Glu Val Glu Arg Met His Asn Asp Thr Trp Val Leu Leu

165 170 175

Asn Gln Arg Arg Gly Phe Leu Cys Asn Gln Ala Pro His Lys His Gly

180 185 190

Phe Leu Glu Gly Arg His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro

195 200 205

Phe Trp Lys Leu Asp Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr

210 215 220

Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile

225 230 235 240

Ser Lys Asn Lys His Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr

245 250 255

Asp Asp Gln Gly Arg Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala

260 265 270

Gly Ala Lys Ile Ser Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp

275 280 285

Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly

290 295 300

Leu Asp Glu His Ser Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu

305 310 315 320

Gln Asn Gln Glu Asn Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

325 330 335

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

340 345 350

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

355 360 365

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

370 375 380

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

385 390 395 400

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

405 410 415

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

420 425 430

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
435 440 445
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
450 455 460
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

465 470 475 480
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
485 490 495
Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
500 505 510
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
515 520 525
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

530 535 540
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
545 550 555 560
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
565 570 575
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
580 585 590
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

595 600 605
Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
610 615 620
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
625 630 635 640
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
645 650 655
Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

660 665 670
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

675 680 685

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

690 695 700

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

705 710 715 720

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

725 730 735

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

740 745 750

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

755 760 765

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

770 775 780

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

785 790 795 800

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

805 810 815

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

820 825 830

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

835 840 845

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

850 855 860

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

865 870 875 880

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

885 890 895

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

900 905 910

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

915 920 925

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

930 935 940

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

945 950 955 960

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

965 970 975

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

980 985 990

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

995 1000 1005

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu

1010 1015 1020

Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

1025 1030 1035

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val

1040 1045 1050

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala

1055 1060 1065

Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val

1070 1075 1080

Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn

1085 1090 1095

Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly

1100 1105 1110

Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile

1115 1120 1125

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn

1130 1135 1140

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn

1145 1150 1155

Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

1160 1165 1170

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

1175 1180 1185

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn

1190 1195 1200

Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys

1205 1210 1215

Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr

1220 1225 1230

Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu

1235 1240 1245

Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu

1250 1255 1260

Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg

1265 1270 1275

Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val

1280 1285 1290

Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys

1295 1300 1305

Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His

1310 1315 1320

Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile

1325 1330 1335

Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr

1340 1345 1350

Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu

1355 1360 1365

Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met

1370 1375 1380

Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg

1385 1390 1395

Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val

1400 1405 1410

Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser

1415 1420 1425

Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly

1430 1435 1440

Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys

1445 1450 1455

Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly

1460 1465 1470

Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys

1475 1480 1485

Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu

1490 1495 1500

Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro

1505 1510 1515

Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp

1520 1525 1530

Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn

1535 1540 1545

Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly

1550 1555 1560

Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu

1565 1570 1575

Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu

1580 1585 1590

Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu

1595 1600 1605

Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala

1610 1615 1620

Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg

1625 1630 1635

Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe

1640 1645 1650

Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

1655 1660 1665

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu

1670 1675 1680

Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr

1685 1690 1695

Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly Gly Ser Thr

1700 1705 1710

Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val

1715 1720 1725

Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val

1730 1735 1740

Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr

1745 1750 1755

Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala

1760 1765 1770

Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly

1775 1780 1785

Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Pro Lys Lys Lys

1790 1795 1800

Arg Lys Val

1805

<210> 178

<211> 1709

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 178

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 Gly Gly Ser Ser Ser Glu Thr Gly Pro Val Ala Val

85 90 95

Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe Glu Val Phe

100 105 110

Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile

115 120 125

Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser Gln Asn Thr

130 135 140

Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg

145 150 155 160

Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp

165 170 175

Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg

180 185 190

Tyr Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg Leu Tyr His His

195 200 205

Ala Asp Pro Arg Asn Arg Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly

210 215 220

Val Thr Ile Gln Ile Met Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg

225 230 235 240

Asn Phe Val Asn Tyr Ser Pro Ser Asn Glu Ala His Trp Pro Arg Tyr

245 250 255

Pro His Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile

260 265 270

Leu Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln

275 280 285

Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu

290 295 300

Pro Pro His Ile Leu Trp Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr

305 310 315 320

Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser

325 330 335

Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr

340 345 350

Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr

355 360 365

Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp

370 375 380

Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg

385 390 395 400

Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe

405 410 415

Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu

420 425 430

Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile

435 440 445

Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr

450 455 460

Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp

465 470 475 480

Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly

485 490 495

His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp
500 505 510
Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu

515 520 525
Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala
530 535 540
Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro
545 550 555 560
Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu
565 570 575
Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala

580 585 590
Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu
595 600 605
Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys
610 615 620
Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr
625 630 635 640
Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp

645 650 655
Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln
660 665 670
Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly
675 680 685
Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys
690 695 700
Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu

705 710 715 720
Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp
725 730 735
Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile

740 745 750

Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu

755 760 765

Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro

770 775 780

Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu

785 790 795 800

Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala

805 810 815

Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu

820 825 830

Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe

835 840 845

Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met

850 855 860

Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp

865 870 875 880

Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu

885 890 895

Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly

900 905 910

Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu

915 920 925

Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp

930 935 940

Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu

945 950 955 960

Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys

965 970 975

Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu

980 985 990

Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr

995 1000 1005

Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe

1010 1015 1020

Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile

1025 1030 1035

Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His

1040 1045 1050

Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu

1055 1060 1065

Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

1070 1075 1080

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

1085 1090 1095

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg

1100 1105 1110

Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu

1115 1120 1125

His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu

1130 1135 1140

Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu

1145 1150 1155

Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro

1160 1165 1170

Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr

1175 1180 1185

Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu

1190 1195 1200

Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn

1205 1210 1215

Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala

1220 1225 1230

Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys

1235 1240 1245

Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln

1250 1255 1260

Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys

1265 1270 1275

Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val

1280 1285 1290

Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile

1295 1300 1305

Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

1310 1315 1320

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

1325 1330 1335

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1340 1345 1350

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1355 1360 1365

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1370 1375 1380

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1385 1390 1395

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1400 1405 1410

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1415 1420 1425

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1430 1435 1440

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1445 1450 1455

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1460 1465 1470

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1475 1480 1485

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1490 1495 1500

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1505 1510 1515

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1520 1525 1530

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1535 1540 1545

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1550 1555 1560

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1565 1570 1575

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1580 1585 1590

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1595 1600 1605

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1610 1615 1620

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1625 1630 1635

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1640 1645 1650

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1655 1660 1665

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1670 1675 1680

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1685 1690 1695

Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1700 1705

<210> 179

<211> 1810

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 179

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile

805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005
Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1145 1150 1155
Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1175 1180 1185
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
1190 1195 1200
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
1205 1210 1215
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1235 1240 1245
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1250 1255 1260
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1265 1270 1275
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1295 1300 1305
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1310 1315 1320
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
1325 1330 1335
Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1355 1360 1365
Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1370 1375 1380
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1415 1420 1425
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1430 1435 1440
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1445 1450 1455
Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545
Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575
Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605
Leu Gly Gly Asp Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Ala
1610 1615 1620
Ser Arg Gly Val Asn Lys Val Ile Leu Val Gly Asn Leu Gly Gln

1625 1630 1635

Asp Pro Glu Val Arg Tyr Met Pro Asn Gly Gly Ala Val Ala Asn

1640 1645 1650

Ile Thr Leu Ala Thr Ser Glu Ser Trp Arg Asp Lys Ala Thr Gly

1655 1660 1665

Glu Met Lys Glu Gln Thr Glu Trp His Arg Val Val Leu Phe Gly

1670 1675 1680

Lys Leu Ala Glu Val Ala Ser Glu Tyr Leu Arg Lys Gly Ser Gln

1685 1690 1695

Val Tyr Ile Glu Gly Gln Leu Arg Thr Arg Lys Trp Thr Asp Gln

1700 1705 1710

Ser Gly Gln Asp Arg Tyr Thr Thr Glu Val Val Val Asn Val Gly

1715 1720 1725

Gly Thr Met Gln Met Leu Gly Gly Arg Gln Gly Gly Gly Ala Pro

1730 1735 1740

Ala Gly Gly Asn Ile Gly Gly Gly Gln Pro Gln Gly Gly Trp Gly

1745 1750 1755

Gln Pro Gln Gln Pro Gln Gly Gly Asn Gln Phe Ser Gly Gly Ala

1760 1765 1770

Gln Ser Arg Pro Gln Gln Ser Ala Pro Ala Ala Pro Ser Asn Glu

1775 1780 1785

Pro Pro Met Asp Phe Asp Asp Asp Ile Pro Phe Ser Gly Gly Ser

1790 1795 1800

Pro Lys Lys Lys Arg Lys Val

1805 1810

<210> 180

<211> 1716

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 180

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu

675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser

690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720

Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
900 905 910

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala

995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu

1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1235 1240 1245
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1250 1255 1260
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1280 1285 1290
Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1295 1300 1305
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1310 1315 1320
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1340 1345 1350
Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1355 1360 1365
Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1370 1375 1380
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1400 1405 1410
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1415 1420 1425
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1430 1435 1440
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr

1610 1615 1620

Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val

1625 1630 1635

Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val

1640 1645 1650

Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr

1655 1660 1665

Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala

1670 1675 1680

Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly

1685 1690 1695

Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Pro Lys Lys Lys
1700 1705 1710

Arg Lys Val
1715

<210> 181

<211> 1722

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 181

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400
Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu

610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu

625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735

Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800

Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895
Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
900 905 910
Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005
Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1175 1180 1185
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
1190 1195 1200
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
1205 1210 1215
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1235 1240 1245
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1250 1255 1260
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1265 1270 1275
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1295 1300 1305
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1310 1315 1320
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
1325 1330 1335
Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1355 1360 1365
Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1370 1375 1380
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605

Leu Gly Gly Asp Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

1610 1615 1620

Ala Thr Pro Glu Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu
1625 1630 1635
Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro
1640 1645 1650

Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile
1655 1660 1665
Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met
1670 1675 1680
Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val
1685 1690 1695
Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly
1700 1705 1710

Gly Ser Pro Lys Lys Lys Arg Lys Val
1715 1720

<210> 182
<211> 1726
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 182

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser

225 230 235 240

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly

245 250 255

Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

260 265 270

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

275 280 285

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

290 295 300

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

305 310 315 320

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

325 330 335

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

340 345 350

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

355 360 365

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

370 375 380

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

385 390 395 400

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

405 410 415

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

420 425 430

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

435 440 445

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

450 455 460

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

465 470 475 480

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

485 490 495

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

500 505 510

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

515 520 525

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

530 535 540

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

545 550 555 560

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

565 570 575

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

580 585 590

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
595 600 605
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

610 615 620
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
625 630 635 640
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
645 650 655
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
660 665 670
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

675 680 685
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
690 695 700
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
705 710 715 720
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
725 730 735
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

740 745 750
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
755 760 765
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
770 775 780
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
785 790 795 800
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

805 810 815
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
820 825 830
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

835 840 845

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

850 855 860

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

865 870 875 880

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

885 890 895

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

900 905 910

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

915 920 925

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

930 935 940

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

945 950 955 960

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

965 970 975

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

980 985 990

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

995 1000 1005

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met

1010 1015 1020

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

1025 1030 1035

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser

1040 1045 1050

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn

1055 1060 1065

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr

1070 1075 1080

Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val

1085 1090 1095

Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp

1100 1105 1110

Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp

1115 1120 1125

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp

1130 1135 1140

Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp

1145 1150 1155

Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

1160 1165 1170

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

1175 1180 1185

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr

1190 1195 1200

Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu

1205 1210 1215

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr

1220 1225 1230

Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr

1235 1240 1245

Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys

1250 1255 1260

Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val

1265 1270 1275

Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr

1280 1285 1290

Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr

1295 1300 1305

Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile

1310 1315 1320

Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg

1325 1330 1335

Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn

1340 1345 1350

Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu

1355 1360 1365

Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys

1370 1375 1380

Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr

1385 1390 1395

Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys

1400 1405 1410

Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile

1415 1420 1425

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu

1430 1435 1440

Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu

1445 1450 1455

Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met

1460 1465 1470

Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu

1475 1480 1485

Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu

1490 1495 1500

Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe

1505 1510 1515

Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile

1520 1525 1530

Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp

1535 1540 1545

Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg

1550 1555 1560

Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu

1565 1570 1575

Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg

1580 1585 1590

Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile

1595 1600 1605

His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser

1610 1615 1620

Gln Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile

1625 1630 1635

Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile

1640 1645 1650

Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro

1655 1660 1665

Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp

1670 1675 1680

Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro

1685 1690 1695

Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys

1700 1705 1710

Met Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1715 1720 1725

<210> 183

<211> 1797

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 183

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240

Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

260 265 270

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

305 310 315 320

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

325 330 335

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

340 345 350

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

355 360 365

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

370 375 380

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

405 410 415

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

420 425 430

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

435 440 445

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu

```
        755                 760                 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
    770                 775                 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785                 790                 795                 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
                805                 810                 815

Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
            820                 825                 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
        835                 840                 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
    850                 855                 860
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865                 870                 875                 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
                885                 890                 895
Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
            900                 905                 910
Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
        915                 920                 925
Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
    930                 935                 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945                 950                 955                 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
                965                 970                 975
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
            980                 985                 990
Lys Val Met Gly Arg His Lys Pro  Glu Asn Ile Val Ile  Glu Met Ala
        995                 1000                1005
```

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110
Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1145 1150 1155
Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
1160 1165 1170
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
1190 1195 1200
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
1205 1210 1215
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1220 1225 1230
Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1250 1255 1260
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1265 1270 1275
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1280 1285 1290
Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1310 1315 1320
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
1325 1330 1335
Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1340 1345 1350
Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1370 1375 1380
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1385 1390 1395
Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1400 1405 1410
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1430 1435 1440
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1445 1450 1455
Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575
Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile
1610 1615 1620
Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu
1625 1630 1635
Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu
1640 1645 1650
Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu
1655 1660 1665

Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp
1670 1675 1680
Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met
1685 1690 1695
Leu Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu

1700 1705 1710

Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro

1715 1720 1725

Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile

1730 1735 1740

Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met

1745 1750 1755

Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val

1760 1765 1770

Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly

1775 1780 1785

Gly Ser Pro Lys Lys Lys Arg Lys Val

1790 1795

<210> 184

<211> 1825

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 184

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser

225 230 235 240

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly

245 250 255

Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

260 265 270

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

275 280 285

Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys

290 295 300

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu

305 310 315 320

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys

325 330 335

Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys

340 345 350

Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu

355 360 365

Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp

370 375 380

Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys

385 390 395 400

Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu

405 410 415

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly

420 425 430

Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu

435 440 445

Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser

450 455 460

Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

465 470 475 480

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly

485 490 495

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe

500 505 510

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys

515 520 525

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp

530 535 540

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

545 550 555 560

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro

565 570 575

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

580 585 590

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
595 600 605
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp

610 615 620
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
625 630 635 640
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
645 650 655
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
660 665 670
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp

675 680 685
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
690 695 700
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
705 710 715 720
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
725 730 735
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile

740 745 750
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
755 760 765
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
770 775 780
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
785 790 795 800
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn

805 810 815
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
820 825 830
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn

835 840 845

Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys

850 855 860

Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val

865 870 875 880

Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu

885 890 895

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys

900 905 910

Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn

915 920 925

Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

930 935 940

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

945 950 955 960

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

965 970 975

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

980 985 990

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

995 1000 1005

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met

1010 1015 1020

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

1025 1030 1035

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser

1040 1045 1050

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn

1055 1060 1065

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr

1070 1075 1080

Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val

1085 1090 1095

Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp

1100 1105 1110

Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp

1115 1120 1125

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp

1130 1135 1140

Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp

1145 1150 1155

Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

1160 1165 1170

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

1175 1180 1185

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr

1190 1195 1200

Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu

1205 1210 1215

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr

1220 1225 1230

Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr

1235 1240 1245

Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys

1250 1255 1260

Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val

1265 1270 1275

Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr

1280 1285 1290

Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr

1295 1300 1305

Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile

1310 1315 1320

Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg

1325 1330 1335

Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn

1340 1345 1350

Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu

1355 1360 1365

Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys

1370 1375 1380

Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr

1385 1390 1395

Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys

1400 1405 1410

Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile

1415 1420 1425

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu

1430 1435 1440

Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu

1445 1450 1455

Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met

1460 1465 1470

Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu

1475 1480 1485

Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu

1490 1495 1500

Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe

1505 1510 1515

Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile

1520 1525 1530

Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp

1535 1540 1545

Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg

1550 1555 1560

Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu

1565 1570 1575

Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg

1580 1585 1590

Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile

1595 1600 1605

His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser

1610 1615 1620

Gln Leu Gly Gly Asp Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

1625 1630 1635

Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu

1640 1645 1650

Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu

1655 1660 1665

Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala

1670 1675 1680

Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp

1685 1690 1695

Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn

1700 1705 1710

Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Gly Gly Ser

1715 1720 1725

Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly

1730 1735 1740

Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu

1745 1750 1755

Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val

1760 1765 1770

His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu

1775 1780 1785

Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln

1790 1795 1800

Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser

1805 1810 1815

Pro Lys Lys Lys Arg Lys Val

1820 1825

<210> 185

<211> 1719

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 185

Met Gly Ser Ser His His His His His His Ser Ser Glu Thr Gly Pro

1 5 10 15

Val Ala Val Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe

20 25 30

Glu Val Phe Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu

35 40 45

Tyr Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser

50 55 60

Gln Asn Thr Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe Thr

65 70 75 80

Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp Phe

85 90 95

Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu Phe

100 105 110

Leu Ser Arg Tyr Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg Leu

115 120 125

Tyr His His Ala Asp Pro Arg Asn Arg Gln Gly Leu Arg Asp Leu Ile

130 135 140

Ser Ser Gly Val Thr Ile Gln Ile Met Thr Glu Gln Glu Ser Gly Tyr

145 150 155 160

Cys Trp Arg Asn Phe Val Asn Tyr Ser Pro Ser Asn Glu Ala His Trp

165 170 175

Pro Arg Tyr Pro His Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr

180 185 190

Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys

195 200 205

Gln Pro Gln Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr

210 215 220

Gln Arg Leu Pro Pro His Ile Leu Trp Ala Thr Gly Leu Lys Ser Gly

225 230 235 240

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys

245 250 255

Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala

260 265 270

Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu

275 280 285

Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu

290 295 300

Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr

305 310 315 320

Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln

325 330 335

Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His

340 345 350

Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg

355 360 365

His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys

370 375 380

Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp

385 390 395 400

Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys
405 410 415
Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser
420 425 430
Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu
435 440 445
Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile

450 455 460
Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala
465 470 475 480
Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala
485 490 495
Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala
500 505 510
Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu

515 520 525
Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu
530 535 540
Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg
545 550 555 560
Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys
565 570 575
Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val

580 585 590
Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser
595 600 605
Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu
610 615 620
Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu
625 630 635 640
Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg

645 650 655

Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu

660 665 670

His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp

675 680 685

Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr

690 695 700

Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg

705 710 715 720

Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp

725 730 735

Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp

740 745 750

Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr

755 760 765

Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr

770 775 780

Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala

785 790 795 800

Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln

805 810 815

Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu

820 825 830

Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His

835 840 845

Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu

850 855 860

Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu

865 870 875 880

Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe

885 890 895

Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp

900 905 910

Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser

915 920 925

Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg

930 935 940

Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp

945 950 955 960

Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His

965 970 975

Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln

980 985 990

Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys

995 1000 1005

Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr

1010 1015 1020

Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

1025 1030 1035

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

1040 1045 1050

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr

1055 1060 1065

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

1070 1075 1080

Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser

1085 1090 1095

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser

1100 1105 1110

Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val

1115 1120 1125

Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys

1130 1135 1140

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg

1145 1150 1155

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln

1160 1165 1170

Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu

1175 1180 1185

Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile

1190 1195 1200

Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp

1205 1210 1215

Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn

1220 1225 1230

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1235 1240 1245

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1250 1255 1260

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1265 1270 1275

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1280 1285 1290

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly

1295 1300 1305

Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly

1310 1315 1320

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys

1325 1330 1335

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val

1340 1345 1350

Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn

1355 1360 1365

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys

1370 1375 1380

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val

1385 1390 1395

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val

1400 1405 1410

Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu

1415 1420 1425

Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

1430 1435 1440

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu

1445 1450 1455

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu

1460 1465 1470

Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe

1475 1480 1485

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu

1490 1495 1500

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1505 1510 1515

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1520 1525 1530

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1535 1540 1545

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1550 1555 1560

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1565 1570 1575

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1580 1585 1590

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1595 1600 1605

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly

1610 1615 1620

Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys

1625 1630 1635

Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val

1640 1645 1650

Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His

1655 1660 1665

Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr

1670 1675 1680

Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp

1685 1690 1695

Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Pro

1700 1705 1710

Lys Lys Lys Arg Lys Val

1715

<210> 186

<211> 1719

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 186

Met Gly Ser Ser His His His His His His Ser Ser Glu Thr Gly Pro

1 5 10 15

Val Ala Val Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe

20 25 30

Glu Val Phe Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu

35 40 45

Tyr Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser

50 55 60

Gln Asn Thr Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe Thr

65 70 75 80

Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp Phe

85 90 95

Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu Phe

100 105 110

Leu Ser Arg Tyr Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg Leu

115 120 125

Tyr His His Ala Asp Pro Arg Asn Arg Gln Gly Leu Arg Asp Leu Ile

130 135 140

Ser Ser Gly Val Thr Ile Gln Ile Met Thr Glu Gln Glu Ser Gly Tyr

145 150 155 160

Cys Trp Arg Asn Phe Val Asn Tyr Ser Pro Ser Asn Glu Ala His Trp

165 170 175

Pro Arg Tyr Pro His Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr

180 185 190

Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys

195 200 205

Gln Pro Gln Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr

210 215 220

Gln Arg Leu Pro Pro His Ile Leu Trp Ala Thr Gly Leu Lys Ser Gly

225 230 235 240

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys

245 250 255

Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala

260 265 270

Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu

275 280 285

Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu

290 295 300

Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr

305 310 315 320

Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln
325 330 335
Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His

340 345 350
Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg
355 360 365
His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys
370 375 380
Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp
385 390 395 400
Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys

405 410 415
Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser
420 425 430
Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu
435 440 445
Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile
450 455 460
Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala

465 470 475 480
Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala
485 490 495
Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala
500 505 510
Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu
515 520 525
Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu

530 535 540
Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg
545 550 555 560
Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys

565 570 575

Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val

580 585 590

Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser

595 600 605

Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu

610 615 620

Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu

625 630 635 640

Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg

645 650 655

Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu

660 665 670

His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp

675 680 685

Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr

690 695 700

Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg

705 710 715 720

Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp

725 730 735

Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Ala Phe Asp

740 745 750

Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr

755 760 765

Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr

770 775 780

Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala

785 790 795 800

Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln

805 810 815

Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu
820 825 830
Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His
835 840 845
Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu

850 855 860
Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu
865 870 875 880
Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe
885 890 895
Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp
900 905 910
Gly Ala Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser

915 920 925
Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg
930 935 940
Asn Phe Met Ala Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp
945 950 955 960
Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His
965 970 975
Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln

980 985 990
Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys
995 1000 1005
Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr
1010 1015 1020
Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
1025 1030 1035
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

1040 1045 1050
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr

1055 1060 1065

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile

1070 1075 1080

Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser

1085 1090 1095

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser

1100 1105 1110

Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val

1115 1120 1125

Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys

1130 1135 1140

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg

1145 1150 1155

Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln

1160 1165 1170

Leu Val Glu Thr Arg Ala Ile Thr Lys His Val Ala Gln Ile Leu

1175 1180 1185

Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile

1190 1195 1200

Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp

1205 1210 1215

Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn

1220 1225 1230

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1235 1240 1245

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1250 1255 1260

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1265 1270 1275

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1280 1285 1290

```
Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly
    1295                 1300                 1305
Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly
    1310                 1315                 1320
Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys
    1325                 1330                 1335
Val Leu  Ser Met Pro Gln Val  Asn Ile Val Lys Lys  Thr Glu Val

    1340                 1345                 1350
Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn
    1355                 1360                 1365
Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys
    1370                 1375                 1380
Tyr Gly  Gly Phe Asp Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val
    1385                 1390                 1395
Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val

    1400                 1405                 1410
Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu
    1415                 1420                 1425
Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val
    1430                 1435                 1440
Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu
    1445                 1450                 1455
Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu

    1460                 1465                 1470
Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe
    1475                 1480                 1485
Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu
    1490                 1495                 1500
Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu Gln His  Lys His Tyr
    1505                 1510                 1515
Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val
```

1520 1525 1530

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1535 1540 1545

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1550 1555 1560

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1565 1570 1575

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1580 1585 1590

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1595 1600 1605

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly

1610 1615 1620

Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys

1625 1630 1635

Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val

1640 1645 1650

Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His

1655 1660 1665

Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr

1670 1675 1680

Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp

1685 1690 1695

Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Pro

1700 1705 1710

Lys Lys Lys Arg Lys Val

1715

<210> 187

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 187

ggtgtttcgt cctttccaca ag 22

<210> 188

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 188

gcttgcagat ggccatggcg gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 189

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 189

tgtcacacat gtagttgtag gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 190

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 190

gaagcgcctg gcagtgtacc gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 191

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 191

gagtccgagc agaagaagaa gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 192

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 192

ggaatccctt ctgcagcacc gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 193

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 193

gaacacaaag catagactgc gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 194

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 194

ggcccagact gagcacgtag agttttagag ctagaaatag caagttaaaa taaggc 56

<210> 195

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 195

ggcactgcgg ctggaggtgg gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 196

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 196

gtcatcttag tcattacctg gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 197

<211> 42

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 197

attattatta ttccgcggat ttatttattt atttatttat tt 42

<210> 198

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 198

attattatta ttagctattt atttatttat ttatttattt 40

<210> 199

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 199

attattatta ttcccggatt tatttattta tttatttatt t 41

<210> 200

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 200

aaaaaaagca ccgactcggt g 21

<210> 201

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 201

taatacgact cactataggc cgcggattta tttatttaag ttttagagct agaaatagca 60

<210> 202

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 202

taatacgact cactataggt ccgcggattt atttatttag ttttagagct agaaatagca 60

<210> 203

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 203

taatacgact cactataggt tccgcggatt tatttattag ttttagagct agaaatagca 60

<210> 204

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 204

taatacgact cactatagga ttccgcggat ttatttattg ttttagagct agaaatagca 60

<210> 205

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 205

taatacgact cactataggt attccgcgga tttatttatg tttttagagc tagaaatagc 60

a 61

<210> 206

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 206

taataacgac tcactatagg ttattccgcg gatttattta gtttagagct agaaatagca 60

<210> 207

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 207

taatacgact cactatagga ttattccgcg gatttatttg ttttagagct agaaatagca 60

<210> 208

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 208

taatacgact cactataggt attattccgc ggatttattg ttttagagct agaaatagca 60

<210> 209

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 209

taatacgact cactatagga ttattatccg cggatttatg ttttagagct agaaatagca 60

<210> 210

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 210

taatacgact cactataggt attatattcc gcggatttag ttttagagct agaaatagca 60

<210> 211

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 211

taatacgact cactataggt tattatattc cgcggatttg ttttagagct agaaatagca 60

<210> 212

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 212

taatacgact cactatagga ttattatatt ccgcggattg ttttagagct agaaatagca 60

<210> 213

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 213

taatacgact cactataggt attattatat tccgcggatg ttttagagct agaaatagca 60

<210> 214

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 214

taatacgact cactatagga ttattattat taccgcggag ttttagagct agaaatagca 60

<210> 215

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 215

taatacgact cactatagga ttattattat tattaccgcg ttttagagct agaaatagca 60

<210> 216

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 216

taatacgact cactatagga tattaattta tttatttaag ttttagagct agaaatagca 60

<210> 217

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 217

taatacgact cactataggg gaggacgtgc gcggccgccg ttttagagct agaaatagca 60

<210> 218

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 218

taatacgact cactataggc tgtggcagtg gcaccagaag ttttagagct agaaatagca 60

<210> 219

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 219

taatacgact cactataggg aagcgcctgg cagtgtaccg ttttagagct agaaatagca 60

<210> 220

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 220

taatacgact cactataggc ctcccggccg gcggtatccg ttttagagct agaaatagca 60

<210> 221
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 221
taatacgact cactataggg cttgcagatg gccatggcgg ttttagagct agaaatagca 60

<210> 222
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 222
taatacgact cactatagga cacatgcagt tgtagtggag ttttagagct agaaatagca 60

<210> 223
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 223
taatacgact cactataggt gtcacacatg tagttgtagg ttttagagct agaaatagca 60

<210> 224
<211> 31
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 224
gtaggtagtt aggatgaatg gaaggttggt a 31

<210> 225
<211> 80
<212> DNA
<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 225

gtccatggat ccagaggtca tccattaaat aaataaatcc gcggggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 226

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 226

gtccatggat ccagaggtca tccataaata aataaatccg cggaagctat accaaccttc 60

cattcatcct aactacctac 80

<210> 227

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 227

gtccatggat ccagaggtca tccataataa ataaatccgc ggaaggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 228

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 228

gtccatggat ccagaggtca tccaaataaa taaatccgcg gaatggctat accaaccttc 60

cattcatcct aaactaccta c 81

<210> 229

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 229

gtccatggat ccagaggtca tccaataaat aaatccgcgg aataggctat accaaccttc 60

cattcatcct aactacctac 80

<210>

230

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 230

gtccatggat ccagaggtca tccaataaat aaatccgcgg aataaggcta tatccaacct 60

tccattcatc ctaactacct ac 82

<210> 231

<211> 79

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 231

gtccatggat ccagaggtca tccaaaataa tccgcggaat aatggctata ccaaccttcc 60

attcatccta actacctac 79

<210> 232

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 232

gtccatggat ccagaggtca tccaaataaa tccgcggaat aataggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 233

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 233

gtccatggat ccagaggtca tccaataaat ccgcggataa taatggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 234

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 234

gtccatggat ccagaggtca tccataaatc cgcggaatat aataggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 235

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 235

gtccatggat ccagaggtca tccaaaatcc gcggaatata ataaggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 236

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 236

gtccatggat ccagaggtca tccaaatccg cggaatataa taatggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 237

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 237

gtccatggat ccagaggtca tccaatccgc ggaatataat aataggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 238

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 238

gtccatggat ccagaggtca tccatccgcg gtaataataa taatggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 239

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 239

gtccatggat ccagaggtca tccagcggta ataataataa taatggctat accaaccttc 60

cattcatcct aactacctac 80

<210> 240

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 240

gtccatggat ccagaggtca tccattaaat aaataaatta atattactat aaccaacctt 60

ccattcatcc taactaccta c 81

<210> 241

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 241

gtaggtagtt aggatgaatg gaaggttggt gtagattatt atcugcggat ttattggatg 60

acctctggat ccatggacat 80

<210> 242

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 242

gcacctcgcc gcggtactgc accaggcggc cgcgcacgtc ctccatgtct accaaccttc 60

cattcatcct aactacctac 80

<210> 243

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 243

cggcgccctc gcgggccccg gcctggtaca ctgccaggcg cttctgcagt accaaccttc 60

cattcatcct aactacctac 80

<210> 244

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 244

ctcttacctg gactctggaa tccattctgg tgccactgcc acagctcctt accaaccttc 60

cattcatcct aactacctac 80

<210> 245

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 245

ggagacgtgc ctgttggaac atcctggata ccgccggccg ggaggagtac taccaacctt 60

ccattcatcc taactaccta c 81

<210> 246

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 246

acccccgccc ggcacccgcg tccgcgccat ggccatctgc aagcagtcat accaaccttc 60

cattcatcct aactacctac 80

<210> 247

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 247

aggttggctc tgactgtacc accatccact acaactgcat gtgtaacagt accaaggttc 60

cattcatcct aactacctac 80

<210> 248

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 248

tggctctgac tgtaccacca tccactacaa ctacatgtgt gacagttcct accaaccttc 60

cattcatcct aactacctac 80

<210> 249

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 249

taatacgact cactataggt tatttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 250

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 250

taatacgact cactatagga tatttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 251

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 251

taatacgact cactataggc tatttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 252

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 252

taatacgact cactataggg tatttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 253

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 253

taatacgact cactataggt aatttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 254

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 254

taatacgact cactataggt catttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 255

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 255

taatacgact cactataggt gatttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 256

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 256

taatacgact cactataggt tttttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 257

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 257

taatacgact cactataggt tctttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 258

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 258

taatacgact cactataggt tgtttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 259

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 259

taatacgact cactataggt taattcgtgg atttatttag ttttagagct agaaatagca 60

<210> 260

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 260

taatacgact cactataggt tacttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 261

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 261

taatacgact cactataggt tagttcgtgg atttatttag ttttagagct agaaatagca 60

<210> 262

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 262

taatacgact cactataggt tatatcgtgg atttatttag ttttagagct agaaatagca 60

<210> 263

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 263

taatacgact cactataggt tatctcgtgg atttatttag ttttagagct agaaatagca 60

<210> 264

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 264

taatacgact cactataggt tatgtcgtgg atttatttag ttttagagct agaaatagca 60

<210> 265

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 265

taatacgact cactataggt tattatcgtg gatttattta gttttagagc tagaaatagc 60

a 61

<210> 266

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 266

taatacgact cactataggt tattctcgtg gatttattta gttttagagc tagaaatagc 60

a 61

<210> 267

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 267

taatacgact cactataggt tattgtcgtg gatttattta gttttagagc tagaaatagc 60

a 61

<210> 268

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 268

taatacgact cactataggt tatttcatgg atttatttag ttttagagct agaaatagca 60

<210> 269

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 269

taatacgact cactataggt tatttcttgg atttatttag ttttagagct agaaatagca 60

<210> 270

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 270

taatacgact cactataggt tatttcctgg atttatttag ttttagagct agaaatagca 60

<210> 271

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 271

taatacgact cactataggt tatttcgagg atttatttag ttttagagct agaaatagca 60

<210> 272

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 272

taatacgact cactataggt tatttcgcgg atttatttag ttttagagct agaaatagca 60

<210> 273

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 273

taatacgact cactataggt tatttcgggg atttatttag ttttagagct agaaatagca 60

<210> 274

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 274

taatacgact cactataggt tatttcgtag atttatttag ttttagagct agaaatagca 60

<210> 275

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 275

taatacgact cactataggt tatttcgttg atttatttag ttttagagct agaaatagca 60

<210> 276

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 276

taatacgact cactataggt tatttcgtcg atttatttag ttttagagct agaaatagca 60

<210> 277

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 277

taatacgact cactataggt tatttcgtga atttatttag ttttagagct agaaatagca 60

<210> 278

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 278

taatacgact cactataggt tatttcgtgt atttatttag ttttagagct agaaatagca 60

<210> 279

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 279

taatacgact cactataggt tatttcgtgc atttatttag ttttagagct agaaatagca 60

<210> 280

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 280

taatacgact cactataggt tatttcgtgg ttttatttag ttttagagct agaaatagca 60

<210> 281

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 281

taatacgact cactataggt tatttcgtgg ctttatttag ttttagagct agaaatagca 60

<210> 282

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 282

taatacgact cactataggt tatttcgtgg gtttatttag ttttagagct agaaatagca 60

<210> 283

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 283

taatacgact cactataggt tatttcgtgg aattatttag ttttagagct agaaatagca 60

<210> 284

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 284

taatacgact cactataggt tatttcgtgg acttatttag ttttagagct agaaatagca 60

<210> 285

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 285

taatacgact cactataggt tatttcgtgg agttatttag ttttagagct agaaatagca 60

<210> 286

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 286

taatacgact cactataggt tcccccccg atttatttag ttttagagct agaaatagca 60

<210> 287

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 287

taatacgact cactataggc gcacccgtgg atttatttag ttttagagct agaaatagca 60

<210> 288

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 288

taatacgact cactataggc tcgcacgtgg atttatttag ttttagagct agaaatagca 60

<210> 289

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 289

taatacgact cactataggc cctcgcgtgg atttatttag ttttagagct agaaatagca 60

<210> 290

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 290

taatacgact cactataggc accctcgtgg atttatttag ttttagagct agaaatagca 60

<210> 291

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 291

taatacgact cactataggt cgcacccgtg gatttattag ttttagagct agaaatagca 60

<210> 292

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 292

taatacgact cactataggc ctcgcacgtg gatttattag ttttagagct agaaatagca 60

<210> 293

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 293

taatacgact cactatagga ccctcgcgtg gatttattag ttttagagct agaaatagca 60

<210> 294

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 294

taatacgact cactataggg caccctcgtg gatttattag ttttagagct agaaatagca 60

<210> 295

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 295

taatacgact cactataggg agtccgagca gaagaagaag ttttagagct agaaatagca 60

<210> 296

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 296

taatacgact cactataggg gaatcccttc tgcagcaccg ttttagagct agaaatagca 60

<210> 297

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 297

taatacgact cactataggg aacacaaagc atagactgcg ttttagagct agaaatagca 60

<210> 298

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 298

taatacgact cactataggg gcccagactg agcacgtgag ttttagagct agaaatagca 60

<210> 299

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 299

taatacgact cactataggg gcactgcggc tggaggtggg ttttagagct agaaatagca 60

<210> 300

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 300

taatacgact cactataggg tcatcttagt cattacctgg ttttagagct agaaatagca 60

<210> 301

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 301

acgtaaacgg ccacaagttc ttatttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 302

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 302

acgtaaacgg ccacaagttc atatttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 303

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 303

acgtaaacgg ccacaagttc ctatttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 304

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 304

acgtaaacgg ccacaagttc gtatttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 305

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 305

acgtaaacgg ccacaagttc taatttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 306

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 306

acgtaaacgg ccacaagttc tcatttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 307

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 307

acgtaaacgg ccacaagttc tgatttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 308

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 308

acgtaaacgg ccacaagttc ttttttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 309

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 309

acgtaaacgg ccacaagttc ttctttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 310

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 310

acgtaaacgg ccacaagttc ttgtttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 311

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 311

acgtaaacgg ccacaagttc ttaattcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 312

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 312

acgtaaacgg ccacaagttc ttacttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 313

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 313

acgtaaacgg ccacaagttc ttagttcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 314

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 314

acgtaaacgg ccacaagttc ttatatcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 315

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 315

acgtaaacgg ccacaagttc ttatctcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 316

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 316

acgtaaacgg ccacaagttc ttatgtcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 317

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 317

acgtaaacgg ccacaagttc ttattacgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 318

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 318

acgtaaacgg ccacaagttc ttattccgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 319

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 319

acgtaaacgg ccacaagttc ttattgcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 320

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 320

acgtaaacgg ccacaagttc ttatttcatg gatttattta tggcatcttc ttcaaggacg 60

<210> 321

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 321

acgtaaacgg ccacaagttc ttatttcttg gatttattta tggcatcttc ttcaaggacg 60

<210> 322

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 322

acgtaaacgg ccacaagttc ttatttcctg gatttattta tggcatcttc ttcaaggacg 60

<210> 323

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 323

acgtaaacgg ccacaagttc ttatttcgaa gatttattta tggcatcttc ttcaaggacg 60

<210> 324

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 324

acgtaaacgg ccacaagttc ttatttcgac gatttattta tggcatcttc ttcaaggacg 60

<210> 325

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 325

acgtaaacgg ccacaagttc ttatttcgag gatttattta tggcatcttc ttcaaggacg 60

<210> 326
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 326
acgtaaacgg ccacaagttc ttatttcgta gatttattta tggcatcttc ttcaaggacg 60

<210> 327
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 327
acgtaaacgg ccacaagttc ttatttcgtt gatttattta tggcatcttc ttcaaggacg 60

<210> 328
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 328
acgtaaacgg ccacaagttc ttatttcgtc gatttattta tggcatcttc ttcaaggacg 60

<210> 329
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 329
acgtaaacgg ccacaagttc ttatttcgtg aatttattta tggcatcttc ttcaaggacg 60

<210> 330
<211> 60
<212> DNA
<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 330

acgtaaacgg ccacaagttc ttatttcgtg tatttattta tggcatcttc ttcaaggacg 60

<210> 331

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 331

acgtaaacgg ccacaagttc ttatttcgtg catttattta tggcatcttc ttcaaggacg 60

<210> 332

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 332

acgtaaacgg ccacaagttc ttatttcgtg gttttattta tggcatcttc ttcaaggacg 60

<210> 333

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 333

acgtaaacgg ccacaagttc ttatttcgtg gctttattta tggcatcttc ttcaaggacg 60

<210> 334

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 334

acgtaaacgg ccacaagttc ttatttcgtg ggtttattta tggcatcttc ttcaaggacg 60

<210> 335

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 335

acgtaaacgg ccacaagttc ttatttcgtg gaattattta tggcatcttc ttcaaggacg 60

<210> 336

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 336

acgtaaacgg ccacaagttc ttatttcgtg gacttattta tggcatcttc ttcaaggacg 60

<210> 337

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 337

acgtaaacgg ccacaagttc ttatttcgtg gagttattta tggcatcttc ttcaaggacg 60

<210> 338

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 338

acgtaaacgg ccacaaagtt tcgcacccgt ggatttattt atggcatctt cttcaaggac 60

g 61

<210> 339

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 339

acgtaaacgg ccacaagttc ctcgcacgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 340

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 340

acgtaaacgg ccacaagtta ccctcgcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 341

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 341

acgtaaacgg ccacaagttg caccctcgtg gatttattta tggcatcttc ttcaaggacg 60

<210> 342

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 342

acgtaaacgg ccacaagtat tcgcacccgt ggatttatta tggcatcttc ttcaaggacg 60

<210> 343

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 343

acgtaaacgg ccacaagtat cctcgcacgt ggatttatta tggcatcttc ttcaaggacg 60

<210> 344

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 344

acgtaaacgg ccacaagtat accctcgcgt ggatttatta tggcatcttc ttcaaggacg 60

<210> 345

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 345

acgtaaacgg ccacaagtat gcaccctcgt ggatttatta tggcatcttc ttcaaggacg 60

<210> 346

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 346

gtaggtagtt aggatgaatg gaaggttggt aggcctgagt ccgagcagaa gaagaagggc 60

tcccatcaca tcaaccggtg 80

<210> 347

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 347

gtaggtagtt aggatgaatg gaaggttggt actcatggaa tcccttctgc agcacctgga 60

tcgcttttcc gagcttctgg 80

<210> 348

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 348

gtaggtagtt aggatgaatg gaaggttggt aaactggaac acaaagcata gactgcgggg 60

cgggccagcc tgaatagctg 80

<210> 349

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 349

gtaggtagtt aggatgaatg gaaggttggt acttggggcc cagactgagc acgtgatggc 60

agaggaaagg aagccctgct 80

<210> 350

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 350

gtaggtagtt aggatgaatg gaaggttggt accggtggca ctgcggctgg aggtgggggt 60

taaagcggag actctggtgc 80

<210> 351

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 351

gtaggtagtt aggatgaatg gaaggttggt atggcagtca tcttagtcat tacctgaggt 60

gttcgttgta actcatataa 80

<210> 352

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 352

gtaggtagtt aggatgaatg gaaggttggt atggcagtca tcttagtcat taccttgagg 60

tgttcgttgt aactcatata a 81

<210> 353

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 353

acactctttc cctacacgac gctcttccga tctnnnnacg taaacggcca caa 53

<210> 354

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 354

tggagttcag acgtgtgctc ttccgatctc gtccttgaag aagatgc 47

<210> 355

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 355

acactctttc cctacacgac gctcttccga tctnnnngta ggtagttagg atgaatggaa 60

<210> 356

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 356

tggagttcag acgtgtgctc ttccgatctc accggttgat gtgatgg 47

<210> 357

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 357

tggagttcag acgtgtgctc ttccgatctc cagaagctcg gaaaagc 47

<210> 358

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 358

tggagttcag acgtgtgctc ttccgatctc agctattcag gctggc 46

<210> 359

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 359

tggagttcag acgtgtgctc ttccgatcta gcagggcttc ctttc 45

<210> 360

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 360

tggagttcag acgtgtgctc ttccgatctg caccagagtc tccg 44

<210> 361

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 361

tggagttcag acgtgtgctc ttccgatctt tatatgagtt acaacgaaca cc 52

<210> 362

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 362

acactctttc cctacacgac gctcttccga tctnnnncag ctcagcctga gtgttga 57

<210> 363

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 363

tggagttcag acgtgtgctc ttccgatctc tcgtgggttt gtggttgc 48

<210> 364

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 364

acactctttc cctacacgac gctcttccga tctnnnncat tgcagagagg cgtatca 57

<210> 365

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 365

tggagttcag acgtgtgctc ttccgatctg gggtcccagg tgctgac 47

<210> 366

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 366

acactctttc cctacacgac gctcttccga tctnnnncca gccccatctg tcaaact 57

<210> 367

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 367

tggagttcag acgtgtgctc ttccgatctt gaatggattc ttggaaacaa tga 53

<210> 368

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 368

acactctttc cctacacgac gctcttccga tctnnnnatg tgggctgcct agaaagg 57

<210> 369

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 369

tggagttcag acgtgtgctc ttccgatctc ccagccaaac ttgtcaacc 49

<210> 370

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 370

acactctttc cctacacgac gctcttccga tctnnnngaa cccaggtagc cagagac 57

<210> 371

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 371

tggagttcag acgtgtgctc ttccgatctt cctttcaacc cgaacggag 49

<210> 372

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 372

acactctttc cctacacgac gctcttccga tctnnnnctc ttctttattt ccagcaatgt 60

<210> 373

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 373

tggagttcag acgtgtgctc ttccgatctg ttttcatgtt ctaaaaatgt atccca 56

<210> 374

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 374

acactctttc cctacacgac gctcttccga tctnnnntac agtactcccc tgccctc 57

<210> 375

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 375

tggagttcag acgtgtgctc ttccgatctg ctgctcacca tcgctatct 49

<210> 376

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 376

acactctttc cctacacgac gctcttccga tctnnnncct catcttgggc ctgtgtt 57

<210> 377

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 377

tggagttcag acgtgtgctc ttccgatcta aatcggtaag aggtgggcc 49

<210> 378

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 378

acactctttc cctacacgac gctcttccga tctnnnngcg gacatggagg acgtg 55

<210> 379

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 379

tggagttcag acgtgtgctc ttccgatctc tgttccacca ggggccc 47

<210> 380

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 380

acactctttc cctacacgac gctcttccga tctnnnntgc ccaatcattg atgctttt 58

<210> 381

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 381

tggagttcag acgtgtgctc ttccgatcta gaaacattta ccatagacta tcacct 56

<210> 382

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 382

acactctttc cctacacgac gctcttccga tctnnnnagt agcctctttc tcaatgtgc 59

<210> 383

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 383

tggagttcag acgtgtgctc ttccgatctg ctttcacaag gatgcagtct 50

<210> 384

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 384

acactctttc cctacacgac gctcttccga tctnnnngag ctagactccg agggga 56

<210> 385

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 385

tggagttcag acgtgtgctc ttccgatctt cctcgtcctg ctctcactt 49

<210> 386

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 386

acactctttc cctacacgac gctcttccga tctnnnnaga ggctgaagag gaagacca 58

<210> 387

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 387

tggagttcag acgtgtgctc ttccgatctg gcccagctgt gcattctat 49

<210> 388

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 388

acactctttc cctacacgac gctcttccga tctnnnncca agagggccaa gtcctg 56

<210> 389

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 389

tggagttcag acgtgtgctc ttccgatctc agcgaggagt gacagcc 47

<210> 390

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 390

acactctttc cctacacgac gctcttccga tctnnnncac tccacctgat ctcgggg 57

<210> 391

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 391

tggagttcag acgtgtgctc ttccgatctc gaggagggag ggagcag 47

<210> 392

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 392

acactctttc cctacacgac gctcttccga tctnnnnacc acaaatgccc aagagac 57

<210> 393

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 393

tggagttcag acgtgtgctc ttccgatctg acacagtcaa gggccgg 47

<210> 394

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 394

acactctttc cctacacgac gctcttccga tctnnnnccc acctttgagg aggcaaa 57

<210> 395

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 395

tggagttcag acgtgtgctc ttccgatctt tccatctgag aagagagtgg t 51

<210> 396

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 396

acactctttc cctacacgac gctcttccga tctnnnngtc ataccttggc ccttcct 57

<210> 397

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 397

tggagttcag acgtgtgctc ttccgatctt ccctaggccc acaccag 47

<210> 398

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 398

acactctttc cctacacgac gctcttccga tctnnnnaac ccactgaaga agcaggg 57

<210> 399

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 399

tggagttcag acgtgtgctc ttccgatctg gtgcttaatc cggctccat 49

<210> 400

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 400
acactctttc cctacacgac gctcttccga tctnnnntcc agtgtttcca tcccgaa 57

<210> 401
<211> 49
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 401
tggagttcag acgtgtgctc ttccgatctc ctctgacctc cacaactct 49

<210> 402
<211> 57
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<220><221> misc_feature
<222> (34)..(37)
<223> n is a, c, g, or t
<400> 402
acactctttc cctacacgac gctcttccga tctnnnnctg ggtacagttc tgcgtgt 57

<210> 403
<211> 49
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 403
tggagttcag acgtgtgctc ttccgatctt cactctgagc atcgccaag 49

<210> 404
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 404

acactctttc cctacacgac gctcttccga tctnnnnggt ttagagccag tgaactagag 60

<210> 405

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 405

tggagttcag acgtgtgctc ttccgatctg caagacaaaa tcctctttat actttg 56

<210> 406

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 406

acactctttc cctacacgac gctcttccga tctnnnnggg aggggacggc cttac 55

<210> 407

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 407

tggagttcag acgtgtgctc ttccgatctg cctctggcga acatggc 47

<210> 408

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 408

acactctttc cctacacgac gctcttccga tctnnnntcc tggttaagag catgggc 57

<210> 409

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 409

tggagttcag acgtgtgctc ttccgatctg attgagtccc cacagcaca 49

<210> 410

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 410

acactctttc cctacacgac gctcttccga tctnnnncca gtgtttccca tccccaa 57

<210> 411

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 411

tggagttcag acgtgtgctc ttccgatctt gacctccaca actggaaaat 50

<210> 412

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 412

acactctttc cctacacgac gctcttccga tctnnnngct tccagaccca cctgaag 57

<210> 413

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 413

tggagttcag acgtgtgctc ttccgatcta ccgaggaaaa ttgcttgtcg 50

<210> 414

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 414

acactctttc cctacacgac gctcttccga tctnnnngtg tggagagtga gtaagcca 58

<210> 415

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 415

tggagttcag acgtgtgctc ttccgatcta cggtaggatg atttcaggca 50

<210> 416

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 416

acactctttc cctacacgac gctcttccga tctnnnncac aaagcagtgt agctcagg 58

<210> 417

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 417

tggagttcag acgtgtgctc ttccgatctt ttttggtact cgagtgttat tcag 54

<210> 418

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 418

acactctttc cctacacgac gctcttccga tctnnnntcc cctgttgacc tggagaa 57

<210> 419

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 419

tggagttcag acgtgtgctc ttccgatctc actgtacttg ccctgacca 49

<210> 420

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 420

acactctttc cctacacgac gctcttccga tctnnnnttg gtgttgacag ggagcaa 57

<210> 421

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 421

tggagttcag acgtgtgctc ttccgatctc tgagatgtgg gcagaaggg 49

<210> 422

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 422

acactctttc cctacacgac gctcttccga tctnnnntga gagggaacag aagggct 57

<210> 423

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 423

tggagttcag acgtgtgctc ttccgatgtg tccaaaggcc caagaacct 49

<210> 424

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 424

acactctttc cctacacgac gctcttccga tctnnnntcc tagcactttg gaaggtcg 58

<210> 425

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 425

tggagttcag acgtgtgctc ttccgatctg ctcatcttaa tctgctcagc c 51

<210> 426

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 426

acactctttc cctacacgac gctcttccga tctnnnnaaa ggagcagctc ttcctgg 57

<210> 427

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 427

tggagttcag acgtgtgctc ttccgatctg tctgcaccat ctcccacaa 49

<210> 428

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 428

acactctttc cctacacgac gctcttccga tctnnnnggc atggcttctg agactca 57

<210> 429

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 429

tggagttcag acgtgtgctc ttccgatctg tctcccttgc actccctgtc ttt 53

<210> 430

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 430

acactctttc cctacacgac gctcttccga tctnnnnttt ggcaatggag gcattgg 57

<210> 431

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 431

tggagttcag acgtgtgctc ttccgatctg aagaggctgc ccatgagag 49

<210> 432

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 432

acactctttc cctacacgac gctcttccga tctnnnnggt ctgaggctcg aatcctg 57

<210> 433

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 433

tggagttcag acgtgtgctc ttccgatctc tgtggcctcc atatccctg 49

<210> 434

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 434

acactctttc cctacacgac gctcttccga tctnnnnttt ccaccagaac tcagccc 57

<210> 435

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 435

tggagttcag acgtgtgctc ttccgatctc ctcggttcct ccacaacac 49

<210> 436

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 436

acactctttc cctacacgac gctcttccga tctnnnncac gggaaggaca ggagaag 57

<210> 437

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 437

tggagttcag acgtgtgctc ttccgatctg caggggaggg ataaagcag 49

<210> 438

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 438

acactctttc cctacacgac gctcttccga tctnnnngtc agtctcggcc cctca 55

<210> 439

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 439

tggagttcag acgtgtgctc ttccgatctg ccactgtaaa gctcttggg 49

<210> 440

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (35)..(38)

<223> n is a, c, g, or t

<400> 440

acactctttc cctacacgaa cgctcttccg atctnnnnag ggtagaggga cagagctg 58

<210> 441

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 441

tggagttcag acgtgtgctc ttccgatctg gaccccacat agtcagtgc 49

<210> 442

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 442

acactctttc cctacacgac gctcttccga tctnnnngct gtcagcccta tctccatc 58

<210> 443

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 443

tggagttcag acgtgtgctc ttccgatctt gggcaattag gacagggac 49

<210> 444

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 444

acactctttc cctacacgac gctcttccga tctnnnngca gcggaggagg tagattg 57

<210> 445

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 445

tggagttcag acgtgtgctc ttccgatctc tcagtacctg gagtcccga 49

<210> 446

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 446

acactctttc cctacacgac gctcttccga tctnnnngac aggctcagga aagctgt 57

<210> 447

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 447

tggagttcag acgtgtgctc ttccgatcta cacaagcctt tctccaggg 49

<210> 448

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 448

acactctttc cctacacgac gctcttccga tctnnnnaat agggggtgag actgggg 57

<210> 449

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 449

tggagttcag acgtgtgctc ttccgatctg cctcagacga gacttgagg 49

<210> 450

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 450

acactctttc cctacacgac gctcttccga tctnnnnggc cagcaggaaa ggaatct 57

<210> 451

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 451

tggagttcag acgtgtgctc ttccgatctt gactgcacct gtagccatg 49

<210> 452

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 452

acactctttc cctacacgac gctcttccga tctnnnntca aggaaatcac cctgccc 57

<210> 453

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 453

tggagttcag acgtgtgctc ttccgatcta acttccttgg tgtgcagct 49

<210> 454

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 454

acactctttc cctacacgac gctcttccga tctnnnnatg ggctcagcta cgtcatg 57

<210> 455

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 455

tggagttcag acgtgtgctc ttccgatcta atcgcagtgt ggtgggcaa 49

<210> 456

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 456

acactctttc cctacacgac gctcttccga tctnnnncgc acatcccttg tctctct 57

<210> 457

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 457

tggagttcag acgtgtgctc ttccgatctc tactggagca caccccaag 49

<210> 458

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 458

acactctttc cctacacgac gctcttccga tctnnnntgg gtcacgtagc tttggtc 57

<210> 459

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 459

tggagttcag acgtgtgctc ttccgatctt gtggccatgt gcaactaa 48

<210> 460

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 460

acactctttc cctacacgac gctcttccga tctnnnncta ctacgtgcca gctcagg 57

<210> 461

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 461

tggagttcag acgtgtgctc ttccgatcta cctcccctcc tcactaacc 49

<210> 462

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 462

acactctttc cctacacgac gctcttccga tctnnnngcc tcagctccat ttcctgt 57

<210> 463

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 463

tggagttcag acgtgtgctc ttccgatcta acctttatgg caccagggg 49

<210> 464

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 464

acactctttc cctacacgac gctcttccga tctnnnngag ctcagcatta gcaggct 57

<210> 465

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 465

tggagttcag acgtgtgctc ttccgatctt tcctggcttt ccgattccc 49

<210> 466

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 466

acactctttc cctacacgac gctcttccga tctnnnngtg caattggagg aggagct 57

<210> 467

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 467

tggagttcag acgtgtgctc ttccgatctc accagctaca ggcagaaca 49

<210> 468

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 468

acactctttc cctacacgac gctcttccga tctnnnncct acccccaaca cagatgg 57

<210> 469

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 469

tggagttcag acgtgtgctc ttccgatctc cacacaactc aggtcctcc 49

<210> 470

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 470

tcatctgtgc ccctccctcc ctggcccagg tgaaggtgtg gttccagaac cggaggacaa 60

agtacaaacg gcagaagctg gaggaggaag ggcctgagtt tgagcagaag aagaagggct 120

cccatcacat caaccggtgg cgcattgcca cgaagcaggc caatggggag gacatcgatg 180

tcacctccaa tgactagggt 200

<210> 471

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 471

accctagtca ttggaggtga catcgatctc ctccccattg gcctgcttcg tggcaatgcg 60

ccaccggttg atgtgatggg agcccttctt cttctgctca aactcaggcc cttcctcctc 120

cagcttctgc cgtttgtact ttgtcctccg gttctggaac cacaccttca cctgggccag 180

ggagggaggg gcacagatga 200

<210> 472

<211> 201

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 472

catgcaatta gtctatttct gctgcaagta agcatgcatt tgtaggcttg atgctttttt 60

tctgcttctc cagccctggc ctgggtcaat ccttggggct tagactgagc acgtgatggc 120

agaggaaagg aagccctgct tcctccagag ggcgtcgcag gacagctttt cctagacagg 180

gggctagtat gtgcagctcc t 201

<210> 473

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 473

aggagctgca catactagcc cctgtctagg aaaagctgtc ctgcgacgcc ctctggagga 60

agcagggctt cctttcctct gccatcacgt gctcagtcta agccccaagg attgacccag 120

gccagggctg gagaagcaga aaaaaagcat caagcctaca aatgcatgct tacttgcagc 180

agaaatagac taattgcatc 200

<210> 474

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 474

ggctgacaaa ggccgggctg ggtggaagga agggaggaag ggcgaggcag agggtccaaa 60

gcaggatgac aggcaggggc accgcggcgc cccggtggca ttgcggctgg aggtgggggt 120

taaagcggag actctggtgc tgtgtgacta cagtgggggc cctgccctct ctgagccccc 180

gcctccaggc ctgtgtgtgt 200

<210> 475

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 475

acacacacag gcctggaggc gggggctcag agagggcagg gcccccactg tagtcacaca 60

gcaccagagt ctccgcttta acccccacct ccagccgcaa tgccaccggg gcgccgcggt 120

gcccctgcct gtcatcctgc tttggaccct ctgcctcgcc cttcctccct tccttccacc 180

cagcccggcc tttgtcagcc 200

<210> 476

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 476

agcaccgagg agctgcgggt gcgcctcgcc tcccacctgc gcaagctgcg taagcggctc 60

ctccgcgatg ccgatgacct gcagaagtgc ctggcagtgt accaggccgg ggcccgcgag 120

ggcgccgagc gcggcctcag cgccatccgc gagcgcctgg ggcccctggt ggaacagggc 180

cgcgtgcggg ccgccactgt 200

<210> 477

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 477

acagtggcgg cccgcacgcg gccctgttcc accaggggcc ccaggcgctc gcggatggcg 60

ctgaggccgc gctcggcgcc ctcgcgggcc ccggcctggt acactgccag gcacttctgc 120

aggtcatcgg catcgcggag gagccgctta cgcagcttgc gcaggtggga ggcgaggcgc 180

acccgcagct cctcggtgct 200

<210> 478

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 478

actcccctgc cctcaacaag atgttttgcc aactggccaa gacctgccct gtgcagctgt 60

gggttgattc cacacccccg cccggcaccc gcgtccgcgc catggccatc tacaagcagt 120

cacagcacat gacggaggtt gtgaggcgct gcccccacca tgagcgctgc tcagatagcg 180

atggtgagca gctggggctg 200

<210> 479

<211> 200

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 479

cagccccagc tgctcaccat cgctatctga gcagcgctca tggtgggggc agcgcctcac 60

aacctccgtc atgtgctgtg actgcttgta gatggccatg gcgcggacgc gggtgccggg 120

cgggggtgtg gaatcaaccc acagctgcac agggcaggtc ttggccagtt ggcaaaacat 180

cttgttgagg gcaggggagt 200

<210> 480

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 480

gagtccgagc agaagaagaa ggg 23

<210> 481

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 481

gagtctaagc agaagaagaa gag 23

<210> 482

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 482

gaggccgagc agaagaaaga cgg 23

<210> 483

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 483

gagtcctagc aggagaagaa gag 23

<210> 484

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 484

ggcccagact gagcacgtga tgg 23

<210> 485

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 485

cacccagact gagcacgtgc tgg 23

<210> 486

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 486

gacacagact gggcacgtga ggg 23

<210> 487

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 487

agctcagact gagcaagtga ggg 23

<210> 488

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 488

gacccctcc accccgcctc cgg 23

<210> 489

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 489

ctacccctcc accccgcctc cgg 23

<210> 490

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 490

attcccccccc accccgcctc agg 23

<210> 491

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 491

acacccccccc accccgcctc agg 23

<210> 492

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 492

tgcccccccc accccacctc tgg 23

<210> 493

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 493

ggaatccctt ctgcagcacc tgg 23

<210> 494

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 494

ggaaccccgt ctgcagcacc agg 23

<210> 495

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 495

ggagtccctc ctacagcacc agg 23

<210> 496

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 496

agaggcccct ctgcagcacc agg 23

<210> 497

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 497

gacccctcc accccgcctc cgg 23

<210> 498

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 498

tccccctcc accccacctc cgg 23

<210> 499

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 499

tgcccacctc accccgcctc tgg 23

<210> 500

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 500

gcccctccca accccacctc tgg 23

<210> 501

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 501

caccccctc accccgcctc agg 23

<210> 502

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 502

acactctttc cctacacgac gctcttccga tctnnnn 37

<210> 503

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 503

tggagttcag acgtgtgctc ttccgatct 29

<210> 504

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (30)..(37)

<223> n is a, c, g, or t

<400> 504

aatgatacgg cgaccaccga gatctacacn nnnnnnnaca ctctttccct acacgac 57

<210> 505

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (25)..(31)

<223> n is a, c, g, or t

<400> 505

caagcagaag acggcatacg agatnnnnnn ngtgactgga gttcagacgt gtgctcttc 59

<210> 506

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 506

acactctttc cctacacgac gctcttccga tctnnnntcc tacaagtaac agtccaagaa 60

<210> 507

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 507

tggagttcag acgtgtgctc ttccgatctt tctgcaactt aacttacgtg aaa 53

<210> 508

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 508

acactctttc cctacacgac gctcttccga tctnnnnacc aagcccattt gtccagg 57

<210> 509

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 509

tggagttcag acgtgtgctc ttccgatctt ccttcttttt gagctttggg c 51

<210> 510

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 510

acactctttc cctacacgac gctcttccga tctnnnntcc ataccagcag cagttcc 57

<210> 511

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 511

tggagttcag acgtgtgctc ttccgatctc tcacctcagc tcctgcac 48

<210> 512

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 512

acactctttc cctacacgac gctcttccga tctnnnnccc actgattcta caccatggt 59

<210> 513

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 513

tggagttcag acgtgtgctc ttccgatctg gagttcccaa ccttttgac a 51

<210> 514

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 514

acactctttc cctacacgac gctcttccga tctnnnnctg gctggagatt cagagacac 59

<210> 515

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 515

tggagttcag acgtgtgctc ttccgatctc tgccccttct gacacacata c 51

<210> 516

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 516

acactctttc cctacacgac gctcttccga tctnnnnacc cctcaaggct tgacatttc 59

<210> 517

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 517

tggagttcag acgtgtgctc ttccgatctt gaaaagttgg gagaggggat g 51

<210> 518

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 518

acactctttc cctacacgac gctcttccga tctnnnnttt gtaccccagt cccctcatc 59

<210> 519

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 519

tggagttcag acgtgtgctc ttccgatctt gaagttacgg ggatgtcact tg 52

<210> 520

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 520

acactctttc cctacacgac gctcttccga tctnnnntta acatccagtc tcccaaacac 60

a 61

<210> 521

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 521

tggagttcag acgtgtgctc ttccgatcta cacacacaca ctactaggac a 51

<210> 522

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 522

acactctttc cctacacgac gctcttccga tctnnnncgc tactacggag cgagaag 57

<210> 523

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 523

tggagttcag acgtgtgctc ttccgatcta caggggcaaa gtgagtgac 49

<210> 524

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 524

aaaaaaagca ccgactcggt gccac 25

<210> 525

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 525

taatacgact cactataggg agtccgagca gaagaagaag ttttagagct agaaatagca 60

<210> 526

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 526

taatacgact cactataggg gaatcccttc tgcagcaccg ttttagagct agaaatagca 60

<210> 527

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 527

taatacgact cactataggg gcccagactg agcacgtgag ttttagagct agaaatagca 60

<210> 528

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 528

taatacgact cactatagga ccccctccac cccgcctcgt tttagagcta gaaatagca 59

<210> 529

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 529

taatacgact cactataggt ctctctctct ctctctctcg ttttagagct agaaatagca 60

<210> 530

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 530

ggtgtttcgt cctttccaca ag 22

<210> 531

<211> 55

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 531

gacccctcc accccgcctc gttttagagc tagaaatagc aagttaaaat aaggc 55

<210> 532

<211> 79

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 532

acgtaaacgg ccacaagttc gcggcgatcg tctctctctc tctctctctc tggatcgcct 60

ggcatcttct tcaaggacg 79

<210> 533

<211> 79

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 533

cgtccttgaa gaagatgcca ggcgatccag agagagagag agagagagac gatcgccgcg 60

aacttgtggc cgtttacgt 79

<210> 534

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 534

acactctttc cctacacgac gctcttccga tctnnnncag ctcagcctga gtgttga 57

<210> 535

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 535

tggagttcag acgtgtgctc ttccgatctc tcgtgggttt gtggttgc 48

<210> 536

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 536

acactctttc cctacacgac gctcttccga tctnnnncat tgcagagagg cgtatca 57

<210> 537

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 537

tggagttcag acgtgtgctc ttccgatctg gggtcccagg tgctgac 47

<210> 538

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 538

acactctttc cctacacgac gctcttccga tctnnnnatg tgggctgcct agaaagg 57

<210> 539

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 539

tggagttcag acgtgtgctc ttccgatctc ccagccaaac ttgtcaacc 49

<210> 540

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 540

acactctttc cctacacgac gctcttccga tctnnnnagt agcctctttc tcaatgtgc 59

<210> 541

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 541

tggagttcag acgtgtgctc ttccgatctg ctttcacaag gatgcagtct 50

<210> 542

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 542

acactctttc cctacacgac gctcttccga tctnnnngag ctagactccg agggga 56

<210> 543

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 543

tggagttcag acgtgtgctc ttccgatctt cctcgtcctg ctctcactt 49

<210> 544

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 544

acactctttc cctacacgac gctcttccga tctnnnnaga ggctgaagag gaagacca 58

<210> 545

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 545

tggagttcag acgtgtgctc ttccgatctg gcccagctgt gcattctat 49

<210> 546

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 546

acactctttc cctacacgac gctcttccga tctnnnnaac ccactgaaga agcaggg 57

<210> 547

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 547

tggagttcag acgtgtgctc ttccgatctg gtgcttaatc cggctccat 49

<210> 548

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 548

acactctttc cctacacgac gctcttccga tctnnnntcc agtgtttcca tcccgaa 57

<210> 549

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 549

tggagttcag acgtgtgctc ttccgatctc ctctgacctc cacaactct 49

<210> 550

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 550

acactctttc cctacacgac gctcttccga tctnnnnctg ggtacagttc tgcgtgt 57

<210> 551

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 551

tggagttcag acgtgtgctc ttccgatctt cactctgagc atcgccaag 49

<210> 552

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 552

acactctttc cctacacgac gctcttccga tctnnnntcc cctgttgacc tggagaa 57

<210> 553

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 553

tggagttcag acgtgtgctc ttccgatctc actgtacttg ccctgacca 49

<210> 554

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 554

acactctttc cctacacgac gctcttccga tctnnnnttg gtgttgacag ggagcaa 57

<210> 555

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 555

tggagttcag acgtgtgctc ttccgatctc tgagatgtgg gcagaaggg 49

<210> 556

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 556

acactctttc cctacacgac gctcttccga tctnnnntga gagggaacag aagggct 57

<210> 557

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 557

tggagttcag acgtgtgctc ttccgatctg tccaaaggcc caagaacct 49

<210> 558

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 558

acactctttc cctacacgac gctcttccga tctnnnngtt cccccgagtc tgcacct 57

<210> 559

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 559

tggagttcag acgtgtgctc ttccgatctc gaacttgcat tcgccgcaa 49

<210> 560

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221

> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 560

acactctttc cctacacgac gctcttccga tctnnnnttc tgccttggca tcgggtg 57

<210> 561

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 561

tggagttcag acgtgtgctc ttccgatctc accataccgc ccctagaact aacattc 57

<210> 562

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (34)..(37)

<223> n is a, c, g, or t

<400> 562

acactctttc cctacacgac gctcttccga tctnnnnaca actgctttcc atggtgtgt 59

<210> 563

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 563

tggagttcag acgtgtgctc ttccgatctt cccagggctt tcgtggaga 49

<210> 564

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 564

gtcaggtcga gggttctgtc agg 23

<210> 565

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 565

cttccaggat gagaacacag agg 23

<210> 566

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 566

caaccactgc tcaaagatgc tgg 23

<210> 567

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 567

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> 568
<211> 1372
<212> PRT
<213> Streptococcus gallolyticus

<400> 568
Met Thr Lys Lys Asn Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser
1 5 10 15
Val Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys
20 25 30
Met Lys Val Leu Gly Asn Thr Asp Lys Lys Tyr Ile Lys Lys Asn Leu
35 40 45
Leu Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg
50 55 60
Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu
65 70 75 80
Arg Tyr Leu Gln Glu Ile Phe Ala Asn Glu Ile Ala Lys Val Asp Glu
85 90 95
Ser Phe Phe Gln Arg Leu Asp Glu Ser Phe Leu Thr Asp Asp Asp Lys
100 105 110
Thr Phe Asp Ser His Pro Ile Phe Gly Asn Lys Ala Glu Glu Asp Ala
115 120 125
Tyr His Gln Lys Phe Pro Thr Ile Tyr His Leu Arg Lys His Leu Ala

130 135 140

Asp Ser Ser Glu Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala

145 150 155 160

His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Glu Leu Asn

165 170 175

Ala Glu Asn Thr Asp Val Gln Lys Ile Phe Ala Asp Phe Val Gly Val

180 185 190

Tyr Asn Arg Thr Phe Asp Asp Ser His Leu Ser Glu Ile Thr Val Asp

195 200 205

Val Ala Ser Ile Leu Thr Glu Lys Ile Ser Lys Ser Arg Arg Leu Glu

210 215 220

Asn Leu Ile Lys Tyr Tyr Pro Thr Glu Lys Lys Asn Thr Leu Phe Gly

225 230 235 240

Asn Leu Ile Ala Leu Ala Leu Gly Leu Gln Pro Asn Phe Lys Thr Asn

245 250 255

Phe Lys Leu Ser Glu Asp Ala Lys Leu Gln Phe Ser Lys Asp Thr Tyr

260 265 270

Glu Glu Asp Leu Glu Glu Leu Leu Gly Lys Ile Gly Asp Asp Tyr Ala

275 280 285

Asp Leu Phe Thr Ser Ala Lys Asn Leu Tyr Asp Ala Ile Leu Leu Ser

290 295 300

Gly Ile Leu Thr Val Asp Asp Asn Ser Thr Lys Ala Pro Leu Ser Ala

305 310 315 320

Ser Met Ile Lys Arg Tyr Val Glu His His Glu Asp Leu Glu Lys Leu

325 330 335

Lys Glu Phe Ile Lys Ala Asn Lys Ser Glu Leu Tyr His Asp Ile Phe

340 345 350

Lys Asp Lys Asn Lys Asn Gly Tyr Ala Gly Tyr Ile Glu Asn Gly Val

355 360 365

Lys Gln Asp Glu Phe Tyr Lys Tyr Leu Lys Asn Ile Leu Ser Lys Ile

370 375 380

Lys Ile Asp Gly Ser Asp Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp
385 390 395 400
Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln

405 410 415
Ile His Leu Gln Glu Met His Ala Ile Leu Arg Arg Gln Gly Asp Tyr
420 425 430
Tyr Pro Phe Leu Lys Glu Lys Gln Asp Arg Ile Glu Lys Ile Leu Thr
435 440 445
Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Val Arg Lys Asp Ser Arg
450 455 460
Phe Ala Trp Ala Glu Tyr Arg Ser Asp Glu Lys Ile Thr Pro Trp Asn

465 470 475 480
Phe Asp Lys Val Ile Asp Lys Glu Lys Ser Ala Glu Lys Phe Ile Thr
485 490 495
Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro
500 505 510
Lys His Ser His Val Tyr Glu Thr Tyr Ala Val Tyr Asn Glu Leu Thr
515 520 525
Lys Ile Lys Tyr Val Asn Glu Gln Gly Lys Glu Ser Phe Phe Asp Ser

530 535 540
Asn Met Lys Gln Glu Ile Phe Asp His Val Phe Lys Glu Asn Arg Lys
545 550 555 560
Val Thr Lys Glu Lys Leu Leu Asn Tyr Leu Asn Lys Glu Phe Pro Glu
565 570 575
Tyr Arg Ile Lys Asp Leu Ile Gly Leu Asp Lys Glu Asn Lys Ser Phe
580 585 590
Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Lys Lys Ile Leu Asp Lys

595 600 605
Ala Phe Leu Asp Asp Lys Val Asn Glu Glu Val Ile Glu Asp Ile Ile
610 615 620
Lys Thr Leu Thr Leu Phe Glu Asp Lys Asp Met Ile His Glu Arg Leu

625 630 635 640

Gln Lys Tyr Ser Asp Ile Phe Thr Ala Asn Gln Leu Lys Lys Leu Glu

645 650 655

Arg Arg His Tyr Thr Gly Trp Gly Arg Leu Ser Tyr Lys Leu Ile Asn

660 665 670

Gly Ile Arg Asn Lys Glu Asn Asn Lys Thr Ile Leu Asp Tyr Leu Ile

675 680 685

Asp Asp Gly Ser Ala Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp

690 695 700

Thr Leu Pro Phe Lys Gln Ile Ile Gln Lys Ser Gln Val Val Gly Asp

705 710 715 720

Val Asp Asp Ile Glu Ala Val Val His Asp Leu Pro Gly Ser Pro Ala

725 730 735

Ile Lys Lys Gly Ile Leu Gln Ser Val Lys Ile Val Asp Glu Leu Val

740 745 750

Lys Val Met Gly Gly Asn Pro Asp Asn Ile Val Ile Glu Met Ala Arg

755 760 765

Glu Asn Gln Thr Thr Asn Arg Gly Arg Ser Gln Ser Gln Gln Arg Leu

770 775 780

Lys Lys Leu Gln Asn Ser Leu Lys Glu Leu Gly Ser Asn Ile Leu Asn

785 790 795 800

Glu Glu Lys Pro Ser Tyr Ile Glu Asp Lys Val Glu Asn Ser His Leu

805 810 815

Gln Asn Asp Gln Leu Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met

820 825 830

Tyr Thr Gly Asp Glu Leu Asp Ile Asp His Leu Ser Asp Tyr Asp Ile

835 840 845

Asp His Ile Ile Pro Gln Ala Phe Ile Lys Asp Asp Ser Ile Asp Asn

850 855 860

Arg Val Leu Thr Ser Ser Ala Lys Asn Arg Gly Lys Ser Asp Asp Val

865 870 875 880

Pro Ser Leu Asp Ile Val Arg Ala Arg Lys Ala Glu Trp Val Arg Leu
885 890 895
Tyr Lys Ser Gly Leu Ile Ser Lys Arg Lys Phe Asp Asn Leu Thr Lys
900 905 910
Ala Glu Arg Gly Gly Leu Thr Glu Ala Asp Lys Ala Gly Phe Ile Lys

915 920 925
Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile
930 935 940
Leu Asp Ala Arg Phe Asn Thr Glu His Asp Glu Asn Asp Lys Val Ile
945 950 955 960
Arg Asp Val Lys Val Ile Thr Leu Lys Ser Asn Leu Val Ser Gln Phe
965 970 975
Arg Lys Asp Phe Glu Phe Tyr Lys Val Arg Glu Ile Asn Asp Tyr His

980 985 990
His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Leu
995 1000 1005
Lys Lys Tyr Pro Lys Leu Ala Ser Glu Phe Val Tyr Gly Glu Tyr
1010 1015 1020
Lys Lys Tyr Asp Ile Arg Lys Phe Ile Thr Asn Ser Ser Asp Lys
1025 1030 1035
Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Leu Met Asn Phe Phe

1040 1045 1050
Lys Thr Lys Val Lys Tyr Ala Asp Gly Thr Val Phe Glu Arg Pro
1055 1060 1065
Ile Ile Glu Thr Asn Ala Asp Gly Glu Ile Ala Trp Asn Lys Gln
1070 1075 1080
Ile Asp Phe Glu Lys Val Arg Lys Val Leu Ser Tyr Pro Gln Val
1085 1090 1095
Asn Ile Val Lys Lys Val Glu Thr Gln Thr Gly Gly Phe Ser Lys

1100 1105 1110
Glu Ser Ile Leu Pro Lys Gly Asp Ser Asp Lys Leu Ile Pro Arg

1115 1120 1125

Lys Thr Lys Lys Val Tyr Trp Asp Thr Lys Lys Tyr Gly Gly Phe

1130 1135 1140

Asp Ser Pro Thr Val Ala Tyr Ser Val Phe Val Val Ala Asp Val

1145 1150 1155

Glu Lys Gly Lys Ala Lys Lys Leu Lys Thr Val Lys Glu Leu Val

1160 1165 1170

Gly Ile Ser Ile Met Glu Arg Ser Phe Phe Glu Glu Asn Pro Val

1175 1180 1185

Glu Phe Leu Glu Asn Lys Gly Tyr His Asn Ile Arg Glu Asp Lys

1190 1195 1200

Leu Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Phe Glu Gly Gly

1205 1210 1215

Arg Arg Arg Leu Leu Ala Ser Ala Ser Glu Leu Gln Lys Gly Asn

1220 1225 1230

Glu Met Val Leu Pro Gly Tyr Leu Val Glu Leu Leu Tyr His Ala

1235 1240 1245

His Arg Ala Asp Asn Phe Asn Ser Thr Glu Tyr Leu Asn Tyr Val

1250 1255 1260

Ser Glu His Lys Lys Glu Phe Glu Lys Val Leu Ser Cys Val Glu

1265 1270 1275

Asp Phe Ala Asn Leu Tyr Val Asp Val Glu Lys Asn Leu Ser Lys

1280 1285 1290

Ile Arg Ala Val Ala Asp Ser Met Asp Asn Phe Ser Ile Glu Glu

1295 1300 1305

Ile Ser Asn Ser Phe Ile Asn Leu Leu Thr Leu Thr Ala Leu Gly

1310 1315 1320

Ala Pro Ala Asp Phe Asn Phe Leu Gly Glu Lys Ile Pro Arg Lys

1325 1330 1335

Arg Tyr Thr Ser Thr Lys Glu Cys Leu Asn Ala Thr Leu Ile His

1340 1345 1350

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys

1355 1360 1365

Leu Gly Glu Glu

1370

<210> 569

<211> 1370

<212> PRT

<213> Streptococcus gordonii

<400> 569

Met Lys Lys Gly Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Phe Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Met

20 25 30

Lys Val Leu Gly Asn Thr Asp Lys Arg Phe Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Glu Gly Thr Thr Ala Glu Ala Arg Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu Arg

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Glu Glu Met Ser Lys Val Asp Ser Ser

85 90 95

Phe Phe His Arg Leu Asp Asp Ser Phe Leu Ile Pro Glu Asp Lys Arg

100 105 110

Glu Ser Lys Tyr Pro Ile Phe Ala Thr Leu Thr Glu Glu Lys Glu Tyr

115 120 125

His Lys Gln Phe Pro Thr Ile Tyr His Leu Arg Lys Gln Leu Ala Asp

130 135 140

Ser Lys Glu Lys Thr Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Tyr Arg Gly His Phe Leu Tyr Glu Glu Ala Phe Asp Ile

165 170 175

Lys Asn Asn Asp Ile Gln Lys Ile Phe Asn Glu Phe Ile Ser Ile Tyr

180 185 190

Asp Asn Thr Phe Glu Gly Ser Ser Leu Ser Gly Gln Asn Ala Gln Val

195 200 205

Glu Ala Ile Phe Thr Asp Lys Ile Ser Lys Ser Ala Lys Arg Glu Arg

210 215 220

Val Leu Lys Leu Phe Pro Asp Glu Lys Ser Thr Gly Leu Phe Ser Glu

225 230 235 240

Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Lys Lys His Phe

245 250 255

Asp Leu Glu Asp Lys Ala Pro Leu Gln Phe Ser Lys Asp Thr Tyr Asp

260 265 270

Glu Asp Leu Glu Asn Leu Leu Gly Gln Ile Gly Asp Asp Phe Thr Asp

275 280 285

Leu Phe Val Ser Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly

290 295 300

Ile Leu Thr Val Thr Asp Pro Ser Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Glu Arg Tyr Glu Asn His Gln Asn Asp Leu Ala Ala Leu Lys

325 330 335

Gln Phe Ile Lys Asn Asn Leu Pro Glu Lys Tyr Asp Glu Val Phe Ser

340 345 350

Asp Gln Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Thr

355 360 365

Gln Glu Thr Phe Tyr Lys Tyr Ile Lys Asn Leu Leu Ser Lys Phe Glu

370 375 380

Gly Thr Asp Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe 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

Gln Glu Met Asn Ala Ile Leu Arg Arg Gln Gly Glu Tyr Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Lys 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 Arg Asp Phe Ala Trp
450 455 460
Leu Thr Arg Asn Ser Asp Glu Ala Ile Arg Pro Trp Asn Phe Glu Glu

465 470 475 480
Ile Val Asp Lys Ala Ser Ser Ala Glu Asp Phe Ile Asn Lys Met Thr
485 490 495
Asn Tyr Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Thr Phe Ala Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Phe Ile Ala Glu Gly Leu Arg Asp Tyr Gln Phe Leu Asp Ser Gly Gln

530 535 540
Lys Lys Gln Ile Val Asn Gln Leu Phe Lys Glu Asn Arg Lys Val Thr
545 550 555 560
Glu Lys Asp Ile Ile His Tyr Leu His Asn Val Asp Gly Tyr Asp Gly
565 570 575
Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ala Ser Leu Ser Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Glu Phe Met Asp Asp

595 600 605
Ala Lys Asn Glu Ala Ile Leu Glu Asn Ile Val His Thr Leu Thr Ile
610 615 620
Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ala Gln Tyr Asp Ser
625 630 635 640
Leu Phe Asp Glu Lys Val Ile Lys Ala Leu Thr Arg Arg His Tyr Thr
645 650 655
Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Cys Asp Lys

660 665 670
Gln Thr Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Lys Ile

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Gly Leu Ser Phe Lys

690 695 700

Glu Ile Ile Gln Lys Ala Gln Val Ile Gly Lys Thr Asp Asp Val Lys

705 710 715 720

Gln Val Val Gln Glu Leu Ser Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met Gly His

740 745 750

Ala Pro Glu Ser Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr

755 760 765

Ala Arg Gly Lys Lys Asn Ser Gln Gln Arg Tyr Lys Arg Ile Glu Asp

770 775 780

Ser Leu Lys Ile Leu Ala Ser Gly Leu Asp Ser Asn Ile Leu Lys Glu

785 790 795 800

Asn Pro Thr Asp Asn Asn Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr

805 810 815

Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Glu Ala Leu Asp Ile

820 825 830

Asn Gln Leu Ser Ser Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe

835 840 845

Ile Lys Asp Asp Ser Leu Asp Asn Arg Val Leu Thr Ser Ser Lys Asp

850 855 860

Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Ile Glu Val Val Gln Lys

865 870 875 880

Arg Lys Ala Phe Trp Gln Gln Leu Leu Asp Ser Lys Leu Ile Ser Glu

885 890 895

Arg Lys Phe Asn Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Asp Glu

900 905 910

Arg Asp Lys Val Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln

915 920 925

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ala Arg Tyr Asn Thr Glu

930 935 940

Val Asn Glu Lys Asp Lys Lys Asn Arg Thr Val Lys Ile Ile Thr Leu

945 950 955 960

Lys Ser Asn Leu Val Ser Asn Phe Arg Lys Glu Phe Arg Leu Tyr Lys

965 970 975

Val Arg Glu Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn

980 985 990

Ala Val Val Ala Lys Ala Ile Leu Lys Lys Tyr Pro Lys Leu Glu Pro

995 1000 1005

Glu Phe Val Tyr Gly Glu Tyr Gln Lys Tyr Asp Leu Lys Arg Tyr

1010 1015 1020

Ile Ser Arg Ser Lys Asp Pro Lys Glu Val Glu Lys Ala Thr Glu

1025 1030 1035

Lys Tyr Phe Phe Tyr Ser Asn Leu Leu Asn Phe Phe Lys Glu Glu

1040 1045 1050

Val His Tyr Ala Asp Gly Thr Ile Val Lys Arg Glu Asn Ile Glu

1055 1060 1065

Tyr Ser Lys Asp Thr Gly Glu Ile Ala Trp Asn Lys Glu Lys Asp

1070 1075 1080

Phe Ala Ile Ile Lys Lys Val Leu Ser Leu Pro Gln Val Asn Ile

1085 1090 1095

Val Lys Lys Arg Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1100 1105 1110

Ile Leu Pro Lys Gly Asn Ser Asp Lys Leu Ile Pro Arg Lys Thr

1115 1120 1125

Lys Asp Ile Leu Leu Asp Thr Thr Lys Tyr Gly Gly Phe Asp Ser

1130 1135 1140

Pro Val Ile Ala Tyr Ser Ile Leu Leu Ile Ala Asp Ile Glu Lys

1145 1150 1155

Gly Lys Ala Lys Lys Leu Lys Thr Val Lys Thr Leu Val Gly Ile

1160 1165 1170

Thr Ile Met Glu Lys Ala Ala Phe Glu Glu Asn Pro Ile Thr Phe

1175 1180 1185

Leu Glu Asn Lys Gly Tyr His Asn Val Arg Lys Glu Asn Ile Leu

1190 1195 1200

Cys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Arg

1205 1210 1215

Arg Leu Leu Ala Ser Ala Lys Glu Leu Gln Lys Gly Asn Glu Ile

1220 1225 1230

Val Leu Pro Val Tyr Leu Thr Thr Leu Leu Tyr His Ser Lys Asn

1235 1240 1245

Val His Lys Leu Asp Glu Pro Gly His Leu Glu Tyr Ile Gln Lys

1250 1255 1260

His Arg Asn Glu Phe Lys Asp Leu Leu Asn Leu Val Ser Glu Phe

1265 1270 1275

Ser Gln Lys Tyr Val Leu Ala Asp Ala Asn Leu Glu Lys Ile Lys

1280 1285 1290

Ser Leu Tyr Ala Asp Asn Glu Gln Ala Asp Ile Glu Ile Leu Ala

1295 1300 1305

Asn Ser Phe Ile Asn Leu Leu Thr Phe Thr Ala Leu Gly Ala Pro

1310 1315 1320

Ala Ala Phe Lys Phe Phe Gly Lys Asp Ile Asp Arg Lys Arg Tyr

1325 1330 1335

Thr Thr Val Ser Glu Ile Leu Asn Ala Thr Leu Ile His Gln Ser

1340 1345 1350

Ile Thr Gly Leu Tyr Glu Thr Trp Ile Asp Leu Ser Lys Leu Gly

1355 1360 1365

Glu Asp

1370

<210> 570

<211> 1056

<212> PRT

<213> Staphylococcus aureus

<400> 570

Gly Ser His Met Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile
1 5 10 15

Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile
20 25 30

Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu
35 40 45

Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg
50 55 60

His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu
65 70 75 80

Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val
85 90 95

Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu
100 105 110

Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu
115 120 125

Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn
130 135 140

Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg
145 150 155 160

Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr
165 170 175

Ser Asp Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala
180 185 190

Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu
195 200 205

Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe
210 215 220

Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys

225 230 235 240

Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala

245 250 255

Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg

260 265 270

Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu

275 280 285

Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys

290 295 300

Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser

305 310 315 320

Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys

325 330 335

Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp

340 345 350

Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln

355 360 365

Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu

370 375 380

Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu

385 390 395 400

Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn

405 410 415

Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp

420 425 430

Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile

435 440 445

Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile

450 455 460

Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu

465 470 475 480

Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu
485 490 495
Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile
500 505 510
Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys
515 520 525
Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile
530 535 540

Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His
545 550 555 560
Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val
565 570 575
Leu Val Lys Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe
580 585 590
Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys
595 600 605

Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr
610 615 620
Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val
625 630 635 640
Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr
645 650 655
Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu
660 665 670

Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg
675 680 685
Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His
690 695 700
Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu
705 710 715 720
Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe

725 730 735

Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu
740 745 750

Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp
755 760 765

Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
770 775 780

Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly
785 790 795 800

Asn Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn
805 810 815

Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met
820 825 830

Tyr His His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu
835 840 845

Gln Tyr Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr
850 855 860

Gly Asn Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile
865 870 875 880

Lys Lys Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile
885 890 895

Thr Asp Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu
900 905 910

Lys Pro Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe
915 920 925

Val Thr Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu
930 935 940

Val Asn Ser Lys Ala Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser
945 950 955 960

Asn Gln Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys
965 970 975

Ile Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu
980 985 990

Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu
995 1000 1005
Glu Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile
1010 1015 1020
Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu
1025 1030 1035
Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile
1040 1045 1050

Lys Lys Gly
1055

<210> 571
<211> 22
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 571
aatctagtaa ataaattgat gt 22

<210> 572
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 572
gaccaacggc taagtgaaga 20

<210> 573
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide

<400> 573

tcctttcttc acggttgcct 20

<210> 574

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 574

ctcctgctca gtgatttcag 20

<210> 575

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 575

tgtcacacat gtagttgtag 20

<210> 576

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 576

cctcccggcc ggcggtatcc 20

<210> 577

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 577

ggcagccgat acccggggca ggg 23

<210> 578

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 578

gggcagccga tacccggggc agg 23

<210> 579

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 579

ttattgtccg aaataaaaga aga 23

<210> 580

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 580

attgtccgaa ataaaagaag agg 23

<210> 581

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 581

ttgtccgaaa taaaagaaga gga 23

<210> 582

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 582

gtccgaaata aaagaagagg aaaaaat 27

<210> 583

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 583

gtccgaaata aaagaagagg aaaaaatt 28

<210> 584

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 584

caggaagcac gaggccactg agg 23

<210> 585

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 585

aacaggaagc acgaggccac tga 23

<210> 586

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 586

aggaagcacg aggccactga gga 23

<210> 587

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 587

ttggattctg gaatccattc tgg 23

<210> 588

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 588

tctgcatcca tcttcacttc aga 23

<210> 589

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 589

cttacagcca gtaatgtcac cga 23

<210> 590

<211> 177

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 590

Ala Ser Arg Gly Val Asn Lys Val Ile Leu Val Gly Asn Leu Gly Gln

1 5 10 15

Asp Pro Glu Val Arg Tyr Met Pro Asn Gly Gly Ala Val Ala Asn Ile

20 25 30

Thr Leu Ala Thr Ser Glu Ser Trp Arg Asp Lys Ala Thr Gly Glu Met

35 40 45

Lys Glu Gln Thr Glu Trp His Arg Val Val Leu Phe Gly Lys Leu Ala

50 55 60

Glu Val Ala Ser Glu Tyr Leu Arg Lys Gly Ser Gln Val Tyr Ile Glu

65 70 75 80

Gly Gln Leu Arg Thr Arg Lys Trp Thr Asp Gln Ser Gly Gln Asp Arg

85 90 95

Tyr Thr Thr Glu Val Val Val Asn Val Gly Gly Thr Met Gln Met Leu

100 105 110

Gly Gly Arg Gln Gly Gly Gly Ala Pro Ala Gly Gly Asn Ile Gly Gly

115 120 125

Gly Gln Pro Gln Gly Gly Trp Gly Gln Pro Gln Gln Pro Gln Gly Gly

130 135 140

Asn Gln Phe Ser Gly Gly Ala Gln Ser Arg Pro Gln Gln Ser Ala Pro

145 150 155 160

Ala Ala Pro Ser Asn Glu Pro Pro Met Asp Phe Asp Asp Asp Ile Pro

165 170 175

Phe

<210> 591

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 591

ttatttcgtg gatttattta 20

<210> 592

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 592

tgcttctcca gccctggcct gg 22

<210> 593

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 593

Lys Arg Arg Thr Thr Ile Ser Ile Ala Ala

1 5 10

<210> 594

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 594

Ala Leu Val Phe Tyr Arg Glu Tyr Ile Gly Arg Leu Lys Gln Ile Lys

1 5 10 15

Phe Lys Phe

<210> 595

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 595

Leu Pro Ile Met Ala Lys Ser Val Leu Asp Leu Tyr Glu Leu

1 5 10

<210> 596

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 596

Leu Leu Arg Leu Gly

1 5

<210> 597

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 597

Thr Asp Lys Glu Ile Asn Pro Val Val Lys Glu Asn Ile Glu Trp

1 5 10 15

<210> 598

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 598

Pro Pro Ser Ile Lys Arg Lys Phe

1 5

<210> 599

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 599

Leu Pro Thr Val Gln Phe Asp Phe Thr

1 5

<210> 600

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 600

Leu Pro Leu Asp Ile Arg Gly Thr Ala Phe Gln Gln Gln Val

1 5 10

<210> 601

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 601

Ala Tyr Val Val Leu Gly Gln Phe

1 5

<210> 602

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 602

Ile Asn Phe Tyr Gly Glu Leu Val

1 5

<210> 603

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 603

Lys Lys Thr Gly Asp Gln Lys

1 5

<210> 604

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 604

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1 5 10 15

<210> 605

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 605

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> 606

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (5)..(120)

<223> May be absent

<400> 606

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> 607

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (6)..(150)

<223> May be absent

<400> 607

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> 608

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (2)..(30)

<223> May be absent

<400> 608

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 Gly
20 25 30

<210> 609

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (6)..(150)

<223> May be absent

<400> 609

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> 610

<211> 90

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (4)..(90)

<223> May be absent

<400> 610

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> 611

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> misc_feature

<222> (1)..(1)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (3)..(60)

<223> May be absent

<400> 611

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> 612

<211> 94

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (8)..(94)

<223> May be absent

<400> 612

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

1 5 10 15

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

20 25 30

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

35 40 45

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

50 55 60

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

65 70 75 80

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

85 90

<210> 613

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (5)..(120)

<223> May be absent

<400> 613

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
20 25 30

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
35 40 45

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
50 55 60

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
65 70 75 80

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
85 90 95

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
100 105 110

Gly Gly Gly Ser Gly Gly Gly Ser
115 120

<210> 614

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 614

Pro Lys Lys Lys Arg Lys Val
1 5

<210> 615

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 615

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> 616

<211> 36

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<220><221> misc_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (4)..(29)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (32)..(35)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (27)..(29)

<223> May be absent

<220><221> MISC_FEATURE

<222> (34)..(35)

<223> May be absent

<400> 616

His Xaa Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Cys Xaa

20 25 30

Xaa Xaa Xaa Cys

35

<210> 617

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 617

Ser Pro Lys Lys Lys Arg Lys Val Glu Ala Ser

1 5 10

<210> 618

<211> 81

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 618

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu u 81

<210> 619

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 619

tcgcacccgt ggatttattt atgg 24

<210> 620

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 620

ttcgcacccg tggatttatt atgg 24

<210> 621

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 621

cctcgcacgt ggatttattt atgg 24

<210> 622

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 622

tcctcgcacg tggatttatt atgg 24

<210> 623

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 623

accctcgcgt ggatttattt atgg 24

<210> 624

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 624

taccctcgcg tggatttatt atgg 24

<210> 625

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 625

gcaccctcgt ggatttattt atgg 24

<210> 626

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 626

tgcaccctcg tggatttatt atgg 24

<210> 627

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 627

gaagcgcctg gcagtgtacc agg 23

<210> 628

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 628

ccactacaac tacatgtgtg aca 23

<210> 629

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 629

ccgcgccatg gccatctgca agc 23

<210> 630

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 630

ggaggacgtg cgcggccgcc tgg 23

<210> 631

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 631

gaagcgcctg gcagtgtacc agg 23

<210> 632

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 632

ctgtggcagt ggcaccagaa tgg 23

<210> 633

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 633

cctcccggcc ggcggtatcc agg 23

<210> 634

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 634

gcttgcagat ggccatggcg cgg 23

<210> 635

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 635

acacatgcag ttgtagtgga tgg 23

<210> 636

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 636

tgtcacacat gtagttgtag tgg 23

<210> 637

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 637

gagttagagc agaagaagaa agg 23

<210> 638

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 638

gagtctaagc agaagaagaa gag 23

<210> 639

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 639

gaggccgagc agaagaaaga cgg 23

<210> 640

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 640

gagtcctagc aggagaagaa gag 23

<210> 641

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 641

gagtccggga aggagaagaa agg 23

<210> 642

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 642

gagccggagc agaagaagga ggg 23

<210> 643

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 643

aagtccgagg agaggaagaa agg 23

<210> 644

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 644

gaatccaagc aggagaagaa gga 23

<210> 645

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 645

acgtctgagc agaagaagaa tgg 23

<210> 646

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 646

ggaaccccgt ctgcagcacc agg 23

<210> 647

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 647

ggagtccctc ctacagcacc agg 23

<210> 648

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 648

agaggcccct ctgcagcacc agg 23

<210> 649

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 649

accatccctc ctgcagcacc agg 23

<210> 650

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 650

ggattgccat ccgcagcacc tgg 23

<210> 651

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 651

tgaatcccat ctccagcacc agg 23

<210> 652

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 652

ggagtccctc ctacagcacc agg 23

<210> 653

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 653

ggagtccctc ctgcagcacc tga 23

<210> 654

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 654

gaacacaatg catagattgc cgg 23

<210> 655

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 655

aaacataaag catagactgc aaa 23

<210> 656

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 656

cacccagact gagcacgtgc tgg 23

<210> 657

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 657

gacacagact gggcacgtga ggg 23

<210> 658

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 658

agctcagact gagcaagtga ggg 23

<210> 659

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 659

agaccagact gagcaagaga ggg 23

<210> 660

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 660

gagccagaat gagcacgtga ggg 23

<210> 661

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 661

tgcactgcgg ccggaggagg tgg 23

<210> 662

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 662

ggctctgcgg ctggaggggg tgg 23

<210> 663

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 663

ggcacgacgg ctggaggtgg ggg 23

<210> 664

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 664

ggcatcacgg ctggaggtgg agg 23

<210> 665

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 665

ggcgctgcgg cgggaggtgg agg 23

<210> 666

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 666

ggcactgaga ctgggggtgg ggg 23

<210> 667

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 667

agcagtgcgg ctagaggtgg tgg 23

<210> 668

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 668

ggcactgcta ctgggggtgg tgg 23

<210> 669

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 669

ggcactgggg ctgggggagg ggg 23

<210> 670

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 670

ggcactgggg ttggaggtgg ggg 23

<210> 671

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 671

Lys Arg Leu Ala Val Tyr Gln

1 5

<210> 672

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 672

Arg Ala Met Ala Ile Cys Lys

1 5

<210> 673

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 673

tcagggtgag catagactgc cgg 23

<210> 674

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 674

gaatcgggag gggagactgc ggg 23

<210> 675

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 675

tgaagtgttg catagactgc agg 23

<210> 676

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 676

ggagagagag catagactgc tgg 23

<210> 677

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 677

ccaaacaaaa catagactgc tgg 23

<210> 678

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 678

caggaagctg gagcacgtga ggg 23

<210> 679

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 679

aaggctgagg gagcacgtga agg 23

<210> 680

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 680

gtcaggggaa gagcacgtga cgg 23

<210> 681

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 681

gttgtgaact gagcacgtga ggg 23

<210> 682

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 682

atatttgctg gagcacgtga agg 23

<210> 683

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 683

gtggctggag gtggaggtgg ggg 23

<210> 684

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 684

gagggaaggg ctggaggtgg agg 23

<210> 685

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 685

cacctgcgca agctgcgtaa gcggc 25

<210> 686

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 686

tcctccgcga tgccgatgac ctgca 25

<210> 687

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 687

aggccggggc ccgcgagggc gccga 25

<210> 688

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 688

gcgcggcctc agcgccatcc gcgag 25

<210> 689

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 689

cctgtgcagc tgtgggttga ttcca 25

<210> 690

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 690

caccccgcc cggcacccgc gtccg 25

<210> 691

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 691

agtcacagca catgacggag gttgt 25

<210> 692

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 692

gaggcgctgc ccccaccatg agcgc 25

<210> 693

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 693

tgccatcacg tgctcaatcg gggcccca 28

<210> 694

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 694

cgccccgcag tctatgcttt gtgttccag 29

<210> 695

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 695

cccgcagtct atgctttgtg ttc 23

<210> 696

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 696

ctcccatcac atcaaccggt 20

<210> 697

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 697

tttcctcccc cccacaggta gaacat 26

<210> 698

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 698

tttccctctg tccacaccct catcctg 27

<210> 699

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 699

tttccccagt cctccacacc ctgaaac 27

<210> 700

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 700

tgcccctccc tccctggccc 20

<210> 701

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 701

agagcccccc ctcaaagaga 20

<210> 702

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 702

ggcccagact gagcacgtga tgg 23

<210> 703

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 703

gagtccgagc agaagaagaa ggg 23

<210> 704

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 704

gaacacaaag catagactgc ggg 23

<210> 705

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 705

gagtccgagc agaagaagaa g 21

<210> 706

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 706

ggaatccctt ctgcagcacc tgg 23

<210> 707

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 707

ggcccagact gagcacgtga 20

<210> 708

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 708

ggcactgcgg ctggaggtgg ggg 23

<210> 709

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 709

gtcaggtcga gggttctgtc 20

<210> 710

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 710

gggccgcagt atcctcactc 20

<210> 711

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 711

ccgccagtcc cagtacggga 20

<210> 712

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 712

caaccactgc tcaaagatgc 20

<210> 713

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 713

cttccaggat gagaacacag 20

<210> 714

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 714

Gly Ser Arg Tyr Pro Gly Glu

1 5

<210> 715

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 715

tgcccctccc tccctggccc agg 23

<210> 716

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 716

ttttcctccc ccccacaggt aga 23

<210> 717

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 717

gccccccctc aaagagaggg tgg 23

<210> 718

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 718

agagcccccc ctcaaagaga ggg 23

<210> 719

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 719

ggcccagact gagcacgtga tgg 23

<210> 720

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 720

gaacacaaag catagactgc ggg 23

<210> 721

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 721

gagtccgagc agaagaagaa ggg 23

<210> 722

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 722

ggaatccctt ctgcagcacc tgg 23

<210> 723

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<220><221> misc_feature

<222> (1)..(4)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (25)..(25)

<223> n is a, c, g, or t

<400> 723

nnnntctctc tctctctctc tctcngg 27

<210> 724

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 724

gcggatgttc caatcagtac gca 23

<210> 725

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 725

cgagcttctg gcggtctcaa gca 23

<210> 726

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 726

tgcttctcca gccctggcct gg 22

<210> 727

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 727

agactgagca cgtgatggca gag 23

<210> 728

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 728

gatgttccaa tcagtacgca gagagt 26

<210> 729

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 729

cttctccagc cctggcctgg gtcaat 26

<210> 730

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 730

gccgtctcca aggtgaaagc ggaa 24

<210> 731

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 731

gtcgccgtct ccaaggtgaa agcg 24

<210> 732

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 732

ggtgcatttt caggaggaag cgat 24

<210> 733

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 733

gccgtctcca aggtgaaagc ggaa 24

<210> 734

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 734

gccgtctcca aggtgaaagc ggaagt 26

<210> 735

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 735

ggcactgcgg ctggaggtgg ggg 23

<210> 736

<211> 83

<212> PRT

<213> Bacillus phage PBS2

<400> 736

Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val

1 5 10 15

Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile

20 25 30

Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu

35 40 45

Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr

50 55 60

Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile

65 70 75 80

Lys Met Leu

<210> 737

<211> 228

<212> PRT

<213> Rattus norvegicus

<400> 737

Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg Arg
1 5 10 15

Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg
20 25 30

Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser
35 40 45

Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val Asn
50 55 60

Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg
65 70 75 80

Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser
85 90 95

Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu Phe
100 105 110

Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln
115 120 125

Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr
130 135 140

Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser Pro
145 150 155 160

Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg Leu
165 170 175

Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu
180 185 190

Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile Ala
195 200 205

Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp Ala
210 215 220

Thr Gly Leu Lys
225

<210> 738

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223>

Synthetic Polynucleotide

<400> 738

tttgtacttt gtcctccggt tctg 24

<210> 739

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 739

tttccagccc gctggccctg taaa 24

<210> 740

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 740

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys

1 5 10 15

Val

<210> 741

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 741

gtaatttcta ctaagtgtag at 22

<210> 742

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 742

tactttgtcc tccggttctg gaa 23

<210> 743

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400

> 743

tactttgtcc tccggttctg 20

<210> 744

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 744

tactttgtcc tccggttct 19

<210> 745

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 745

tactttgtcc tccggttc 18

<210> 746

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 746

tactttgtcc tccggtt 17

<210> 747

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 747

cagcccgctg gccctgtaaa gga 23

<210> 748

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 748

cagcccgctg gccctgtaaa 20

<210> 749

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 749

cagcccgctg gccctgtaa 19

<210> 750

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 750

cagcccgctg gccctgta 18

<210> 751

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 751

cagcccgctg gccctgt 17