(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0091858
(43) 공개일자 2020년07월31일

(51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) *C12N 15/79* (2006.01)
C12N 9/22 (2006.01) *C12Q 1/6809* (2018.01)
C12Q 1/6816 (2018.01) *C12Q 1/70* (2006.01)
(52) CPC특허분류
C12N 15/102 (2013.01)
C12N 15/79 (2013.01)
(21) 출원번호 10-2020-7012989
(22) 출원일자(국제) 2018년10월31일
심사청구일자 없음
(85) 번역문제출일자 2020년05월06일
(86) 국제출원번호 PCT/US2018/058545
(87) 국제공개번호 WO 2019/089820
국제공개일자 2019년05월09일
(30) 우선권주장
62/580,395 2017년11월01일 미국(US)

(71) 출원인
더 리젠츠 오브 더 유니버시티 오브 캘리포니아
미국 캘리포니아주 94607-5200 오크랜드 5층 프랭클린 스트리트 1111
(72) 발명자
다우드나 제니퍼 에이.
미국 캘리포니아 94705 버클리 비센테 로드 164
버스테인 데이비드
미국 캘리포니아 94720 버클리 버클리 웨이 2151
(뒷면에 계속)
(74) 대리인
특허법인와이에스장

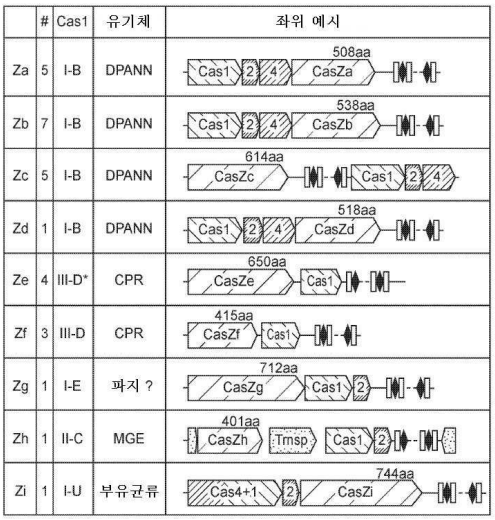
전체 청구항 수 : 총 86 항

(54) 발명의 명칭 **CasZ 조성물 및 사용 방법**

(57) 요 약

하기 중 하나 이상을 포함하는 조성물 및 방법이 제공된다: (1) "CasZ" 단백질(또한 CasZ 폴리펩타이드로서 지칭됨), CasZ 단백질을 암호화하는 핵산, 및/또는 CasZ 단백질(및/또는 이를 암호화하는 핵산)을 포함하는 변형된 숙주 세포; (2) CasZ 단백질에 결합하고 이에 대해 서열 특이성을 제공하는 CasZ 가이드 RNA, CasZ 가이드 RNA를 암호화하는 핵산, 및/또는 CasZ 가이드 RNA(및/또는 이를 암호화하는 핵산)를 포함하는 변형된 숙주 세포; 및 (3) CasZ 전사촉진 비암호화 RNA(trancRNA)(본 명세서에서 "CasZ trancRNA"로서 지칭됨), CasZ trancRNA를 암호화하는 핵산, 및/또는 CasZ trancRNA(및/또는 이를 암호화하는 핵산)를 포함하는 변형된 숙주 세포.

대 표 도 - 도2



(52) CPC특허분류
C12N 9/22 (2013.01)
C12Q 1/6809 (2018.05)
C12Q 1/6816 (2018.05)
C12Q 1/70 (2013.01)
C12N 2310/20 (2017.05)
C12Q 2563/107 (2013.01)
(72) 발명자
첸 재니스 에스.
미국 캘리포니아 94720 버클리 스탠리 홀 731 엠에스 # 3220
해링턴 루카스 비.
미국 캘리포니아 94720 버클리 스탠리 홀 731 엠에스 # 3220
파에즈-에스피노 데이비드
미국 캘리포니아 94598 월넛 크릭 미첼 드라이브 2800
밴필드 질리안 에프.
미국 캘리포니아 94708 버클리 힐 로드 160

명 세 서

청구범위

청구항 1

CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법으로서,

상기 표적 핵산을,

(a) CasZ 폴리펩타이드;

(b) 상기 표적 핵산의 표적 서열에 혼성화되는 가이드 서열을 포함하고, 상기 CasZ 폴리펩타이드에 결합하는 영역을 포함하는 CasZ 가이드 RNA

를 포함하는 조작된 그리고/또는 비천연 유래 복합체와 접촉시키는 단계를 포함하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 2

제1항에 있어서, 상기 방법은 상기 표적 핵산의 변형, 상기 표적 핵산으로부터 전사의 조절, 또는 표적 핵산과 연관된 폴리펩타이드의 변형을 초래하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 3

제2항에 있어서, 상기 표적 핵산은 절단에 의해 변형되는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 이중 가닥 DNA, 단일 가닥 DNA, RNA, 게놈 DNA, 및 염색체외 DNA로부터 선택된, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드에 결합하는 상기 가이드 서열 및 상기 영역은 서로에 대해 이종성인, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 접촉시키는 단계는 게놈 편집을 초래하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 7

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 접촉시키는 단계는 박테리아 세포 외에서 그리고 고세균 세포 외에서 일어나는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 8

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 접촉시키는 단계는 세포 외부의 시험관내에서 일어나는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 9

제1항 내지 제7항 중 어느 한 항에 있어서, 상기 접촉시키는 단계는 표적 세포 내에서 일어나는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 10

제9항에 있어서, 상기 접촉시키는 단계는,

(a) 상기 CasZ 폴리펩타이드, 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산; 및

(b) 상기 CasZ 가이드 RNA, 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산

중 적어도 하나를 상기 표적 세포에 도입하는 것을 포함하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 11

제10항에 있어서, 상기 CasZ 폴리펩타이드를 암호화하는 핵산은 상기 표적 세포에서의 발현을 위해 코돈 최적화된 비천연 서열인, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 12

제9항 내지 제11항 중 어느 한 항에 있어서, 상기 표적 세포는 진핵 세포인, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 13

제9항 내지 제12항 중 어느 한 항에 있어서, 상기 표적 세포는 시험관내 배양물에 있는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 14

제9항 내지 제12항 중 어느 한 항에 있어서, 상기 표적 세포는 생체내에 있는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 15

제9항 내지 제12항 중 어느 한 항에 있어서, 상기 표적 세포는 생체외에 있는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 16

제12항에 있어서, 상기 진핵 세포는 식물 세포, 진균 세포, 단세포 진핵 유기체, 포유류 세포, 파충류 세포, 곤충 세포, 조류 세포, 어류 세포, 기생충 세포, 절지동물 세포, 무척추동물의 세포, 척추동물의 세포, 설치류 세포, 마우스 세포, 래트 세포, 영장류 세포, 비-인간 영장류 세포, 및 인간 세포로 이루어진 군으로부터 선택되는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 17

제9항 내지 제16항 중 어느 한 항에 있어서, 상기 접촉시키는 단계는 DNA 공여자 주형을 상기 표적 세포 내로 도입하는 것을 더 포함하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 18

제1항 내지 제17항 중 어느 한 항에 있어서, 상기 방법은 상기 표적 핵산을 CasZ 전사촉진(transactivating) 비암호화 RNA(trancRNA)와 접촉시키는 단계를 포함하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 19

제9항 내지 제17항 중 어느 한 항에 있어서, 상기 접촉시키는 단계는 CasZ 전사촉진 비암호화 RNA(trancRNA) 및/또는 상기 CasZ trancRNA를 암호화하는 핵산을 표적 세포 내로 도입하는 것을 포함하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 20

제18항 또는 제19항에 있어서, 상기 trancRNA는 표 2의 trancRNA 서열과 70% 이상의 동일성을 갖는 뉴클레오타

이드 서열을 포함하는, CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법.

청구항 21

조작된 그리고/또는 비천연 유래 복합체를 포함하는 조성물로서,

(a) CasZ 폴리펩타이드, 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산; 및

(b) CasZ 가이드 RNA, 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산을 포함하되, 상기 CasZ 가이드 RNA는 표적 핵산의 표적 서열에 상보성인 가이드 서열을 포함하고, 상기 CasZ 폴리펩타이드에 결합할 수 있는 영역을 포함하는, 조성물.

청구항 22

제21항에 있어서, CasZ 전사촉진 비암호화 RNA(trancRNA), 또는 상기 CasZ trancRNA를 암호화하는 핵산을 더 포함하는, 조성물.

청구항 23

조작된 그리고/또는 비천연 유래 복합체를 포함하는 키트로서,

(a) CasZ 폴리펩타이드, 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산;

(b) CasZ 가이드 RNA, 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산을 포함하되, 상기 CasZ 가이드 RNA는 표적 핵산의 표적 서열에 상보성인 가이드 서열을 포함하고, 상기 CasZ 폴리펩타이드에 결합할 수 있는 영역을 포함하는, 키트.

청구항 24

제23항에 있어서, CasZ 전사촉진 비암호화 RNA(trancRNA), 또는 상기 CasZ trancRNA를 암호화하는 핵산을 더 포함하는, 키트.

청구항 25

유전자 변형된 진핵 세포로서,

(a) CasZ 폴리펩타이드, 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산;

(b) CasZ 가이드 RNA, 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산; 및

(c) CasZ 전사촉진 비암호화 RNA(trancRNA), 또는 상기 CasZ trancRNA를 암호화하는 핵산을 포함하되,

상기 CasZ 가이드 RNA는 표적 핵산의 표적 서열에 상보성인 가이드 서열을 포함하고, 상기 CasZ 폴리펩타이드에 결합할 수 있는 영역을 포함하는, 유전자 변형된 진핵 세포.

청구항 26

제1항 내지 제25항 중 어느 하나에 있어서,

(a) 상기 CasZ 폴리펩타이드를 암호화하는 상기 핵산은,

(i) 상기 CasZ 폴리펩타이드를 암호화하고, (ii) 이종성 프로모터에 작동 가능하게 연결된, 뉴클레오타이드 서열을 포함하는 것;

(b) 상기 CasZ 가이드 RNA를 암호화하는 상기 핵산은,

(i) 상기 CasZ 가이드 RNA를 암호화하고, (ii) 이종성 프로모터에 작동 가능하게 연결된, 뉴클레오타이드 서열을 포함하는 것; 그리고

(c) 상기 CasZ trancRNA를 암호화하는 상기 핵산은,

(i) 상기 CasZ trancRNA를 암호화하고, (ii) 이종성 프로모터에 작동 가능하게 연결된, 뉴클레오타이드 서열을 포함하는 것

중 적어도 하나를 특징으로 하는, 조성물, 키트 또는 진핵 세포.

청구항 27

제1항 내지 제26항 중 어느 한 항에 있어서, 환자의 치료 방법에서 사용하기 위한, 조성물, 키트 또는 진핵 세포.

청구항 28

제1항 내지 제27항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드를 암호화하는 핵산, 상기 CasZ 가이드 RNA를 암호화하는 핵산, 및 상기 CasZ trancRNA를 암호화하는 핵산 중 적어도 하나는 재조합 발현 벡터인, 방법, 조성물, 키트 또는 진핵 세포.

청구항 29

제1항 내지 제28항 중 어느 한 항에 있어서, 상기 CasZ 가이드 RNA 및/또는 상기 CasZ trancRNA는 변형된 핵염기, 변형된 골격 또는 비천연 뉴클레오타이드간 결합, 변형된 당 모이어티, 잠금 핵산, 펩타이드 핵산 및 데옥시리보뉴클레오타이드 중 하나 이상을 포함하는, 방법, 조성물, 키트 또는 진핵 세포.

청구항 30

제1항 내지 제29항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드는 대응하는 야생형 CasZ 단백질에 비해 감소된 뉴클레아제 활성을 갖는 변이체 CasZ 폴리펩타이드인, 방법, 조성물, 키트 또는 진핵 세포.

청구항 31

제1항 내지 제30항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드, 상기 CasZ 폴리펩타이드를 암호화하는 핵산, 상기 CasZ 가이드 RNA, 상기 CasZ 가이드 RNA를 암호화하는 상기 핵산, 상기 CasZ trancRNA, 및 상기 CasZ trancRNA를 암호화하는 핵산 중 적어도 하나는 이종성 모이어티에 접합된, 방법, 조성물, 키트 또는 진핵 세포.

청구항 32

제31항에 있어서, 상기 이종성 모이어티는 이종성 폴리펩타이드인, 방법, 조성물, 키트 또는 진핵 세포.

청구항 33

제1항 내지 제32항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드는 대응하는 야생형 CasZ 단백질에 비해 감소된 뉴클레아제 활성을 갖고, 이종성 폴리펩타이드에 융합된, 방법, 조성물, 키트 또는 진핵 세포.

청구항 34

제33항에 있어서, 상기 이종성 폴리펩타이드는 (i) DNA 변형 활성을 갖고, (ii) 전사를 증가시키거나 감소시키는 능력을 나타내고/내거나 (iii) DNA와 연관된 폴리펩타이드를 변형시키는 효소 활성을 갖는, 방법, 조성물, 키트 또는 진핵 세포.

청구항 35

제1항 내지 제34항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드는 도 1 또는 도 7의 CasZ 단백질과 70% 이상의 동일성을 갖는 아미노산 서열을 포함하는, 방법, 조성물, 키트 또는 진핵 세포.

청구항 36

제1항 내지 제35항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드에 결합하는 상기 가이드 서열 및 상기 영역은 서로에 대해 이종성인, 방법, 조성물, 키트 또는 진핵 세포.

청구항 37

샘플 내 표적 DNA를 검출하는 방법으로서,

(a) 상기 샘플을,

(i) CasZ 폴리펩타이드;

(ii) 상기 CasZ 폴리펩타이드에 결합하는 영역, 및 상기 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA; 및

(iii) 단일 가닥이며 상기 가이드 RNA의 상기 가이드 서열과 혼성화되지 않는 검출기 DNA

와 접촉시키는 단계; 및

(b) 상기 CasZ 폴리펩타이드에 의한 상기 단일 가닥 검출기 DNA의 절단에 의해 생성되는 검출 가능한 신호를 측정하여, 상기 표적 DNA를 검출하는 단계

를 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 38

제37항에 있어서, 상기 표적 DNA는 단일 가닥인, 샘플 내 표적 DNA를 검출하는 방법.

청구항 39

제37항 또는 제38항에 있어서, 상기 표적 DNA는 이중 가닥인, 샘플 내 표적 DNA를 검출하는 방법.

청구항 40

제37항 내지 제39항 중 어느 한 항에 있어서, 상기 표적 DNA는 바이러스 DNA인, 샘플 내 표적 DNA를 검출하는 방법.

청구항 41

제37항 내지 제40항 중 어느 한 항에 있어서, 상기 표적 DNA는 파포바바이러스, 헤파드나바이러스, 헤르페스 바이러스, 아데노바이러스, 폭스바이러스 또는 파보바이러스 DNA인, 샘플 내 표적 DNA를 검출하는 방법.

청구항 42

제37항 내지 제41항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드는 도 1 내지 도 7에 중 어느 하나에 제시된 CasZ 아미노산 서열에 대해 적어도 85%의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 43

제37항 내지 제41항 중 어느 한 항에 있어서, 상기 CasZ 폴리펩타이드는 Cas14a 폴리펩타이드인, 샘플 내 표적 DNA를 검출하는 방법.

청구항 44

제37항 내지 제43항 중 어느 한 항에 있어서, 상기 샘플은 세포 용해물로부터의 DNA 분자를 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 45

제37항 내지 제44항 중 어느 한 항에 있어서, 상기 샘플은 세포를 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 46

제37항 내지 제45항 중 어느 한 항에 있어서, 상기 접촉시키는 단계는 시험관내, 생체외 또는 생체내의 세포 내에서 수행되는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 47

제46항에 있어서, 상기 세포는 진핵 세포인, 샘플 내 표적 DNA를 검출하는 방법.

청구항 48

제37항 내지 제47항 중 어느 한 항에 있어서, 상기 표적 DNA는 10aM만큼 낮은 농도에서 검출될 수 있는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 49

제37항 내지 제48항 중 어느 한 항에 있어서, 상기 샘플에 존재하는 상기 표적 DNA의 양을 결정하는 단계를 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 50

제49항에 있어서, 상기 결정하는 단계는,

상기 검출 가능한 신호를 측정하여 시험 측정치를 생성하는 것;

기준 샘플 또는 세포에 의해 생성된 검출 가능한 신호를 측정하여 기준 측정치를 생성하는 것; 및

상기 시험 측정치를 상기 기준 측정치와 비교하여 상기 샘플에 존재하는 표적 DNA의 양을 결정하는 것을 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 51

제37항 내지 제50항 중 어느 한 항에 있어서, 검출 가능한 신호를 측정하는 단계는 금 나노입자 기반 검출, 형광 편광, 콜로이드 상 전이/분산, 전기 화학적 검출 및 반도체-기반 센싱 중 하나 이상을 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 52

제37항 내지 제51항 중 어느 한 항에 있어서, 상기 단일 가닥 검출기 DNA는 형광-방출 염료쌍을 포함하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 53

제52항에 있어서, 상기 형광-방출 염료쌍은 상기 단일 가닥 검출기 DNA의 절단되기 전에 검출 가능한 신호의 양을 생성하고, 검출 가능한 신호의 상기 양은 상기 단일 가닥 검출기 DNA의 절단 후에 감소되는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 54

제52항에 있어서, 상기 단일 가닥 검출기 DNA는 절단되기 전에 제1 검출 가능한 신호를 생성하고, 상기 단일 가닥 검출기 DNA의 절단 후에 제2 검출 가능한 신호를 생성하는, 샘플 내 표적 DNA를 검출하는 방법.

청구항 55

제52항 내지 제54항 중 어느 한 항에 있어서, 상기 형광-방출 염료쌍은 형광 공명 에너지 전달(FRET) 쌍인, 샘플 내 표적 DNA를 검출하는 방법.

청구항 56

제18항에 있어서, 검출 가능한 신호의 양은 상기 단일 가닥 검출기 DNA의 절단 후에 증가되는, 방법.

청구항 57

제52항 또는 제56항에 있어서, 상기 형광-방출 염료쌍은 퀀처/플루오르(quencher/fluor) 쌍인, 방법.

청구항 58

제52항 내지 제57항 중 어느 한 항에 있어서, 상기 단일 가닥 검출기 DNA는 둘 이상의 형광-방출 염료쌍을 포함하는, 방법.

청구항 59

제58항에 있어서, 상기 둘 이상의 형광-방출 염료쌍은 형광 공명 에너지 전달(fluorescence resonance energy transfer: FRET) 쌍 및 퀀처/플루오르 쌍을 포함하는, 방법.

청구항 60

제37항 내지 제59항 중 어느 한 항에 있어서, 상기 단일 가닥 검출기 DNA는 변형된 핵염기, 변형된 당 모이어티, 및/또는 변형된 핵산 결합을 포함하는, 방법.

청구항 61

제37항 내지 제60항 중 어느 한 항에 있어서, 상기 방법은 상기 샘플에서 핵산을 증폭시키는 단계를 포함하는, 방법.

청구항 62

제61항에 있어서, 상기 증폭시키는 단계는 등온 증폭을 포함하는, 방법.

청구항 63

제62항에 있어서, 상기 등온 증폭은 재조합효소 중합효소 증폭(recombinase polymerase amplification: RPA)을 포함하는, 방법.

청구항 64

제61항 내지 제63항 중 어느 한 항에 있어서, 상기 증폭시키는 단계는 단계 (a)의 접촉시키는 단계 전에 시작하는, 방법.

청구항 65

제61항 내지 제63항 중 어느 한 항에 있어서, 상기 증폭시키는 단계는 단계 (a)의 접촉시키는 단계와 함께 시작하는, 방법.

청구항 66

샘플 내 표적 DNA를 검출하기 위한 키트로서,

(a) 가이드 RNA가 CasZ 폴리펩타이드에 결합하는 영역 및 표적 DNA에 상보성인 가이드 서열을 포함하는, 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 핵산; 및

(b) 단일 가닥이고 상기 가이드 RNA의 상기 가이드 서열과 혼성화되지 않는, 표지된 검출기 DNA를 포함하는, 키트.

청구항 67

제66항에 있어서, CasZ 폴리펩타이드를 더 포함하는, 키트.

청구항 68

제67항에 있어서, 상기 CasZ 폴리펩타이드는 도 1 내지 도 7에 중 어느 하나에 제시된 CasZ 아미노산 서열에 대해 적어도 85%의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, 키트.

청구항 69

제67항에 있어서, 상기 CasZ 폴리펩타이드는 Cas14a 폴리펩타이드인, 키트.

청구항 70

제66항 내지 제69항 중 어느 한 항에 있어서, 상기 단일 가닥 검출기 DNA는 형광-방출 염료쌍을 포함하는, 키트.

청구항 71

제70항에 있어서, 상기 형광-방출 염료쌍은 FRET 쌍인, 키트.

청구항 72

제70항에 있어서, 상기 형광-방출 염료쌍은 퀀처/플루오르 쌍인, 키트.

청구항 73

제70항 내지 제72항 중 어느 한 항에 있어서, 상기 단일 가닥 검출기 DNA는 둘 이상의 형광-방출 염료쌍을 포함하는, 키트.

청구항 74

제73항에 있어서, 상기 둘 이상의 형광-방출 염료쌍은 제1 검출 가능한 신호를 생성하는 제1 형광-방출 염료쌍 및 제2 검출 가능한 신호를 생성하는 제2 형광-방출 염료쌍을 포함하는, 키트.

청구항 75

제66항 내지 제74항 중 어느 한 항에 있어서, 핵산 증폭 성분을 더 포함하는, 키트.

청구항 76

제75항에 있어서, 상기 핵산 증폭 성분은 재조합효소 중합효소 증폭(RPA)을 위한 성분인, 키트.

청구항 77

단일 가닥 DNA(ssDNA)를 절단하는 방법으로서,

핵산의 집단을.

(i) CasZ 폴리펩타이드; 및

(ii) 상기 CasZ 폴리펩타이드에 결합하는 영역 및 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA

와 접촉시키는 단계를 포함하되, 상기 집단은 표적 DNA 및 복수의 비표적 ssDNA를 포함하고,

상기 CasZ 폴리펩타이드는 상기 복수의 비-표적 ssDNA를 절단하는, 방법.

청구항 78

제77항에 있어서, 상기 접촉시키는 단계는 시험관내, 생체외 또는 생체내의 세포 내에서 수행되는, 방법.

청구항 79

제78항에 있어서, 상기 세포는 진핵 세포인, 방법.

청구항 80

제79항에 있어서, 상기 진핵 세포는 식물 세포인, 방법.

청구항 81

제78항 내지 제80항 중 어느 한 항에 있어서, 상기 비-표적 ssDNA는 상기 세포에 대해 외래(foreign)인, 방법.

청구항 82

제81항에 있어서, 상기 비표적 ssDNA는 바이러스 DNA인, 방법.

청구항 83

제77항 내지 제82항 중 어느 한 항에 있어서, 상기 표적 DNA는 단일 가닥인, 방법.

청구항 84

제77항 내지 제82항 중 어느 한 항에 있어서, 상기 표적 DNA는 단일 가닥인, 방법.

청구항 85

제77항 내지 제84항 중 어느 한 항에 있어서, 상기 표적 DNA는 바이러스 DNA인, 방법.

청구항 86

제77항 내지 제84항 중 어느 한 항에 있어서, 상기 표적 DNA는 파포바바이러스, 헤파드나바이러스, 헤르페스 바이러스, 아데노바이러스, 폭스바이러스 또는 파보바이러스 DNA인, 방법.

발명의 설명

기 술 분 야

[0001] **상호 참조**

[0002] 본 출원은 2017년 11월 1일자로 출원된 미국 가출원 특허 제62/580,395호의 유익을 주장하며, 이 기초출원은 그의 전문이 본 명세서에 참조에 의해 원용된다.

[0003] **텍스트 파일로서 제공된 서열목록의 참고에 의한 편입**

[0004] 서열목록은 2018년 10월 30일자에 생성되고, 용량이 536 KB인 텍스트 파일 "BERK-374WO_SEQ_LISTING_ST25.txt"로서 본 명세서에 제공된다. 텍스트 파일의 내용은 그 전체가 본 명세서에 참고로 편입된다.

배 경 기 술

[0005] DNA 서열분석 시대 전의 과학에 알려지지 않은 경로의 예인 CRISPR-Cas 시스템은 이제 파지 및 바이러스에 대해 획득된 면역을 박테리아 및 고세균에 부여하는 것으로 이해된다. 광대한 연구는 이 생화학 시스템을 발견하지 못하였다. CRISPR-Cas 시스템은 외래(foreign) DNA 또는 RNA의 획득, 표적화 및 절단에 연루된 Cas 단백질, 및 Cas 단백질을 그들의 표적에 가이드하는 짧은 스페이서 서열에 측접하는 직접 반복부를 포함하는 CRISPR 어레이로 이루어진다. RNA에 결합된 단일 Cas 단백질이 표적화된 서열에 대한 결합 및 절단을 초래하는 클래스 2 CRISPR-Cas는 최신식의 형태이다. 이들 최소 시스템의 프로그램가능한 특성은 그들의 용도를 게놈 조작 분야의 혁명을 일으키는 다재다능한 기술로서 용이하게 하였다.

발명의 내용

[0006] 본 개시내용은 하기 중 하나 이상을 포함하는 조성물 및 방법을 제공한다: (1) "CasZ" 단백질(또한 CasZ 폴리펩타이드로서 지칭됨), CasZ 단백질을 암호화하는 핵산, 및/또는 CasZ 단백질(및/또는 이를 암호화하는 핵산)을 포함하는 변형된 숙주 세포; (2) CasZ 단백질에 결합하고 이에 대해 서열 특이성을 제공하는 CasZ 가이드 RNA, CasZ 가이드 RNA를 암호화하는 핵산, 및/또는 CasZ 가이드 RNA(및/또는 이를 암호화하는 핵산)를 포함하는 변형된 숙주 세포; 및 (3) CasZ 전사촉진(transactivating) 비암호화 RNA(trancRNA)(본 명세서에서 "CasZ trancRNA"로서 지칭됨), CasZ trancRNA를 암호화하는 핵산, 및/또는 CasZ trancRNA(및/또는 이를 암호화하는 핵산)를 포함하는 변형된 숙주 세포.

도면의 간단한 설명

[0007] **도 1**은 천연 유래 CasZ 단백질 서열의 예를 도시한 도면.

도 2는 CasZ 단백질에 추가로 Cas1 단백질을 포함하는 CasZ 좌위의 개략적 표현을 도시한 도면.

도 3은 다른 클래스 2 CRISPR/Cas 효과기 단백질 서열에 관해 CasZ 서열의 계통수를 도시한 도면.

도 4는 다른 클래스 2 CRISPR/Cas 좌위로부터의 Cas1 서열에 관해 CasZ 좌위로부터의 Cas1 서열의 계통수를 도시한 도면.

도 5는 CasZ 좌위로부터의 trancRNA의 발현을 입증하는 전사체 RNA 맵핑 데이터를 도시한 도면. trancRNA는

CasZ 반복부 어레이에 인접하지만, 반복부 서열을 포함하지 않고, 반복부 서열에 대해 상보성이 아니다. 다음의 좌위에 대한 RNA 맵핑 데이터를 나타낸다: CasZa3, CasZb4, CasZc5, CasZd1 및 CasZe3. 작은 반복 정렬 화살표는 CRISPR 어레이의 반복부를 나타내고(가이드 RNA-암호화 서열의 존재를 나타냄); 반복부 어레이 밖의 그리고 인접한 피크는 고도로 전사된 trancRNA를 나타낸다. 도 5(계속 1) 뉴클레오타이드 서열(상부에서 하부로: 서열번호 312 내지 331). 도 5(계속 3) 뉴클레오타이드 서열(상부에서 하부로: 서열번호 161 내지 177).

도 6은 CasZc(상부) 및 CasZb(하부)에 대한 PAM 고갈 분석을 이용하여 분석한 바와 같은 PAM 선호도에 대한 결과를 도시한 도면.

도 7은 본 명세서에 기재된 Cas14 단백질의 서열을 도시한 도면.

도 8, 패널 A 내지 D는 CRISPR-Cas14 게놈 좌위의 구성 및 계통발생을 도시한 도면.

도 9는 Cas14 오솔로그의 계통발생 분석을 도시한 도면.

도 10은 공지된 CRISPR 시스템으로부터의 Cas1에 대한 최대 가능성 계통수(Maximum likelihood tree)를 도시한 도면.

도 11, 패널 A 내지 B는 CRISPR-Cas14 시스템에 의한 새로운 스페이서의 획득을 도시한 도면.

도 12, 패널 A 내지 D는 CRISPR-Cas14a가 적극적으로 적합하게 되고, tracrRNA를 암호화하는 것을 도시한 도면.

도 13, 패널 a 내지 b는 CRISPR-Cas14 좌위에 대한 메타전사체를 도시한 도면.

도 14, 패널 A 내지 B는 RNA 가공 및 CRISPR-Cas14에 의한 이종성 발현을 도시한 도면.

도 15, 패널 a 내지 d는 Cas14a1 및 SpCas9에 의한 플라스미드 고갈을 도시한 도면.

도 16, 패널 a 내지 D는 CRISPR-Cas14a가 RNA-가이드된 DNA-엔도뉴클레아제라는 것을 도시한 도면.

도 17, 패널 A 내지 E는 Cas14a1에 의한 ssDNA의 분해를 도시한 도면.

도 18은 다양한 가이드 RNA 성분에 의한 ssDNA의 Cas14a1 절단 역학을 도시한 도면.

도 19, 패널 A 내지 F는 Cas14a1 가이드 RNA 성분의 최적화를 도시한 도면.

도 20, 패널 a 내지 E는 CRISPR-Cas14a에 의한 고충실도 ssDNA DNP 검출을 도시한 도면. 도 20, 패널 c는 뉴클레오타이드 서열(상부에서 하부로: 서열번호 367 내지 370)을 제공한 도면.

도 21, 패널 A 내지 F는 Cas14a1 절단율에 대한 다양한 활성체의 영향을 도시한 도면.

도 22, 패널 A 내지 B는 CRISPR-Cas14 시스템의 다양성을 도시한 도면.

도 23, 패널 A 내지 C는 이종성 숙주에서의 Cas14a1 매개 간섭 시험을 도시한 도면. 이콜라이(*E. coli*)에서의 간섭을 시험하기 위한 Cas14a1 및 LbCas12a 작제물의 다이어그램.

도 24는 본 발명에서 사용되는 플라스미드의 Cas14 뉴클레오타이드 서열을 도시한 도면.

도 25, 패널 A 내지 E는 도 24에 개시된 플라스미드 각각의 서열 맵을 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0008] **정의**

[0009] 본 명세서에서 사용되는 바와 같은 "이종성"은 각각 천연 핵산 또는 단백질에서 발견되지 않는 뉴클레오타이드 또는 폴리뉴클레오타이드 서열을 의미한다. 예를 들어, CasZ 폴리펩타이드에 대해, 이종성 폴리펩타이드는 CasZ 폴리펩타이드 이외의 단백질로부터의 아미노산 서열을 포함한다. 일부 경우에, 하나의 종으로부터의 CasZ 단백질 부분은 상이한 종으로부터의 CasZ 단백질 부분에 융합된다. 따라서 각각의 종으로부터의 CasZ 서열은 서로에 대해 이종성인 것으로 고려될 수 있었다. 다른 예로서, CasZ 단백질(예를 들어, dCasZ 단백질)은 비-CasZ 단백질(예를 들어, 히스톤 데아세틸라제)로부터의 활성 도메인에 융합될 수 있고, 활성 도메인의 서열은 이종성 폴리펩타이드로 고려될 수 있었다(이는 CasZ 단백질에 대해 이종성이다).

[0010] 본 명세서에서 상호 호환적으로 사용되는 용어 "폴리뉴클레오타이드" 및 "핵산"은 리보뉴클레오타이드 또는 데옥시뉴클레오타이드 중 하나의 임의의 길이 뉴클레오타이드의 중합체 형태를 지칭한다. 따라서, 이 용어는 단일

-, 이중-, 또는 다중-가닥 DNA 또는 RNA, 게놈 DNA, cDNA, DNA-RNA 혼성체, 또는 퓨린 및 피리미딘 염기 또는 다른 천연, 화학적 또는 생화학적으로 변형된, 비천연 또는 유도체화된 뉴클레오타이드 염기를 포함하는 중합체를 포함하지만, 이들로 제한되지 않는다. 용어 "폴리뉴클레오타이드" 및 "핵산"은 기재 중인 실시형태에 적용 가능한, 단일-가닥(예컨대 센스 또는 안티센스) 및 이중-가닥 폴리뉴클레오타이드를 포함하는 것으로 이해되어야 한다.

[0011] 용어 "폴리펩타이드", "펩타이드" 및 "단백질"은 본 명세서에서 상호 호환적으로 사용되고, 유전적으로 암호화된 그리고 비유전적으로 암호화된 아미노산, 화학적으로 또는 생화학적으로 변형 또는 유도체화된 아미노산, 및 변형된 펩타이드 골격을 갖는 폴리펩타이드를 포함할 수 있는 임의의 길이의 아미노산의 중합체 형태를 지칭한다. 상기 용어는 N-말단의 메티오닌 잔기가 있거나 또는 없는, 이종성 아미노산 서열과의 융합 단백질, 이종성 및 상동성 리더 서열과의 융합; 면역학적으로 태그된 단백질 등을 포함하지만, 이들로 제한되지 않는 융합 단백질을 포함한다.

[0012] 핵산, 단백질, 세포 또는 유기체에 적용되는 본 명세서에서 사용되는 용어 "천연 유래"는 천연에서 발견되는 핵산, 세포, 단백질 또는 유기체를 지칭한다.

[0013] 본 명세서에서 사용되는 용어 "단리된"은 폴리뉴클레오타이드, 폴리펩타이드 또는 세포가 자연적으로 생기는 환경과 상이한 환경에 있는 폴리뉴클레오타이드, 폴리펩타이드 또는 세포를 설명하는 것을 의미한다. 단리된 유전자 변형된 숙주 세포는 유전자 변형된 숙주 세포의 혼합 집단으로 제공될 수 있다.

[0014] 본 명세서에 사용되는 바와 같은 용어 "외인성 핵산"은 천연에서 주어진 박테리아, 유기체 또는 세포에서 정상적으로 또는 자연적으로 발견되지 않고/않거나 생성되지 않는 핵산을 지칭한다. 본 명세서에 사용되는 바와 같은 용어 "내인성 핵산"은 천연에서 주어진 박테리아, 유기체 또는 세포에서 정상적으로 또는 자연적으로 발견되고/되거나 생성된 핵산을 지칭한다. "내인성 핵산"은 또한 "천연 핵산" 또는 주어진 박테리아, 유기체 또는 세포에 대해 "천연"인 핵산으로서 지칭된다.

[0015] 본 명세서에서 사용되는 "재조합체"는 특정 핵산(DNA 또는 RNA)이 천연 시스템에서 발견되는 내인성 핵산과 구별 가능한 구조적 암호 또는 비-암호 서열을 갖는 작제물을 초래하는 클로닝, 제한 및/또는 결찰 단계의 다양한 조합의 생성물이라는 것을 의미한다. 일반적으로, 구조적 암호 서열을 암호화하는 DNA 서열은 세포에 또는 무세포 전사 및 번역 시스템에 함유된 재조합 전사 단위로부터 발현될 수 있는 합성 핵산을 제공하기 위해 cDNA 단편 및 짧은 올리고뉴클레오타이드 링커로부터, 또는 일련의 합성 올리고뉴클레오타이드로부터 조립될 수 있다. 이러한 서열은 전형적으로 진핵 유전자에 존재하는 내부 비-번역 서열 또는 인트론에 의해 중단되지 않는 오픈 리딩 프레임 형태로 제공될 수 있다. 적절한 서열을 포함하는 게놈 DNA는 또한 재조합 유전자 또는 전사 단위의 형태로 사용될 수 있다. 비번역 DNA의 서열은 오픈 리딩 프레임으로부터의 5' 또는 3'에 존재할 수 있으며, 여기서 이러한 서열은 암호 영역의 조작 또는 발현을 방해하지 않고, 정말로 다양한 메커니즘에 의해 목적으로 하는 생성물의 생성물을 조작하도록 작용할 수 있다(" DNA 조절 서열" 참조, 이하).

[0016] 따라서, 예를 들어, 용어 "재조합" 폴리뉴클레오타이드 또는 "재조합" 핵산은 천연 유래가 아닌 것, 예를 들어, 인간 개입을 통해 서열의 2개의 다르게 분리된 세그먼트의 인공 조합에 의해 만들어진 것을 지칭한다. 이 인공 조합은 종종 화학적 합성 수단에 의해, 또는 핵산의 단리된 세그먼트의 인공 조작에 의해, 예를 들어, 유전자 조작 기법에 의해 수행된다. 이는 보통 코돈을 동일한 또는 보존적 아미노산을 암호화하는 풍부한 코돈으로 대체하는 한편, 전형적으로 서열 인식 부위를 도입하거나 또는 제거하기 위해 행한다. 대안적으로, 목적으로 하는 기능의 조합을 생성하기 위해 목적으로 하는 기능의 핵산 세그먼트를 함께 결합하도록 수행된다. 이 인공 조합은 종종 화학적 합성 수단에 의해, 또는 핵산의 단리된 세그먼트의 인공 조작에 의해, 예를 들어, 유전자 조작 기법에 의해 수행된다.

[0017] 유사하게, 용어 "재조합" 폴리펩타이드는 천연 유래가 아니며, 예를 들어, 인간 개입을 통해 아미노 서열의 2개의 다르게 분리된 세그먼트의 인공 조합물에 의해 제조된 폴리펩타이드를 지칭한다. 따라서, 예를 들어, 이종성 아미노산 서열을 포함하는 폴리펩타이드는 재조합이다.

[0018] "작제물" 또는 "벡터"는 특정 뉴클레오타이드 서열(들)의 발현 및/또는 증식의 목적을 위해 생성된, 또는 다른 재조합 뉴클레오타이드 서열의 작제물에서 사용될 재조합 핵산, 일반적으로 재조합 DNA를 의미한다.

[0019] 본 명세서에서 상호 호환적으로 사용되는 용어 "DNA 조절 서열", "제어 요소" 및 "조절 요소"는 암호 서열의 발현 및/또는 숙주 세포에서 암호화된 폴리펩타이드의 생성을 제공하고/하거나 조절하는 전사 및 번역 제어 서열, 예컨대 프로모터, 인핸서, 폴리아데닐화 신호, 종결자, 단백질 분해 신호 등을 지칭한다.

[0020] 용어 "형질전환"은 본 명세서에서 "유전자 변형"과 상호 호환적으로 사용되고, 세포 내로 새로운 핵산(예를 들어, 세포에 외인성인 DNA)의 도입 후 세포에 유도된 영구한 또는 일시적 유전자 변화를 지칭한다. 유전자 변화("변형")는 숙주 세포의 게놈 내로 새로운 핵산의 혼입에 의해, 또는 에피솜 요소로서 새로운 핵산의 일시적 또는 안정한 유지에 의해 달성될 수 있다. 세포가 진핵 세포인 경우에, 영구한 유전자 변화는 일반적으로 세포 게놈 내로 새로운 DNA의 도입에 의해 달성된다. 원핵세포에서, 염색체 내로 또는 염색체외 요소, 예컨대 플라스미드 및 발현 벡터를 통해 영구적 변화가 도입될 수 있는데, 이는 재조합 숙주 세포에서 그들의 유지에 도움을 주는 하나 이상의 선택 가능한 마커를 함유할 수 있다. 유전자 변형의 적합한 방법은 바이러스 감염, 형질감염, 접합, 원형질체 융합, 전기천공법, 입자총 기법, 인산칼슘 침전, 직접 마이크로주사 등을 포함한다. 방법의 선택은 일반적으로 형질감염 중인 세포의 유형 및 형질감염이 일어나는 환경(즉, 시험관내, 생체외 또는 생체내)에 의존한다. 이들 방법의 일반적 논의는 문헌[Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995]에서 찾을 수 있다.

[0021] "작동 가능하게 연결된"은 이렇게 기재된 성분이 그들의 의도된 방식으로 작용하도록 허용하는 관계인 병치를 지칭한다. 예를 들어, 프로모터가 그의 전사 또는 발현에 영향을 미친다면 프로모터는 암호 서열에 작동 가능하게 연결된다. 본 명세서에서 사용되는 용어 "이종성 프로모터" 및 "이종성 제어 영역"은 천연에서 특정 핵산과 정상적으로 결합되지 않은 프로모터 및 다른 제어 영역을 지칭한다. 예를 들어, "암호 영역에 대해 이종성인 전사 제어 영역"은 천연에서 암호 영역과 정상적으로 결합되지 않은 전사 제어 영역이다.

[0022] 본 명세서에서 사용되는 "숙주 세포"는 진핵 또는 원핵 세포가 핵산에 대한 수용자(예를 들어, 발현 벡터)로서 사용될 수 있거나 또는 사용되고, 핵산에 의해 유전자 변형된 본래의 세포의 자손을 포함하는, 생체내 또는 시험관내 진핵 세포, 원핵 세포 또는 단세포 독립체로서 배양시킨 다세포 유기체로부터의 세포(예를 들어, 세포주)를 나타낸다. 단일 세포의 자손은 단일 세포의 자손은 천연, 우연한 또는 의도적인 돌연변이에 기인하여, 본래의 모체와 형태가 또는 게놈 또는 총 DNA 상보체와 완전히 동일할 필요는 없다는 것이 이해된다. "재조합 숙주 세포"(또한 "유전자 변형된 숙주 세포"로서 지칭됨)는 이종성 핵산, 예를 들어, 발현 벡터가 도입된 숙주 세포이다. 예를 들어, 대상 원핵 숙주 세포는 이종성 핵산, 예를 들어, 원핵 숙주 세포에서 정상적으로 발견되지 않는 원핵 숙주 세포 또는 재조합 핵산에 대해 외래인(그리고 천연에서 정상적으로 발견되지 않는) 외인성 핵산의 적합한 원핵 숙주 세포 내로의 도입에 의한 유전자 변형된 원핵 숙주 세포(예를 들어, 박테리아)이고; 그리고 대상 진핵 숙주 세포는 이종성 핵산, 예를 들어, 진핵 숙주 세포에 정상적으로 발견되지 않는 진핵 숙주 세포, 또는 재조합 핵산에 대해 외래인 외인성 핵산의 적합한 진핵 숙주 세포 내로 도입에 의한 유전자 변형된 진핵 숙주 세포이다.

[0023] 용어 "보존적 아미노산 치환"은 유사한 측쇄를 갖는 아미노산 잔기의 단백질에서의 상호 호환성을 지칭한다. 예를 들어, 지방족 측쇄를 갖는 아미노산의 그룹은 글리신, 알라닌, 발린, 류신, 및 아이소류신으로 이루어지며; 지방족-하이드록실 측쇄를 갖는 아미노산의 그룹은 세린 및 트레오닌으로 이루어지고; 아마이드-함유 측쇄를 갖는 아미노산의 그룹은 아스파라긴 및 글루타민으로 이루어지며; 방향족 측쇄를 갖는 아미노산 그룹은 페닐알라닌, 타이로신 및 트립토판으로 이루어지고; 염기성 측쇄를 갖는 아미노산의 그룹은 라이신, 알기닌 및 히스티딘으로 이루어지며; 그리고 황-함유 측쇄를 갖는 아미노기의 그룹은 시스테인 및 메티오닌으로 이루어진다. 예시적인 보존적 아미노산 치환 그룹은 발린-류신-아이소류신, 페닐알라닌-타이로신, 라이신-알기닌, 알라닌-발린, 및 아스파라긴-글루타민이다.

[0024] 폴리뉴클레오타이드 또는 폴리펩타이드는 다른 폴리뉴클레오타이드 또는 폴리펩타이드에 대한 특정 "서열 동일성" 백분율을 갖는데, 이는 정렬될 때, 염기 또는 아미노산의 백분율이 동일하고, 두 서열을 비교할 때 동일한 상대적 위치에 있다는 것을 의미한다. 서열 유사성은 다수의 상이한 방식으로 결정될 수 있다. 서열 동일성을 결정하기 위해, 서열은 ncbi.nlm.nih.gov/BLAST에서 월드 와이드 웹에 대해 이용 가능한 BLAST를 포함하는 방법 및 컴퓨터 프로그램을 이용하여 정렬될 수 있다. 예를 들어, 문헌[Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10] 참조. 다른 정렬 알고리즘은 옥스퍼드 몰레큘러 그룹, 인코포레이티드(Oxford Molecular Group, Inc.)의 전액 출자된 자회사인 미국 위스콘신주 메디슨에 소재한 제네틱스 컴퓨팅 그룹(Genetics Computing Group: GCG) 패키지에서 이용 가능한 FASTA이다. 정렬을 위한 다른 기법은 문헌[Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA]에 기재되어 있다. 특정 관심 대상 중에서 서열 내 갭을 허용하는 정렬 프로그램이다. 스미스-워터만(Smith-Waterman)은 서열 정렬에서 갭을 허용하는 알고리즘 중 하나의 유형이다. 문헌[*Meth. Mol. Biol.* 70: 173-187 (1997)] 참조. 또한, 니들만(Needleman) 및 분쉬(Wunsch) 정렬 방법을 이용하는 GAP 프로그램은 서열을 정렬하기 위해 이용될 수 있다. 문

헌[*J. Mol. Biol.* 48: 443-453 (1970)] 참조.

[0025] 본 명세서에서 사용되는 용어 "치료", "치료하는" 등은 목적으로 하는 약학적 및/또는 생리적 효과를 얻는 것을 지칭한다. 효과는 질환 또는 이의 증상을 완전히 또는 부분적으로 방지하는 것에 관해 예방적일 수 있고/있거나 질환 및/또는 질환에 기인하는 유해한 효과에 대한 부분적 또는 완전한 치유에 관해 치유적일 수 있다. 본 명세서에서 사용되는 "치료"는 포유류에서, 예를 들어, 인간에서의 질환의 임의의 치료를 아우르며, (a) 질환의 성향이 있을 수 있지만, 그것을 갖는 것으로 아직 진단되지 않은 대상체에서 질환이 생기는 것을 예방하는 것; (b) 질환을 저해하는 것, 즉, 그의 발생을 저지하는 것; 및 (c) 질환을 완화시키는 것, 즉, 질환의 퇴행을 야기하는 것을 포함한다.

[0026] 본 명세서에서 상호 호환적으로 사용되는 용어 "개체", "대상체", "숙주" 및 "환자"는 뮤린, 유인원, 비-인간 영장류, 인간, 포유류 농장 동물, 포유류 스포츠 동물 및 포유류 반려동물을 포함하지만, 이들로 제한되지 않는 개개 유기체, 예를 들어, 포유류를 지칭한다.

[0027] 본 발명을 추가로 기재하기 전에, 본 발명은 기재한 특정 실시형태로 제한되지 않으며, 그렇게 해서 물론 다를 수 있다는 것이 이해되어야 한다. 또한 본 명세서에 사용되는 용어는 단지 특정 실시형태를 기재할 목적을 위한 것이며, 제한하는 것으로 의도되지 않는데, 본 발명의 범주는 첨부된 청구범위에 의해서만 제한될 것이기 때문이다.

[0028] 수치의 범위가 제공되는 경우에, 문맥에서 달리 분명하게 나타내지 않는 한, 해당 범위 및 해당 언급된 범위에서의 임의의 다른 언급된 또는 개재된 값의 상한과 하한 사이에서 하한 단위의 1/10까지의 각각의 사이의 값은 본 발명 내에 포함된다는 것이 이해된다. 이들 더 작은 범위의 상한 및 하한은 더 작은 범위에 독립적으로 포함될 수 있고, 또한 본 발명 내에 포함되며, 언급된 범위 내에서 임의의 구체적으로 제외된 제한이 적용된다. 언급된 범위가 제한 중 하나 또는 둘 다를 포함하는 경우에, 포함된 제한 중 하나 또는 둘 다를 제외한 범위가 또한 본 발명에 포함된다.

[0029] 달리 정의되지 않는 한, 본 명세서에서 사용되는 모든 기술적 및 과학적 용어는 본 발명이 속하는 기술분야의 당업자에 의해 통상적으로 이해되는 것과 동일한 의미를 가진다. 본 명세서에 기재된 것과 유사하거나 또는 동일한 임의의 방법 및 물질이 또한 본 발명의 실행 또는 시험에서 사용될 수 있지만, 바람직한 방법 및 물질이 이제 기재된다. 본 명세서에서 언급된 모든 간행물은 간행물과 함께 인용된 방법 및/또는 물질을 개시하고 기재하기 위해 본 명세서에 참고로 편입된다.

[0030] 본 명세서에 사용되고 첨부하는 청구범위에 사용되는 바와 같은 단수 형태는 달리 분명하게 나타내지 않는 한 복수의 대상을 포함하는 것임을 주목하여야 한다. 따라서, 예를 들어, "CasZ 폴리펩타이드"에 대한 언급은 복수의 이러한 폴리펩타이드를 포함하고, "가이드 RNA"에 대한 언급은 당업자에게 공지된 하나 이상의 가이드 RNA 및 이의 동등물 등에 대한 언급을 포함한다. 청구범위는 임의의 선택적 요소를 제외하도록 작성될 수 있다는 것을 추가로 주목한다. 이렇게 해서, 이러한 언급은 청구범위 요소의 열거와 관련하여 "유일하게", "단지" 등의 사용 또는 "음성" 제한의 사용을 위한 선행 기준으로서 작용하는 것으로 의도된다.

[0031] 명확함을 위해 별개의 실시형태와 관련하여 기재된 본 발명의 특정 특징은 또한 단일 실시형태에서 조합하여 제공될 수 있다는 것이 인식된다. 대조적으로, 간결함을 위해 단일 실시형태와 관련하여 기재된 본 발명의 다양한 특징은 또한 별개로 또는 임의의 적합한 하위-조합으로 제공될 수 있다. 본 발명에 관한 실시형태의 모든 조합은 본 발명에 의해 구체적으로 포함되고, 각각의 그리고 모든 조합이 개별적으로 그리고 명확하게 개시되는 것과 같이 본 명세서에 개시된다. 추가로, 다양한 실시형태의 모든 하위 조합 및 이의 요소는 또한 본 발명에 의해 구체적으로 포함되고, 각각의 그리고 모든 이러한 하위 조합이 개별적으로 그리고 명확하게 본 명세서에 개시되는 것과 같이 본 명세서에 개시된다.

[0032] 본 명세서에서 논의되는 간행물은 본 출원의 출원일 전의 그들의 개시내용에 대해서만 제공된다. 본 명세서의 어떤 것도 본 발명이 선행 발명 때문에 이러한 간행물보다 선행한다는 자격이 부여되지 않는다는 용인으로서 해석되어서는 안 된다. 추가로, 제공된 간행물의 날짜는 독립적으로 확인될 필요가 있을 수 있는 실제 간행물과 상이할 수 있다.

[0033] **발명을 실시하기 위한 구체적인 내용**

[0034] 본 개시내용은 하기 중 하나 이상을 포함하는 조성물 및 방법을 제공한다: (1) "CasZ" 단백질(또한 CasZ 폴리펩타이드로서 지칭됨), CasZ 단백질을 암호화하는 핵산, 및/또는 CasZ 단백질(및/또는 이를 암호화하는 핵산)을 포함하는 변형된 숙주 세포; (2) CasZ 단백질에 결합하고 이에 대해 서열 특이성을 제공하는 CasZ 가이드 RNA,

CasZ 가이드 RNA를 암호화하는 핵산, 및/또는 CasZ 가이드 RNA(및/또는 이를 암호화하는 핵산)를 포함하는 변형된 숙주 세포; 및 (3) CasZ 전사촉진 비암호화 RNA(trancRNA)(본 명세서에서 "CasZ trancRNA"로서 지칭됨), CasZ trancRNA를 암호화하는 핵산, 및/또는 CasZ trancRNA(및/또는 이를 암호화하는 핵산)를 포함하는 변형된 숙주 세포.

[0035] **조성물**

[0036] **CRISPR/CasZ 단백질, 가이드 RNA 및 trancRNA**

[0037] 클래스 2 CRISPR-Cas 시스템은 단일 멀티도메인 단백질을 포함하는 효과기 분자를 특징으로 한다. CasZ 시스템에서, CRISPR/Cas 엔도뉴클레아제(예를 들어, CasZ 단백질)는 대응하는 가이드 RNA(예를 들어, CasZ 가이드 RNA)와 상호작용(결합)하여 표적 핵산 분자 내에서 가이드 RNA와 표적 서열 사이의 염기쌍을 통해 표적 핵산 내 특정 부위로 표적화된 리보뉴클레오단백질(RNP) 복합체를 형성한다. 가이드 RNA는 표적 핵산의 서열(표적 부위)에 상보성인 뉴클레오타이드 서열(가이드 서열)을 포함한다. 따라서, CasZ 단백질은 CasZ 가이드 RNA와 복합체를 형성하고, 가이드 RNA는 가이드 서열을 통해 RNP 복합체에 서열 특이성을 제공한다. 복합체의 CasZ 단백질은 부위-특이적 활성을 제공한다. 다시 말해서, CasZ 단백질은 가이드 RNA와의 결합 때문에 표적 핵산 내의 표적 부위(예를 들어, 표적 염색체 핵산 내의 표적 뉴클레오타이드 서열; 또는 표적 염색체외 핵산, 예를 들어, 에피솜 핵산, 미니서클 핵산, 미토콘드리아 핵산, 엽록체 핵산 내의 표적 뉴클레오타이드 서열 등)에 가이드된다(예를들어, 표적 부위에서 안정화된다).

[0038] 본 개시내용은 CasZ 폴리펩타이드(및/또는 CasZ 폴리펩타이드를 암호화하는 핵산)(예를 들어, CasZ 폴리펩타이드가 자연적으로 존재하는 CasZ 단백질, 틈내기효소 CasZ 단백질, dCasZ 단백질, 키메라 CasZ 단백질 등일 수 있는 경우에)(CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질)을 포함하는 조성물을 제공한다. 본 개시내용은 CasZ 가이드 RNA(및/또는 CasZ 가이드 RNA를 암호화하는 핵산)를 포함하는 조성물을 제공한다. 예를 들어, 본 개시내용은 (a) CasZ 폴리펩타이드 (및/또는 CasZ 폴리펩타이드를 암호화하는 핵산) 및 (b) CasZ 가이드 RNA(및/또는 CasZ 가이드 RNA를 암호화하는 핵산)를 포함하는 조성물을 제공한다. 본 개시내용은: (a) CasZ 폴리펩타이드; 및 (b) CasZ 가이드 RNA를 포함하는 핵산/단백질 복합체(RNP 복합체)를 제공한다. 본 개시내용은 CasZ trancRNA를 포함하는 조성물을 제공한다. 본 개시내용은 CasZ trancRNA 및: (a) CasZ 단백질, 및 (b) CasZ 가이드 RNA(예를 들어, CasZ trancRNA 및 CasZ 단백질을 포함함), CasZ trancRNA 및 CasZ 가이드 RNA, 또는 CasZ trancRNA 및 CasZ 단백질 및 CasZ 가이드 RNA 중 하나 이상을 포함하는 조성물을 제공한다. 본 개시내용은: (a) CasZ 폴리펩타이드; 및 (b) CasZ 가이드 RNA; 및 (c) CasZ trancRNA를 포함하는 핵산/단백질 복합체(RNP 복합체)를 제공한다. 본 개시내용은 CasZ 단백질 및: (a) CasZ trancRNA, 및 (b) CasZ 가이드 RNA 중 하나 이상을 포함하는 조성물을 제공한다.

[0039] ***CasZ 단백질***

[0040] CasZ 폴리펩타이드(이 용어는 용어 "CasZ 단백질", "Cas14", "Cas14 폴리펩타이드" 또는 "Cas14 단백질"과 상호 호환적으로 사용됨)는 표적 핵산과 결합된 표적 핵산 및/또는 폴리펩타이드에 결합하고/하거나 변형할 수 있다(예를 들어, 절단, 틈내기, 메틸화, 데메틸화 등)(예를 들어, 히스톤 꼬리의 메틸화 또는 아세틸화)(예를 들어, 일부 경우에, CasZ 단백질은 활성을 갖는 융합 상대를 포함하고, 일부 경우에 CasZ 단백질은 뉴클레아제 활성을 제공한다). 일부 경우에, CasZ 단백질은 천연 유래 단백질이다(예를 들어, 원핵 세포에서 자연적으로 생긴다). 다른 경우에, CasZ 단백질은 천연 유래 폴리펩타이드가 아니다(예를 들어, CasZ 단백질은 변이체 CasZ 단백질, 키메라 단백질 등이다). CasZ 단백질은 CasZ 단백질의 주요 아미노산 서열에 대해 인접하지 않지만, 일단 단백질이 생성되고 폴딩된다면 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III, 또한 본 명세서에서 서브도메인으로서 지칭됨)을 포함한다. 천연 유래 CasZ 단백질은 표적화된 핵산(예를 들어, 이중 가닥 DNA(dsDNA)) 내 특정 서열에서 절단을 촉매하는 엔도뉴클레아제로서 작용한다. 서열 특이성은 표적 DNA 내에서 표적 서열에 혼성화되는 회합된 가이드 RNA에 의해 제공된다. 천연 유래 CasZ 가이드 RNA는 crRNA이며, 여기서 crRNA는 (i) 표적 DNA에서 표적 서열에 혼성화되는 가이드 서열 및 (ii) CasZ 단백질에 결합하는 단백질 결합 세그먼트를 포함한다.

[0041] 일부 실시형태에서, 대상 방법 및/또는 조성물의 CasZ 단백질은 천연 유래(야생형) 단백질이다(또는 이로부터 유래된다). 천연 유래 CasZ 단백질(예를 들어, CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZk, CasZl)의 예는 도 1에 도시된다. 일부 경우에, 대상 CasZ 단백질은 CasZa 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZb 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZc 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZd 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZe 단백질이다. 일부 경우

에, 대상 CasZ 단백질은 CasZf 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZg 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZh 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZi 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZj 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZk 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZl 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZe, CasZf, CasZg 또는 CasZh 단백질이다. 일부 경우에, 대상 CasZ 단백질은 CasZj, CasZk 또는 CasZl 단백질이다.

[0042] 이 새로 발견된 단백질(CasZ)은 이전에 확인된 CRISPR-Cas 엔도뉴클레아제에 비해 짧고, 따라서 대안으로서 이 단백질의 사용은 단백질을 암호화하는 뉴클레오타이드가 상대적으로 짧다는 이점을 제공한다는 것을 주목하는 것은 중요하다. 이는, 예를 들어, 연구 및/또는 임상 적용을 위한 세포, 예컨대 진핵 세포(예를 들어, 시험관내, 생체외, 생체내 포유류 세포, 인간 세포, 마우스 세포)에 대한 전달을 위해 CasZ 단백질을 암호화하는 핵산이 바람직한 경우에, 예를 들어, 바이러스 벡터(예를 들어, AAV 벡터)를 사용하는 경우에 유용하다. 추가로, 이들의 자연적 맥락에서, CasZ-암호화 DNA 서열은 또한 Cas1 단백질을 갖는 좌위에 존재한다.

[0043] 일부 경우에, 대상 CasZ 단백질은 길이가 900개 이하의 아미노산(예를 들어, 850개 이하의 아미노산, 800개 이하의 아미노산, 750개 이하의 아미노산 또는 700개 이하의 아미노산)이다. 일부 경우에, 대상 CasZ 단백질은 길이가 850개 이하의 아미노산(예를 들어, 850개 이하의 아미노산)이다. 일부 경우에, 대상 CasZ 단백질은 길이가 800개 이하의 아미노산(예를 들어, 750개 이하의 아미노산)이다. 일부 경우에, 대상 CasZ 단백질은 길이가 700개 이하의 아미노산이다. 일부 경우에, 대상 CasZ 단백질은 길이가 650개 이하의 아미노산이다.

[0044] 일부 경우에, 대상 CasZ 단백질은 길이가 350 내지 900개의 아미노산(예를 들어, 350 내지 850, 350 내지 800, 350 내지 750, 350 내지 700, 400 내지 900, 400 내지 850, 400 내지 800, 400 내지 750, 또는 400 내지 700개의 아미노산)의 범위에 있다.

[0045] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZa)은 길이가 350 내지 750개의 아미노산(예를 들어, 350 내지 700, 350 내지 550, 450 내지 550, 450 내지 750, 450 내지 650, 또는 450 내지 550개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZa)은 길이가 450 내지 750개의 아미노산(예를 들어, 500 내지 700개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZa)은 길이가 350 내지 700개의 아미노산(예를 들어, 350 내지 650, 350 내지 600, 또는 350 내지 550개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZa)은 길이가 500 내지 700개의 아미노산의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZa)은 길이가 450 내지 550개의 아미노산의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZa)은 길이가 350 내지 550개의 아미노산의 범위에 있다.

[0046] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZb)은 길이가 350 내지 700개의 아미노산(예를 들어, 350 내지 650 또는 350 내지 620개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZb)은 길이가 450 내지 700개의 아미노산(예를 들어, 450 내지 650, 500 내지 650 또는 500 내지 620개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZb)은 길이가 500 내지 650개의 아미노산(예를 들어, 500 내지 620개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZb)은 길이가 500 내지 620개의 아미노산의 범위에 있다.

[0047] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZc)은 길이가 600 내지 800개의 아미노산(예를 들어, 600 내지 650 또는 700 내지 800개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZc)은 길이가 600 내지 650개의 아미노산의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZc)은 길이가 700 내지 800개의 아미노산의 범위에 있다.

[0048] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZd)은 길이가 400 내지 650개의 아미노산(예를 들어, 400 내지 600, 400 내지 550, 500 내지 650, 500 내지 600 또는 500 내지 550개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZd)은 길이가 500 내지 600개의 아미노산의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZd)은 길이가 500 내지 550개의 아미노산의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZd)은 길이가 400 내지 550개의 아미노산의 범위에 있다.

[0049] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZe)은 길이가 450 내지 700개의 아미노산(예를 들어, 450 내지 650, 450 내지 615, 475 내지 700, 475 내지 650, 또는 475 내지 615개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZe)은 길이가 450 내지 675개의 아미노산의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZe)은 길이가 475 내지 675개의 아미노산의 범위에 있다.

[0050] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZf)은 길이가 400 내지 550개의 아미노산(예를 들어, 400 내지

520, 400 내지 500, 400 내지 475, 415 내지 550, 415 내지 520, 415 내지 500 또는 415 내지 475개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZf)은 길이가 400 내지 475개의 아미노산(예를 들어, 400 내지 450개의 아미노산)의 범위에 있다.

[0051] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZg)은 길이가 500 내지 750개의 아미노산(예를 들어, 550 내지 750 또는 500 내지 700개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZg)은 길이가 700 내지 750개의 아미노산의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZg)은 길이가 550 내지 600개의 아미노산의 범위에 있다.

[0052] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZh)은 길이가 380 내지 450개의 아미노산(예를 들어, 380 내지 420, 400 내지 450 또는 400 내지 420개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZh)은 길이가 400 내지 420개의 아미노산의 범위에 있다.

[0053] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZi)은 길이가 700 내지 800개의 아미노산(예를 들어, 700 내지 750, 720 내지 800 또는 720 내지 750개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZi)은 길이가 720 내지 780개의 아미노산의 범위에 있다.

[0054] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZj)은 길이가 600 내지 750개의 아미노산(예를 들어, 600 내지 700 또는 650 내지 700개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZj)은 길이가 400 내지 420개의 아미노산의 범위에 있다.

[0055] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZk)은 길이가 450 내지 600개의 아미노산(예를 들어, 450 내지 580, 480 내지 600, 480 내지 580, 또는 500 내지 600개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZk)은 길이가 480 내지 580개의 아미노산의 범위에 있다.

[0056] 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZl)은 길이가 350 내지 500개의 아미노산(예를 들어, 350 내지 450, 380 내지 450, 350 내지 420, 또는 380 내지 420개의 아미노산)의 범위에 있다. 일부 경우에, 대상 CasZ 단백질(예를 들어, CasZl)은 길이가 380 내지 420개의 아미노산의 범위에 있다.

[0057] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 하나 이상의 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 350 내지 800개의 아미노산(예를 들어, 350 내지 800, 350 내지 750, 350 내지 700, 350 내지 550, 450 내지 550, 450 내지 750, 450 내지 650, 또는 450 내지 550개의 아미노산)의 범위에 있다.

[0058] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZb 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZb 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZb 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또

는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZb 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZb 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZb 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZb 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 350 내지 700개의 아미노산(예를 들어, 350 내지 650, 또는 350 내지 620개의 아미노산)의 범위에 있다.

[0059] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZc 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZc 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZc 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZc 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZc 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZc 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZc 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 600 내지 800개의 아미노산(예를 들어, 600 내지 650, 또는 700 내지 800개의 아미노산)의 범위에 있다.

[0060] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZd 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZd 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZd 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZd 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZd 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZd 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZd 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 400 내지 650개의 아미노산(예를 들어, 400 내지 600, 400 내지 550, 500 내지 650, 500 내지 600 또는 500 내지 550개의 아미노산)의 범위에 있다.

[0061] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe 단

백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 450 내지 700개의 아미노산(예를 들어, 450 내지 650, 450 내지 615, 475 내지 700, 475 내지 650 또는 475 내지 615개의 아미노산)의 범위에 있다.

[0062] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZf 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZf 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZf 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZf 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZf 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZf 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZf 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 400 내지 750개의 아미노산(예를 들어, 400 내지 700, 700 내지 650, 400 내지 620, 400 내지 600, 400 내지 550, 400 내지 520, 400 내지 500, 400 내지 475, 415 내지 550, 415 내지 520, 415 내지 500, 또는 415 내지 475개의 아미노산)의 범위에 있다.

[0063] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZg 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZg 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZg 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZg 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZg 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZg 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZg 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 500 내지 750개의 아미노산(예를 들어, 500 내지 750개의 아미노산(예를 들어, 550 내지 750개의 아미노산))의 범위에 있다.

[0064] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZh 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZh 단백질과 50% 이상의 서열 동일성(예를 들어,

60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZh 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZh 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZh 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZh 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZh 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 380 내지 450개의 아미노산(예를 들어, 380 내지 420, 400 내지 450, 또는 400 내지 420개의 아미노산)의 범위에 있다.

[0065] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZi 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZi 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZi 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZi 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZi 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZi 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZi 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 700 내지 800개의 아미노산(예를 들어, 700 내지 750, 720 내지 800, 또는 720 내지 750개의 아미노산)의 범위에 있다.

[0066] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZj 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZj 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZj 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZj 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZj 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZj 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZj 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 600 내지 750개의 아미노산(예를 들어, 600 내지 700, 또는 650 내지 700개의 아미노산)의 범위에 있다.

[0067] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZk 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를

들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZk 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZk 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZk 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZk 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZk 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZk 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 450 내지 600개의 아미노산(예를 들어, 450 내지 580, 480 내지 600, 480 내지 580 또는 500 내지 600개의 아미노산)의 범위에 있다.

[0068] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZ1 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZ1 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZ1 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZ1 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZ1 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서)(예를 들어, 일부 경우에 CasZ 단백질이 dCasZ이 되도록) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZ1 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZ1 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 450 내지 600개의 아미노산(예를 들어, 450 내지 580, 480 내지 600, 480 내지 580 또는 500 내지 600개의 아미노산)의 범위에 있다.

[0069] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe, CasZf, CasZg 또는 CasZh 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe, CasZf, CasZg 또는 CasZh 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe, CasZf, CasZg 또는 CasZh 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe, CasZf, CasZg 또는 CasZh 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe, CasZf, CasZg 또는 CasZh 단백질 서열을 갖는 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe, CasZf, CasZg 또는 CasZh 단백질을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZe, CasZf, CasZg 또는 CasZh 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 350 내지 900개의 아미노산(예를 들어, 350 내지 850, 350 내지 800, 400 내지 900, 400 내지 850, 또는 400 내지 800개의 아미노산)의 범위에 있다.

[0070] 일부 경우에, (대상 조성물 및/또는 방법의) 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질 서열을 갖는 아미노산 서열을 포함한다. 일부 경우에, (예를 들어, 하나 이상의 촉매적 아미노산 위치에서) 서열이 단백질의 천연 유래 촉매적 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)을 포함하는 경우를 제외하고, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질을 갖는 아미노산 서열을 포함한다. 일부 경우에, 대상 CasZ 단백질은 도 1 또는 도 7의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK 또는 CasZl 단백질과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함하고, 길이가 350 내지 900개의 아미노산(예를 들어, 350 내지 850, 350 내지 800, 400 내지 900, 400 내지 850, 또는 400 내지 800개의 아미노산)의 범위에 있다.

[0071] ***CasZ 변이체***

[0072] 변이체 CasZ 단백질은 대응하는 야생형 CasZ 단백질의 아미노산 서열에 비교할 때 적어도 하나의 아미노산만큼 상이한 아미노산 서열을 가진다(예를 들어, 결실, 삽입, 치환, 융합을 가진다). 이중 가닥 표적 핵산 중 한 가닥을 절단하지만, 다른 한 가닥은 절단하지 않는 CasZ단백질은 본 명세서에서 "틈내기효소"(예를 들어, "틈내기효소 CasZ")로서 지칭된다. 뉴클레아제 활성이 실질적으로 없는 CasZ 단백질은 본 명세서에서 사멸 CasZ 단백질("dCasZ")로서 지칭된다(뉴클레아제 활성이 이하에 상세하게 기재되는 키메라 CasZ 단백질의 경우에 이종성 폴리펩타이드(융합 상대)에 의해 제공될 수 있다는 경고가 있음). 본 명세서에 기재된 임의의 CasZ 변이체 단백질(예를 들어, 틈내기효소 CasZ, dCasZ, 키메라 CasZ)에 대해, CasZ 변이체는 상기 기재된 것과 동일한 매개변수(예를 들어, 존재하는 도메인, 동일성 백분율, 길이 등)를 갖는 CasZ 단백질 서열을 포함할 수 있다.

[0073] ***변이체 - 촉매 활성***

[0074] 일부 경우에, CasZ 단백질은, 예를 들어, 천연 유래 촉매적으로 활성인 서열에 대해 돌연변이된 변이체 CasZ 단백질이고, 대응하는 천연 유래 서열에 비교할 때 감소된 절단 활성을 나타낸다(예를 들어, 90%, 이하, 80% 이하, 70% 이하, 60% 이하, 50% 이하, 40% 이하 또는 30% 이하의 절단 활성을 나타낸다). 일부 경우에, 이러한 변이체 CasZ 단백질은 촉매적으로 '사멸' 단백질이고(절단 활성이 실질적으로 없음) 'dCasZ'로서 지칭될 수 있다. 일부 경우에, 변이체 CasZ 단백질은 틈내기효소이다(이중 가닥 표적 핵산, 예를 들어, 이중 가닥 표적 DNA의 단지 하나의 가닥을 절단함). 본 명세서에서 더 상세하게 기재되는 바와 같이, 일부 경우에, CasZ 단백질(일부 경우에, 야생형 절단 활성을 갖는 CasZ 단백질 및 일부 경우에 감소된 절단 활성을 갖는 변이체 CasZ, 예를 들어, dCasZ 또는 틈내기효소 CasZ)은 융합 단백질(키메라 CasZ 단백질)을 형성하기 위해 관심 대상의 활성(예를 들어, 관심 대상의 촉매 활성)을 갖는 이종성 폴리펩타이드에 융합된다(접합된다).

[0075] CasZ의 촉매적 잔기는 CasZi.1에 따라 넘버링될 때 D405, E586 및 D684를 포함한다(예를 들어, 도 1 참조). 따라서, 일부 경우에, CasZ 단백질은 감소된 활성을 가지고, 상기 기재된 아미노산 중 하나 이상(또는 임의의 CasZ 단백질의 하나 이상의 대응하는 아미노산)이 돌연변이된다(예를 들어, 알라닌으로 치환된다). 일부 경우에, 변이체 CasZ 단백질은 촉매적으로 '사멸' 단백질이고(촉매적으로 비활성임), 'dCasZ'로서 지칭된다. dCasZ 단백질은 활성을 제공하는 융합 상대에 융합될 수 있고, 그리고 일부 경우에, dCasZ(예를 들어, 촉매 활성을 제공하지만, 진핵 세포에서 발현될 때 NLS를 가질 수 있는 융합 상대가 없는 하나)는 표적 DNA에 결합할 수 있고, 영상화를 위해 사용될 수 있고/있거나(예를 들어, 단백질은 태그/표지될 수 있음), 표적 DNA로부터의

전사로부터 RNA 중합효소를 차단할 수 있다. 일부 경우에, 변이체 CasZ 단백질은 틈내기효소이다(이중 가닥 표적 핵산, 예를 들어, 이중 가닥 표적 DNA의 단지 하나의 가닥을 절단함).

[0076] ***변이체 - 키메라 CasZ(즉, 융합 단백질)***

[0077] 상기 언급한 바와 같이, 일부 경우에, CasZ 단백질(일부 경우에, 야생형 절단 활성을 갖는 CasZ 단백질 및 일부 경우에 감소된 절단 활성을 갖는 변이체 CasZ, 예를 들어, dCasZ 또는 틈내기효소 CasZ)은 융합 단백질(키메라 CasZ 단백질)을 형성하기 위해 관심 대상의 활성(예를 들어, 관심 대상의 촉매 활성)을 갖는 이종성 폴리펩타이드에 융합된다(접합된다). CasZ 단백질이 융합될 수 있는 이종성 폴리펩타이드는 본 명세서에서 '융합 상대'로서 지칭된다.

[0078] 일부 경우에, 융합 상대는 표적 DNA의 전사를 조절할 수 있다(예를 들어, 전사를 저해, 전사를 증가시킬 수 있다). 예를 들어, 일부 경우에 융합 상대는 전사를 저해하는 단백질(또는 단백질로부터의 도메인)(예를 들어, 전사 리프레서, 전사 저해제 단백질의 보충을 통해 작용하는 단백질, 표적 DNA의 변형, 예컨대 메틸화, DNA 변형제의 보충, 표적 DNA와 회합된 히스톤의 조절, 히스톤 변형제, 예컨대 히스톤의 아세틸화 및/또는 메틸화를 변형시키는 것의 보충 등)이다. 일부 경우에 융합 상대는 전사를 증가시키는 단백질(또는 단백질로부터의 도메인)(예를 들어, 전사 활성체, 전사 활성체 단백질의 보충을 통해 작용하는 단백질, 표적 DNA의 변형, 예컨대 탈메틸화의 변형, DNA 변형제의 보충, 표적 DNA와 회합된 히스톤의 조절, 히스톤 변형제, 예컨대 히스톤의 아세틸화 및/또는 메틸화를 변형시키는 것의 보충 등)이다.

[0079] 일부 경우에, 키메라 CasZ 단백질은 표적 핵산을 변형시키는 효소 활성(예를 들어, 뉴클레아제 활성, 예컨대 FokI 뉴클레아제 활성, 메틸트랜스퍼라제 활성, 데메틸라제 활성, DNA 수선 활성, DNA 손상 활성, 탈아미노화 활성, 디스무타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 재조합효소 활성, 중합효소 활성, 리가제 활성, 헬리카제 활성, 광분해효소 활성 또는 글리코실라제 활성)을 갖는 이종성 폴리펩타이드를 포함한다.

[0080] 일부 경우에, 키메라 CasZ 단백질은 표적 핵산과 회합된 폴리펩타이드(예를 들어, 히스톤)를 변형하는 효소 활성(예를 들어, 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 데유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화(SUMOylating) 활성, 탈수모일화(deSUMOylating) 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성 또는 탈미리스토일화 활성)을 갖는 이종성 폴리펩타이드를 포함한다.

[0081] 전사 증가에서 사용될 수 있는 단백질(또는 이의 단편)의 예는 전사 활성체, 예컨대 VP16, VP64, VP48, VP160, p65 서브도메인(예를 들어, NFkB로부터 유래), 및 EDLL의 활성화 도메인 및/또는 TAL 활성화 도메인(예를 들어, 식물의 활성에 대해); 히스톤 라이신 메틸트랜스퍼라제, 예컨대 SET1A, SET1B, MLL1 내지 5, ASH1, SYMD2, NSD1 등; 히스톤 라이신 데메틸라제, 예컨대 JHDM2a/b, UTX, JMJD3 등; 히스톤 아세틸트랜스퍼라제, 예컨대 GCN5, PCAF, CBP, p300, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, SRC1, ACTR, P160, CLOCK 등; 및 DNA 데메틸라제, 예컨대 염색체 10번과 11번 사이의 전좌(Ten-Eleven Translocation: TET) 다이옥시게나제 1(TET1CD), TET1, DME, DML1, DML2, ROS1 등을 포함하지만, 이들로 제한되지 않는다.

[0082] 전사 감소에서 사용될 수 있는 단백질(또는 이의 단편)의 예는 전사 리프레서, 예컨대 쿠루펠(Kruppel) 관련 박스(KRAB 또는 SKD); KOX1 억제 도메인; Mad mSIN3 상호작용 도메인(SID); ERF 리프레서 도메인(ERD), SRDX 억제 도메인(예를 들어, 식물에서의 억제를 위해) 등; 히스톤 라이신 메틸트랜스퍼라제, 예컨대 Pr-SET7/8, SUV4-20H1, RIZ1 등; 히스톤 라이신 데메틸라제, 예컨대 JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY 등; 히스톤 라이신 데아세틸라제, 예컨대 HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 등; DNA 메틸라제, 예컨대 HhaI DNA m5c-메틸트랜스퍼라제(M.HhaI), DNA 메틸트랜스퍼라제 1(DNMT1), DNA 메틸트랜스퍼라제 3a(DNMT3a), DNA 메틸트랜스퍼라제 3b(DNMT3b), METI, DRM3(식물), ZMET2, CMT1, CMT2 (식물) 등; 및 주변 보충 요소, 예컨대 라민(Lamin) A, 라민 B 등을 포함하지만, 이들로 제한되지 않는다.

[0083] 일부 경우에, 융합 상대는 표적 핵산(예를 들어, ssRNA, dsRNA, ssDNA, dsDNA)을 변형시키는 효소 활성을 가진다. 융합 상대에 의해 제공될 수 있는 효소 활성의 예는 뉴클레아제 활성, 예컨대 제한 효소(예를 들어, FokI 뉴클레아제)에 의해 제공되는 것, 메틸트랜스퍼라제 활성, 예컨대 메틸트랜스퍼라제(예를 들어, HhaI DNA m5c-메틸트랜스퍼라제(M.HhaI), DNA 메틸트랜스퍼라제 1(DNMT1), DNA 메틸트랜스퍼라제 3a (DNMT3a), DNA 메틸트랜스퍼라제 3b(DNMT3b), METI, DRM3(식물), ZMET2, CMT1, CMT2(식물) 등)에 의해 제공되는 것; 데메틸라제 활성,

예컨대 데메틸라제(예를 들어, 염색체 10번과 11번 사이의 전좌(TET) 다이옥시게나제 1(TET1CD), TET1, DME, DML1, DML2, ROS1 등)에 의해 제공되는 것, DNA 수선 활성, DNA 손상 활성, 탈아미노화 활성, 예컨대 데아미나제(예를 들어, 사이토신 데아미나제 효소, 예컨대 래트 APOBEC1)에 의해 제공되는 것, 디스무타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 예컨대 인테그라제 및/또는 레솔바제(예를 들어, Gin 인버타제, 예컨대 Gin 인버타제의 과활성 돌연변이체, GinH106Y; 인간 면역결핍 바이러스 1형 인테그라제(IN); Tn3 레솔바제; 등)에 의해 제공되는 것, 트랜스포사제 활성, 재조합효소 활성, 예컨대 재조합효소(예를 들어, Gin 재조합효소의 촉매 도메인)에 의해 제공되는 것, 중합효소 활성, 리가제 활성, 헬리카제 활성, 광분해효소 활성, 및 글리코실라제 활성)를 포함하지만, 이들로 제한되지 않는다.

[0084] 일부 경우에, 융합 상대는 표적 핵산(예를 들어, ssRNA, dsRNA, ssDNA, dsDNA)과 회합되는 단백질(예를 들어, 히스톤, RNA 결합 단백질, DNA 결합 단백질 등)을 변형시키는 효소 활성을 가진다. 융합 상대에 의해 제공될 수 있는 (표적 핵산과 회합된 단백질을 변형시키는) 효소 활성의 예는 메틸트랜스퍼라제 활성, 예컨대 히스톤 메틸트랜스퍼라제(HMT)에 의해 제공되는 것(예를 들어, 얼룩 3-9 상동체 1의 억제자(SUV39H1, 또한 KMT1A로서 알려짐), 진정염색성 히스톤 라이신 메틸트랜스퍼라제 2(G9A, 또한 KMT1C 및 EHMT2로서 알려짐), SUV39H2, ESET/SETDB1 등, SET1A, SET1B, MLL1 내지 5, ASH1, SYMD2, NSD1, DOT1L, Pr-SET7/8, SUV4-20H1, EZH2, RIZ1), 데메틸라제 활성, 예컨대 히스톤 데메틸라제(예를 들어, 라이신 데메틸라제 1A(또한 LSD1로서 알려진 KDM1A), JHDM2a/b, JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, UTX, JMJD3 등)에 의해 제공되는 것, 아세틸트랜스퍼라제 활성, 예컨대 히스톤 아세틸라제 트랜스퍼라제에 의해 제공되는 것(예를 들어, 촉매 코어/인간 아세틸트랜스퍼라제 p300, GCN5, PCAF, CBP, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, HBO1/MYST2, HMOF/MYST1, SRC1, ACTR, P160, CLOCK 등의 단편), 데아세틸라제 활성, 예컨대 히스톤 데아세틸라제(예를 들어, HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 등)에 의해 제공되는 것, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화 활성, 탈수모일화 활성, 리보실화 활성, 데리보실화 활성, 미리스토일화 활성, 및 데미리스토일화 활성을 포함하지만, 이들로 제한되지 않는다.

[0085] 적합한 융합 상대의 추가적인 예는 다이하이드로엽산 환원효소(DHFR) 탈안정화 도메인(예를 들어, 화학적으로 제어 가능한 키메라 키메라 CasZ 단백질을 생성), 및 엽록체 수송 펩타이드이다. 적합한 엽록체 수송 펩타이드는 하기를 포함하지만, 이들로 제한되지 않는다:

MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGR VKCMQVWPPIGKKKFETLSYLPPLTRDSRA (서열번호 101); MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS (SEQ ID NO: 102); MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQVWPPIE KKKFETLSYLPDLTDSGGRVNC (서열번호 103); MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRPL KVMSSVSTAC (서열번호 104); MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRP LKVMSSVSTAC (서열번호 105); MAQINNMAQGIQTLNPNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKDSIFMQLFCSFRIS ASVATAC (서열번호 106); MAALVTSQLATSGTVLSVTDRFRRPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPHRFDRR CLSMVV (서열번호 107); MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQQRSV QRGSRRFPSVVVC (서열번호 108); MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQC (서열번호 109); MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISRSAAA A (서열번호 110); 및 MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTINGAAAT TNGASAASS (서열번호 111).

[0086]

[0087] 일부 경우에, 본 개시내용의 CasZ 융합 폴리펩타이드는 a) 본 개시내용의 CasZ 폴리펩타이드; 및 b) 엽록체 수송 펩타이드를 포함한다. 따라서, 예를 들어, CRISPR-CasZ 복합체는 엽록체에 표적화될 수 있다. 일부 경우에, 이 표적화는 엽록체 수송 펩타이드(CTP) 또는 색소체 수송 펩타이드로 불리는 N-말단 연장의 존재에 의해 달성될 수 있다. 발현된 폴리펩타이드가 식물 색소체(예를 들어, 엽록체)에서 구획화된다면, 박테리아 공급원으로부터의 염색체 이식유전자는 발현된 폴리펩타이드를 암호화하는 서열에 융합된 CTP 서열을 암호화하는 서열을 가져야 한다. 따라서, 엽록체에 대한 외인성 폴리펩타이드의 국소화는 종종 외인성 폴리펩타이드를 암호화하는 폴리뉴클레오타이드의 5' 영역에 CTP 서열을 암호화하는 폴리뉴클레오타이드 서열을 작동 가능하게 연결함으로써 1 달성된다. CTP는 색소체 내로 전좌 동안 가공 단계에서 제거된다. 그러나, 가공 효율은 펩타이드의 CTP의 서열 그리고 NH 2 말단에서 근처의 서열에 의해 영향받을 수 있다. 기재된 엽록체에 대한 표적화를 위한 다른 선택은 마이스 cab-m7 신호 서열(미국 특허 제7,022,896호, WO 97/41228) 완두콩 글루타티온 환원효소 신호 서열(WO 97/41228) 및 미국 특허 제2009029861호에 기재된 CTP이다.

[0088] 일부 경우에, 본 개시내용의 CasZ 융합 폴리펩타이드는 a) 본 개시내용의 CasZ 폴리펩타이드; 및 b) 엔도솜 탈출 펩타이드를 포함할 수 있다. 일부 경우에, 엔도솜 탈출 폴리펩타이드는 아미노산 서열 GLFXALLXLLXSLWXLLLXA(서열번호 112)를 포함하되, 각각의 X는 라이신, 히스티딘 및 알기닌으로부터 독립적으로 선택된다. 일부 경우에, 엔도솜 탈출 폴리펩타이드는 아미노산 서열 GLFHALLHLLHSLWHLLLHA (서열번호 113)를 포함한다.

[0089] (부위 특이적 표적 변형, 전사의 조절 및/또는 표적 단백질 변형, 예를 들어, 히스톤 변형에 대해) Cas9, 아연 핑거 및/또는 TALE 단백질과의 융합과 관련하여 사용되는 상기 융합 상대의 일부(그리고 더 많은) 예는, 예를 들어, 문헌[Nomura et al, J Am Chem Soc. 2007 Jul 18;129(28):8676-7; Rivenbark et al., Epigenetics. 2012 Apr;7(4):350-60; Nucleic Acids Res. 2016 Jul 8;44(12):5615-28; Gilbert et al., Cell. 2013 Jul 18;154(2):442-51; Kearns et al., Nat Methods. 2015 May;12(5):401-3; Mendenhall et al., Nat Biotechnol. 2013 Dec;31(12):1133-6; Hilton et al., Nat Biotechnol. 2015 May;33(5):510-7; Gordley et al., Proc Natl Acad Sci U S A. 2009 Mar 31;106(13):5053-8; Akopian et al., Proc Natl Acad Sci U S A. 2003 Jul 22;100(15):8688-91; Tan et., al., J Virol. 2006 Feb;80(4):1939-48; Tan et al., Proc Natl Acad Sci U S A. 2003 Oct 14;100(21):11997-2002; Papworth et al., Proc Natl Acad Sci U S A. 2003 Feb 18;100(4):1621-6; Sanjana et al., Nat Protoc. 2012 Jan 5;7(1):171-92; Beerli et al., Proc Natl Acad

Sci U S A. 1998 Dec 8;95(25):14628-33; Snowden et al., Curr Biol. 2002 Dec 23;12(24):2159-66; Xu et.al., Xu et al., Cell Discov. 2016 May 3;2:16009; Komor et al., Nature. 2016 Apr 20;533(7603):420-4; Chaikind et al., Nucleic Acids Res. 2016 Aug 11; Choudhury at. al., Oncotarget. 2016 Jun 23; Du et al., Cold Spring Harb Protoc. 2016 Jan 4; Pham et al., Methods Mol Biol. 2016;1358:43-57; Balboa et al., Stem Cell Reports. 2015 Sep 8;5(3):448-59; Hara et al., Sci Rep. 2015 Jun 9;5:11221; Piatek et al., Plant Biotechnol J. 2015 May;13(4):578-89; Hu et al., Nucleic Acids Res. 2014 Apr;42(7):4375-90; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; 및 Maeder et al., Nat Methods. 2013 Oct;10(10):977-9] 참조.

[0090] 추가적인 적합한 이종성 폴리펩타이드는 표적 핵산의 증가된 전사 및/또는 번역을 직접적으로 그리고/또는 간접적으로 제공하는 폴리펩타이드(예를 들어, 전사 활성체 또는 이의 단편, 전사 활성체를 보충하는 단백질 또는 이의 단편, 소분자/약물-반응성 전사 및/또는 번역 조절제, 번역-조절 단백질 등)를 포함하지만, 이들로 제한되지 않는다. 증가된 또는 감소된 전사를 달성하기 위한 이종성 폴리펩타이드의 비제한적 예는 전사 활성체 및 전사 리프레서 도메인을 포함한다. 일부 이러한 경우에, 키메라 CasZ 폴리펩타이드는 표적 핵산에서 특정 위치(즉, 서열)에 대해 가이드 핵산(가이드 RNA)에 의해 표적화되고, (예를 들어, 표적 핵산을 변형시키거나 또는 표적 핵산과 회합된 폴리펩타이드를 변형시키는 융합 서열이 사용될 때) 프로모터(전사 활성체 기능을 선택적으로 저해)에 대한 RNA 중합효소 결합을 차단하고/하거나 국소 염색질 상태를 변형시키는 것과 같은 좌위-특이적 조절을 발휘한다. 일부 경우에, 변화는 일시적이다(예를 들어, 전사 억제 또는 활성화). 일부 경우에, 변화는 유전성이다(예를 들어, 표적 핵산에 대해 또는 표적 핵산과 회합된 단백질, 예를 들어, 뉴클레오솜 히스톤에 대해 후성적 변형이 만들어질 때).

[0091] ssRNA 표적 핵산을 표적화할 때 사용하기 위한 이종성 폴리펩타이드의 비제한적 예는 스플라이싱 인자(예를 들어, RS 도메인); 단백질 번역 성분(예를 들어, 번역 개시, 신장 및/또는 방출 인자; 예를 들어, eIF4G); RNA 메틸라제; RNA 편집 효소(예를 들어, RNA 탈아미나제, 예를 들어, RNA 상에서 작용하는 아데노신 데아미나제(ADAR)(A 내지 I 및/또는 C 내지 U 편집 효소를 포함)); 헬리카제; RNA-결합 단백질 등을 포함한다(그러나 이들로 제한되지 않음). 이종성 폴리펩타이드는 전체 단백질을 포함할 수 있거나 또는 일부 경우에 단백질(예를 들어, 기능성 도메인)의 단편을 포함할 수 있다는 것이 이해된다.

[0092] 대상 키메라 CasZ 폴리펩타이드의 이종성 폴리펩타이드는 일시적이든 또는 비가역적이든, 직접적이든 또는 간접적이든; 엔도뉴클레아제(예를 들어, SMG5 및 SMG6과 같은 단백질로부터의 RNase III, CRR22 DYW 도메인, Dicer, 및 PIN(PilT N-말단) 도메인); RNA 절단의 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, CPSF, CstF, CFIm 및 CFIIm); 엑소뉴클레아제(예를 들어, XRN-1 또는 엑소뉴클레아제 T); 데아데닐라제(예를 들어, HNT3); 넌센스 매개된 RNA 붕괴를 초래하는 단백질 및 단백질 도메인(예를 들어, UPF1, UPF2, UPF3, UPF3b, RNP S1, Y14, DEK, REF2, 및 SRm160); RNA 안정화를 초래하는 단백질 및 단백질 도메인(예를 들어, PABP) ; 번역 억제를 초래하는 단백질 및 단백질 도메인(예를 들어, Ago2 및 Ago4); 번역 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, Staufen); 번역 조절을 초래하는(예를 들어, 초래할 수 있는) 단백질 및 단백질 도메인(예를 들어, 번역 인자, 예컨대 개시 인자, 신장 인자, 방출 인자 등, 예를 들어, eIF4G); RNA의 폴리아데닐화를 초래하는 단백질 및 단백질 도메인(예를 들어, PAP1, GLD-2, 및 Star- PAP) ; RNA의 폴리유리딘일화를 초래하는 단백질 및 단백질 도메인(예를 들어, CI D1 및 말단의 유리딜레이트 트랜스퍼라제); RNA 국소화를 초래하는 단백질 및 단백질 도메인(예를 들어, IMP1, ZBP1, She2p, She3p 및 Bicaudal-D로부터의); RNA의 핵 보유를 초래하는 단백질 및 단백질 도메인(예를 들어, Rrp6); RNA의 핵 수송을 초래하는 단백질 및 단백질 도메인(예를 들어, TAP, NXF1, THO, TREX, REF 및 Aly) ; RNA 스플라이싱의 억제를 초래하는 단백질 및 단백질 도메인(예를 들어, PTB, Sam68 및 hnRNP A1) ; RNA 스플라이싱의 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, 세린/알기닌-풍부(SR) 도메인); 전사 효율 감소를 초래하는 단백질 및 단백질 도메인(예를 들어, FUS (TLS)); 및 전사 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, CDK7 및 HIV Tat)을 포함하는 군으로부터 선택된 효과기 도메인을 포함하지만, 이들로 제한되지 않는, ssRNA와 상호작용할 수 있는 임의의 도메인(본 개시내용의 목적을 위해, 분자내 및/또는 분자간 2차 구조, 예를 들어, 이중-가닥 RNA 듀플렉스, 예컨대 헤어핀, 줄기-루프 등을 포함함)일 수 있다. 대안적으로, 효과기 도메인은 엔도뉴클레아제; RNA 절단을 자극할 수 있는 단백질 및 단백질 도메인; 엑소뉴클레아제; 데아데닐라제; 넌센스 매개된 RNA 붕괴 활성을 갖는 단백질 및 단백질 도메인; RNA를 안정화시킬 수 있는 단백질 및 단백질 도메인; 번역을 억제할 수 있는 단백질 및 단백질 도메인; 번역을 자극할 수 있는 단백질 및 단백질 도메인; 번역을 조절할 수 있는 단백질 및 단백질 도메인(예를 들어, 번역 인자, 예컨대 개시 인자, 신장 인자, 방출 인자 등, 예를 들어, eIF4G); RNA의 폴리아데닐화를 가능하게 하는 단백질 및 단백질 도메인; RNA의 폴리유리딘일화를 가능하게 하는 단백질 및 단백질 도메인; RNA 국소화 활성을

갖는 단백질 및 단백질 도메인; RNA의 핵 보유를 가능하게 하는 단백질 및 단백질 도메인; RNA 핵 수송 활성을 갖는 단백질 및 단백질 도메인; RNA 스플라이싱의 억제를 가능하게 하는 단백질 및 단백질 도메인; RNA 스플라이싱의 자극을 가능하게 하는 단백질 및 단백질 도메인; 전사 효율을 감소시킬 수 있는 단백질 및 단백질 도메인; 및 전사를 자극할 수 있는 단백질 및 단백질 도메인을 포함하는 군으로부터 선택될 수 있다. 다른 적합한 이종성 폴리펩타이드는 본 명세서에 전문이 참고로 편입된 WO2012068627에서 더 상세하게 기재되는 PUF RNA-결합 도메인이다.

[0093] 키메라 CasZ 폴리펩타이드에 대한 이종성 폴리펩타이드로서 사용될 수 있는(전체로 또는 이의 단편으로서) 일부 RNA 스플라이싱 인자는 별개의 서열-특이적 RNA 결합 분자 및 스플라이싱 효과기 도메인과 함께 모듈 조직화를 가진다. 예를 들어, 세린/알기닌-풍부(SR) 단백질 패밀리의 구성원은 엑손 포함을 촉진하는 프레-mRNA 및 C-말단의 RS 도메인에서 엑손 스플라이싱 인핸서(ESE)에 결합하는 N-말단의 RNA 인식 모티프(RRM)를 함유한다. 다른 예로서, hnRNP 단백질 hnRNP A1는 그의 RRM 도메인을 통해 엑손 스플라이싱 사일런서(ESS)에 결합하고 C-말단의 글리신-풍부 도메인을 통해 엑손 포함을 저해한다. 일부 스플라이싱 인자는 두 대안의 부위 사이의 조절 서열에 대한 결합에 의해 스플라이스 부위(ss)의 대안의 사용을 조절할 수 있다. 예를 들어, ASF/SF2는 ESE를 인식하고 인트론 근위 부위의 사용을 촉진시킬 수 있는 반면, hnRNP ESS에 결합하고 A1은 인트론 원위 부위의 사용에 대해 스플라이싱을 이동시킬 수 있다. 이러한 인자의 한 가지 적용은 내인성 유전자, 특히 질환 관련 유전자의 대안의 스플라이싱을 조절하는 ESF를 생성하는 것이다. 예를 들어, Bcl-x 프레-mRNA는 반대 기능의 단백질을 암호화하기 위해 2개의 대안의 5' 스플라이스 부위를 갖는 2개의 스플라이싱 아이소폼을 생성한다. 긴 스플라이싱 아이소폼 Bcl-xL은 장기간 생존된 유사분열후(postmitotic) 세포에서 발현된 강력한 세포자멸사 저해제이고, 다수의 암세포, 세포자멸사 신호에 대한 보호 세포에서 상향조절된다. 짧은 아이소폼 Bcl-xS는 세포자멸사전 아이소폼이고, 높음 전환율을 갖는 세포(예를 들어, 발생 중인 림프구)에서 고수준으로 발현된다. 2개의 Bcl-x 스플라이싱 아이소폼의 비는 코어 엑손 영역 또는 엑손 연장 영역 중 하나에(즉, 두 대안의 5' 스플라이스 부위 사이에) 위치된 다중 $c\acute{\omega}$-요소에 의해 조절된다. 더 많은 예에 대해, 본 명세서에 전문이 참고로 편입된 WO2010075303 참조.

[0094] 추가적인 적합한 융합 상대는 경계 요소인 단백질(또는 이의 단편)(예를 들어, CTCF), 주변 보충을 제공하는 단백질 및 이의 단편(예를 들어, 라민 A, 라민 B 등), 단백질 도킹 요소(예를 들어, FKBP/FRB, Pil1/Aby1 등)를 포함하지만, 이들로 제한되지 않는다.

[0095] 대상 키메라 CasZ 폴리펩타이드에 대한 다양한 추가적인 적합한 이종성 폴리펩타이드(또는 이의 단편)의 예는 다음의 요소에 기재된 것을 포함하지만, 이들로 제한되지 않는다(이들 간행물은 다른 CRISPR 엔도뉴클레아제, 예컨대 Cas9와 관련되지만, 기재된 융합 상대는 또한 CasZ 대신 사용될 수 있다): PCT 특허 출원: WO2010075303, WO2012068627, 및 WO2013155555, 그리고, 예를 들어 미국 특허 제8,906,616호; 제8,895,308호; 제8,889,418호; 제8,889,356호; 제8,871,445호; 제8,865,406호; 제8,795,965호; 제8,771,945; 8,697,359호; 및 미국 특허 출원 공개 제20140068797호; 제20140170753호; 제20140179006호; 제20140179770호; 제20140186843호; 제20140186919호; 제20140186958호; 제20140189896호; 제20140227787호; 제20140234972호; 제20140242664호; 제20140242699호; 제20140242700호; 제20140242702호; 제20140248702호; 제20140256046호; 제20140273037호; 제20140273226호; 제20140273230호; 제20140273231호; 제20140273232호; 제20140273233호; 제20140273234호; 제20140273235호; 제20140287938호; 제20140295556호; 제20140295557호; 제20140298547호; 제20140304853호; 제20140309487호; 제20140310828호; 제20140310830호; 제20140315985호; 제20140335063호; 제20140335620호; 제20140342456호; 제20140342457호; 제20140342458호; 제20140349400호; 제20140349405호; 제20140356867호; 제20140356956호; 제20140356958호; 제20140356959호; 제20140357523호; 제20140357530호; 제20140364333호; 및 제20140377868호에서 찾을 수 있으며; 이들 모두는 본 명세서에 그들의 전문이 참고로 편입된다.

[0096] 일부 경우에, 이종성 폴리펩타이드(융합 상대)는 세포하 국소화를 제공하고, 즉, 이종성 폴리펩타이드는 핵에 대한 표적화를 위한 세포하 국소화 서열(예를 들어, 핵에 대한 국소화를 위한 핵 국소화 신호(NLS), 핵 밖에서 융합 단백질을 유지하기 위한 서열, 예를 들어, 핵 수송 서열(NES), 세포질에 보유된 융합 단백질을 유지하기 위한 서열, 미토콘드리아에 대한 표적화를 위한 미토콘드리아 국소화 신호, 엽록체, ER 체류 신호 등에 대한 표적화를 위한 엽록체 국소화 신호)을 함유한다. 일부 실시형태에서, CasZ 융합 폴리펩타이드는 NLS를 포함하지 않으며, 따라서 단백질은 핵에 대해 표적화되지 않는다(이는, 예를 들어, 표적 핵산이 사이토졸에 존재하는 RNA일 때 유리할 수 있다). 일부 실시형태에서, 이종성 폴리펩타이드는 수송 및/또는 정제의 용이함을 위한 태그(즉, 이종성 폴리펩타이드는 검출 가능한 표지임)(예를 들어, 형광 단백질, 예를 들어, 녹색 형광 단백질(GFP),

YFP, RFP, CFP, mCherry, tdTomato 등; 히스티딘 태그, 예를 들어, 6XHis 태그; 혈구응집소(HA) 태그; FLAG 태그; Myc 태그; 등)을 제공할 수 있다.

[0097] 일부 경우에 CasZ 단백질(예를 들어, 야생형 CasZ 단백질, 변이체 CasZ 단백질, 키메라 CasZ 단백질, dCasZ 단백질, 키메라 CasZ 단백질, 여기서 CasZ 부분은 감소된 뉴클레아제 활성을 가짐 - 예컨대 융합상대에 융합된 dCasZ 단백질 등)은 핵 국소화 신호(NLS)(예를 들어, 일부 경우에 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다(이에 융합된다). 따라서, 일부 경우에, CasZ 폴리펩타이드는 1개 이상의 NLS(예를 들어, 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다. 일부 경우에, 1개 이상의 NLS(2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)는 N-말단 및/또는 C-말단에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, 1개 이상의 NLS(2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)는 N-말단에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, 1개 이상의 NLS(2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)는 C-말단에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, 1개 이상의 NLS(3개 이상, 4개 이상 또는 5개 이상의 NLS)는 N-말단과 C-말단 둘 다에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, NLS는 N-말단에 위치되고, NLS는 C-말단에 위치된다.

[0098] 일부 경우에 CasZ 단백질(예를 들어, 야생형 CasZ 단백질, 변이체 CasZ 단백질, 키메라 CasZ 단백질, dCasZ 단백질, 키메라 CasZ 단백질, 여기서 CasZ 부분은 감소된 뉴클레아제 활성 - 예컨대 융합 상대에 융합된 dCasZ 단백질 등을 가짐)은 1 내지 10개의 NLS(예를 들어, 1 내지 9, 1 내지 8, 1 내지 7, 1 내지 6, 1 내지 5, 2 내지 10, 2 내지 9, 2 내지 8, 2 내지 7, 2 내지 6 또는 2 내지 5개의 NLS)를 포함한다(이에 융합된다). 일부 경우에 CasZ 단백질(예를 들어, 야생형 CasZ 단백질, 변이체 CasZ 단백질, 키메라 CasZ 단백질, dCasZ 단백질, 키메라 CasZ 단백질, 여기서 CasZ 부분은 감소된 뉴클레아제 활성 - 예컨대 융합 상대에 융합된 dCasZ 단백질 등을 가짐)은 2 내지 5개의 NLS(예를 들어, 2 내지 4 또는 2 내지 3개의 NLS)를 포함한다(이에 융합된다).

[0099] NLS의 비제한적 예는 아미노산 서열 PKKKRKV (서열번호 114)를 갖는 SV40 바이러스 거대 T-항원의 NLS; 뉴클레오플라스민으로부터의 NLS(예를 들어, 서열 KRPAATKKAGQAKKKK(서열번호 115)을 갖는 뉴클레오플라스민 이분(bipartite) NLS)로부터 유래된 NLS 서열; 아미노산 서열 PAAKRVKLD (서열번호 116) 또는 RQRRNELKRSP(서열번호 117)를 갖는 c-myc NLS; 서열 NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(서열번호 118)을 갖는 hRNPA1 M9 NLS; 임포틴-알파로부터의 IBB 도메인의 서열 RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (서열번호 119); 근종 T 단백질의 서열 VSRKRPRP(서열번호 120) 및 PPKKARED (서열번호 121); 인간 p53의 서열 PQPKKKPL (서열번호 122); 마우스 c-abl IV의 서열 SALIKKKKKMAP (서열번호 123); 인플루엔자 바이러스 NS1의 서열 DRLRR (서열번호 124) 및 PKQKKRK (서열번호 125); 간염 바이러스 델타 항원의 서열 RKLKKKIKKL (서열번호 126); 마우스 Mx1 단백질의 서열 REKKKFLKRR (서열번호 127); 인간 폴리(ADP-리보스) 중합효소의 서열 KRKGDEVDGVDEVAKKKSKK (서열번호 128); 및 스테로이드 호르몬 수용체(인간) 글루코코르티코이드의 서열 RKCLQAGMNLEARKTKK (서열번호 129)를 포함한다. 일반적으로, NLS(또는 다중 NLS)는 진핵 세포의 핵에서 검출 가능한 양으로 CasZ 단백질의 축적을 유도하기에 충분한 강도를 가진다. 핵에서 축적의 검출은 임의의 적합한 기법에 의해 수행될 수 있다. 예를 들어, 검출 가능한 마커는 세포 내의 위치가 시각화될 수 있도록 CasZ 단백질에 융합될 수 있다. 세포 핵은 또한 세포로부터 단리될 수 있고, 이어서, 이의 내용은 단백질을 검출하기 위한 임의의 적합한 방법, 예컨대 면역조직화학, 웨스턴 블롯, 또는 효소 활성 분석에 의해 분석될 수 있다. 핵에서의 축적은 또한 간접적으로 결정될 수 있다.

[0100] 일부 경우에, CasZ 융합 폴리펩타이드는 "단백질 형질도입 도메인" 또는 PTD(또한 CPP - 세포 침투성 펩타이드로서 알려짐)를 포함하는데, 이는 지질 이중층, 마이셀, 세포막, 세포소기관 막 또는 소수포 막을 가로지르는 것을 용이하게 하는 폴리펩타이드, 폴리뉴클레오타이드, 탄수화물 또는 유기 또는 무기 화합물을 지칭한다. 거대 분자 및/또는 나노분자의 범위일 수 있는 다른 분자에 부착된 PTD는 분자가 막을 가로지르는 것, 예를 들어, 세포외 공간으로부터 세포내 공간 또는 사이토졸을 세포소기관 내로 이동하는 것을 용이하게 한다. 일부 실시형태에서, PTD는 아미노 말단 폴리펩타이드에 공유적으로 연결된다(예를 들어, 융합 단백질을 생성하기 위해 야생형 CasZ에 연결되거나, 또는 융합 단백질을 생성하기 위해 변이체 CasZ 단백질, 예컨대 dCasZ, 틈내기효소 CasZ, 또는 키메라 CasZ 단백질에 연결됨). 일부 실시형태에서, PTD는 폴리펩타이드의 카복실 말단에 공유적으로 연결된다(예를 들어, 융합 단백질을 생성하기 위해 야생형 CasZ에 연결되거나, 또는 융합 단백질을 생성하기 위해 변이체 CasZ 단백질, 예컨대 dCasZ, 틈내기효소 CasZ, 또는 키메라 CasZ 단백질에 연결됨). 일부 경우에, PTD는 적합한 삽입 부위에서 CasZ 융합 폴리펩타이드에서 내부로 삽입된다(즉, CasZ 융합 폴리펩타이드의 N- 또는 C-말단에서가 아님). 일부 경우에, 대상 CasZ 융합 폴리펩타이드는 1개 이상의 PTD(예를 들어, 2개 이상, 3

개 이상, 4개 이상의 PTD)를 포함한다(이에 접합되고, 융합된다). 일부 경우에, PTD는 핵 국소화 신호(NLS)(예를 들어, 일부 경우에 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다. 따라서, 일부 경우에, CasZ 융합 폴리펩타이드는 1개 이상의 NLS(예를 들어, 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다. 일부 경우에, PTD는 핵산(예를 들어, CasZ 가이드 핵산, CasZ 가이드 핵산을 암호화하는 폴리뉴클레오타이드, CasZ 융합 폴리펩타이드를 암호화하는 폴리뉴클레오타이드, 공여자 폴리뉴클레오타이드 등)에 공유적으로 연결된다. PTD의 예는 최소 운데카펩타이드 단백질 형질도입 도메인(YGRKKRRQRRR; 서열번호 130를 포함하는 HIV-1 TAT의 잔기 47 내지 57에 대응); 세포 내로 유입을 지시하기에 충분한 다수의 알기닌을 포함하는 폴리알기닌 서열(예를 들어, 3, 4, 5, 6, 7, 8, 9, 10 또는 10 내지 50개의 알기닌); VP22 도메인(Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); 초파리 안테나페디아(Antennapedia) 단백질 형질도입 도메인(Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); 절단된 인간 칼시토닌 펩타이드(Trehin et al. (2004) *Pharm. Research* 21:1248-1256); 폴리라이신(Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR (서열번호 131); 트랜스포탄 GWTLNSAGYLLGKINLKALAALAKKIL (서열번호 132); KALAWEAKLAKALAKALAKHLAKALAKALKCEA (서열번호 133); 및 RQIKIWFQNRRMKWKK (서열번호 134)을 포함하지만, 이들로 제한되지 않는다. 예시적인 PTD는 YGRKKRRQRRR (서열번호 130), RKKRRQRRR (서열번호 135); 3개의 알기닌 잔기 내지 50개의 알기닌 잔기의 알기닌 동종중합체를 포함하지만, 이들로 제한되지 않고; 예시적인 PTD 도메인 아미노산 서열은 다음 중 어떤 것을 포함하지만, 이들로 제한되지 않는다: YGRKKRRQRRR (서열번호 130); RKKRRQRR (서열번호 136); YARAAARQARA (서열번호 137); THRLPRRRRRR (서열번호 138); 및 GGRRARRRRRR (서열번호 139). 일부 실시형태에서, PTD는 활성화 가능한 CPP(ACPP)이다(Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPP는 매칭 절단 가능한 링커를 통해 다음이온(예를 들어, Glu9 또는 " E9")에 연결된 다양이온 CPP(예를 들어, Arg9 또는 " R9")를 포함하는데, 이는 순전하를 거의 0으로 감소시키고, 이에 의해 세포에 대한 접착 및 흡수를 저해한다. 링커의 절단 시, 다음이온이 방출되어, 다알기닌 및 그의 고유한 부착성을 국소로 가리움 해제하며, 따라서 막을 가로지르도록 ACPP를 "활성화시킨다".

[0101] ***링커(예를 들어, 융합 상대에 대해)***

[0102] 일부 예에서, 대상 CasZ 단백질은 링커 폴리펩타이드(예를 들어, 하나 이상의 링커 폴리펩타이드)를 통해 융합 상대에 융합된다. 링커 폴리펩타이드는 임의의 다양한 아미노산 서열을 가질 수 있다. 다른 화학적 결합이 제외되지는 않지만, 단백질은 일반적으로 가요성 특성의 스페이서 펩타이드에 의해 결합될 수 있다. 적합한 링커는 길이가 4개의 아미노산 내지 40개의 아미노산, 또는 길이가 4개의 아미노산 내지 25개의 아미노산인 폴리펩타이드를 포함한다. 이들 링커는 단백질을 결합하기 위해 합성, 링커-암호화 올리고뉴클레오타이드를 이용함으로써 생성될 수 있거나, 또는 융합 단백질을 암호화하는 핵산 서열에 의해 암호화될 수 있다. 일정 정도의 가요성이 있는 펩타이드 링커가 사용될 수 있다. 연결 펩타이드는 사실상 임의의 아미노산 서열을 가질 수 있고, 바람직한 링커는 일반적으로 가요성 펩타이드를 초래하는 서열을 가질 것임을 유념한다. 작은 아미노산, 예컨대 글리신 및 알라닌의 사용은 가요성 펩타이드를 생성하는 용도를 가진다. 이러한 서열의 생성은 당업자에게 일상적이다. 다양한 상이한 링커가 상업적으로 입수 가능하고, 사용에 적합한 것으로 고려된다.

[0103] 링커 폴리펩타이드의 예는 글리신 중합체$(G)_n$, 글리신-세린 중합체(예를 들어, $(GS)_n$, $GSGGS_n$ (서열번호 140), $GGSGGS_n$ (서열번호 141) 및 $GGGS_n$(서열번호 142)을 포함, 여기서 n은 적어도 하나의 정수임), 글리신-알라닌 중합체, 알라닌-세린 중합체를 포함한다. 예시적인 링커는 GGSG (서열번호 143), GGSGG (서열번호 144), GSGSG (서열번호 145), GSGGG (서열번호 146), GGGSG (서열번호 147), GSSSG(서열번호 148) 등을 포함하지만, 이들로 제한되지 않는 아미노산 서열을 포함할 수 있다. 당업자는 링커가 가요성 링커뿐만 아니라 덜 가요성인 구조를 부여하는 하나 이상의 부분을 포함할 수 있도록 임의의 목적으로 하는 요소에 접합된 펩타이드의 설계가 모두 또는 부분적으로 가요성인 링커를 포함할 수 있다는 것을 인식할 것이다.

[0104] ***검출 가능한 표지***

[0105] 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드는 검출 가능한 표지를 포함한다(예를 들어, 이에 부착/융합될 수 있다). 적합한 검출 가능한 표지 및/또는 검출 가능한 신호를 제공할 수 있는 모이어티는 효소, 방사성 동위원소, 특정 결합쌍의 구성원; 형광단; 형광 단백질; 양자점 등을 포함할 수 있지만, 이들로 제한되지 않는다.

[0106] 적합한 형광 단백질은 녹색 형광 단백질(GFP) 또는 이의 변이체, GFP의 청색 형광 변이체(BFP), GFP의 사이안 형광 변이체(CFP), GFP의 황색 형광 변이체(YFP), 증강된 GFP(EGFP), 증강된 CFP(ECFP), 증강된 YFP(EYFP), GFPS65T, 에메랄드, 토파즈(TYFP), 비너스, 시트린, 엠시트린(mCitrine), GFPuv, 불안정 EGFP (dEGFP), 불안정 ECFP (dECFP), 불안정 EYFP (dEYFP), mCFPm, 세룰리언, T-사파이어, CyPet, YPet, mKO, HcRed, t-HcRed,

DsRed, DsRed2, DsRed-단량체, J-Red, 이량체2, t-이량체2(12), mRFP1, 포실로포린, 레닐라 GFP, 몬스터 GFP, paGFP, Kaede 단백질 및 킨들링 단백질, 피코빌리단백질 및 피코빌리단백질 접합체(B-피코에리트린, R-피코에리트린 및 알로피코사이아닌을 포함)를 포함하지만, 이들로 제한되지 않는다. 형광 단백질의 다른 예는 mHoneydew, mBanana, mOrange, dTomato, tdTomato, mTangerine, mStrawberry, mCherry, mGrape1, mRaspberry, mGrape2, mPlum(Shaner et al. (2005) *Nat. Methods* 2:905-909) 등을 포함한다. 예를 들어, 문헌[Matz et al. (1999) *Nature Biotechnol.* 17:969-973]에 기재된 바와 같은 화충류(Anthozoan) 종으로부터의 임의의 다양한 형광 및 착색 단백질이 사용에 적합하다.

[0107] 적합한 효소는 겨자무과산화효소(HRP), 알칼리성 포스파타제(AP), 베타-갈락토시다제(GAL), 글루코스-6-인산 탈수소효소, 베타-N-아세틸글루코사미니다제, β글루쿠로니다제, 인버타제, 잔틴 옥시다제, 반딧불이 루시퍼라제, 글루코스 옥시다제(GO) 등을 포함하지만, 이들로 제한되지 않는다.

[0108] ***프로토스페이서 인접 모티프(PAM)***

[0109] 천연 CasZ 단백질은 DNA-표적화 RNA와 표적 DNA 사이의 상보성 영역에 의해 정해지는 표적 서열에서 표적 DNA에 결합한다. 다수의 CRISPR 엔도뉴클레아제에 대한 경우에서와 같이, 이중 가닥 표적 DNA의 부위-특이적 결합(및/또는 절단)은 (i) 가이드 RNA와 표적 DNA 사이의 염기쌍 상보성; 그리고 (ii) 표적 DNA 내 짧은 모티프[프로토스페이서 인접 모티프(PAM)로서 지칭됨]에 의해 결정된 위치에서 생긴다.

[0110] 일부 경우에, CasZ 단백질에 대한 PAM은 표적 DNA의 비상보성 가닥(또한 비표적 가닥으로서 지칭됨; 상보성 가닥은 가이드 RNA의 가이드 서열에 혼성화되는 한편, 비-상보성 가닥은 가이드 RNA에 직접 혼성화되지 않으며, 비-상보성 가닥의 역상보체임)의 표적 서열의 5' 바로 옆이다. 일부 경우에(예를 들어, CasZc에 대해), 비-상보성 가닥의 PAM 서열은 5'-TTA-3'이다. 일부 경우에(예를 들어, CasZb에 대해), 비-상보성 가닥의 PAM 서열은 5'-TTTN-3'이다. 일부 경우에(예를 들어, CasZb에 대해), 비-상보성 가닥의 PAM 서열은 5'-TTTA-3'이다.

[0111] 일부 경우에, 상이한 CasZ 단백질(즉, 다양한 종으로부터의 CasZ 단백질)은 상이한 CasZ 단백질의 다양한 효소 특징을 활용하기 위해(예를 들어, 상이한 PAM 서열 선호도에 대해; 증가된 또는 감소된 효소 활성에 대해; 세포 독성의 증가된 또는 감소된 수준에 대해; NHEJ, 상동성-직접 수선, 단일 가닥 파손, 이중 가닥 파손 등 사이의 균형 변화를 위해; 짧은 총 서열의 이점을 취하기 위해; 등) 다양한 제공된 방법으로 사용하는 것이 유리할 수 있다. 상이한 종으로부터의 CasZ 단백질은 표적 DNA에서 상이한 PAM 서열을 필요로 할 수 있다. 따라서, 선택의 특정 CasZ 단백질에 대해, PAM 서열 선호도는 상기 기재한 서열(들)과 상이할 수 있다. 적절한 PAM 서열의 확인을 위한 다양한 방법(인실리코 및/또는 습식 실험실 방법을 포함)은 당업계에 공지되어 있고, 일상적이며, 임의의 편리한 방법이 사용될 수 있다.

[0112] ***CasZ 가이드 RNA***

[0113] 리보뉴클레오단백질 복합체(RNP)를 형성하고, 표적 핵산(예를 들어, 표적 DNA) 내의 특정 위치에 대해 복합체를 표적화하는 CasZ 단백질에 결합하는 핵산은 본 명세서에서 "CasZ 가이드 RNA"로서 또는 단순히 "가이드 RNA"로서 지칭된다. 일부 경우에, CasZ 가이드 RNA가 RNA 염기에 추가로 DNA 염기를 포함하도록 혼성 DNA/RNA가 만들어질 수 있지만, 용어 "CasZ 가이드 RNA"는 여전히 본 명세서의 이러한 분자를 포함하기 위해 사용된다는 것이 이해되어야 한다.

[0114] CasZ 가이드 RNA는 2개의 세그먼트, 표적화 세그먼트 및 단백질-결합 세그먼트를 포함하는 것으로 언급될 수 있다. CasZ 가이드 RNA의 표적화 세그먼트는 표적 핵산(예를 들어, 표적 ssRNA, 표적 ssDNA, 이중 가닥 표적 DNA의 상보성 가닥 등) 내의 특정 서열(표적 부위)에 상보성인(그리고 따라서 이에 혼성화되는) 뉴클레오타이드 서열(가이드 서열)을 포함한다. 단백질-결합 세그먼트(또는 "단백질-결합 서열")는 CasZ 폴리펩타이드에 상호작용한다(이에 결합한다). 대상 CasZ 가이드 RNA의 단백질의 단백질-결합 세그먼트는 이중 가닥 RNA 듀플렉스(dsRNA 듀플렉스)을 형성하기 위해 서로 혼성화하는 뉴클레오타이드의 2개의 상보적 신장부를 포함한다. 표적 핵산(예를 들어, 게놈 DNA)의 부위-특이적 결합 및/또는 절단은 CasZ 가이드 RNA(CasZ 가이드 RNA의 가이드 서열)와 표적 핵산 사이의 염기쌍 상보성에 의해 결정되는 위치(예를 들어, 표적 좌위의 표적 서열)에서 일어날 수 있다.

[0115] CasZ 가이드 RNA 및 CasZ 단백질, 예를 들어, 융합 CasZ 폴리펩타이드는 복합체를 형성한다(예를 들어, 비공유 상호작용을 통해 결합한다). CasZ 가이드 RNA는 가이드 서열(표적 핵산의 서열에 상보성인 뉴클레오타이드 서열)을 포함하는 표적화 세그먼트를 포함함으로써 복합체에 표적 특이성을 제공한다. 복합체의 CasZ 단백질은 부위-특이적 활성(예를 들어, CasZ 단백질에 의해 제공된 절단 활성 및/또는 키메라 CasZ 단백질의 경우에 융합

상대에 의해 제공된 활성)을 제공한다. 다시 말해서, CasZ 단백질은 CasZ 가이드 RNA와 그의 회합 때문에 표적 핵산 서열(예를 들어, 표적 서열)로 가이드된다.

[0116] CasZ 가이드 RNA의 "표적화 서열"로서도 지칭되는 "가이드 서열"은 PAM 서열을 고려할 수 있다는 것을 제외하고 (예를 들어, 본 명세서에 기재된 바와 같이), CasZ 가이드 RNA가 임의의 목적으로 하는 표적 핵산의 임의의 목적으로 하는 서열에 대해 CasZ 단백질(예를 들어, 천연 유래 CasZ 단백질, 융합 CasZ 폴리펩타이드(키메라 CasZ) 등)를 표적화할 수 있도록 변형될 수 있다. 따라서, 예를 들어, CasZ 가이드 RNA는 진핵 세포 내 핵산, 예를 들어, 바이러스 핵산, 진핵 핵산(예를 들어, 진핵 염색체, 염색체 서열, 진핵 RNA 등) 등에서의 서열에 대해 상보성인(예를 들어, 혼성화할 수 있는) 가이드 서열을 가질 수 있다.

[0117] 일부 경우에, CasZ 가이드 RNA는 길이가 30개 이상의 뉴클레오타이드(nt)(예를 들어, 35개 이상의 nt, 40개 이상의 nt, 45개 이상의 nt, 50개 이상의 nt, 55개 이상의 nt 또는 60개 이상의 nt)이다. 일부 실시형태에서, CasZ 가이드 RNA는 길이가 40개 이상의 뉴클레오타이드(nt)(예를 들어, 45개 이상의 nt, 50개 이상의 nt, 55개 이상의 nt 또는 60개 이상의 nt)를 가진다. 일부 경우에, CasZ 가이드 RNA는 길이가 30개의 뉴클레오타이드(nt) 내지 100 nt(예를 들어, 30 내지 90, 30 내지 80, 30 내지 75, 30 내지 70, 30 내지 65, 40 내지 100, 40 내지 90, 40 내지 80, 40 내지 75, 40 내지 70, 또는 40 내지 65 nt)이다. 일부 경우에, CasZ 가이드 RNA는 길이가 40개의 뉴클레오타이드(nt) 내지 100개의 nt(예를 들어, 40 내지 90, 40 내지 80, 40 내지 75, 40 내지 70, 또는 40 내지 65개의 nt)이다.

[0118] ***CasZ 가이드 RNA의 가이드 서열***

[0119] 대상 CasZ 가이드 RNA는 표적 핵산 내 서열(표적 부위)에 대해 상보성은 뉴클레오타이드 서열인 가이드 서열(즉, 표적화 서열)을 포함한다. 다시 말해서, CasZ 가이드 RNA의 가이드 서열은 혼성화(즉, 염기쌍)를 통해 서열 특이적 방식으로 표적 핵산 (예를 들어, 이중 가닥 DNA (dsDNA), 단일 가닥 DNA(ssDNA), 단일 가닥 RNA(ssRNA) 또는 이중 가닥 RNA(dsRNA))과 상호작용할 수 있다. CasZ 가이드 RNA의 가이드 서열은 표적 핵산(예를 들어, 진핵 표적 핵산, 예컨대 게놈 DNA) 내에서 임의의 목적으로 하는 표적 서열(예를 들어, PAM을 고려하는 동안, 예를 들어, dsDNA 표적을 표적화할 때)에 혼성화하도록 변형(예를 들어, 유전자 조작에 의해)/설계될 수 있다.

[0120] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 60% 이상(예를 들어, 65% 이상, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 100% 이상이다.

[0121] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 표적 핵산의 7개의 인접한 가장 3'의 뉴클레오타이드에 대해 100%이다.

[0122] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 60% 이상 (예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 80% 이상 (예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 90% 이상 (예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상)의 인접한 뉴클레오타이드에 대해 100%이다.

[0123] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의(예를 들어, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 60% 이상 (예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의(예를 들어, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 80% 이상 (예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99%

이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의 (예를 들어, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 90% 이상 (예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의(예를 들어, 20개 이상, 21개 이상, 22개 이상)의 인접한 뉴클레오타이드에 대해 100%이다.

[0124] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타이드에 대해 60% 이상(예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타이드에 대해 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타이드에 대해 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타이드에 대해 100% 이상이다.

[0125] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 60% 이상(예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 100% 이상이다.

[0126] 일부 경우에, 가이드 서열은 길이가 17 내지 30개의 뉴클레오타이드(nt)(예를 들어, 17-25, 17-22, 17-20, 19-30, 19-25, 19-22, 19-20, 20-30, 20-25 또는 20-22 nt) 범위이다. 일부 경우에, 가이드 서열은 길이가 17 내지 25개의 뉴클레오타이드(nt)(예를 들어, 17-22, 17-20, 19-25, 19-22, 19-20, 20-25 또는 20-22 nt) 범위이다. 일부 경우에, 가이드 서열은 길이가 17개 이상의 nt(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상 또는 22개 이상의 nt; 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt 등)를 가진다. 일부 경우에, 가이드 서열은 길이가 19개 이상의 nt(예를 들어, 20개 이상, 21개 이상 또는 22개 이상의 nt; 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt 등)를 가진다. 일부 경우에, 가이드 서열은 길이가 17 nt이다. 일부 경우에, 가이드 서열은 길이가 18 nt이다. 일부 경우에, 가이드 서열은 길이가 19 nt이다. 일부 경우에, 가이드 서열은 길이가 20 nt이다. 일부 경우에, 가이드 서열은 길이가 21 nt이다. 일부 경우에, 가이드 서열은 길이가 22 nt이다. 일부 경우에, 가이드 서열은 길이가 23 nt이다.

[0127] ***CasZ 가이드 RNA의 단백질-결합 세그먼트***

[0128] 대상 CasZ 가이드 RNA의 단백질-결합 세그먼트는 CasZ 단백질과 상호작용한다. CasZ 가이드 RNA는 결합된 CasZ 단백질을 상기 언급한 가이드 서열을 통해 표적 핵산 내의 특정 뉴클레오타이드 서열로 가이드한다. CasZ 가이드 RNA의 단백질-결합 세그먼트는 서로에 대해 상보성이고 이중 가닥 RNA 듀플렉스(dsRNA 듀플렉스)을 형성하도록 혼성화되는 뉴클레오타이드의 2개의 신장부를 포함한다. 따라서, 단백질-결합 세그먼트는 dsRNA 듀플렉스를 포함한다.

[0129] 일부 경우에, dsRNA 듀플렉스 영역은 5-25개의 염기쌍(bp)(예를 들어, 5-22, 5-20, 5-18, 5-15, 5-12, 5-10, 5-8, 8-25, 8-22, 8-18, 8-15, 8-12, 12-25, 12-22, 12-18, 12-15, 13-25, 13-22, 13-18, 13-15, 14-25, 14-22, 14-18, 14-15, 15-25, 15-22, 15-18, 17-25, 17-22 또는 17-18 bp, 예를 들어, 5 bp, 6 bp, 7 bp, 8 bp, 9 bp, 10 bp 등)의 범위를 포함한다. 일부 경우에, dsRNA 듀플렉스 영역은 6 내지 15개의 염기쌍(bp)(예를 들어, 6-12, 6-10 또는 6-8 bp, 예를 들어, 6 bp, 7 bp, 8 bp, 9 bp, 10 bp 등)의 범위를 포함한다. 일부 경우에, 듀플렉스 영역은 5개 이상의 bp(예를 들어, 6개 이상, 7개 이상 또는 8개 이상의 bp)를 포함한다. 일부 경우에, 듀플렉스 영역은 6개 이상의 bp(예를 들어, 7개 이상 또는 8개 이상의 bp)를 포함한다. 일부 경우에, 듀플렉스 영역의 모든 뉴클레오타이드가 짝지어지지는 않으며, 따라서 듀플렉스 형성 영역은 벌지(bulge)를 포함할 수 있다. 본 명세서의 용어 "벌지"는 이중 가닥 듀플렉스에 기여하지 않지만, 기여하는 뉴클레오타이드에 의해 5' 및 3'을 둘러싸는 뉴클레오타이드의 신장부(하나의 뉴클레오타이드 또는 다중 뉴클레오타이드일 수 있음)를 의미하기 위해 사용되며, 이러한 벌지는 듀플렉스 영역의 부분으로 고려된다. 일부 경우에, dsRNA는 1개

이상의 벌지(예를 들어, 2개 이상, 3개 이상, 4개 이상의 벌지)를 포함한다. 일부 경우에, dsRNA 듀플렉스는 2개 이상의 벌지(예를 들어, 3개 이상, 4개 이상의 벌지)를 포함한다. 일부 경우에, dsRNA 듀플렉스는 1 내지 5개의 벌지(예를 들어, 1-4, 1-3, 2-5, 2-4 또는 2-3개의 벌지)를 포함한다.

[0130] 따라서, 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 70% 내지 100% 상보성(예를 들어, 75% 내지 100%, 80% 내지 10%, 85% 내지 100%, 90% 내지 100%, 95% 내지 100% 상보성)을 가진다. 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 70% 내지 100% 상보성(예를 들어, 75% 내지 100%, 80% 내지 10%, 85% 내지 100%, 90% 내지 100%, 95% 내지 100% 상보성)을 가진다. 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 85% 내지 100% 상보성(예를 들어, 90% 내지 100%, 95% 내지 100% 상보성)을 가진다. 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 70% 내지 95% 상보성(예를 들어, 75% 내지 95%, 80% 내지 95%, 85% 내지 95%, 90% 내지 95% 상보성)을 가진다.

[0131] 다시 말해서, 일부 실시형태에서, dsRNA 듀플렉스는 서로 70% 내지 100% 상보성(예를 들어, 75% 내지 100%, 80% 내지 10%, 85% 내지 100%, 90% 내지 100%, 95% 내지 100% 상보성)을 갖는 뉴클레오타이드의 2개 신장부를 포함한다. 일부 경우에, dsRNA 듀플렉스는 서로 85% 내지 100% 상보성(예를 들어, 90% 내지 100%, 95% 내지 100% 상보성)을 갖는 뉴클레오타이드의 2개 신장부를 포함한다. 일부 경우에, dsRNA 듀플렉스는 서로 70% 내지 95% 상보성(예를 들어, 75% 내지 95%, 80% 내지 95%, 85% 내지 95%, 90% 내지 95% 상보성)을 갖는 뉴클레오타이드의 2개 신장부를 포함한다.

[0132] 대상 CasZ 가이드 RNA의 이중가닥 영역은 천연 유래 이중가닥 영역에 대해 1개 이상의 (1, 2, 3, 4, 5개 등) 돌연변이를 포함할 수 있다. 예를 들어, 일부 경우에 염기쌍은 유지될 수 있는 한편 각각의 세그먼트로부터의 염기쌍에 기여하는 뉴클레오타이드는 상이할 수 있다. 일부 경우에, 대상 CasZ 가이드 RNA의 이중가닥 영역은 (천연 유래 CasZ 가이드 RNA의) 천연 유래 이중가닥 영역에 비교할 때 더 짝지어진 염기, 덜 짝지어진 염기, 더 작은 벌지, 더 큰 벌지, 소수의 벌지, 더 많은 벌지, 또는 이들의 임의의 편리한 조합을 포함한다.

[0133] 다양한 Cas9 가이드 RNA 및 cpf1 가이드 RNA의 예는 당업계에서 찾을 수 있고, 그리고 일부 경우에 Cas9 가이드 RNA 내로 도입된 것과 유사한 변형이 또한 본 개시내용의 CasZ 가이드 RNA 내로 도입될 수 있다(예를 들어, dsRNA 듀플렉스 영역에 대한 돌연변이, 다른 단백질과의 상호작용을 제공하기 위해 추가된 안정성을 위한 5' 또는 3' 말단의 연장 등). 예를 들어, 문헌[Jinek et al., Science. 2012 Aug 17;337(6096):816-21; Chylinski et al., RNA Biol. 2013 May;10(5):726-37; Ma et al., Biomed Res Int. 2013;2013:270805; Hou et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9; Jinek et al., Elife. 2013;2:e00471; Pattanayak et al., Nat Biotechnol. 2013 Sep;31(9):839-43; Qi et al, Cell. 2013 Feb 28;152(5):1173-83; Wang et al., Cell. 2013 May 9;153(4):910-8; Auer et al., Genome Res. 2013 Oct 31; Chen et al., Nucleic Acids Res. 2013 Nov 1;41(20):e19; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; Cho et al., Genetics. 2013 Nov;195(3):1177-80; DiCarlo et al., Nucleic Acids Res. 2013 Apr;41(7):4336-43; Dickinson et al., Nat Methods. 2013 Oct;10(10):1028-34; Ebina et al., Sci Rep. 2013;3:2510; Fujii et. al, Nucleic Acids Res. 2013 Nov 1;41(20):e187; Hu et al., Cell Res. 2013 Nov;23(11):1322-5; Jiang et al., Nucleic Acids Res. 2013 Nov 1;41(20):e188; Larson et al., Nat Protoc. 2013 Nov;8(11):2180-96; Mali et. at., Nat Methods. 2013 Oct;10(10):957-63; Nakayama et al., Genesis. 2013 Dec;51(12):835-43; Ran et al., Nat Protoc. 2013 Nov;8(11):2281-308; Ran et al., Cell. 2013 Sep 12;154(6):1380-9; Upadhyay et al., G3 (Bethesda). 2013 Dec 9;3(12):2233-8; Walsh et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15514-5; Xie et al., Mol Plant. 2013 Oct 9; Yang et al., Cell. 2013 Sep 12;154(6):1370-9; Briner et al., Mol Cell. 2014 Oct 23;56(2):333-9]; 및 미국 특허 및 특허 출원: 8,906,616호; 제8,895,308호; 제8,889,418호; 제8,889,356호; 제8,871,445호; 제8,865,406호; 제8,795,965호; 제8,771,945호; 제8,697,359호; 제20140068797호; 제20140170753호; 제20140179006호; 제20140179770호; 제20140186843호; 제20140186919호; 제20140186958호; 제20140189896호; 제20140227787호; 제20140234972호; 제20140242664호; 제20140242699호; 제20140242700호; 제20140242702호; 제20140248702호; 제20140256046호; 제20140273037호; 제20140273226호; 제20140273230호; 제20140273231호; 제20140273232호; 제20140273233호; 제20140273234호; 제20140273235호; 제20140287938호; 제20140295556호; 제20140295557호; 제20140298547호; 제20140304853호; 제20140309487호; 제20140310828호; 제20140310830호; 제20140315985호; 제20140335063호; 제20140335620호; 제20140342456호; 제20140342457호; 제20140342458호; 제20140349400호; 제20140349405호; 제20140356867호; 제20140356956호; 제20140356958호; 제20140356959호; 제20140357523호; 제20140357530호; 제20140364333호; 및 20140377868호를

참조하며; 이들 모두는 본 명세서에 그들의 전문이 참고로 편입된다.

[0134] CasZ 가이드 RNA는 dsRNA 듀플렉스를 형성하기 위해 단백질-결합 세그먼트에 혼성화되는 뉴클레오타이드의 가이드 서열과 2개의 신장부("듀플렉스-형성 세그먼트")을 둘 다 포함한다. 주어진 CasZ 가이드 RNA의 특정 서열은 crRNA가 발견된 종의 특징일 수 있다. 적합한 CasZ 가이드 RNA의 예는 본 명세서에 제공된다.

[0135] *가이드 RNA 서열의 예*

[0136] 자연적으로 존재하는 CasZ 단백질(예를 들어, 도 1 및 도 7 참조)에 대한 crRNA의 반복부 서열(CasZ 가이드 RNA의 비-가이드 서열)은 표 1 및 표 3에 나타낸다.

표 1

CasZ 단백질에 대한 crRNA 반복부 서열

CasZ 단백질	반복부 서열	서열번호
Za.1	GTTGCATTCCTTCATTCGTCTATTCGGGTTCTGCAAC	51
Za.2	GTTGCATTCCTTCATTCGTCTATCCGGGTTCTGCAAG	52
Za.3	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC	53
Za.4	CTATCATATTCAGAACAAAGGGATTAAGGAATGCAAC	54
Za.5	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55
Za.6	GTCTACAACTCATTGATAGAAATCAATGAGTTAGACA	56
Za.7	GTTATAAAGGCGGGGATCGCGACCGAGCGATTGAAAG	57
Zb.1	GTTGCATTCCTTAATTCATTTTCTCAATATCGGAAAC	58
Zb.2	GTTGCAGAAATAGAATAAAGGAATTAAGGAATGCAAC	59
Zb.3	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55
Zb.4	ATTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	61
Zb.5	GTTTCAGCGCACGAATTAACGAGATGAGAGATGCAACT	62
Zb.6	CTTGCAGAAGCTGAATAGACGAATCAAGGAATGCAAC	63
Zb.7	CACTTGCAGGCCTTGAATAGAGGAGTTAAGGAATGCAAC	64
Zb.8	GTCTCCATGACTGAAAAGTCGTGGCCGAATTGAAAC	65
Zb.9	GTTGCAGCGCCCGAACTGACGAGACGAGAGATGCAAC	66
Zb.10	GTTGCGCGAATAGAATAAAGGAATTAAGGAATGCAAC	67
Zb.11	AGTTGCATTCCTTAATCCCTCTGTTCAGTTTGTGCAAT	68
Zc.1	GTTGCATTCCTAGTTTCTCTAATTAGCACTGTGCAAC	69
Zc.2	GTTGCGGCGCGCGAATAAACGAGACTAGGAATGCAAC	70
Zc.3	ACTAGTTGCATTCCTTAATCCCTTTGTTCTGAATATGCTAG	71
Zc.4	CTTTCATATTCAGAACAAAGGGATTAAGGAATGCAAC	72
Zc.5	GTTGCAGTCCTTAACCCCTAGTTTCTGAATATGAAAGAT	73
Zc.6	GTTGCAGCCCCCGAACTAACGAGATGAGAGATGCAAC	74
Zc.7	CTTGCAGAACAATCATATATGACTAATCAGACTGCAAC	75
Zd.1	GTTGCACTCACCGGTGCTCACGACGTAGGGATGCAAC	76
Zd.2	GTCCCTACTCGCTAGGGAAACTAATTGAATGGAAAC	77
Ze.1	GTTGCATTCGGGTGCAAAACAGGGAGTAGAGTGTAAC	78
Ze.2	CTTCCAAACTCGAGCCAGTGGGGAGAGAAGTGGCA	79
Ze.3	CCTGTAGACCGGTCTCATTCTGAGAGGGGTATGCAACT	80
Ze.4	GTCTCGAGACCCTACAGATTTTGGAGAGGGGTGGGAC	81
Ze.4b	GTCCCACCCCTCTCCAAAATCTGTAGGGTCTCGAGAC	82
Zf.1	GTAGCAGGACTCTCCTCGAGAGAAACAGGGGTATGCT	83
Zf.2	GTACAATACCTCTCCTTTAAGAGAGGGAGGGGTACGCTAC	84
Zf.3	CCCCCTCGTTTCCTTCAGGGGATTCCTTTCC	85
Zg.1	GGTTCCCCCGGGCGCGGGTGGGGTGGCG	86
Zg.2	GGCTGCTCCGGGTGCGCGTGGAGCGAGG	87

[0137]

CasZ 단백질	반복부 서열	서열번호
Zh.1	GTTTTATACCCTTTAGAATTTAAACTGTCTAAAAG	88
Zi.1	ATTGCACCGGCCAACGCAAATCTGATTGATGGACAC	89
Zi.2	GCCGCAGCGGCCGACGCGGCCCTGATCGATGGACAC	90
Zj.1	GTCGAAATGCCCGCGCGGGGGCGTCGTACCCGCGAC	91
Zk.1	GGCTAGCCCGTGCGCGCAGGGACGAGTGG	92
Zk.2	GCCCGTGCGCGCAGGGACGAGTGG	93
Zk.3	GTTGCAGCGGCCGACGGAGCGCGAGCGTGGATGCCAC	94
Zk.4	CCATCGCCCCGCGCGCACGTGGATGAGCC	95
Zl.1	CTTTAGACTTCTCCGGAAGTCGAATTAATGGAAAC	96
Zl.2	GGGCGCCCCGCGCGAGCGGGGGTTGAAG	97
Za.8	CTTGCAGAACCCGGATAGACGAATGAAGGAATGCAAC	295
Zb.12	CTTGCAGGCCTTGAATAGAGGAGTTAAGGAATGCAAC	296
Zb.13	GTTGCACAGTGCTAATTAGAGAAACTAGGAATGCAAC	297
Zb.14	CTAGCATATTCAGAACAAAGGGATTAAGGAATGCAAC	298
Zb.15	CTTTCATATTCAGAAACTAGGGGTTAAGGACTGCAAC	299
Zc.8	GTTGCATCCCTACGTCGTGAGCACCGGTGAGTGCAAC	300
Ze.5	GGAAAGGAATCCCCTGAAGGAAACGAGGGGG	301
Zg.3	GTGTCCATCAATCAGATTTGCGTTGGCCGGTGCAAT	302
Zb.16	GTTTCAGCGCACGAATTAACGAGATGAGAGATGCAAC	303
Zj.2	CTTTTAGACAGTTTAAATTCTAAAGGGTATAAAAC	307

[0138]

[0139] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0140] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh 또는 CasZi crRNA crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, 또는 CasZi crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, 또는 CasZi crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh 또는 CasZi crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0141] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZa crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0142] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZb crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또

는 표 3의 CasZb crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZb crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZb crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0143] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZc crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZc crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZc crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZc crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0144] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZd crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZd crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZd crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZd crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0145] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZe crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZe crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZe crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZe crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0146] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZf crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZf crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZf crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZf crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0147] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZg crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZg crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZg crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZg crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0148] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZh crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또

는 표 3의 CasZh crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZh crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZh crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0149] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZi crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZi crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZi crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZi crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0150] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0151] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZk crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZk crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZk crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZk crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0152] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZl crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZl crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZl crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZl crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0153] 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj, CasZl 또는 CasZk crRNA 서열을 (예를 들어, 가이드 서열에 추가로, 예를 들어, 단백질-결합 영역의 부분으로서) 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj, CasZl 또는 CasZk crRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj, CasZl 또는 CasZk crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ 가이드 RNA는 표 1 또는 표 3의 CasZj, CasZl 또는 CasZk crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0154] ***CasZ 전사촉진 비암호화 RNA(trancRNA)***

[0155] 본 개시내용의 조성물 및 방법은 CasZ 전사촉진 비암호화 RNA("trancRNA"; 또는 본 명세서에서 "CasZ trancRNA"로서 지칭됨)를 포함한다. 일부 경우에, trancRNA는 본 개시내용의 CasZ 폴리펩타이드 및 CasZ 가이드 RNA와 복합체를 형성한다. trancRNA는 CasZ 좌위에 존재하는 뉴클레오타이드 서열에 의해 암호화된 고도로 전사된 RNA로서 확인될 수 있다. trancRNA를 암호화하는 서열은 보통 *cas* 유전자와 CasZ 좌위의 어레이(반복부) 사이에 위치된다(예를 들어, 반복부 서열에 인접하여 위치될 수 있다). 이하의 예는 CasZ trancRNA의 검출을 입증한다. 일부 경우에, CasZ trancRNA는 CasZ 폴리펩타이드와의 공동-면역침강 반응을 보인다(복합체를 형성한다). 일부 경우에, CasZ trancRNA의 존재는 시스템의 기능에 필요하다. trancRNA와 관련된 데이터(예를 들어, 천연 유래 어레이 상에서 이들의 발현 및 이들의 위치)는 이하의 실시예 부문에 제시한다.

[0156] 일부 경우에, CasZ trancRNA는 길이가 60개의 뉴클레오타이드(nt) 내지 270 nt(예를 들어, 60 내지 260, 70 내지 270, 70 내지 260, 또는 75 내지 255개의 nt)이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZa trancRNA)는 길이가 60 내지 150 nt(예를 들어, 60 내지 140, 60 내지 130, 65 내지 150, 65 내지 140, 65 내지 130, 70 내지 150, 70 내지 140, 또는 70 내지 130 nt)이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZa trancRNA)는 길이가 70 내지 130 nt이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZa trancRNA)는 길이가 약 80 nt이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZa trancRNA)는 길이가 약 90 nt이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZa trancRNA)는 길이가 약 120 nt이다.

[0157] 일부 경우에, CasZ trancRNA(예를 들어, CasZb trancRNA)는 길이가 85 내지 240 nt(예를 들어, 85 내지 230, 85 내지 220, 85 내지 150, 85 내지 130, 95 내지 240, 95 내지 230, 95 내지 220, 95 내지 150, 또는 95 내지 130개의 nt)이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZb trancRNA)는 길이가 95 내지 120 nt이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZb trancRNA)는 길이가 약 105 nt이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZb trancRNA)는 길이가 약 115 nt이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZb trancRNA)는 길이가 약 215 nt이다.

[0158] 일부 경우에, CasZ trancRNA(예를 들어, CasZc trancRNA)는 길이가 80 내지 275 nt(예를 들어, 85 내지 260 nt)이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZc trancRNA)는 길이가 80 내지 110 nt(예를 들어, 85 내지 105 nt)이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZc trancRNA)는 길이가 235 내지 270 nt(예를 들어, 240 내지 260 nt)이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZc trancRNA)는 길이가 약 95 nt이다. 일부 경우에, CasZ trancRNA(예를 들어, CasZc trancRNA)는 길이가 약 250 nt이다.

[0159] ***예시적인 trancRNA 서열***

[0160] 자연적으로 존재하는 CasZ 단백질에 대한 trancRNA 서열의 예를 표 2에 나타낸다.

표 2

CasZ trancRNA 서열

CasZ 단백질	trancRNA 서열	서열 번호
Za.1	CGATTCCTCCCTACAGTAGTTAGGTATAGCCGAAAGGTAG AGACTAAATCTGTAGTTGGAGTGGGCCGCTTGCATCGGCC	151
Za.2	TCGTCTCGAGGGTTACCAAAATTGGCACTTCTCGACTTTAG GCCGATGCAAGCGGCCCACTCCACTACAGATTTAGTCTCTA CCTTGCGGCTATACCTAACTTACTGTAGGGAGGAATCGTG	152
Za.3	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGT CCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTG CATCAGCCTAA	153
Zb.2	CAGAATAATACTGACTTACTAAGATATCTTGAGGGTATACC CGAAAAGATTGGCGTTGTTGCAACGCAATAAGATGTAAAT CTGAAAAGGTTTGGAATCATATAAATAATTTTA	154
Zb.4	AAGCCAAGATATGGAATGCCATTGTAATATTATGGTGTTG ACTTAGTTTAGATTTAAACAATCTTCGATGGCTATATGCGG AAGGTTTGGCGTCGTTGTAACGC	155
Zb.6	CAGTGTGCATAGCTATAACACTACGCAAAGACTGCTAAAG AGCGATGTGCTCTATCGCAGTCTCACCTTTAATGGACTTAC GGATCTTTTGGAGCACTAAGCTCCGCTGCGGTGCAACACC GCCCTTTTCTTGCCTCTGCTTGCCCTTTCCGGTTATTATAGC CGGGAGAGTGCGGAAGATTACCGCTCTAGCTCGCAGCATG TTACTGAGTC	156
Zc.3	GCAAGTCATTCGGGGACACTTTTTGTTATTTAAAGTGTTTT AGATAAATCAGTGTCATGCTGAATAACGACCCGACCTATA AATAACATAATCC	157
Zc.5	GTCCTTAAGGTACTACACATTACATGTGAACGTGGAGCTA ATAATAGAAATATTATTAGACTACACCTTATTAATAACGGT AGGAGATCTATATGGTCTTGAATGGAATAGTAATTGTGAA ATTATAATTTCTGTTCTTAGCTACTTAAGATGGCTCGTTGC AAGCCACTCGGGGGCTCTCTTGAAGTCAAAGAGCTTTAGA CAAATCAGTGTCAAACTGAATAACGACCCGACCATGACTT CATAATCCCG	158

[0161]

[0162] 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa trancRNA 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa trancRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa trancRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함하고, 길이가 60 내지 150 nt(예를 들어, 60 내지 140, 60 내지 130, 65 내지 150, 65 내지 140, 65 내지 130, 70 내지 150, 70 내지 140, 또는 70 내지 130 nt)이다.

[0163] 일부 경우에, 대상 CasZ trancRNA는 상기 CasZb trancRNA 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZb trancRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZb trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZb trancRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZb trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함하고, 길이가 85 내지 240 nt(예를 들어, 85 내지 230, 85 내지 220, 85 내지 150, 85 내지 130, 95 내지 240, 95 내지 230, 95 내지 220, 95 내지 150, 또는 95 내지 130 nt)이다.

[0164] 일부 경우에, 대상 CasZ trancRNA는 상기 CasZc trancRNA 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZc trancRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93%

이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZc trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZc trancRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZc trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함하고, 길이가 80 내지 110 nt(예를 들어, 85 내지 105 nt) 또는 235 내지 270 nt(예를 들어, 240 내지 260 nt)이다.

[0165] 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa, CasZb 또는 CasZc trancRNA 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa, CasZb 또는 CasZc trancRNA 서열과 70% 이상의 동일성(예를 들어, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa, CasZb 또는 CasZc trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa, CasZb 또는 CasZc trancRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasZ trancRNA는 상기 CasZa, CasZb 또는 CasZc trancRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함하고, 길이가 60개의 뉴클레오타이드(nt) 내지 270 nt(예를 들어, 60 내지 260, 70 내지 270, 70 내지 260, 또는 75 내지 255 nt)이다.

[0166] 일부 경우에, CasZ trancRNA는 변형된 뉴클레오타이드를 포함한다(예를 들어, 메틸화된다). 일부 경우에, CasZ trancRNA는: i) 염기 변형 또는 치환; ii) 골격 변형; iii) 변형된 뉴클레오타이드간 결합; 및 iv) 변형된 당 모이어티 중 하나 이상을 포함한다. 가능한 핵산 변형을 이하에 기재한다.

[0167] **CasZ 시스템**

[0168] 본 개시내용은 CasZ 시스템을 개시한다. 본 개시내용의 CasZ 시스템은 하기 중 하나 이상을 포함할 수 있다: (1) CasZ 전사촉진 비암호화 RNA(trancRNA)(본 명세서에서 "CasZ trancRNA"로서 지칭됨) 또는 CasZ trancRNA를 암호화하는 핵산(예를 들어, 발현 벡터); (2) CasZ 단백질(예를 들어, 야생형 단백질, 변이체, 촉매적으로 손상된 변이체, CasZ 융합 단백질 등) 또는 CasZ 단백질을 암호화하는 핵산(예를 들어, RNA, 발현 벡터 등); 및 (3) CasZ 가이드 RNA(CasZ 단백질에 결합하고 이에 서열 특이성을 제공함, 예를 들어, 진핵 게놈의 표적 서열에 결합할 수 있는 가이드 RNA) 또는 CasZ 가이드 RNA를 암호화하는 핵산)(예를 들어, 발현 벡터). CasZ 시스템은 (임의의 조합에서)(1), (2) 및 (3) 중 하나 이상을 포함하는 숙주 세포를 포함할 수 있으며, 예를 들어, 일부 경우에, 숙주 세포는 trancRNA 및/또는 trancRNA를 암호화하는 핵산을 포함한다. 일부 경우에, CasZ 시스템은 (예를 들어, 상기에 추가로) 공여자 주형 핵산을 포함한다. 일부 경우에, CasZ 시스템은 하나 이상의 핵산의 시스템(예를 들어, 상기의 임의의 조합을 암호화하는 하나 이상의 발현 벡터)이다.

[0169] **핵산**

[0170] 본 개시내용은: CasZ trancRNA 서열, CasZ trancRNA를 암호화하는 뉴클레오타이드 서열, CasZ 폴리펩타이드(예를 들어, 야생형 CasZ 단백질, 틈내기효소 CasZ 단백질, dCasZ 단백질, 키메라 CasZ 단백질/CasZ 융합 단백질 등)를 암호화하는 뉴클레오타이드 서열, CasZ 가이드 RNA 서열, CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 폴리뉴클레오타이드(공여자 주형, 공여자 DNA) 서열 중 하나 이상을 포함하는 하나 이상의 핵산을 제공한다. 일부 경우에, 대상 핵산(예를 들어, 하나 이상의 핵산)은 재조합 발현 벡터(예를 들어, 플라스미드, 바이러스 벡터, 미니서클 DNA 등). 일부 경우에, CasZ trancRNA를 암호화하는 뉴클레오타이드 서열, CasZ 단백질을 암호화하는 뉴클레오타이드 서열 및/또는 CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열은 프로모터(예를 들어, 유도성 프로모터), 예를 들어, (예를 들어, 원핵 세포, 진핵 세포, 식물 세포, 동물 세포, 포유류 세포, 영장류 세포, 설치류 세포, 인간 세포 등)에서 작동 가능한 것에 작동 가능하게 연결된다.

[0171] 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 코돈 최적화된다. 이런 최적화 유형은 의도된 숙주 유기체 또는 세포의 코돈 선호도를 모방하는 한편, 동일한 단백질을 암호화하는 CasZ-암호화 뉴클레오타이드 서열의 돌연변이를 수반한다. 따라서, 코돈은 변할 수 있지만, 암호화된 단백질은 변하지 않고 남아있다. 예를 들어, 의도된 표적 세포가 인간 세포라면, 인간 코돈 최적화된 CasZ-암호화 뉴클레오타이드 서열이 사용될 수 있었다. 다른 비제한적 예로서, 의도된 숙주 세포가 마우스 세포였다면, 마우스 코돈-최

적화된 CasZ-암호화 뉴클레오타이드 서열이 생성될 수 있었다. 다른 비제한적 예로서, 의도된 숙주 세포가 식물 세포였다면, 식물 코돈-최적화된 CasZ-암호화 뉴클레오타이드 서열이 생성될 수 있었다. 다른 비제한적 예로서, 의도된 숙주 세포가 곤충 세포였다면, 곤충 코돈-최적화된 CasZ-암호화 뉴클레오타이드 서열이 생성될 수 있었다.

[0172] 본 개시내용은 (일부 경우에 상이한 재조합 발현 벡터에서, 그리고 일부 경우에 동일한 재조합 발현 벡터에서): CasZ trancRNA 서열, CasZ trancRNA를 암호화하는 뉴클레오타이드 서열, CasZ 폴리펩타이드(예를 들어, 야생형 CasZ 단백질, 틈내기효소 CasZ 단백질, dCasZ 단백질, 키메라 CasZ 단백질/CasZ 융합 단백질 등)를 암호화하는 뉴클레오타이드 서열, CasZ 가이드 RNA 서열, CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 폴리뉴클레오타이드(공여자 주형, 공여자 DNA) 서열을 포함하는 하나 이상의 재조합 발현 벡터를 제공한다. 일부 경우에, 대상 핵산(예를 들어, 하나 이상의 핵산)은 재조합 발현 벡터(예를 들어, 플라스미드, 바이러스 벡터, 미니서클 DNA 등). 일부 경우에, CasZ trancRNA를 암호화하는 뉴클레오타이드 서열, CasZ 단백질을 암호화하는 뉴클레오타이드 서열 및/또는 CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열은 프로모터(예를 들어, 유도성 프로모터), 예를 들어, (예를 들어, 원핵 세포, 진핵 세포, 식물 세포, 동물 세포, 포유류 세포, 영장류 세포, 설치류 세포, 인간 세포 등)에서 작동 가능한 것에 작동 가능하게 연결된다.

[0173] 적합한 발현 벡터는 바이러스 발현 벡터(예를 들어, 백시니아 바이러스; 폴리오 바이러스; 아데노바이러스에 기반한 바이러스 벡터(예를 들어, 문헌[Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999]; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 및 WO 95/00655); 아데노-연관 바이러스(AAV)(예를 들어, 문헌[Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; 및 Flotte et al., PNAS (1993) 90:10613-10617] 참조); SV40; 단순포진 바이러스; 인간 면역결핍 바이러스(예를 들어, 문헌[Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999]); 레트로바이러스 벡터(예를 들어, 뮤린 백혈병 바이러스, 비장 괴사 바이러스, 및 레트로바이러스, 예컨대 라우스 육종 바이러스, 하비 육종 바이러스, 조류 백혈병 바이러스, 렌티바이러스, 인간 면역결핍 바이러스, 골수증식성 육종 바이러스 및 유방암 바이러스로부터 유래된 벡터); 등을 포함한다. 일부 경우에, 본 개시내용의 재조합 발현 벡터는 재조합 아데노-연관 바이러스(AAV) 벡터이다. 일부 경우에, 본 개시내용의 재조합 발현 벡터는 재조합 렌티바이러스 벡터이다. 일부 경우에, 본 개시내용의 재조합 발현 벡터는 재조합 레트로바이러스 벡터이다.

[0174] 이용되는 숙주/벡터 시스템에 따라서, 다수의 적합한 전사 및 번역 제어 요소(구성적 및 유도성 프로모터, 전사 인핸서 요소, 전사 종결자 등을 포함)가 발현 벡터에서 사용될 수 있다.

[0175] 일부 실시형태에서, CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열은 제어 요소, 예를 들어, 전사 제어 요소, 예컨대 프로모터에 작동 가능하게 연결된다. 일부 실시형태에서, CasZ 단백질 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 제어 요소, 예를 들어, 전사 제어 요소, 예컨대 프로모터에 작동 가능하게 연결된다.

[0176] 전사 제어 요소는 프로모터일 수 있다. 일부 경우에, 프로모터는 구성적으로 활성인 프로모터이다. 일부 경우에, 프로모터는 조절 가능한 프로모터이다. 일부 경우에, 프로모터는 유도성 프로모터이다. 일부 경우에, 프로모터는 조직-특이적 프로모터이다. 일부 경우에, 프로모터는 세포 유형-특이적 프로모터이다. 일부 경우에, 전사 제어 요소(예를 들어, 프로모터)는 표적화된 세포 유형 또는 표적화된 세포 집단에서 기능성이다. 예를 들어, 일부 경우에, 전사 제어 요소는 진핵 세포, 예를 들어, 조혈모세포(예를 들어, 이동 말초 혈액(mobilized peripheral blood: mPB) CD34(+) 세포, 골수(BM) CD34(+) 세포 등)에서 기능성일 수 있다.

[0177] 진핵 프로모터(진핵 세포에서 프로모터 기능성)인 비제한적 예는 EF1α, 거대세포바이러스(CMV) 급초기, 단순포진 바이러스(HSV) 티미딘 키나제, 초기 및 후기 SV40, 레트로바이러스로부터의 긴 말단 반복부(LTR) 및 마우스 메탈로티오네인-I로부터의 프로모터를 포함한다. 적절한 벡터 및 프로모터의 선택은 당업자 수준 내에서 매우 용이하다. 발현 벡터는 또한 번역 개시 및 전사 종결자에 대한 리보솜 결합 부위를 함유할 수 있다. 발현 벡터는 또한 발현을 증폭시키기 위한 적절한 서열을 포함할 수 있다. 발현 벡터는 또한 CasZ 단백질에 융합되어, 키메라 CasZ 폴리펩타이드를 생성할 수 있는 단백질 태그(예를 들어, 6xHis 태그, 혈구응집소 태그, 형광 단백질

등)를 암호화하는 뉴클레오타이드 서열을 포함할 수 있다.

[0178] 일부 경우에, CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열 및/또는 CasZ 융합 폴리펩타이드는 유도성 프로모터에 작동 가능하게 연결된다. 일부 경우에, CasZ 가이드 RNA 및/또는 CasZ 융합 단백질을 암호화하는 뉴클레오타이드 서열은 구성적 프로모터에 작동 가능하게 연결된다.

[0179] 프로모터는 구성적으로 활성인 프로모터(즉, 구성적으로 활성인/"온(ON)" 상태인 프로모터)일 수 있고, 유도성 프로모터(즉, 상태가 활성/"온" 또는 비활성/"오프"인 프로모터, 외부 자극, 예를 들어, 특정 온도, 화합물 또는 단백질의 존재에 의해 제어됨)일 수 있으며, 공간적으로 제한된 프로모터(즉, 전사 제어 요소, 인핸서 등)(예를 들어, 조직 특이적 프로모터, 세포 유형 특이적 프로모터 등)일 수 있고, 그리고 일시적으로 제한된 프로모터일 수 있다(즉, 프로모터는 배아 발생의 특정 단계 동안에 또는 생물학적 과정(예를 들어, 마우스에서의 모낭 주기)의 특정 단계 동안에 "온" 상태 또는 "오프" 상태이다.

[0180] 적합한 프로모터는 바이러스로부터 유래될 수 있고, 따라서 바이러스 프로모터로서 지칭될 수 있거나, 또는 그들은 원핵 또는 진핵 유기체를 포함하는 임의의 유기체로부터 유래될 수 있다. 적합한 프로모터는 임의의 RNA 중합효소(예를 들어, pol I, pol II, pol III)에 의해 발현을 유도하는 데 사용될 수 있다. 예시적인 프로모터는 SV40 초기 프로모터, 마우스 유방 종양 바이러스 긴 말단 반복부(LTR) 프로모터; 아데노바이러스 주요 후기 프로모터(Ad MLP); 단순포진 바이러스(HSV) 프로모터, 거대세포바이러스(CMV) 프로모터, 예컨대 CMV 급초기 프로모터 영역(CMVIE), 라우스 육종 바이러스(RSV) 프로모터, 인간 U6 소핵 프로모터(U6)(Miyagishi et al., Nature Biotechnology 20, 497 - 500 (2002)), 증강된 U6 프로모터(예를 들어, Xia et al., Nucleic Acids Res. 2003 Sep 1;31(17)), 인간 H1 프로모터(H1) 등을 포함하지만, 이들로 제한되지 않는다.

[0181] 일부 경우에, CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열은 진핵 세포에서 작동 가능한 프로모터(예를 들어, U6 프로모터, 증강된 U6 프로모터, H1 프로모터 등)(의 제어 하에서) 작동 가능하게 연결된다. 당업자에 의해 이해될 바와 같이, U6 프로모터(예를 들어, 진핵 세포에서), 또는 다른 PolIII 프로모터를 이용하여 핵산(예를 들어, 발현 벡터)로부터 RNA(예를 들어, 가이드 RNA)를 발현시킬 때, 행에 몇 개의 T이 있다면(RNA에서 U에 대한 암호) RNA는 돌연변이될 필요가 있을 수 있다. 이는 DNA 내 T(예를 들어, 5개의 T)의 문자열이 중합효소 III(PolIII)에 대한 종결자로서 작용할 수 있기 때문이다. 따라서, 진핵 세포 내 가이드 RNA의 전사를 보장하기 위해, 때때로 T의 실행을 제거하는 가이드 RNA를 암호화하는 서열을 변형할 필요가 있을 수 있다. 일부 경우에, CasZ 단백질(예를 들어, 야생형 CasZ 단백질, 틈내기효소 CasZ 단백질, dCasZ 단백질, 키메라 CasZ 단백질 등)을 암호화하는 뉴클레오타이드 서열은 진핵 세포에서 작동 가능한 프로모터(예를 들어, CMV 프로모터, EF1α 프로모터, 에스트로겐 수용체-조절 프로모터 등)에 작동 가능하게 연결된다.

[0182] 유도성 프로모터의 예는 T7 RNA 중합효소 프로모터, T3 RNA 중합효소 프로모터, 아이소프로필-베타-D-티오갈락토피라노사이드(IPTG)-조절된 프로모터, 락토스 유도된 프로모터, 열 충격 프로모터, 테트라사이클린-조절 프로모터, 스테로이드-조절 프로모터, 금속-조절 프로모터, 에스트로겐 수용체-조절 프로모터 등을 포함하지만, 이들로 제한되지 않는다. 따라서 유도성 프로모터는 독시사이클린; 에스트로겐 및/또는 에스트로겐 유사체; IPTG 등을 포함하지만, 이들로 제한되지 않는 분자에 의해 조절될 수 있다.

[0183] 사용에 적합한 유도성 프로모터는 본 명세서에 기재되거나 또는 당업자에게 공지된 임의의 유도성 프로모터를 포함한다. 유도성 프로모터의 예는 화학적으로/생화학적으로-조절된 그리고 물리적으로 조절된 프로모터, 예컨대 알코올-조절된 프로모터, 테트라사이클린-조절된 프로모터(예를 들어, 테트라사이클린 리프레서 단백질(tetR), 테트라사이클린 오퍼레이터 서열(tetO) 및 테트라사이클린 트랜스작용인자 융합 단백질(tTA)을 포함하는, 안하이드로테트라사이클린(aTc)-반응성 프로모터 및 다른 테트라사이클린-반응성 프로모터 시스템), 스테로이드-조절된 프로모터(예를 들어, 래트 글루코코르티코이드 수용체에 기반한 프로모터, 인간 에스트로겐 수용체, 나방 엑디손 수용체, 및 스테로이드/레티노이드/갑상선 수용체 슈퍼패밀리로부터의 프로모터), 금속-조절된 프로모터(예를 들어, 효모, 마우스 및 인간으로부터의 메탈로티오네인으로부터 유래된 프로모터(금속 이온에 결합하고 격리시키는 단백질) 유전자), 발병-조절된 프로모터(예를 들어, 살리실산, 에틸렌 또는 벤조티아다이아졸(BTH)에 의해 유도), 온도/열-유도성 프로모터(예를 들어, 열 충격 프로모터), 및 광-조절된 프로모터(예를 들어, 식물 세포로부터의 광 반응성 프로모터)를 포함하지만, 이들로 제한되지 않는다.

[0184] 일부 경우에, 다세포 유기체에서, 프로모터가 특정 세포의 서브세트에서 활성(즉, "온")이 되도록 프로모터는 공간적으로 제한된 프로모터(즉, 세포 유형 특이적 프로모터, 조직 특이적 프로모터 등)이다. 공간적으로 제한된 프로모터는 또한 인핸서, 전사 제어 요소, 제어 서열 등으로서 지칭될 수 있다. 프로모터가 표적화된 숙주 세포(예를 들어, 진핵 세포; 원핵 세포)에서 기능성이라면, 임의의 편리한 공간적으로 제한된 프로모터가 사용

될 수 있다.

[0185] 일부 경우에, 프로모터는 가역적 프로모터이다. 가역적 유도성 프로모터를 포함하는 적합한 가역적 프로모터는 당업계에 공지되어 있다. 이러한 가역적 프로모터는 다수의 유기체, 예컨대 진핵생물 및 원핵생물로부터 단리 및 유래될 수 있다. 제2 유기체, 예를 들어, 제1 원핵생물 및 제2 진핵생물, 제1 진핵생물 및 제2 원핵생물에서 사용하기 위한 제1 유기체로부터 유래된 가역적 프로모터의 변형은 당업계에 잘 공지되어 있다. 이러한 가역적 프로모터, 및 이러한 가역적 프로모터에 기반할 뿐만 아니라 추가적인 제어 단백질을 포함하는 시스템은 알코올 조절된 프로모터(예를 들어, 알코올 탈수소효소 I(alcA) 유전자 프로모터, 알코올 트랜스작용인자 단백질(AlcR)에 반응성인 프로모터 등), 테트라사이클린 조절된 프로모터, (예를 들어, 테트라사이클린활성체(TetActivator), 테트라사이클린온(TetON), 테트라사이클린오프(TetOFF) 등을 포함하는 프로모터 시스템), 스테로이드 조절된 프로모터(예를 들어, 래트 글루코코르티코이드 수용체 프로모터 시스템, 인간 에스트로겐 수용체 프로모터 시스템, 레티노이드 프로모터 시스템, 갑상선 프로모터 시스템, 엑디손 프로모터 시스템, 미페프리스톤 프로모터 시스템 등), 금속 조절된 프로모터(예를 들어, 메탈로티오네인 프로모터 시스템 등), 발병-관련 조절 프로모터(예를 들어, 살리실산 조절 프로모터, 에틸렌 조절 프로모터, 벤조티아다이아졸 조절 프로모터 등), 온도 조절 프로모터(예를 들어, 열 충격 유도성 프로모터(예를 들어, HSP-70, HSP-90, 대두 열 충격 프로모터 등), 광 조절 프로모터, 합성 유도성 프로모터 등을 포함하지만, 이들로 제한되지 않는다.

[0186] 핵산(예를 들어, DNA 또는 RNA)(예를 들어, 공여자 폴리뉴클레오타이드 서열을 포함하는 핵산, CasZ 단백질 및/또는 CasZ 가이드 RNA 및/또는 CasZ trancRNA 등을 암호화하는 하나 이상의 핵산)을 숙주 세포 내로 도입하는 방법은 당업계에 공지되어 있고, 핵산(예를 들어, 발현 작제물)을 세포 내로 도입하기 위해 임의의 편리한 방법이 사용될 수 있다. 적합한 방법은, 예를 들어, 바이러스 감염, 형질감염, 리포펙션, 전기천공법, 인산칼슘 침전, 폴리에틸렌이민(PEI)-매개 형질감염, DEAE-덱스트란 매개 형질감염, 리포좀-매개 형질감염, 입자총 기법, 인산칼슘 침전, 직접 미량주사법, 나노입자-매개 핵산 전달 등을 포함한다.

[0187] 재조합 발현 벡터를 세포 내로 도입하는 것은 임의의 배양 배지에서 그리고 세포 생존을 촉진시키는 임의의 배양 조건 하에 생길 수 있다. 표적 세포 내로 재조합 발현 벡터를 도입하는 것은 생체내 또는 생체외에서 수행될 수 있다. 표적 세포 내로 재조합 발현 벡터를 도입하는 것은 시험관외에서 수행될 수 있다.

[0188] 일부 경우에, CasZ 단백질은 RNA로서 제공될 수 있다. RNA는 직접 화학 합성에 의해 제공될 수 있거나 또는 DNA(예를 들어, CasZ 단백질을 암호화)로부터 시험관외에서 전사될 수 있다. 일단 합성되면, RNA는 핵산을 세포 내로 도입하기 위한 임의의 잘 공지된 기법(예를 들어, 미량주사법, 전기천공법, 형질감염 등)에 의해 세포 내로 도입될 수 있다.

[0189] 핵산은 잘 개발된 형질감염 기법(예를 들어, 문헌[Angel and Yanik (2010) PLoS ONE 5(7): e11756] 참조), 및 퀴아젠사(Qiagen)로부터의 상업적으로 입수 가능한 TransMessenger(등록상표) 시약, 스템젠트사(Stemgent)로부터의 Stemfect(상표명) RNA 형질감염 키트, 및 미루스 바이오 엘엘씨(Mirus Bio LLC)로부터의 TransIT(등록상표)-mRNA 형질감염 키트를 이용하여 세포에 제공될 수 있다. 또한 문헌[Beumer et al. (2008) PNAS 105(50):19821-19826] 참조.

[0190] 벡터는 표적 숙주 세포에 직접 제공될 수 있다. 다시 말해서, 벡터가 세포에 의해 취해지도록, 세포는 대상 핵산을 포함하는 벡터(예를 들어, 공여자 주형 서열을 갖고 CasZ 가이드 RNA를 암호화하는 재조합 발현 벡터; CasZ 단백질을 암호화하는 재조합 발현 벡터 등)와 접촉된다. 세포를 플라스미드인 핵산 벡터와 접촉시키는 방법은 당업계에 잘 공지되어 있는 전기천공법, 염화칼슘 형질감염, 미량주사법, 및 리포펙션을 포함한다. 바이러스 벡터 전달을 위해, 세포는 대상 바이러스 발현 벡터를 포함하는 바이러스 입자와 접촉될 수 있다.

[0191] 레트로바이러스, 예를 들어, 렌티바이러스는 본 개시내용의 방법에서 사용하기에 적합하다. 통상적으로 사용되는 레트로바이러스 벡터는 "결함성"이며, 즉, 증식성 감염에 필요한 바이러스 단백질을 생성할 수 없다. 오히려, 벡터의 복제는 패키징 세포주에서의 성장을 필요로 한다. 관심 대상을 포함하는 바이러스 입자를 생성하기 위해, 핵산을 포함하는 레트로바이러스 핵산은 패키징 세포주에 의해 바이러스 캡시드 내로 패키징된다. 상이한 패키징 세포주는캡시드 내로 혼입될 상이한 외피 단백질(동종숙주성(ecotropic), 양쪽성 또는 이종향성(xenotropic))을 제공하는데, 이 외피 단백질은 세포에 대한 바이러스 입자의 특이성(뮤린 및 래트에 대한 동종숙주; 인간, 개 및 마우스를 포함하는 대부분의 포유류 세포 유형에 대해 양쪽성; 및 뮤린 세포를 제외한 대부분의 포유류 세포에 대해 이종향성)을 결정한다. 적절한 패키징 세포주를 사용하여 세포가 패키징된 바이러스 입자에 의해 표적화되는 것을 보장할 수 있다. 대상 벡터 발현 벡터를 패키징 세포주에 도입하고 패키징 세포주에 의해 생성된 바이러스 입자를 수집하는 방법은 당업계에 잘 공지되어 있다. 핵산은 또한 직접 미량주사법(예

를 들어, RNA의 주사)에 의해 도입될 수 있다.

[0192] CasZ 가이드 RNA 및/또는 CasZ 폴리펩타이드를 암호화하는 핵산을 표적 숙주 세포에 제공하기 위해 사용되는 벡터는 발현을 유도하기 위한, 즉, 관심 대상의 핵산의 전사 활성화를 위한 적합한 프로모터를 포함할 수 있다. 다시 말해서, 일부 경우에, 관심 대상의 핵산은 프로모터에 작동 가능하게 연결될 것이다. 이는 편재하는 작용 프로모터, 예를 들어, CMV-β-액틴 프로모터, 또는 유도성 프로모터, 예컨대 특정 세포 집단에서 활성이거나 또는 테트라사이클린과 같은 약물의 존재에 반응하는 프로모터를 포함할 수 있다. 전사 활성화에 의해, 전사는 10배, 100배만큼, 더 보통으로는 1000배만큼 표적 세포에서 기준 수준 초과로 증가될 것이 의도된다. 추가로, CasZ 가이드 RNA 및/또는 CasZ 단백질을 암호화하는 핵산을 세포에 제공하기 위해 사용되는 벡터는 CasZ 가이드 RNA 및/또는 CasZ 단백질을 취한 세포를 확인하기 위해 표적 세포에서 선택 가능한 마커를 암호화하는 핵산 서열을 포함할 수 있다.

[0193] CasZ 폴리펩타이드, 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산은 일부 경우에 RNA이다. 따라서, CasZ 융합 단백질은 RNA로서 세포 내로 도입될 수 있다. 세포 내로 RNA를 도입하는 방법은 당업계에 공지되어 있고, 예를 들어, 직접 미량주사법, 형질감염, 또는 DNA의 도입을 위해 사용되는 임의의 다른 방법을 포함할 수 있다. CasZ 단백질은 대신에 폴리펩타이드로서 세포에 제공될 수 있다. 이러한 폴리펩타이드는 생성물의 용해도를 증가시키는 폴리펩타이드에 선택적으로 융합될 수 있다. 도메인은 정해진 프로테아제 절단 부위, 예를 들어, TEV 프로테아제에 의해 절단되는 TEV 서열을 통해 폴리펩타이드에 연결될 수 있다. 링커는 또한 하나 이상의 가요성 서열, 예를 들어, 1 내지 10개의 글리신 잔기를 포함할 수 있다. 일부 실시형태에서, 융합 단백질의 절단은 생성물의 용해도를 유지하는 완충제에서, 예를 들어, 0.5 내지 2 M 유레아의 존재 하에, 용해도 등을 증가시키는 폴리펩타이드 및/또는 폴리뉴클레오타이드의 존재 하에 수행된다. 관심 대상의 도메인은 엔도솜 용해성 도메인, 예를 들어, 인플루엔자 HA 도메인; 및 생성에 도움을 주는 다른 폴리펩타이드, 예를 들어, IF2 도메인, GST 도메인, GRPE 도메인 등을 포함한다. 폴리펩타이드는 개선된 안정성을 위해 제형화될 수 있다. 예를 들어, 펩타이드는 페길화될 수 있으며(PEGylated), 여기서 폴리에틸렌옥시기는 혈류에서 증강된 수명을 제공한다.

[0194] 추가적으로 또는 대안적으로, 본 개시내용의 CasZ 폴리펩타이드는 세포에 의한 흡수를 촉진하기 위해 폴리펩타이드 침투 도메인에 융합될 수 있다. 다수의 침투 도메인이 당업계에 공지되어 있고, 펩타이드, 펩타이드 모방체 및 비-펩타이드 담체를 포함하는 본 개시내용의 비통합 폴리펩타이드에서 사용될 수 있다. 예를 들어, 침투 펩타이드는 아미노산 서열 RQIKIWFQNRRMKWKK (서열번호 134)를 포함하는, 페너트라틴(*penetratin*)으로서 지칭되는 노랑초파리 전사 인자 안테나피디어(*Antennapaedia*)의 세 번째 알파 나선으로부터 유래될 수 있다. 다른 예로서, 침투 펩타이드는, 예를 들어, 천연 유래 tat 단백질의 아미노산 49 내지 57을 포함할 수 있는 HIV-1 tat 염기성 영역 아미노산 서열을 포함한다. 다른 침투 도메인은 폴리-알기닌 모티프, 예를 들어, HIV-1 rev 단백질, 노나-알기닌, 옥타-알기닌 등의 아미노산 34 내지 56의 영역을 포함한다. (예를 들어, 전좌 펩타이드 및 펩토이드의 교시에 대해 본 명세서에 구체적으로 참고로 편입된 문헌[Futaki et al. (2003) Curr Protein Pept Sci. 2003 Apr; 4(2): 87-9 및 446; 및 Wender et al. (2000) Proc. Natl. Acad. Sci. U.S.A 2000 Nov. 21; 97(24):13003-8]; 공개된 미국 특허 출원 제20030220334호; 제20030083256호; 제20030032593호; 및 제20030022831호를 참조). 노나-알기닌(R9) 서열은 특성규명된 더 효율적인 PTD 중 하나이다(Wender et al. 2000; Uemura et al. 2002). 융합이 이루어지는 부위는 폴리펩타이드의 생물학적 활성, 분비 또는 결합 특징을 최적화하기 위해 선택될 수 있다. 최적의 부위는 일상적인 실험에 의해 결정될 것이다.

[0195] 본 개시내용의 CasZ 폴리펩타이드는 시험관내에서 또는 진핵세포에 의해 또는 원핵 세포에 의해 생성될 수 있고, 비폴딩, 예를 들어, 열 변성, 다이티오트레이톨 환원 등에 의해 추가로 가공될 수 있고, 당업계에 공지된 방법을 이용하여 추가로 재폴딩될 수 있다.

[0196] 1차 서열을 변경시키지 않는 관심 대상의 변형은 폴리펩타이드의 화학적 유도체화, 예를 들어, 아실화, 아세틸화, 카복실화, 아마이드화 등을 포함한다. 또한 글리코실화의 변형, 예를 들어, 합성 및 가공 동안 또는 추가 가공 단계에서 폴리펩타이드의 글리코실화 패턴을 변형시킴으로써; 예를 들어, 글리코실화, 예컨대 포유류 글리코실화 또는 탈글리코실화 효소에 영향을 미치는 효소에 폴리펩타이드를 노출시킴으로써 만들어지는 것이 포함된다. 또한 포스포릴화된 아미노산 잔기, 예를 들어, 포스포타이로신, 포스포세린 또는 포스포트레오닌을 갖는 서열이 포함된다.

[0197] 또한 단백질 분해에 대한 그들의 내성을 개선시키기 위해, 표적 서열 특이성을 변화시키기 위해, 용해도 특성을 최적화하기 위해, 단백질 활성(예를 들어, 전사 조절 활성, 효소 활성 등)을 변경하기 위해 또는 그들을 더 적

합하게 제공하기 위해 보통의 분자 생물학 기법 및 합성 화학을 이용하여 변형된 핵산(예를 들어, CasZ 가이드 RNA를 암호화, CasZ 융합 단백질을 암호화 등) 및 단백질(예를 들어, 야생형 단백질 또는 변이체 단백질로부터 유래된 CasZ 융합 단백질)이 본 개시내용의 실시형태에 포함하기에 적합하다. 이러한 폴리펩타이드의 유사체는 천연 유래 L-아미노산, 예를 들어, D-아미노산 또는 비천연 유래 합성 아미노산 이외의 잔기를 함유하는 것을 포함한다. D-아미노산은 일부 또는 모든 아미노산 잔기로 치환될 수 있다.

[0198] 본 개시내용의 CasZ 폴리펩타이드는 당업계에 공지된 통상적인 방법을 이용하여 시험관내 합성에 의해 제조될 수 있다. 다양한 상업적 합성 장치, 예를 들어, 어플라이드 바이오시스템즈, 인코포레이티드(Applied Biosystems, Inc.), 베크만(Beckman) 등에 의한 자동화된 합성기를 이용 가능하다. 합성기를 이용함으로써, 천연 유래 아미노산은 비천연 아미노산으로 치환될 수 있다. 특정 서열 및 제조 방식은 편리함, 경제성, 필요한 순도 등에 의해 결정될 것이다.

[0199] 원한다면, 합성 동안 또는 발현 동안에 다양한 기가 펩타이드 내로 도입되는데, 이는 다른 분자에 또는 표면에 대한 연결을 허용한다. 따라서 티오에터, 금속 이온 복합체에 대한 연결을 위한 히스티딘, 아마이드 또는 에스터를 형성하기 위한 카복실기 또는 아마이드를 형성하기 위한 아미노기 등을 만들기 위해 시스테인이 사용될 수 있다.

[0200] 본 개시내용의 CasZ 폴리펩타이드는 또한 단리될 수 있고, 재조합 합성의 통상적인 방법에 따라 정제될 수 있다. 발현 숙주의 용해물이 제조될 수 있고, 용해물은 고성능 액체 크로마토그래피(HPLC), 배제 크로마토그래피, 겔 전기영동, 친화도 크로마토그래피 또는 다른 정제 기법을 이용하여 정제된다. 대부분에 대해, 사용되는 조성물은 생성물의 제조 및 그의 정제 방법과 관련된 오염물질에 대해 20중량% 이상의 목적으로 하는 생성물, 더 보통으로는 75중량% 이상, 바람직하게는 95중량% 이상, 그리고 치료적 목적을 위해, 보통 99.5중량% 이상을 포함할 것이다. 보통, 백분율은 총 단백질을 기준으로 할 것이다. 따라서, 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드, 또는 CasZ 융합 폴리펩타이드는 적어도 80% 순수, 적어도 85% 순수, 적어도 90% 순수, 적어도 95% 순수, 적어도 98% 순수 또는 적어도 99% 순수(예를 들어, 오염물질, 비-CasZ 단백질 또는 다른 거대분자 등이 없음)이다.

[0201] 표적 핵산(예를 들어, 게놈 DNA)에 대한 절단 또는 임의의 목적으로 하는 변형, 또는 표적 핵산, CasZ 가이드 RNA 및/또는 CasZ 폴리펩타이드 및/또는 CasZ trancRNA, 및/또는 공여자 주형 서열과 관련된 폴리펩타이드에 대한 임의의 목적으로 하는 변형을 유도하기 위해, 그들이 핵산으로서 도입되든 또는 폴리펩타이드로서 도입되든, 약 30분 내지 약 24시간, 예를 들어, 1시간, 1.5시간, 2시간, 2.5시간, 3시간, 3.5시간 4시간, 5시간, 6시간, 7시간, 8시간, 12시간, 16시간, 18시간, 20시간, 또는 약 30분 내지 약 24시간의 임의의 다른 기간 동안 세포에 제공될 수 있는데, 이는 대략 매일 내지 약 4일마다, 예를 들어, 1.5일마다, 2일마다, 3일마다의 빈도로, 또는 대략 매일 내지 약 4일마다의 임의의 다른 빈도로 반복될 수 있다. 제제(들)는 대상 세포에 1회 이상, 예를 들어, 1회, 2회, 3회, 3회 초과로 제공될 수 있고, 세포는 각각의 접촉 사건 후 일정한 시간 동안, 예를 들어, 16 내지 24시간 동안 제제(들)와 함께 인큐베이션되도록 허용되며, 이 시간 후에 배지는 신선한 배지로 대체되고, 세포는 추가로 배양된다.

[0202] 2 이상의 상이한 표적화 복합체가 세포(예를 들어, 동일 또는 상이한 표적 핵산 내에서 상이한 서열에 상보성인 2개의 상이한 CasZ 가이드 RNA)에 제공되는 경우에, 복합체는 동시에(예를 들어, 2개의 폴리펩타이드 및/또는 핵산으로서) 제공될 수 있거나, 또는 동시에 전달될 수 있다. 대안적으로, 그들은 연속적으로, 예를 들어, 제1의 제공된 표적화 복합체, 다음에 제2 표적화 복합체 등 또는 그 반대로 제공될 수 있다.

[0203] DNA 벡터의 표적 세포의 전달을 개선시키기 위해, DNA는 손상으로부터 보호될 수 있고, 세포 내로 그의 유입은, 예를 들어, 리포플렉스(lipoplex) 및 폴리플렉스(polyplex)를 이용함으로써 용이하게 된다. 따라서, 일부 경우에, 본 개시내용의 핵산(예를 들어, 본 개시내용의 재조합 발현 벡터)는 마이셀 또는 리포좀과 같이 조직화된 구조에서 지질로 뒤덮일 수 있다. 조직화된 구조가 DNA와 함께 복합체화될 때, 이는 리포플렉스로 불린다. 3가지 유형의 지질, 음이온(음으로 하전됨), 중성 또는 양이온성(양으로 하전됨)이 있다. 양이온성 지질을 이용하는 리포플렉스는 유전자 전달을 위한 증명된 효용을 가진다. 양전하 지질은 그들의 양전하에 기인하여 음으로 하전된 DNA와 자연적으로 복합체화된다. 또한 그들의 하전의 결과로서, 그들은 세포막과 상호작용한다. 이어서, 리포플렉스의 내포작용이 일어나고, DNA는 세포질 내로 방출된다. 양이온성 지질은 또한 세포에 의해 DNA의 분해에 대해 보호한다.

[0204] DNA와 중합체의 복합체는 폴리플렉스(polyplex)로 불린다. 대부분의 폴리플렉스는 양이온성 중합체로 이루어지고, 그들의 생성은 이온 상호작용에 의해 조절된다. 폴리플렉스와 리포플렉스의 작용 방법 간의 한 가지 큰 차

이는 폴리플렉스는 세포질 내로 그들의 DNA 부하를 방출할 수 없고, 이를 위하여, 엔도솜-용해제(내포작용 동안 만들어진 엔도솜을 용해시키기 위함), 예컨대 비활성화 아데노바이러스와의 공동형질감염이 일어나야 한다는 것이다. 그러나, 이는 항상 그런 것은 아니며; 중합체, 예컨대 폴리에틸렌이민은 키토산 및 트라이메틸키토산과 같이 엔도솜 붕괴의 그 자체의 방법을 가진다.

[0205] 구체 형태를 갖는 상당히 분지형인 거대분자인 덴드리머는 또한 줄기 세포를 유전적으로 변형하기 위해 사용될 수 있다. 덴드리머 입자의 표면은 그의 특성을 변경시키도록 작용화될 수 있다. 특히, 양이온성 덴드리머(즉, 양성 표면 전하를 갖는 것)를 작제하는 것이 가능하다. 유전자 물질, 예컨대 DNA 플라스미드의 존재 하에 있을 때, 전하 상보성은 양이온성 덴드리머와 핵산의 일시적 회합을 야기한다. 그의 목적에 도달되었을 때, 덴드리머-핵산 복합체는 내포작용에 의해 세포 내로 흡수될 수 있다.

[0206] 일부 경우에, 개시내용의 핵산(예를 들어, 발현 벡터)은 관심 대상의 가이드 서열에 대한 삽입 부위를 포함한다. 예를 들어, 핵산은 관심 대상의 가이드 서열에 대한 삽입 부위를 포함할 수 있으며, 여기서 삽입 부위는 가이드 서열이 목적으로 하는 표적 서열(예를 들어, 가이드 RNA의 CasZ 결합 양상에 기여하는 서열, 예를 들어, CasZ 가이드 RNA의 dsRNA 듀플렉스(들)에 기여하는 서열 - 가이드 RNA의 이 부분은 또한 가이드 RNA의 '스캐폴드' 또는 '불변 영역'으로서 지칭될 수 있음)에 혼성화하도록 변할 때 변하지 않는 CasZ 가이드 RNA의 일부를 암호화하는 뉴클레오타이드 서열에 바로 인접한다. 따라서, 일부 경우에, 가이드 RNA의 가이드 서열을 암호화하는 부분이 삽입 서열(삽입 부위)라는 것을 제외하고, 대상 핵산(예를 들어, 발현 벡터)는 CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함한다. 삽입 부위는 목적으로 하는 서열의 삽입을 위해 사용되는 임의의 뉴클레오타이드 서열이다. 다양한 기법과 함께 사용하기 위한 "삽입 부위"는 당업자에게 공지되어 있고, 임의의 편리한 삽입 부위가 사용될 수 있다. 삽입 부위는 핵산 서열을 조작하기 위한 임의의 방법에 대한 것일 수 있다. 예를 들어, 일부 경우에 삽입 부위는 다중 클로닝 부위(MCS)(예를 들어, 하나 이상의 제한 효소 인식 서열을 포함하는 부위), 결찰 독립적 클로닝을 위한 부위, 재조합 기반 클로닝을 위한 부위(예를 들어, *att* 부위에 기반한 재조합), CRISPR/Cas(예를 들어, Cas9) 기반 기법에 의해 인식되는 뉴클레오타이드 서열 등이다.

[0207] 삽입 부위는 임의의 바람직한 길이일 수 있고, 삽입 부위 유형에 의존할 수 있다(예를 들어, 부위가 하나 이상의 제한 효소 인식 서열을 포함하는지의 여부(및 얼마나 많이 포함하는지), 부위가 CRISPR/Cas 단백질에 대한 표적 부위를 포함하는지의 여부 등에 의존할 수 있다). 일부 경우에, 대상 핵산의 삽입 부위는 길이가 3개 이상의 뉴클레오타이드(nt)이다(예를 들어, 길이가 5개 이상, 8개 이상, 10개 이상, 15개 이상, 17개 이상, 18개 이상, 19개 이상, 20개 이상 또는 25개 이상 또는 30개 이상의 nt). 일부 경우에, 대상 핵산의 삽입 부위 길이는 길이가 2 내지 50개의 뉴클레오타이드(nt)(예를 들어, 2 내지 40 nt, 2 내지 30개의 nt, 2 내지 25개의 nt, 2 내지 20개의 nt, 5 내지 50개의 nt, 5 내지 40개의 nt, 5 내지 30개의 nt, 5 내지 25개의 nt, 5 내지 20개의 nt, 10 내지 50개의 nt, 10 내지 40개의 nt, 10 내지 30개의 nt, 10 내지 25개의 nt, 10 내지 20개의 nt, 17 내지 50개의 nt, 17 내지 40개의 nt, 17 내지 30개의 nt, 17 내지 25 nt) 범위이다. 일부 경우에, 대상 핵산의 삽입 부위 길이는 길이가 5 내지 40 nt이다.

[0208] ***핵산 변형***

[0209] 일부 실시형태에서, 새로운 또는 증강된 특징(예를 들어, 개선된 안정성)을 갖는 핵산을 제공하기 위해 대상 핵산(예를 들어, CasZ 가이드 RNA 또는 trancRNA)는 하나 이상의 변형, 예를 들어, 염기 변형, 골격 변형 등을 가진다. 뉴클레오사이드는 염기-당 조합이다. 뉴클레오사이드의 염기 부분은 정상적으로 복소환식 염기이다. 이러한 복소환식 염기의 2가지 가장 통상적인 부류는 퓨린 및 피리미딘이다. 뉴클레오타이드는 뉴클레오사이드의 당 부분에 공유적으로 연결된 인산염기를 추가로 포함하는 뉴클레오사이드이다. 펜토퓨라노실 당을 포함하는 해당 뉴클레오사이드에 대해, 인산염기는 당의 2', 3' 또는 5' 하이드록실 모이어티에 연결될 수 있다. 올리고뉴클레오타이드를 형성함에 있어서, 인산염기는 서로 인접한 뉴클레오사이드에 공유적으로 연결되어 선형 중합체 화합물을 형성한다. 결국, 이 선형 중합체 화합물의 각각의 말단은 추가로 결합되어 원형 화합물을 형성할 수 있지만, 그러나, 선형 화합물이 적합하다. 추가로, 선형 화합물은 내부 뉴클레오타이드 염기 상보성을 가질 수 있으며, 따라서 완전히 또는 부분적으로 이중-가닥 화합물을 생성하기 위한 방식으로 폴딩될 수 있다. 올리고뉴클레오타이드 내에서, 인산염기는 통상적으로 올리고뉴클레오타이드의 뉴클레오사이드간 골격을 형성하는 것으로 지칭된다. RNA 및 DNA의 정상 결합 또는 골격은 3' - 5' 포스포다이에스터 결합이다.

[0210] 적합한 핵산 변형은 하기를 포함하지만, 이들로 제한되지 않는다: 2'O메틸 변형된 뉴클레오타이드, 2' 플루오로 변형된 뉴클레오타이드, 잠금 핵산(locked nucleic acid: LNA) 변형된 뉴클레오타이드, 펩타이드 핵산(peptide nucleic acid: PNA) 변형된 뉴클레오타이드, 포스포로티오에이트 결합을 갖는 뉴클레오타이드 및 5' 캡(예를 들

어, 7-메틸구아닐레이트 캡(m7G)). 추가적인 상세한 설명 및 추가적인 변형을 이하에 기재한다.

[0211] 2'-O-메틸 변형된 뉴클레오타이드(또한 2'-O-메틸 RNA로서 지칭됨)는 tRNA에서 발견되는 RNA 및 전사후 변형으로서 생기는 다른 작은 RNA의 천연 유래 변형이다. 2'-O-메틸 RNA를 함유하는 올리고뉴클레오타이드가 직접적으로 합성될 수 있다. 이 변형은 RNA:RNA 듀플렉스의 Tm을 증가시키지만 RNA:DNA 안정성의 작은 변화만을 초래한다. 이는 단일-가닥 리보뉴클레아제에 의한 공격에 대해 안정하며 전형적으로 DNA보다 DNase에 대해 5 내지 10배 덜 민감하다. 이는 표적 전령에 대한 안정성 및 결합 친화도를 증가시키기 위한 수단으로서 안티센스 올리고에서 통상적으로 사용된다.

[0212] 2' 플루오로 변형된 뉴클레오타이드(예를 들어, 2' 플루오로 염기)는 결합 친화도(Tm)를 증가시키는 플루오린 변형된 리보스를 가지며, 또한 천연 RNA에 비해 일부 상대적 뉴클레아제 내성을 부여한다. 이들 변형은 혈청 또는 다른 생물학적 유체에서의 안정성을 개선시키기 위해 리보자임 및 siRNA에서 통상적으로 사용된다.

[0213] LNA 염기는 RNA A-형 나선 듀플렉스 기하학을 선호하는 C3'-엔도 위치에서 염기를 잠그는 리보스 골격에 대한 변형을 가진다. 이 변형은 Tm을 상당히 증가시키며, 또한 매우 뉴클레아제 내성이다. 다중 LNA 삽입은 3'-말단을 제외한 임의의 위치에서 올리고에 위치될 수 있다. 적용은 혼성화 프로브에 대한 안티센스 올리고로부터 SNP 검출 및 대립유전자 특이적 PCR까지의 범위로 기재되었다. LNA에 의해 부여되는 Tm의 큰 증가에 기인하여, 그들은 또한 프라이머 이량체 형성뿐만 아니라 자기-헤어핀 형성의 증가를 야기할 수 있다. 일부 경우에, 단일 올리고 내로 혼입된 LNA의 수는 10개 이하의 염기이다.

[0214] 포스포로티오에이트(PS) 결합(즉, 포스포로티오에이트 결합)은 핵산(예를 들어, 올리고)의 인산염 골격에서 황 원자를 비-브리징 산소로 치환한다. 이 변형은 뉴클레아제 분해에 내성이 있는 뉴클레오타이드간 결합을 제공한다. 포스포로티오에이트 결합은 엑소뉴클레아제 분해를 저해하기 위해 올리고의 5'- 또는 3'-말단에서 마지막 3 내지 5개 뉴클레오타이드 사이에 도입될 수 있다. 올리고 내에서(예를 들어, 전체 올리고를 전체적으로) 포스포로티오에이트 결합을 포함하는 것은 마찬가지로 엔도뉴클레아제에 의한 공격을 감소시킬 수 있다.

[0215] 일부 실시형태에서, 대상 핵산은 2'-O-메틸 변형된 뉴클레오타이드인 하나 이상의 뉴클레오타이드이다. 일부 실시형태에서, 대상 핵산(예를 들어, 가이드 RNA, tranc RNA 등)은 하나 이상의 2' 플루오로 변형된 뉴클레오타이드를 가진다. 일부 실시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 하나 이상의 LNA 염기를 가진다. 일부 실시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 포스포로티오에이트 결합에 의해 연결된 하나 이상의 뉴클레오타이드를 가진다(즉, 대상 핵산은 하나 이상의 포스포로티오에이트 결합을 가진다). 일부 실시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 5' 캡(예를 들어, 7-메틸구아닐레이트 캡(m7G))을 가진다. 일부 실시형태에서, 대상 핵산(예를 들어, 가이드 RNA, tranc RNA 등)은 변형된 뉴클레오타이드의 조합을 가진다. 예를 들어, 대상 핵산(예를 들어, 가이드 RNA, tranc RNA 등)은 다른 변형을 갖는 하나 이상의 뉴클레오타이드(예를 들어, 2'-O-메틸 뉴클레오타이드 및/또는 2' 플루오로 변형된 뉴클레오타이드 및/또는 LNA 염기 및/또는 포스포로티오에이트 결합)를 갖는 것에 추가로 5' 캡(예를 들어, 7-메틸구아닐레이트 캡(m7G))을 가질 수 있다.

[0216] ***변형된 골격 및 변형된 뉴클레오사이드간 결합***

[0217] 변형을 함유하는 적합한 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA)의 예는 변형된 골격 또는 비천연 뉴클레오사이드간 결합을 함유하는 핵산을 포함한다. 변형된 골격을 갖는 핵산은 골격에 인 원자를 보유하는 것 및 골격에 인 원자를 갖지 않는 것을 포함한다.

[0218] 인 원자를 함유하는 적합한 변형된 올리고뉴클레오타이드 골격은, 예를 들어, 포스포로티오에이트, 카이랄 포스포로티오에이트, 포스포로다이티오에이트, 포스포트라이에스터, 아미노알킬포스포트라이에스터, 메틸 및 다른 알킬 포스포네이트(3'-알킬렌 포스포네이트, 5'-알킬렌 포스포네이트 및 카이랄 포스포네이트를 포함), 포스피네이트, 포스포르아미데이트(3'-아미노 포스포르아미데이트 및 아미노알킬포스포르아미데이트, 포스포로다이아미데이트, 티오노포스포르아미데이트를 포함), 티오노알킬포스포네이트, 티오노알킬포스포트라이에스터, 셀레노포스페이트 및 정상 3'-5' 결합을 갖는 보라노포스페이트, 이들의 2'-5'연결된 유사체, 및 반전된 극성을 갖는 것을 포함하되, 하나 이상의 뉴클레오타이드간 결합은 3' - 3', 5' - 5' 또는 2' - 2' 결합이다. 반전된 극성을 갖는 적합한 올리고뉴클레오타이드는 가장 3'의 뉴클레오타이드간 결합에서 단일 3' - 3' 결합, 즉, 염기성일 수 있는 단일 반전 뉴클레오사이드 잔기를 포함한다(뉴클레오염기는 그 대신에 상실되거나 하이드록실기를 가짐). 다양한 염(예컨대, 칼륨 또는 나트륨), 혼합된 염 및 유리 산 형태가 또한 포함된다.

[0219] 일부 실시형태에서, 대상 핵산은 하나 이상의 포스포로티오에이트 및/또는 헤테로원자 뉴클레오사이드간 결합, 특히 $-CH_2-NH-O-CH_2-$, $-CH_2-N(CH_3)-O-CH_2-$(메틸렌(메틸이미노) 또는 MMI 골격으로서 알려짐), $-CH_2-O-N(CH_3)-CH_2-$

, $-CH_2-N(CH_3)-N(CH_3)-CH_2-$ 및 $-O-N(CH_3)-CH_2-CH_2-$를 포함한다(천연 포스포다이에스터 뉴클레오타이드간 결합은 $-O-P(=O)(OH)-O-CH_2-$로서 나타냄). MMI 유형 뉴클레오사이드간 결합은 상기 언급된 미국 특허 제5,489,677호에 개시되어 있으며, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입된다. 적합한 아마이드 뉴클레오사이드간 결합은 미국 특허 제5,602,240호에 개시되어 있으며, 이의 개시내용은 전문이 본 명세서에 참고로 편입된다.

[0220] 또한, 예를 들어, 미국 특허 제5,034,506호에 기재된 바와 같은 몰폴리노 골격 구조를 갖는 핵산이 적합하다. 예를 들어, 일부 실시형태에서, 대상 핵산은 리보스 고리 대신에 6-원 몰폴리노 고리를 포함한다. 일부 이들 실시형태에서, 포스포로다이아미데이트 또는 다른 비-포스포다이에스터 뉴클레오사이드간 결합은 포스포다이에스터 결합을 대체한다.

[0221] 인 원자를 포함하지 않는 적합한 변형된 폴리뉴클레오타이드 골격은 짧은 쇄알킬 또는 사이클로알킬 뉴클레오사이드간 결합, 혼합된 헤테로원자 및 알킬 또는 사이클로알킬 뉴클레오사이드간 결합 또는 하나 이상의 짧은 쇄 헤테로원자 또는 복소환식 뉴클레오사이드간 결합에 의해 형성된 골격을 가진다. 이들은 몰폴리노 결합을 갖는 것(뉴클레오사이드의 당 부분으로부터 부분적으로 형성됨); 실록산 골격; 설파이드, 설폭사이드 및 설폰 골격; 폼아세틸 및 티오폼아세틸 골격; 메틸렌 폼아세틸 및 티오폼아세틸 골격; 리보아세틸 골격; 알켄 함유 골격; 설파메이트 골격; 메틸렌이미노 및 메틸렌하이드라지노 골격; 설포네이트 및 설폰아마이드 골격; 아마이드 골격; 및 혼합된 N, O, S 및 CH_2 성분 부분을 갖는 다른 것을 포함한다.

[0222] ***모방체***

[0223] 대상 핵산은 핵산 모방체일 수 있다. 폴리뉴클레오타이드에 적용되는 용어 "모방체"는 폴리뉴클레오타이드를 포함하는 것으로 의도되되, 푸라노스 고리만 또는 푸라노스 고리와 뉴클레오타이드 결합 둘 다는 비-푸라노스 기로 대체되며, 푸라노스 고리만의 대체는 또한 본 명세서에서 당 대용물로서 지칭된다. 복소환식 염기 모이어티 또는 변형된 복소환식 염기 모이어티는 적절한 표적 핵산과의 혼성화를 위해 유지된다. 우수한 혼성화 특성을 갖는 것을 나타난 하나의 이러한 핵산, 폴리뉴클레오타이드 모방체는 펩타이드 핵산(PNA)으로서 지칭된다. PNA에서, 폴리뉴클레오타이드의 당-골격은 골격, 특히 아미노에틸글리신 골격을 함유하는 아마이드로 대체된다. 뉴클레오타이드가 보유되며, 골격의 아마이드 부분의 아자 질소 원자에 직접 또는 간접적으로 결합된다.

[0224] 우수한 혼성화 특성을 갖는 것으로 보고된 하나의 폴리뉴클레오타이드 모방체는 펩타이드 핵산(PNA)이다. PNA 화합물의 골격은 PNA에 아마이드 함유 골격을 제공하는 2 이상의 연결된 아미노에틸 글리신 단위이다. 복소환식 염기 모이어티는 골격의 아마이드 부분의 아자 질소 원자에 직접 또는 간접적으로 결합된다. PNA 화합물의 제조를 기재하는 대표적인 미국 특허는 미국 특허 제5,539,082호; 제5,714,331호; 및 제5,719,262호를 포함하지만, 이들로 제한되지 않으며, 이의 개시내용은 본 명세서에 전문이 참고로 포함된다.

[0225] 연구된 다른 부류의 폴리뉴클레오타이드 모방체는 몰폴리노 고리에 부착된 복소환식 염기를 갖는 연결된 몰폴리노 단위(몰폴리노 핵산)에 기반한다. 몰폴리노 핵산에서 몰폴리노 단량체 단위를 연결하는 다수의 연결기가 보고되었다. 비이온성 올리고머 화합물을 제공하기 위한 연결기의 한 가지 부류가 선택되었다. 비-이온성 몰폴리노계 올리고머 화합물은 세포 단백질과의 원치않는 상호작용을 가질 가능성이 더 적다. 몰폴리노계 폴리뉴클레오타이드는 세포 단백질과의 원치않는 상호작용을 형성할 가능성이 더 적은 올리고뉴클레오타이드의 비이온성 모방체이다(Dwaine A. Braasch and David R. Corey, *Biochemistry*, 2002, 41(14), 4503-4510). 몰폴리노계 폴리뉴클레오타이드는 미국 특허 제5,034,506호에 개시되어 있고, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입된다. 단량체 서브유닛을 결합하는 다양한 상이한 연결기를 갖는 폴리뉴클레오타이드의 몰폴리노 부류 내의 다양한 화합물이 제조되었다.

[0226] 추가적인 부류의 폴리뉴클레오타이드 모방체는 사이클로헥센일 핵산(CeNA)으로서 지칭된다. DNA/RNA 분자에 정상적으로 존재하는 푸라노스 고리는 사이클로헥센일 고리로 대체된다. CeNA DMT 보호된 포스포르아미다이트 단량체가 제조되었고, 고전적 포스포르아미다이트 화학 다음의 올리고머 화합물 합성을 위해 사용되었다. CeNA로 변형된 특정 위치를 갖는 완전히 변형된 CeNA 올리고머 화합물 및 올리고뉴클레오타이드가 제조되고, 연구되었다(문헌[Wang et al., *J. Am. Chem. Soc.*, 2000, 122, 8595-8602]을 참조하며, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨). 일반적으로 DNA 쇄에 대한 CeNA 단량체의 편입은 DNA/RNA 혼성체의 그의 안정성을 증가시킨다. CeNA 올리고아데닐레이트는 천연 복합체와 유사한 안정성을 갖는 RNA 및 DNA 상보체와의 복합체를 형성하였다. 천연 핵산 구조에 CeNA 구조를 혼입하는 연구는 용이한 입체배좌 적용에 의해 진행되는 NMR 및 원이색법에 의해 나타낸다.

[0227] 추가적인 변형은 잠금 핵산(LNA)을 포함하며, 이때 2'-하이드록실기는 당 고리의 4' 탄소 원자에 연결됨으로써,

2'-C,4'-C-옥시메틸렌 결합을 형성하고, 이에 의해 이환식 당 모이어티를 형성한다. 결합은 2' 산소 원자와 4' 탄소 원자를 브리징하는 기인 메틸렌(-CH_2-)일 수 있되, n은 1 또는 2이다(문헌[Singh et al., *Chem. Commun.*, 1998, 4, 455-456], 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨). LNA 및 LNA 유사체는 상보성 DNA 및 RNA에 의한 매우 높은 듀플렉스 열 안정성(Tm=+3 내지 +10℃), 3'-핵산외부 분해 및 양호한 용해도 특성에 대한 안정성을 나타낸다. LNA를 함유하는 강력한 그리고 비독성의 안티센스 올리고뉴클레오타이드가 기재되었다(예를 들어, 문헌[Wahlestedt et al., *Proc. Natl. Acad. Sci.* U.S.A., 2000, 97, 5633-5638] 참조, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨).

[0228] LNA 단량체 아데닌, 사이토신, 구아닌, 5-메틸-사이토신, 티민 및 유라실의 합성 및 제조는, 그들의 올리고머, 및 핵산 인식 특성과 함께 기재되었다(예를 들어, 문헌[Koshkin et al., *Tetrahedron*, 1998, 54, 3607-3630] 참조, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨). LNA 및 이의 제조는 또한 WO 98/39352 및 WO 99/14226뿐만 아니라 미국 특허 출원 제20120165514호, 제20100216983호, 제20090041809호, 제20060117410호, 제20040014959호, 제20020094555호 및 제20020086998호에 기재되어 있고, 이의 개시내용은 그들의 전문이 본 명세서에 참고로 포함된다.

[0229] ***변형된 당 모이어티***

[0230] 대상 핵산은 또한 하나 이상의 치환된 당 모이어티를 포함할 수 있다. 적합한 폴리뉴클레오타이드는 OH; F; O-, S-, 또는 N-알킬; O-, S- 또는 N-알켄일; O-, S- 또는 N-알킨일; 또는 O-알킬-O-알킬로부터 선택된 당 치환체기를 포함하되, 알킬, 알켄일 및 알킨일은 치환된 또는 비치환된 C_1 내지 C_{10} 알킬 또는 C_2 내지 C_{10} 알켄일 및 알킨일일 수 있다. $O((CH_2)_nO)_mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$ 및 $O(CH_2)_nON((CH_2)_nCH_3)_2$가 특히 적합하며, n 및 m은 1 내지 약 10이다. 다른 적합한 폴리뉴클레오타이드는 C_1 내지 C_{10} 저급 알킬, 치환된 저급 알킬, 알켄일, 알킨일, 알카릴, 아르알킬, O-알카릴 또는 O-아르알킬, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, 헤테로사이클로알킬, 헤테로사이클로알카릴, 아미노알킬아미노, 폴리알킬아미노, 치환된 실릴, RNA 절단기, 리포터기, 삽입제(intercalator), 올리고뉴클레오타이드의 약물동력학 특성을 개선시키기 위한 기, 또는 올리고뉴클레오타이드의 약물동력학 특성을 개선시키기 위한 기 및 유사한 특성을 갖는 다른 치환체로부터 선택된 당 치환체기를 포함한다. 적합한 변형은 2'-메톡시에톡시(2'-O-$CH_2CH_2OCH_3$, 또한 2'-O-(2-메톡시에틸) 또는 2'-MOE로서 알려짐)(Martin et al., *Helv. Chim. Acta*, 1995, 78, 486-504, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨) 즉, 알콕시알콕시기를 포함한다. 추가적인 적합한 변형은 본 명세서의 이하의 실시예에 기재된 바와 같은 2'-DMAOE로서도 알려진 2'-다이메틸아미노옥시에톡시, 즉, $O(CH_2)_2ON(CH_3)_2$ 기, 및 2'-다이메틸아미노에톡시에톡시(또한 당업계에서 2'-O-다이메틸-아미노-에톡시-에틸 또는 2'-DMAEOE로서 알려짐), 즉, 2'-O-CH_2-O-CH_2-$N(CH_3)_2$를 포함한다.

[0231] 다른 적합한 당 치환체기는 메톡시(-O-CH_3), 아미노프로폭시(--O CH_2 CH_2 CH_2NH_2), 알릴(-CH_2-CH=CH_2), -O-알릴(--O-- CH_2-CH=CH_2) 및 플루오로(F)를 포함한다. 2'-당 치환체기는 아라비노(위) 위치 또는 리보(아래) 위치일 수 있다. 적합한 2'-아라비노 변형은 2'-F이다. 유사한 변형은 또한 올리고머 화합물 상의 다른 위치에서, 특히 3' 말단의 뉴클레오사이드 상의 당의 3' 위치에서 또는 2'-5' 연결된 올리고뉴클레오타이드 및 5' 말단 뉴클레오타이드의 5' 위치에서 만들어질 수 있다. 올리고머 화합물은 또한 펜토퓨라노실 당 대신 당 모방체, 예컨대 사이클로뷰틸 모이어티를 가질 수 있다.

[0232] ***염기 변형 및 치환***

[0233] 대상 핵산은 또한 핵염기(종종 본 명세서에서 단순히 "염기"로서 지칭됨) 변형 또는 치환을 포함할 수 있다. 본 명세서에서 사용되는 바와 같은, "비변형" 또는 "천연" 핵염기는 퓨린 염기 아데닌(A) 및 구아닌(G), 및 피리미딘 염기 티민(T), 사이토신(C) 및 유라실(U)을 포함한다. 변형된 핵염기는 다른 합성 및 천연 핵염기, 예컨대 5-메틸사이토신(5-me-C), 5-하이드록시메틸 사이토신, 잔틴, 하이포잔틴, 2-아미노아데닌, 아데닌 및 구아닌의 6-메틸 및 기타 알킬 유도체, 아데닌 및 구아닌의 2-프로필 및 기타 알킬 유도체, 2-티오유라실, 2-티오티민 및 2-티오사이토신, 5-할로유라실 및 사이토신, 5-프로핀일(-C=C-CH_3) 유라실 및 피리미딘 염기의 사이토신 및 다른 알킨일 유도체, 6-아조 유라실, 사이토신 및 티민, 5-유라실 (슈도유라실), 4-티오유라실, 8-할로, 8-아미노, 8-티올, 8-티오알킬, 8-하이드록실 및 다른 8-치환된 아데닌 및 구아닌, 5-할로, 특히 5-브로모, 5-트라이플루오로메틸 및 다른 5-치환된 유라실 및 사이토신, 7-메틸구아닌 및 7-메틸아데닌, 2-F-아데닌, 2-아미

노-아데닌, 8-아자구아닌 및 8-아자아데닌, 7-데아자구아닌 및 7-데아자아데닌 및 3-데아자구아닌 및 3-데아자아데닌을 포함한다. 추가로 변형된 핵염기는 삼환식 피리미딘, 예컨대 페녹사진 사이티딘(1H-피리미도(5,4-b)(1,4)벤족사진-2(3H)-온), 페노티아진 사이티딘(1H-피리미도(5,4-b)(1,4)벤조티아진-2(3H)-온), G-클램프, 예컨대 치환된 페녹사진 사이티딘(예를 들어, 9-(2-아미노에톡시)-H-피리미도(5,4-(b)(1,4)벤족사진-2(3H)-온), 카바졸 사이티딘(2H-피리미도(4,5-b)인돌-2-온), 피리도인돌 사이티딘(H-피리도(3',2':4,5)피롤로(2,3-d)피리미딘-2-온)을 포함한다.

[0234] 복소환식 염기 모이어티는 또한 퓨린 또는 피리미딘 염기가 다른 헤테로사이클, 예를 들어, 7-데아자-아데닌, 7-데아자구아노신, 2-아미노피리딘 및 2-피리돈으로 대체되는 것을 포함할 수 있다. 추가로 핵염기는 미국 특허 제3,687,808호에 개시된 것, 문헌[*The Concise Encyclopedia Of Polymer Science And Engineering*, 페이지 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990]에 개시된 것, 문헌[Englisch et al., *Angewandte Chemie, International Edition*, 1991, 30, 613]에 의해 개시된 것, 및 문헌[Sanghvi, Y. S., *Chapter 15, Antisense Research and Applications*, 페이지 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993]에 의해 개시된 것을 포함하며; 이의 개시내용은 본 명세서에 그들의 전문이 참고로 편입된다. 특정 이들 핵염기는 올리고머 화합물의 결합 친화도를 증가시키는 데 유용하다. 이들은 5-치환된 피리미딘, 6-아자피리미딘 및 N-2, N-6 및 O-6 치환된 퓨린(2-아미노프로필아데닌, 5-프로핀일유라실 및 5-프로핀일사이토신을 포함)을 포함한다. 5-메틸사이토신 치환은 0.6 내지 1.2℃만큼 핵산 듀플렉스 안정성을 증가시키는 것으로 나타났고(Sanghvi et al., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278; 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨), 예를 들어, 2'-O-메톡시에틸 당 변형과 조합될 때, 적합한 염기 치환이다.

[0235] ***접합체***

[0236] 대상 핵산의 다른 가능한 변형은 올리고뉴클레오타이드의 활성, 세포 분포 또는 세포 흡수를 향상시키는 하나 이상의 모이어티 또는 접합체를 폴리뉴클레오타이드에 화학적으로 연결하는 것을 수반한다. 이들 모이어티 또는 접합체는 작용기, 예컨대 1차 또는 2차 하이드록실기에 공유 결합된 접합체기를 포함할 수 있다. 접합체기는 삽입제, 리포터 분자, 폴리아민, 폴리아마이드, 폴리에틸렌 글리콜, 폴리에터, 올리고머의 약력학적 특성을 향상시키는 기, 및 올리고머의 약물동력학적 특성을 향상시키는 기를 포함하지만, 이들로 제한되지 않는다. 적합한 접합기는 콜레스테롤, 지질, 인지질, 바이오틴, 페나진, 엽산, 페난트리딘, 안트라퀴논, 아크리딘, 플루오레세인, 로다민, 쿠마린 및 염료를 포함하지만, 이들로 제한되지 않는다. 약력학적 특성을 향상시키는 기는 흡수를 개선시키고/시키거나, 분해에 대한 내성을 향상시키고/시키거나 표적 핵산과의 서열-특이적 혼성화를 강화시키는 기를 포함한다. 약물동태학적 특성을 향상시키는 기는 대상 핵산의 흡수, 분포, 대사 또는 배설을 개선시키는 기를 포함한다.

[0237] 접합 모이어티는 지질 모이어티, 예컨대 콜레스테롤 모이어티(Letsinger et al., *Proc. Natl. Acad. Sci.* USA, 1989, 86, 6553-6556), 콜산(Manoharan et al., *Bioorg. Med. Chem. Let.*, 1994, 4, 1053-1060), 티오에터, 예를 들어, 헥실-S-트라이틸티올(Manoharan et al., *Ann. N.Y. Acad. Sci.*, 1992, 660, 306-309; Manoharan et al., Bioorg. *Med. Chem. Let.*, 1993, 3, 2765-2770), 티오콜레스테롤(Oberhauser et al., *Nucl. Acids Res.*, 1992, 20, 533-538), 지방족 쇄, 예를 들어, 도데칸다이올 또는 운데실 잔기(Saison-Behmoaras et al., *EMBO J.*, 1991, 10, 1111-1118; Kabanov et al., *FEBS Lett.*, 1990, 259, 327-330; Svinarchuk et al., *Biochimie*, 1993, 75, 49-54), 인지질, 예를 들어, 다이-헥사데실-rac-글리세롤 또는 트라이에틸암모늄 1,2-다이-O-헥사데실-rac-글리세로-3-H-포스포네이트(Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654; Shea et al., *Nucl. Acids Res.*, 1990, 18, 3777-3783), 폴리아민 또는 폴리에틸렌 글리콜 쇄(Manoharan et al., *Nucleosides & Nucleotides*, 1995, 14, 969-973), 또는 아다만탄 아세트산(Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654), 팔미틸 모이어티(Mishra et al., *Biochim. Biophys. Acta*, 1995, 1264, 229-237), 또는 옥타데실아민 또는 헥실아미노-카보닐-옥시콜레스테롤 모이어티(Crooke et al., *J. Pharmacol. Exp. Ther.*, 1996, 277, 923-937)를 포함하지만, 이들로 제한되지 않는다.

[0238] 접합체는 "단백질 형질도입 도메인" 또는 PTD(또한 CPP - 세포 침투성 펩타이드로서 알려짐)를 포함하는데, 이는 지질 이중층, 마이셀, 세포막, 세포소기관 막 또는 소수포 막을 가로지르는 것을 용이하게 하는 폴리펩타이드, 폴리뉴클레오타이드, 탄수화물 또는 유기 또는 무기 화합물을 지칭할 수 있다. 거대 분자 및/또는 나노분자의 범위일 수 있는 다른 분자에 부착된 PTD는 분자가 막을 가로지르는 것, 예를 들어, 세포외 공간으로부터 세포내 공간 또는 사이토졸을 세포소기관(예를 들어, 핵) 내로 이동하는 것을 용이하게 한다. 일부 경우에, PTD는 외인성 폴리뉴클레오타이드의 3' 말단에 공유적으로 연결된다. 일부 경우에, PTD는 외인성 폴리뉴클레오타이드

의 5' 말단에 공유적으로 연결된다. 예시적인 PTD는 최소 운데카펩타이드 단백질 형질도입 도메인(YGRKKRRQRRR; 서열번호 130를 포함하는 HIV-1 TAT의 잔기 47 내지 57에 대응); 세포 내로 유입을 지시하기에 충분한 다수의 알기닌을 포함하는 폴리알기닌 서열(예를 들어, 3, 4, 5, 6, 7, 8, 9, 10 또는 10 내지 50개의 알기닌); VP22 도메인(Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); 초파리 안테나페디아(Antennapedia) 단백질 형질도입 도메인(Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); 절단된 인간 칼시토닌 펩타이드(Trehin et al. (2004) *Pharm. Research* 21:1248-1256); 폴리라이신(Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR 서열번호 131); 트랜스포탄 GWTLNSAGYLLGKINLKALAALAKKIL 서열번호 132); KALAWEAKLAKALAKALAKHLAKALAKALKCEA 서열번호 133); 및 RQIKIWFQNRRMKWKK 서열번호 134)을 포함하지만, 이들로 제한되지 않는다. 예시적인 PTD는 YGRKKRRQRRR 서열번호 130), RKKRRQRRR 서열번호 135); 3개의 알기닌 잔기 내지 50개의 알기닌 잔기의 알기닌 동종중합체를 포함하지만, 이들로 제한되지 않고; 예시적인 PTD 도메인 아미노산 서열은 다음 중 어떤 것을 포함하지만, 이들로 제한되지 않는다: YGRKKRRQRRR 서열번호 130); RKKRRQRR 서열번호 136); YARAAARQARA 서열번호 137); THRLPRRRRRR 서열번호 138); 및 GGRRARRRRRR 서열번호 139). 일부 경우에, PTD는 활성화 가능한 CPP(ACPP)이다(Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPP는 매칭 절단 가능한 링커를 통해 다음이온(예를 들어, Glu9 또는 "E9")에 연결된 다양이온 CPP(예를 들어, Arg9 또는 "R9")를 포함하는데, 이는 순전하를 거의 0으로 감소시키고, 이에 의해 세포에 대한 접착 및 흡수를 저해한다. 링커의 절단 시, 다음이온이 방출되고, 다알기닌 및 그의 고유한 부착성을 국소로 가리움 해제하며, 따라서 막을 가로지르도록 ACPP를 "활성화시킨다".

[0239] ***표적 세포 내로의 성분의 도입***

[0240] CasZ 가이드 RNA(또는 이를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산) 및/또는 CasZ 폴리펩타이드(또는 이를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산) 및/또는 CasZ trancRNA(또는 이를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산) 및/또는 공여자 폴리뉴클레오타이드(공여자 주형)가 임의의 다양한 잘 공지된 방법에 의해 숙주 세포 내로 도입될 수 있다.

[0241] 임의의 다양한 화합물 및 방법은 본 개시내용의 CasZ 시스템의 표적세포에 전달하는 데 사용될 수 있다. 비제한적 예로서, 본 개시내용의 CasZ 시스템은 지질과 조합될 수 있다. 다른 비제한적 예로서, 본 개시내용의 CasZ 시스템은 입자와 조합될 수 있거나 또는 입자 내로 제형화될 수 있다.

[0242] 숙주 세포 내로 핵산을 도입하는 방법은 당업계에 공지되어 있고, 임의의 편리한 방법은 대상 핵산(예를 들어, 발현 작제물/벡터)을 표적 세포(예를 들어, 원핵 세포, 진핵 세포, 식물 세포, 동물 세포, 포유류 세포, 인간 세포 등) 내로 도입하는 데 사용될 수 있다. 적합한 방법은, 예를 들어, 바이러스 감염, 형질감염, 접합, 프로토플라스트 융합, 리포펙션, 전기천공법, 인산칼슘 침전, 폴리에틸렌이민(PEI)-매개 형질감염, DEAE-덱스트란 매개 형질감염, 리포좀-매개 형질감염, 입자총 기법, 인산칼슘 침전, 직접 미량주사법, 나노입자-매개 핵산 전달(예를 들어, 문헌[Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023] 참조)등을 포함한다.

[0243] 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드(예를 들어, 야생형 단백질, 변이체 단백질, 키메라/융합 단백질, dCasZ, etc.)는 CasZ 폴리펩타이드를 암호화하는 핵산(예를 들어, mRNA, DNA, 플라스미드, 발현 벡터, 바이러스 벡터 등)으로서 제공된다. 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드는 단백질로서(예를 들어, 관련된 가이드 RNA 없이 또는 회합 가이드 RNA와 함께, 즉, 리보핵단백질로서) 직접 제공된다. 본 개시내용의 CasZ 폴리펩타이드는 임의의 편리한 방법에 의해 세포 내로 도입될 수 있으며(세포에 제공되며); 이러한 방법은 당업자에게 공지되어 있다. 예시적 예로서, 본 개시내용의 CasZ 폴리펩타이드는 (예를 들어, CasZ 가이드 RNA 또는 CasZ 가이드 RNA를 암호화하는 핵산과 함께 또는 이들 없이, 그리고 공여자 폴리뉴클레오타이드와 함께 또는 이것 없이, 그리고 CasZ trancRNA와 함께 그리고 이것 없이) 세포 내로 직접 주사될 수 있다. 다른 예로서, 본 개시내용의 CasZ 폴리펩타이드 및 CasZ 가이드 RNA(RNP)의 수행된 복합체는 (예를 들어, 주사를 통해, 뉴클레오펙션을 통해; 하나 이상의 성분에 접합된, 예를 들어, CasZ 단백질에 접합된, 가이드 RNA 등에 접합된, CasZ trancRNA에 접합된, 본 개시내용의 CasZ 폴리펩타이드 및 가이드 RNA에 접합된 단백질 형질도입 도메인(PTD)을 통해) 세포(예를 들어, 진핵 세포) 내로 도입될 수 있다.

[0244] 일부 경우에, 핵산(예를 들어, CasZ 가이드 RNA 및/또는 이를 암호화하는 핵산, CasZ 단백질을 암호화하는 핵산, CasZ trancRNA 및/또는 이를 암호화하는 핵산 등) 및/또는 폴리펩타이드(예를 들어, CasZ 폴리펩타이드; CasZ 융합 폴리펩타이드)는 입자에서 세포(예를 들어, 표적 숙주 세포)에 전달되거나, 입자와 결합된다. 일부 경우에, 본 개시내용의 CasZ 시스템은 입자 내 세포에 전달되거나 또는 입자와 회합된다. 용어 "입자" 및 나노

입자"는 적절하다면, 상호 호환적으로 사용될 수 있다. 예를 들어, 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및/또는 CasZ 가이드 RNA, 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 mRNA, 및 가이드 RNA를 포함하는 재조합 발현 벡터는 입자 또는 지질 외피를 이용하여 동시에 전달될 수 있고; 예를 들어, CasZ 폴리펩타이드 및/또는 CasZ 가이드 RNA 및/또는 trancRNA는, 예를 들어, 복합체(예를 들어, 리보뉴클레오단백질(RNP) 복합체)로서, 예를 들어, 지질 또는 리피도이드 및 친수성 중합체, 예를 들어, 양이온성 지질 및 친수성 중합체를 포함하는 전달 입자를 통해 전달될 수 있되, 예를 들어, 양이온성 지질은 1,2-다이올레오일-3-트라이메틸암모늄-프로판(DOTAP) 또는 1,2-다이테트라데칸오일-sn-글리세로-3-포스포콜린(DMPC)을 포함하고/하거나 친수성 중합체는 에틸렌 글리콜 또는 폴리에틸렌 글리콜(PEG)을 포함하고; 그리고/또는 입자는 추가로 콜레스테롤을 포함한다(예를 들어, 제형 1로부터의 입자=DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0; 제형 번호 2=DOTAP 90, DMPC 0, PEG 10, 콜레스테롤 0; 제형 번호 3=DOTAP 90, DMPC 0, PEG 5, 콜레스테롤 5). 예를 들어, 입자는 CasZ 폴리펩타이드 및 CasZ 가이드RNA가, 예를 들어, 1:1 몰비에서, 예를 들어, 실온에서, 예를 들어, 30분 동안, 예를 들어, 멸균, 무 뉴클레아제 1×인산염-완충 식염수(PBS) 중에서 함께 혼합되고; 그리고 별개로, 제형에 대해 적용 가능하다면 DOTAP, DMPC, PEG 및 콜레스테롤은 알코올, 예를 들어, 100% 에탄올 중에서 용해되고; 그리고, 두 용액은 함께 혼합되어 복합체를 함유하는 입자를 형성하는 다단계 공정을 이용하여 형성될 수 있다.

[0245] 본 개시내용의 CasZ 폴리펩타이드(또는 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 mRNA; 또는 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터) 및/또는 CasZ 가이드 RNA(또는 CasZ 가이드 RNA를 암호화하는 하나 이상의 발현 벡터와 같은 핵산)는 입자 또는 지질 외피를 이용하여 동시에 전달될 수 있다. 예를 들어, 인지질 이중층 껍질로 뒤덮인 폴리(β 아미노 에스터)(PBAE) 코어를 갖는 생분해성 코어 껍질 구조의 나노입자를 사용할 수 있다. 일부 경우에, 자가 조립 생접착성 중합체에 기반한 입자/나노입자가 사용되고; 이러한 입자/나노입자는, 예를 들어, 뇌에 대해, 펩타이드의 경구 전달, 펩타이드의 정맥내 전달 및 펩타이드의 비강 전달에 적용될 수 있다. 다른 실시형태, 예컨대 경구 흡수 및 소수성 약물의 눈 전달이 또한 상정된다. 보호되고 질환 부위로 전달되는 조작된 중합체를 수반하는 분자 외피 기법이 사용될 수 있다. 약 5 mg/kg의 용량은 다양한 인자, 예를 들어, 표적 조직에 따라서 단일 또는 다회 용량으로 사용될 수 있다.

[0246] 리피도이드(Lipidoid) 화합물(예를 들어, 미국 특허 출원 제20110293703호에 기재됨)은 폴리뉴클레오타이드의 투여에서 유용하며, 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasZ 시스템을 전달하는 데 사용될 수 있다. 일 양상에서, 아미노 알코올 리피도이드 화합물은 세포 또는 대상에 전달될 제제와 조합되어 마이크로입자, 나노입자, 리포좀 또는 마이셀을 형성한다. 아미노알코올 리피도이드 화합물은 다른 아미노알코올 리피도이드 화합물, 중합체(합성 또는 천연), 계면활성제, 콜레스테롤, 탄수화물, 단백질, 지질 등과 조합되어 입자를 형성할 수 있다. 이어서, 이들 입자는 선택적으로 약제학적 부형제와 조합되어 약제학적 조성물을 형성할 수 있다.

[0247] 폴리(베타-아미노 알코올)(PBAA)은 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 미국 특허 공개 제20130302401호는 조합적 중합을 이용하여 제조된 폴리(베타-아미노 알코올)(PBAA) 부류에 관한 것이다.

[0248] 당-기반 입자는, 예를 들어, 참고로 WO2014118272(본 명세서에 참고로 편입됨) 및 문헌[Nair, J K et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961]에 기재된 바와 같은 GalNAc이 사용될 수 있고, 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하기 위해 사용될 수 있다.

[0249] 일부 경우에, 지질 나노입자(LNP)는 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 음으로 하전된 중합체, 예컨대 RNA는 낮은 pH 값(예를 들어, pH 4)에서 LNP에 부하될 수 있고, 여기서 이온화 가능한 지질은 양전하를 나타낸다. 그러나, 생리적 pH 값에서, LNP는 더 긴 순환 시간에 적합한 낮은 표면 전하를 나타낸다. 이온화 가능한 양이온성 지질의 4가지 종, 즉, 1,2-다이리놀레일-3-다이메틸암모늄-프로판(DLinDAP), 1,2-다이리놀레일옥시-3-N,N-다이메틸아미노프로판(DLinDMA), 1,2-다이리놀레일옥시-케토-N,N-다이메틸-3-아미노프로판(DLinKDMA), 및 1,2-다이리놀레일-4-(2-다이메틸아미노에틸)-[1,3]-다이옥솔란(DLinKC2-DMA)에 중점을 두었다. LNP의 제조는, 예를 들어, 문헌[Rosin et al. (2011) Molecular Therapy 19:1286-2200)]에 기재되어 있다. 양이온성 지질 1,2-다이리놀레일-

3-다이메틸암모늄-프로판(DLinDAP), 1,2-다이리놀레일옥시-3-N,N-다이메틸아미노프로판(DLinDMA), 1,2-다이리놀레일옥시케토-N,N-다이메틸-3-아미노프로판(DLinK-DMA), 1,2-다이리놀레일-4-(2-다이메틸아미노에틸)-[1,3]-다이옥솔란 (DLinKC2-DMA), (3-o-[2''-(메톡시폴리에틸렌글리콜 2000) 숙신오일]-1,2-다이미리스토일-sn-글리콜(PEG-S-DMG), 및 R-3-[(.오메가.-메톡시-폴리(에틸렌 글리콜)2000) 카바모일]-1,2-다이미리스틸옥실프로필-3-아민(PEG-C-DOMG)이 사용될 수 있다. 핵산(예를 들어, CasZ 가이드 RNA; 본 개시내용의 핵산 등)은 DLinDAP, DLinDMA, DLinK-DMA 및 DLinKC2-DMA(40:10:40:10 몰비로 양이온성 지질:DSPC:CHOL: PEGS-DMG 또는 PEG-C-DOMG)를 함유하는 LNP에서 캡슐화될 수 있다. 일부 경우에, 0.2% SP-DiOC18이 혼입된다.

[0250] 구형 핵산(SNA(상표명)) 작제물 및 다른 나노입자(특히 금 나노입자)는 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 예를 들어, 문헌[Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19): 7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) 및 Mirkin, et al., Small, 10:186-192] 참조.

[0251] RNA를 갖는 자기-조립 나노입자는 폴리에틸렌 글리콜(PEG)의 원위 말단에 부착된 Arg-Gly-Asp(RGD)펩타이드 리간드로 페길화된 폴리에틸렌이민(PEI)에 의해 작제될 수 있다.

[0252] 일반적으로, "나노입자"는 직경이 1000nm 미만인 임의의 입자를 지칭한다. 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA), 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 500nm 이하, 예를 들어, 25nm 내지 35nm, 35nm 내지 50nm, 50nm 내지 75nm, 75nm 내지 100nm, 100nm 내지 150nm, 150nm 내지 200nm, 200nm 내지 300nm, 300nm 내지 400nm, 또는 400nm 내지 500nm이다. 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 25nm 내지 200nm이다. 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 100nm 이하이고, 일부 경우에, 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 35nm 내지 60nm이다.

[0253] 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA), 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 상이한 형태, 예를 들어, 고체 나노입자(예를 들어, 금속, 예컨대 은, 금, 철, 티타늄), 비-금속, 지질계 고체, 중합체), 나노입자의 현탁액 또는 이들의 조합물로 제공될 수 있다. 금속, 유전체 및 반도체 나노입자뿐만 아니라 혼성 구조(예를 들어, 코어-껍질 나노입자)가 제조될 수 있다. 반도체 물질로 만들어진 나노입자는 또한, 그들이 전자 에너지 준위의 양자화가 일어나기에 충분히 작다면(전형적으로 10nm 미만), 표지된 양자점일 수 있다. 이러한 나노규모 입자는 약물 담체 또는 영상화제로서의 생의학적 적용분야에서 사용되고, 본 개시내용에서 유사한 목적을 위해 채택될 수 있다.

[0254] 반고체 및 연질 나노입자는 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 적합하다. 반고체 특성의 원형 나노입자는 리포좀이다.

[0255] 일부 경우에, 엑소좀은 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 엑소좀은 RNA 및 단백질을 수송하고, RNA를 뇌 및 다른 표적 기관에 전달할 수 있는 내인성 나노소포체이다.

[0256] 일부 경우에, 리포좀은 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 리포좀은 내부 수성 구획을 둘러싸는 단층 또는 다층 지질 이

중층 및 상대적으로 불침투성의 외부 친유성 인지질 이중층으로 구성된 구체 소수포 구조이다. 리포좀은 몇몇 상이한 유형의 지질로부터 이루어질 수 있지만, 그러나 인지질은 리포좀을 생성하는 데 가장 통상적으로 사용된다. 지질막이 수성 용액과 혼합될 때 리포좀 제형은 자발적이지만, 이는 또한 균질기, 초음파분해기 또는 압출 장치를 이용함으로써 진탕 형태로 힘을 적용함으로써 촉진될 수 있다. 이 구조 및 특성을 변형시키기 위래 몇몇 다른 첨가제가 리포좀에 첨가될 수 있다. 예를 들어, 리포좀 구조를 안정화시키기 위해 그리고 리포좀 내부 내용물의 누출을 방지하기 위해 콜레스테롤 또는 스핑고마이엘린 중 하나가 리포좀 혼합물에 첨가될 수 있다. 리포좀 제형은 주로 천연 인지질 및 지질, 예컨대 1,2-다이스테아릴-sn-글리세로-3-포스파티딜 콜린(DSPC), 스핑고마이엘린, 난황 포스파티딜콜린 및 모노시알로강글리오사이드를 포함할 수 있다.

[0257] 안정한 핵산-지질 입자(SNALP)는 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. SNALP 제형은 2:40:10:48 몰 백분율 비로 지질 3-N-[(메톡시폴리(에틸렌 글리콜) 2000) 카바모일]-1,2-다이미리스틸옥시-프로필아민(PEG-C-DMA), 1,2-다이리놀레일옥시-N,N-다이메틸-3-아미노프로판(DLinDMA), 1,2-다이스테아로일-sn-글리세로-3-포스포콜린 (DSPC) 및 콜레스테롤을 함유할 수 있다. SNALP 리포좀은 콜레스테롤/D-Lin-DMA/DSPC/PEG-C-DMA의 25:1 지질/siRNA 비 및 48/40/10/2 몰비를 이용하여 다이스테아로일포스파티딜콜린(DSPC), 콜레스테롤 및 siRNA와 함께 D-Lin-DMA 및 PEG-C-DMA를 제형화함으로써 제조될 수 있다. 얻어진 SNALP 리포좀은 크기가 약 80 내지 100nm일 수 있다. SNALP는 합성 콜레스테롤(시그마-알드리치(Sigma-Aldrich), 미국 미주리주 세인트루이스에 소재), 다이팔미토일포스파티딜콜린(아반티 폴라 인코포레이티드(Avanti Polar Lipids), 미국 앨리버마주 알라바스터에 소재), 3-N-[(w-메톡시 폴리(에틸렌 글리콜)2000)카바모일]-1,2-다이미리스틸옥시프로필아민 및 양이온성 1,2-다이리놀레일옥시-3-N,N다이메틸아미노프로판을 포함할 수 있다. SNALP는 합성 콜레스테롤(시그마-알드리치), 1,2-다이스테아로일-sn-글리세로-3-포스포콜린 (DSPC; 아반티 폴라 리피트 인코포레이티드), PEG-cDMA, 및 1,2-다이리놀레일옥시-3-(N;N-다이메틸)아미노프로판(DLinDMA)을 포함할 수 있다.

[0258] 다른 양이온성 지질, 예컨대 아미노 지질 2,2-다이리놀레일-4-다이메틸아미노에틸-[1,3]-다이옥솔란(DLin-KC2-DMA)은 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 다음의 지질 조성으로 수행된 소수포가 상정될 수 있다: 각각 몰비 40/10/40/10로 아미노 지질, 다이스테아로일포스파티딜콜린 (DSPC), 콜레스테롤 및 (R)-2,3-비스(옥타데실옥시)프로필-1-(메톡시 폴리(에틸렌 글리콜)2000)프로필카바메이트(PEG-지질), 및 대략 0.05 (w/w)의 FVII siRNA/총 지질비. 70 내지 90nm 범위의 좁은 입자 크기 분포 및 0.11.+-.0.04 (n=56)의 낮은 다분산지수를 보장하기 위해, 입자는 가이드 RNA를 첨가하기 전에 80nm 막을 통해 3회까지 압출될 수 있다. 상당히 강한 아미노 지질 16을 함유하는 입자가 사용될 수 있으며, 이때 4가지 지질 성분 16, DSPC, 콜레스테롤 및 PEG-지질 (50/10/38.5/1.5)의 몰비는 생체내 활성을 향상시키도록 추가로 최적화될 수 있다.

[0259] 지질은 본 개시내용의 CasZ 시스템 또는 이의 성분(들) 또는 이를 암호화하는 핵산으로 제형화되어 지질 나노입자(LNP)를 형성할 수 있다. 적합한 지질은 DLin-KC2-DMA4, C12-200 및 코지질 다이스테로일포스파티딜 콜린, 콜레스테롤 및 PEG-DMG를 포함하지만, 이들로 제한되지 않으며, 자발적 소수포 제형화 절차를 이용하여 본 개시내용의 CasZ 시스템, 또는 이의 성분으로 제형화될 수 있다. 성분 몰비는 약 50/10/38.5/1.5(DLin-KC2-DMA 또는 C12-200/다이스테로일포스파티딜 콜린/콜레스테롤/PEG-DMG)일 수 있다.

[0260] 미국 특허 공개 제20130252281호 및 제20130245107호 및 제20130244279호에서 추가로 기재되는 것과 같이 PLGA 마이크로스피어에서 캡슐화된 본 개시내용의 CasZ 시스템, 또는 이의 성분이 전달될 수 있다.

[0261] 초전하(Supercharged) 단백질이 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의 CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 과급된 단백질은 흔치않게 높은 양성 또는 음성 이론적 순전하를 갖는 조작된 또는 천연 유래 단백질의 부류이다. 초음성과 초양성으로 하전된 단백질은 둘 다 열적으로 또는 화학적으로 유도된 응집을 견뎌내는 능력을 나타낸다. 초양성으로 하전된 단백질은 또한 포유류 세포를 침투할 수 있다. 이들 단백질, 예컨대 플라스미드 DNA, RNA 또는 다른 단백질과 화물의 회합은 시험관내 그리고 생체내 둘 다에서 포유류 세포 내로 이들 거대분자의 기능적 전달을 가능하게 할 수 있다.

[0262] 세포 침투성 펩타이드(CPP)는 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA) 또는 본 개시내용의

CasZ 시스템을 표적 세포에 전달하는 데 사용될 수 있다. CPP는 전형적으로 양으로 하전된 아미노산, 예컨대 라이신 또는 알기닌의 낮은 상대적 존재비를 함유하거나 또는 극성/하전된 아미노산 및 비극성, 소수성 아미노산의 교번의 패턴을 함유하는 서열을 갖는 아미노산 조성물을 가진다.

[0263] 이식 가능한 장치는 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA)(예를 들어, CasZ 가이드 RNA, CasZ 가이드 RNA를 암호화하는 핵산, CasZ 폴리펩타이드를 암호화하는 핵산, 공여자 주형 등), 또는 본 개시내용의 CasZ 시스템을 표적 세포(예를 들어, 생체내 표적 세포, 여기서 표적 세포는 순환에서의 표적 세포, 조직 내 표적 세포, 기관 내 표적 세포임)에 전달하는 데 사용될 수 있다. 본 개시내용의 CasZ 폴리펩타이드, 본 개시내용의 CasZ 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasZ 가이드 RNA 및/또는 CasZ trancRNA), 또는 본 개시내용의 CasZ 시스템을 표적 세포(예를 들어, 생체내 표적 세포, 여기서 표적 세포는 순환에서의 표적 세포, 조직 내 표적 세포, 기관 내 표적 세포 등임)에 전달하는 데 사용하기에 적합한 이식 가능한 장치는 CasZ 폴리펩타이드, CasZ 융합 폴리펩타이드, RNP, 또는 CasZ 시스템(또는 이의 성분, 예를 들어, 본 개시내용의 핵산)을 포함하는 용기(예를 들어, 저장소, 매트릭스 등)를 포함할 수 있다.

[0264] 적합한 이식 가능한 장치는 중합체 기질, 예컨대 매트릭스(예를 들어, 장치 몸체로서 사용됨), 및 일부 경우에 추가적인 스캐폴드 물질, 예컨대 금속 또는 추가적인 중합체, 및 가시성 및 이미징을 향상시키기 위한 물질을 포함할 수 있다. 이식 가능한 전달 장치는 국소로 그리고 장기간에 걸쳐 방출을 제공하는 데 유리할 수 있고, 여기서 전달될 폴리펩타이드 및/또는 핵산은 표적 부위, 예를 들어, 세포외 기질(ECM), 종양을 둘러싸는 혈관, 질병 조직 등으로 직접 방출된다. 적합한 이식 가능한 전달 장치는 복강과 같은 내강에 대한 전달 및/또는 임의의 다른 유형의 투여에서 사용하기에 적합한 장치를 포함하며, 이때 생체 안정성 및/또는 분해성 및/또는 생체 흡수성 중합체 기질(예를 들어, 선택적으로 매트릭스일 수 있음)을 포함하는 약물 전달 시스템은 고정 또는 부착되지 않는다. 일부 경우에, 적합한 이식 가능한 약물 전달 장치는 분해성 중합체를 포함하되, 주요 방출 메커니즘은 벌크 침식(bulk erosion)이다. 일부 경우에, 적합한 이식 가능한 약물 전달 장치는 비분해성 또는 느리게 분해되는 중합체를 포함하되, 주요 방출 메커니즘은 벌크 침식보다는 확산성이며, 따라서 외부 부분은 막으로서 작용하고, 그의 내부 부분은 약물 저장소로서 작용하는데, 이는 사실상 장기간 동안(예를 들어, 약 1주 내지 약 몇 개월) 주변에 의해 영향받지 않는다. 상이한 중합체와 상이한 방출 메커니즘의 조합이 또한 선택적으로 사용될 수 있다. 농도 구배는 총 방출 기간의 상당한 기간 동안 효과적으로 유지될 수 있고, 따라서 확산 속도는 효과적으로 일정하다("제로 방식(zero mode)" 확산으로 칭해짐). 용어 "일정한"은 치료적 유효성의 하부 역치 초과로 유지되지만, 여전히 선택적으로 초기 버스트 확산 속도를 특징으로 할 수 있고/있거나, 예를 들어, 특정 정도로 증가 및 감소하면서 변동할 수 있는 확산 속도를 의미한다. 확산 속도는 장기간 동안 그렇게 유지될 수 있고, 이는 치료적으로 유효한 기간, 예를 들어, 유효한 침묵 기간을 최적화하기 위해 특정 수준에 대해 일정하게 고려될 수 있다.

[0265] 일부 경우에, 이식 가능한 전달 시스템은, 천연에서 화학물질이든 또는 대상체의 신체에서 효소 및 다른 인자로부터의 공격에 기인하든 뉴클레오타이드계 치료제가 분해되는 것을 막도록 설계된다.

[0266] 장치의 이식을 위한 부위 또는 표적 부위는 최대 치료 효능을 위해 선택될 수 있다. 예를 들어, 전달 장치는 종양 환경, 또는 종양과 관련된 혈액 공급 내에서 또는 이의 근위에 이식될 수 있다. 표적 위치는, 예를 들어 하기일 수 있다: 1) 기저핵, 백질 및 회백질에서 파킨슨 또는 알츠하이머병에서와 같이 퇴행성 부위에서의 뇌; 2) 근위축성측색경화증(ALS)의 경우에 척추; 3) 자궁경부; 4) 활성 및 만성 염증성 관절염; 5) 건선의 경우에 피부; 7) 알레르기 효과에 대한 교감신경성 및 감각 신경 부위; 7) 뼈; 8) 급성 또는 만성 감염 부위; 9) 질내; 10) 내이--청각계, 내이의 미로, 전정계; 11) 기관내; 12) 심장내; 관상동맥, 심외막; 13) 비뇨기관 또는 방광; 14) 담도계; 15) 신장, 표, 비장을 포함하지만, 이들로 제한되지 않는 실질세포; 16) 림프절; 17) 침샘; 18) 덴탈검(dental gum); 19) 관절강내(관절내)호; 제20) 안구내; 21) 뇌조직; 22) 뇌실; 23) 복강을 포함하는 강(예를 들어, 제한되는 것은 아니지만, 난소암에 대해); 24) 식도내; 및 25) 직장내; 및 26) 혈관구조내.

[0267] 삽입, 예컨대 이식 방법은 선택적으로 변형 없이 또는 대안적으로 선택적으로 이러한 방법에서 중요하지 않은 변형만으로 다른 유형의 조직 이식을 위해 그리고/또는 삽입을 위해 그리고/또는 조직 샘플링을 위해 선택적으로 이미 사용될 수 있다. 이러한 방법은 선택적으로 브라키테라피 방법, 생검, 초음파와 함께 그리고/또는 초음파 없이 내시경검사, 예컨대 뇌 조직 내로 주촉성 방법, 관절, 복부 기관, 방광벽 및 체강 내로 복강경의 이식을 포함하는 복강경검사를 포함하지만, 이들로 제한되지 않는다.

[0268] **변형된 숙주 세포**

[0269] 본 개시내용은 본 개시내용의 CasZ 폴리펩타이드를 포함하는 변형된 세포 및/또는 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 제공한다. 본 개시내용은 본 개시내용의 CasZ 폴리펩타이드를 포함하는 변형된 세포를 제공하며, 여기서 변형된 세포는 본 개시내용의 CasZ 폴리펩타이드를 정상적으로 포함하지 않는 세포이다. 본 개시내용은 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 포함하는 변형된 세포(예를 들어, 유전자 변형된 세포)를 제공한다. 본 개시내용은 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 mRNA로 유전자 변형된 유전자 변형 세포를 제공한다. 본 개시내용은 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터로 유전자 변형된 유전자 변형 세포를 제공한다. 본 개시내용은: a) 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; 및 b) 본 개시내용의 CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터로 유전자 변형된 유전자 변형 세포를 제공한다. 본 개시내용은: a) 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; b) 본 개시내용의 CasZ 가이드 RNA를 암호화하는 뉴클레오타이드 서열; 및 c) 공여자 주형을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터로 유전자 변형된 유전자 변형 세포를 제공한다.

[0270] 본 개시내용의 CasZ 폴리펩타이드 및/또는 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 및/또는 본 개시내용의 CasZ 가이드 RNA(또는 이를 암호화하는 핵산) 및/또는 CasZ trancRNA(또는 이를 암호화하는 핵산)에 대한 수용자로서 작용하는 세포는, 예를 들어, 시험관내 세포; 생체내 세포; 생체외 세포; 1차 세포; 암 세포; 동물 세포; 식물 세포; 조류 세포; 진균 세포 등을 포함하는 임의의 다양한 세포일 수 있다. 본 개시내용의 CasZ 폴리펩타이드 및/또는 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 및/또는 본 개시내용의 CasZ 가이드 RNA에 대한 수용자로서 작용하는 세포는 "숙주 세포" 또는 "표적 세포"로서 지칭된다. 숙주 세포 또는 표적 세포는 본 개시내용의 CasZ 시스템의 수용자일 수 있다. 숙주 세포 또는 표적 세포는 본 개시내용의 CasZ RNP 시스템의 수용자일 수 있다. 숙주 세포 또는 표적 세포는 본 개시내용의 CasZ 시스템의 단일 성분의 수용자일 수 있다.

[0271] 세포(표적 세포)의 비제한적 예는 원핵 세포, 진핵 세포, 박테리아 세포, 고세균 세포, 단세포 진핵 유기체의 세포, 원생동물 세포, 식물로부터의 세포(예를 들어, 식물 작물, 과일, 채소, 곡물, 대두, 옥수수, 마이스, 밀, 종자, 토마토, 벼, 카사바, 사탕수수, 호박, 건초, 감자, 목화, 대마초, 담배, 현화 식물, 구과 식물, 겉씨식물, 속씨식물, 양치식물, 석송, 뿔이끼류, 우산이끼류, 선태식물, 쌍떡잎식물, 외떡잎식물 등), 조류 세포, (예를 들어, 보트리오코쿠스 브라우니(*Botryococcus braunii*), 칼라마이도모나스 레인하르티(*Chlamydomonas reinhardtii*), 난노클로롭시스 가디타나(*Nannochloropsis gaditana*), 클로렐라 피레노이도사(*Chlorella pyrenoidosa*), 모자반속(*Sargassum patens*, *C. agardh*) 등), 해초(예를 들어, 켈프) 진균 세포(예를 들어, 효모 세포, 버섯으로부터의 세포), 동물 세포, 무척추동물(예를 들어, 과실파리, 강장동물, 극피동물, 선충 등)으로부터의 세포, 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포, 포유류(예를 들어, 유제류(예를 들어, 돼지, 소, 염소, 양)로부터의 세포; 설치류(예를 들어, 래트, 마우스); 비-인간 영장류; 인간; 고양이과(예를 들어, 고양이); 갯과(예를 들어, 개); 등) 등을 포함한다. 일부 경우에, 세포는 천연 유기체로부터 유래되지 않은 세포이다(예를 들어, 세포는 합성으로 만들어진 세포일 수 있고; 또한 인공 세포로서 지칭된다).

[0272] 세포는 시험관내 세포일 수 있다(예를 들어, 배양물 내 세포, 예를 들어, 확립된 배양 세포주). 세포는 생체외 세포일 수 있다(개체로부터의 배양 세포). 세포는 생체내 세포(예를 들어, 개체 내 세포)일 수 있다. 세포는 단리된 세포일 수 있다. 세포는 유기체 내부의 세포일 수 있다. 세포는 유기체일 수 있다. 세포는 세포 배양물 내 세포일 수 있다(예를 들어, 시험관내 세포 배양물). 세포는 세포 수집물 중 하나일 수 있다. 세포는 원핵 세포이거나 또는 원핵 세포로부터 유래될 수 있다. 세포는 박테리아 세포일 수 있거나 또는 박테리아 세포로부터 유래될 수 있다. 세포는 고세균 세포일 수 있거나 또는 고세균 세포로부터 유래될 수 있다. 세포는 진핵 세포일 수 있거나 또는 진핵 세포로부터 유래될 수 있다. 세포는 식물 세포일 수 있거나 또는 식물 세포로부터 유래될 수 있다. 세포는 동물 세포일 수 있거나 또는 동물 세포로부터 유래될 수 있다. 세포는 무척추동물 세포일 수 있거나 또는 무척추동물 세포로부터 유래될 수 있다. 세포는 척추동물 세포일 수 있거나 또는 척추동물 세포로부터 유래될 수 있다. 세포는 포유류 세포일 수 있거나 또는 포유류 세포로부터 유래될 수 있다. 세포는 설치류 세포일 수 있거나 또는 설치류 세포로부터 유래될 수 있다. 세포는 인간 세포일 수 있거나 또는 인간 세포로부터 유래될 수 있다. 세포는 미생물 세포일 수 있거나 또는 미생물 세포로부터 유래될 수 있다. 세포는 진균 세포일 수 있거나 또는 진균 세포로부터 유래될 수 있다. 세포는 곤충 세포일 수 있다. 세포는 절지동물 세포일 수 있다. 세포는 원생동물 세포일 수 있다. 세포는 유충 세포일 수 있다.

[0273] 적합한 세포는 줄기 세포(예를 들어, 배아 줄기(ES) 세포, 유도 만능 줄기(iPS) 세포; 생식세포(예를 들어, 난모세포, 정자, 난원세포, 정원세포 등); 체세포, 예를 들어, 섬유아세포, 희소돌기신경교, 교세포, 조혈모세포, 뉴런, 근육 세포, 골세포, 간세포, 췌장 세포 등을 포함한다.

[0274] 적합한 세포는 인간 배아 줄기 세포, 태아 심장 근육 세포, 근육섬유아세포, 간충직 줄기 세포, 자가이식 확장된 심장 근육 세포, 지방세포, 전분화 세포, 만능 세포, 혈액 줄기 세포, 근아세포, 성체 줄기 세포, 골수세포, 간충직 세포, 배아 줄기 세포, 실질 세포, 상피세포, 내피세포, 중피세포, 섬유아세포, 조골세포, 연골세포, 외인성 세포, 내인성 세포, 줄기 세포, 조혈모세포, 골수 유래 전구체 세포, 심근세포, 골격 세포, 태아 세포, 미분화 세포, 다능 전구체 세포, 단분화능 전구체 세포, 단핵구, 심장 근아세포, 골격 근아세포, 대식세포, 모세혈관 내피세포, 이종 세포, 동종 세포 및 출산후 줄기 세포를 포함한다.

[0275] 일부 경우에, 세포는 면역세포, 뉴런, 상피세포 및 내피세포 또는 줄기 세포이다. 일부 경우에, 면역세포는 T 세포, B 세포, 단핵구, 자연살해 세포, 수지상 세포 또는 대식세포이다. 일부 경우에, 면역세포는 세포독성 T 세포이다. 일부 경우에, 면역세포는 헬퍼 T 세포이다. 일부 경우에, 면역세포는 조절 T 세포(Treg)이다.

[0276] 일부 경우에, 세포는 줄기 세포이다. 줄기 세포는 성체 줄기 세포를 포함한다. 성체 줄기 세포는 또한 체세포 줄기 세포로서 지칭된다.

[0277] 성체 줄기 세포는 분화된 조직 내에 존재하지만, 자기 재생 특성 및 다중 세포 유형, 보통 줄기 세포가 발견된 전형적인 조직의 세포 유형이 생기게 하는 능력을 보유한다. 체세포 줄기 세포의 수많은 예는 당업자에게 공지되어 있으며, 근육 줄기 세포; 조혈모세포; 상피 줄기 세포; 신경 줄기 세포; 간충직 줄기 세포; 유방 줄기 세포; 장 줄기 세포; 중배엽 줄기 세포; 내피 줄기 세포; 후각 줄기 세포; 신경 능선 줄기 세포 등을 포함한다.

[0278] 관심 대상의 줄기 세포는 포유류 줄기 세포를 포함하며, 용어 "포유류"는 인간; 비-인간 영장류; 가축 및 농장 동물; 및 동물원, 실험실, 운동 또는 반려 동물, 예컨대 개, 말, 고양이, 소, 마우스, 래트, 토끼 등을 포함하는 포유류로서 분류된 임의의 동물을 지칭한다. 일부 경우에, 줄기 세포는 인간 줄기 세포이다. 일부 경우에, 줄기 세포는 설치류(예를 들어, 마우스; 래트) 줄기 세포이다. 일부 경우에, 줄기 세포는 비-인간 영장류 줄기 세포이다.

[0279] 줄기 세포는 하나 이상의 줄기 세포 마커, 예를 들어, SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17 및 PPARGC1A를 발현시킬 수 있다.

[0280] 일부 경우에, 줄기 세포는 조혈모세포(HSC)이다. HSC는 골수, 혈액, 제대혈, 태아간 및 난황낭으로부터 단리될 수 있는 중배엽-유래 세포이다. HSC는 $CD34^+$ 및 $CD3^-$으로서 특성규명된다. HSC는 생체내 적혈구, 호중성-대식세포, 거핵구 및 림프계 조혈세포 계통을 재생시킬 수 있다. 시험관내에서, HSC는 적어도 일부의 자기-재생 세포 분열을 겪도록 유도될 수 있고, 생체내에서 알 수 있는 바와 같이 동일한 계통으로 분화되도록 유도될 수 있다. 이렇게 해서, HSC는 적혈구 세포, 거핵구, 호중구, 대식세포 및 림프구 세포 중 하나 이상으로 분화되도록 유도될 수 있다.

[0281] 다른 경우에, 줄기 세포는 신경 줄기 세포(NSC)이다. 신경 줄기 세포(NSC)는 뉴런, 및 교세포(희소돌기신경교 및 성상교세포를 포함)로 분화할 수 있다. 신경 줄기 세포는 다회 분열할 수 있는 다능성 줄기 세포이고, 특정 조건 하에서 신경아세포 또는 아교모세포, 예를 들어, 각각 뉴런 및 교세포 중 하나 이상의 유형이 되도록 기여하는 세포일 수 있는 신경 줄기 세포 또는 신경 전구체 세포인 딸 세포를 생성할 수 있다. NSC를 얻는 방법은 당업계에 공지되어 있다.

[0282] 다른 경우에, 줄기 세포는 간충직 줄기 세포(MSC)이다. MSC는 본래 배아 중배엽으로부터 유래되고, 성체 골수로부터 단리되며, 근육, 뼈, 연골, 지방, 골수 기질 및 힘줄을 형성하도록 분화될 수 있다. MSC를 단리시키는 방법은 당업계에 공지되어 있으며; MSC를 얻기 위한 임의의 공지된 방법이 사용될 수 있다. 예를 들어, 인간 MSC의 단리를 기재하는 미국 특허 제5,736,396호 참조.

[0283] 세포는 일부 경우에 식물 세포이다. 식물 세포는 외떡잎식물의 세포일 수 있다. 세포는 쌍떡잎식물의 세포일 수 있다.

[0284] 일부 경우에, 세포는 식물 세포이다. 예를 들어, 세포는 주요 농작물, 예를 들어, 보리, 콩류(건조 식용), 카놀라, 옥수수, 목화(피마), 목화(업랜드), 아마씨, 건초(알팔파), 건초(비-알팔파), 귀리, 땅콩, 벼, 수수, 대두, 사탕무, 사탕수수, 해바라기(오일), 해바라기(비-오일), 고구마, 담배(벌리), 담배(연관처리(Flue-cured)), 토마토, 밀(듀럼), 밀(봄), 밀(겨울) 등의 세포일 수 있다. 다른 예로서, 세포는, 예를 들어, 알팔파, 알로에 잎,

애로루트, 벗풀, 아티초크, 아스파라거스, 죽순, 바나나꽃, 콩나물, 콩류, 비이트톱, 근대, 비타멜론, 청경채, 브로콜리, 브로콜리 라베(엽경채), 방울다다기 양배추, 양배추, 방울 양배추, 가재발 선인장(노팔), 호리병박 열매, 카르둔, 당근, 콜리플라워, 셀러리, 차요테, 중국 아티초크(두루미냉이), 중국 양배추, 중국 셀러리, 중국 차이브, 채심, 국화잎(통호), 콜라드 그린, 옥수수대, 단옥수수, 오이, 대컨, 단데리온 그린, 타로토란, 다우 무에(dau mue)(피 팁(pea tip)), 동과(겨울멜론), 가지, 엔다이브, 에스카롤, 선수 장식 양치식물, 갓류 식물, 치커리, 갓 (중국 겨자), 가일론(gailon), 갈랑가(시암, 태국 생강), 마늘, 생강 뿌리, 우엉, 나물, 하노버 샐러드 나물, 후아우존틀, 예루살렘 아티초크, 히카마, 케일 나물, 콜라비, 명아주(꼘리떼), 상추(비브), 상추(보스턴), 상추(보스턴 레드), 상추(녹색잎), 상추(아이스버그), 상추(롤라로사), 상추(오크리프 - 녹색), 상추(오크리프 - 적색), 상추(가공처리), 상추(적색잎), 상추(로메인), 상추(루비 로메인), 상추(러시안 레드 겨자), 린쿡, 로 복, 긴콩류, 연근, 마하, 용설란(아가베) 잎, 말랑가, 메스꿀랭 믹스, 겨자과, 모압(수세미오이), 무(moo), 모쿠아(moqua)(동과), 버섯, 겨자, 참마, 오크라, 옹초이, 그린 어니언, 오포(opo)(롱 스쿼시), 장식용 옥수수, 관상용 호박, 파슬리, 파스닙, 완두콩, 후추(벨형), 후추, 호박, 라디키오, 무순, 무, 레이프 그린(rape greens), 레이프 그린, 대황, 로메인(베이비 레드), 루타베가, 함초(씨빈), 신쿠아(앵글드/줄기가 선 수세미), 시금치, 스쿼시, 짚 더미, 사탕수수, 고구마, 근대, 타마린드, 타로, 타로잎, 타로 슈트, 탓소이, 테페구아헤(tepeguaje)(구아헤), 빔바, 토마틸로, 토마토, 토마토 (체리), 토마토(포도형), 토마토(자두형), 강황, 순무 어린잎 나물, 순무, 마름, 얌피, 얌스(네임스), 유초이, 유카(카사바) 등을 포함하지만, 이들로 제한되지 않는 채소 작물의 세포이다.

[0285] 세포는 일부 경우에 절지동물 세포이다. 예를 들어, 세포는, 예를 들어, 협각류, 다지류, 헥시포디아(*Hexipodia*), 거미강, 곤충강, 돌좀목, 좀류, 구시아강, 하루살이목, 잠자리목, 잠자리하목, 실잠자리아목, 신시류, 외시류, 강도래목, 흰개미붙이목, 메뚜기, 민벌레목, 집게벌레목, 바퀴목, 귀뚜라미붙이목, 갈르와벌레과, 만토파스마토데과, 대벌레목, 바퀴목, 흰개미목, 사마귀목, 책좀, 다듬이벌레목, 총채벌레목, 이목, 반시목, 내시류 또는 완전변태류, 막시류, 초시류, 부채벌레목, 약대벌레목, 뱀잠자리목, 풀잠자리목, 밑드리목, 버룩목, 파리목, 날도래류 또는 인시목의 아목, 과, 아과, 군, 하위군, 또는 종의 세포일 수 있다.

[0286] 세포는 일부 경우에 곤충 세포이다. 예를 들어, 일부 경우에, 세포는 모기, 메뚜기, 노린재류, 파리, 벼룩, 벌, 말벌, 개미, 이, 나방 또는 딱정벌레의 세포이다.

[0287] **키트**

[0288] 본 개시내용은 본 개시내용의 CasZ 시스템 또는 본 개시내용의 CasZ 시스템의 성분을 포함하는 키트를 제공한다.

[0289] 본 개시내용의 키트는 CasZ 시스템(예를 들어, 상기 참조)에 열거한 바와 같은 임의의 조합을 포함할 수 있다. 본 개시내용의 키트는 a) 본 개시내용의 CasZ 시스템, 또는 본 개시내용의 CasZ 시스템의 상기 기재한 바와 같은 성분; 및 b) 1종 이상의 추가적인 시약, 예를 들어, i) 완충제; ii) 프로테아제 저해제; iii) 뉴클레아제 저해제; iv) 검출 가능한 표지를 개발하거나 또는 시각화하는 데 필요한 시약; v) 양성 및/또는 음성 대조군 표적 DNA; vi) 양성 및/또는 음성 대조군 CasZ 가이드 RNA; vii) CasZ trancRNA 등을 포함할 수 있다. 본 개시내용의 키트는 a) 상기 기재한 바와 같은 본 개시내용의 CasZ 시스템의 성분, 또는 본 개시내용의 CasZ 시스템; 및 b) 치료제를 포함할 수 있다.

[0290] 본 개시내용의 키트는 a) 표적 핵산에서 표적 뉴클레오타이드 서열에 혼성화되는 CasZ 가이드 RNA의 일부를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 삽입하기 위한 삽입 부위; 및 b) CasZ 가이드 RNA의 CasZ-결합 부분을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터를 포함할 수 있다. 본 개시내용의 키트는 a) 표적 핵산에서 표적 뉴클레오타이드 서열에 혼성화되는 CasZ 가이드 RNA의 일부를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 삽입하기 위한 삽입 부위; b) CasZ 가이드 RNA의 CasZ-결합 부분을 암호화하는 뉴클레오타이드 서열 및 c) 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터를 포함할 수 있다. 본 개시내용의 키트는 CasZ trancRNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터를 포함할 수 있다.

[0291] **ssDNA의 검출**

[0292] 본 개시내용의 CasZ(Cas14) 폴리펩타이드는, 일단 표적 DNA(이중 또는 단일 가닥)의 검출에 의해 활성화되면, 비표적화된 단일 가닥 DNA(ssDNA)를 무차별적으로 절단할 수 있다. 일단 가이드 RNA가 표적 DNA의 표적 서열에 혼성화할 때(즉, 샘플이 표적 DNA, 예를 들어, 표적 ssDNA를 포함할 때) 생기는 가이드 RNA에 의해 CasZ(Cas1

4)가 활성화되면, 단백질은 ssDNA를 무차별적으로 절단하는 뉴클레아제가 된다(즉, 뉴클레아제는 비-표적 ssDNA, 즉, 가이드 RNA의 가이드 서열이 혼성화하지 않는 ssDNA를 절단한다). 따라서, 표적 DNA가 샘플에 존재할 때(예를 들어, 일부 경우에 역치 초과), 결과는 임의의 편리한 검출 방법을 이용하여(예를 들어, 표지된 단일 가닥 검출기 DNA를 이용하여) 검출될 수 있는 샘플 내 ssDNA의 절단이다. 일부 경우에, CasZ 폴리펩타이드는 CasZ 가이드 RNA에 추가로, 활성화를 위해 tranc RNA를 필요로 한다.

[0293] 샘플 내 표적 DNA(이중 가닥 또는 단일 가닥)를 검출하기 위한 조성물 및 방법이 제공된다. 일부 경우에, 단일 가닥(ssDNA)이고 가이드 RNA의 가이드 서열과 혼성화되지 않는 검출기 DNA가 사용된다(즉, 검출기 ssDNA는 비표적 ssDNA이다). 이러한 방법은 (a) 샘플을 (i) CasZ 폴리펩타이드; (ii) CasZ 폴리펩타이드에 결합하는 영역, 및 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA; 및 (iii) 단일 가닥이며 가이드 RNA의 가이드 서열과 혼성화되지 않는 검출기 DNA와 접촉시키는 단계; 및 (b) CasZ 폴리펩타이드에 의한 단일 가닥 검출기 DNA의 절단에 의해 생성되는 검출 가능한 신호를 측정하여, 표적 DNA를 검출하는 단계를 포함할 수 있다. 일부 경우에, 상기 방법은 (a) 샘플을 (i) CasZ 폴리펩타이드; (ii) CasZ 폴리펩타이드에 결합하는 영역, 및 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA; 및 (iii) CasZ tranc RNA; 및 (iv) 단일 가닥이며 가이드 RNA의 가이드 서열과 혼성화되지 않는 검출기 DNA와 접촉시키는 단계; 및 (b) CasZ 폴리펩타이드에 의한 단일 가닥 검출기 DNA의 절단에 의해 생성되는 검출 가능한 신호를 측정하여, 표적 DNA를 검출하는 단계를 포함한다. 상기 언급한 바와 같이, 일단 샘플이 가이드 RNA에 혼성화되는 표적 DNA를 포함할 때(즉, 샘플이 표적화된 표적 DNA를 포함함) 생기는 가이드 RNA에 의해 대상 CasZ 폴리펩타이드 단백질이 활성화되면, CasZ 폴리펩타이드는 활성화되고, 샘플에 존재하는 ssDNA(비-표적 ssDNA를 포함)를 비특이적으로 절단하는 엔도리보뉴클레아제로서 작용한다. 따라서, 표적화된 표적 DNA가 샘플에 존재할 때(예를 들어, 일부 경우에 역치 초과), 결과는 임의의 편리한 검출 방법을 이용하여(예를 들어, 표지된 검출기 ssDNA를 이용하여) 검출될 수 있는 샘플 내 ssDNA의 절단(비표적 ssDNA를 포함함)이다.

[0294] 또한 단일 가닥 DNA(ssDNA)(예를 들어, 비-표적 ssDNA)를 절단하기 위한 조성물 및 방법이 제공된다. 이러한 방법은 핵산의 집단을: (i) CasZ 폴리펩타이드; 및 (ii) CasZ 폴리펩타이드에 결합하는 영역 및 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA와 접촉시키는 단계를 포함하되, 상기 집단은 표적 DNA 및 복수의 비표적 ssDNA를 포함하고, CasZ 폴리펩타이드는 상기 복수의 비-표적 ssDNA를 절단한다. 이러한 방법은 핵산의 집단을: (i) CasZ 폴리펩타이드; 및 (ii) CasZ 폴리펩타이드에 결합하는 영역 및 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA, 및 (iii) CasZ tranc RNA와 접촉시키는 단계를 포함하되, 상기 집단은 표적 DNA 및 복수의 비표적 ssDNA를 포함하고, CasZ 폴리펩타이드는 상기 복수의 비-표적 ssDNA를 절단한다. 이러한 방법은, 예를 들어, 세포에서 외래 ssDNA(예를 들어, 바이러스 DNA)를 절단하는 데 사용될 수 있다.

[0295] 대상 방법의 접촉시키는 단계는 2가 금속 이온을 포함하는 조성물에서 수행될 수 있다. 접촉시키는 단계는 무세포 환경에서, 예를 들어, 세포 밖에서 수행될 수 있다. 접촉시키는 단계는 세포 내에서 수행될 수 있다. 접촉시키는 단계는 시험관내에서 수행될 수 있다. 접촉시키는 단계는 생체외에서 수행될 수 있다. 접촉시키는 단계는 생체내에서 수행될 수 있다.

[0296] 가이드 RNA는 RNA로서 또는 가이드 RNA를 암호화하는 핵산(예를 들어, DNA, 예컨대 재조합 발현 벡터)으로서 제공될 수 있다. tranc RNA는 가이드 RNA를 암호화하는 핵산(예를 들어, DNA, 예컨대 재조합 발현 벡터)으로서 제공될 수 있다. CasZ 폴리펩타이드는 단백질 그 자체로서 또는 단백질을 암호화하는 핵산(예를 들어, mRNA, DNA, 예컨대 재조합 발현 벡터)으로서 제공될 수 있다. 일부 경우에, 2 이상의(예를 들어, 3 이상, 4 이상, 5 이상 또는 6 이상의) 가이드 RNA가 제공될 수 있다. 일부 경우에, i) CasZ 가이드 RNA; 및 ii) tranc RNA(또는 단일-분자 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산)을 포함하는 단일-분자 RNA가 사용된다.

[0297] 일부 경우에 (예를 들어, 샘플을 가이드 RNA 및 CasZ 폴리펩타이드와 접촉시킬 때; 또는 샘플을 가이드 RNA, CasZ 폴리펩타이드 및 tranc RNA와 접촉시킬 때), 샘플은 측정하는 단계 전에 2시간 이하(예를 들어, 1.5시간 이하, 1시간 이하, 40분 이하, 30분 이하, 20분 이하, 10분 이하, 또는 5분 이하 또는 1분 이하) 동안 접촉된다. 예를 들어, 일부 경우에 샘플은 측정하는 단계 전에 40분 이하 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 20분 이하 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 10분 이하 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 5분 이하 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 1분 이하 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 50초 내지 60초 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 40초 내지 50초 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 30초 내지 40초 동안 접촉된다. 일부 경우에 샘플은 측정하는 단계 전에 20초 내지 30초 동안 접촉된다. 일부 경우에 샘플

은 측정하는 단계 전에 10초 내지 20초 동안 접촉된다.

[0298] 일부 경우에, 표적 DNA를 검출하기 위한 본 개시내용의 방법은 a) 샘플을 검출기 DNA의 트랜스 절단을 제공하는 조건 하에 가이드 RNA, CasZ 폴리펩타이드 및 검출기 DNA와 접촉시키는 단계로서, 샘플은 2시간 이하(예를 들어, 1.5시간 이하, 1시간 이하, 40분 이하, 30분 이하, 20분 이하, 10분 이하 또는 5분 이하, 또는 1분 이하) 동안 접촉되는, 상기 접촉시키는 단계; b) 검출기 RNA의 트랜스 절단을 제공하지 않는 조건 하에 일정 시간 동안 단계 (a)로부터의 샘플을 유지하는 단계; 및 c) 단계 (b)의 일정 시간 후에, CasZ 폴리펩타이드에 의해 단일 표준 검출기 DNA의 절단에 의해 생성된 검출 가능한 신호를 측정함으로써, 표적 DNA를 검출하는 단계를 포함한다. 검출기 DNA의 트랜스 절단을 제공하는 조건은 온도 조건, 예컨대 17℃ 내지 약 39℃(예를 들어, 약 37℃)를 포함한다. 검출기 DNA의 트랜스 절단을 제공하지 않는 조건은 10℃ 이하, 5℃ 이하, 4℃ 이하 또는 0℃의 온도를 포함한다.

[0299] 일부 경우에, 표적 DNA를 검출하기 위한 본 개시내용의 방법은 a) 샘플을 검출기 DNA의 트랜스 절단을 제공하는 조건 하에 가이드 RNA, tranc RNA, CasZ 폴리펩타이드 및 검출기 DNA와 접촉시키는 단계(또는 샘플을 i) 가이드 RNA 및 tranc RNA를 포함하는 단일-분자 RNA; i) CasZ 폴리펩타이드; 및 iii) 검출기 DNA와 접촉시키는 단계)로서, 샘플은 2시간 이하(예를 들어, 1.5시간 이하, 1시간 이하, 40분 이하, 30분 이하, 20분 이하, 10분 이하 또는 5분 이하, 또는 1분 이하) 동안 접촉되는, 상기 접촉시키는 단계; b) 검출기 RNA의 트랜스 절단을 제공하지 않는 조건 하에 일정 시간 동안 단계 (a)로부터의 샘플을 유지하는 단계; 및 c) 단계 (b)의 일정 시간 후에, CasZ 폴리펩타이드에 의해 단일 표준 검출기 DNA의 절단에 의해 생성된 검출 가능한 신호를 측정함으로써, 표적 DNA를 검출하는 단계를 포함한다. 검출기 DNA의 트랜스 절단을 제공하는 조건은 온도 조건, 예컨대 17℃ 내지 약 39℃(예를 들어, 약 37℃)를 포함한다. 검출기 DNA의 트랜스 절단을 제공하지 않는 조건은 10℃ 이하, 5℃ 이하, 4℃ 이하 또는 0℃의 온도를 포함한다.

[0300] 일부 경우에, 단일-가닥 검출기 DNA의 절단에 의해 생성된 검출 가능한 신호는 60분 이하 동안 생성된다. 예를 들어, 일부 경우에, 단일-가닥 검출기 DNA의 절단에 의해 생성된 검출 가능한 신호는 60분 이하, 45분 이하, 30분 이하, 15분 이하, 10분 이하 또는 5분 이하 동안 생성된다. 예를 들어, 일부 경우에, 단일-가닥 검출기 DNA의 절단에 의해 생성된 검출 가능한 신호는 1분 내지 60분, 예를 들어, 1분 내지 5분, 5분 내지 10분, 10분 내지 15분, 15분 내지 30분, 30분 내지 45분, 또는 45분 내지 60분의 일정 기간 동안 생성된다. 일부 경우에, 검출 가능한 신호가 생성된 후에(예를 들어, 60분 이하 동안 생성됨), 검출 가능한 신호의 생성은, 예를 들어, 샘플의 온도를 낮추는 것(예를 들어, 온도를 10℃ 이하, 5℃ 이하, 4℃ 이하 또는 0℃로 낮춤)에 의해, 샘플에 저해제를 첨가하는 것에 의해, 샘플을 동결건조시키는 것에 의해, 샘플을 40℃ 초과로 가열하는 것 등에 의해 중단될 수 있다. 측정하는 단계는 검출 가능한 신호의 생성이 중단된 후의 임의의 시간에 일어날 수 있다. 예를 들어, 측정하는 단계는 검출 가능한 신호의 생성이 중단된 후 5분 내지 48시간에 일어날 수 있다. 예를 들어, 측정하는 단계는 검출 가능한 신호의 생성이 중단된 후 5분 내지 15분, 15분 내지 30분, 30분 내지 60분, 1시간 내지 4시간, 4시간 내지 8시간, 8시간 내지 12시간, 12시간 내지 24시간, 24시간 내지 36시간 또는 36시간 내지 48시간에 일어날 수 있다. 측정하는 단계는 검출 가능한 신호의 생성이 중단된 후 48시간 초과에 일어날 수 있다.

[0301] 샘플 내 표적 DNA(단일-가닥 또는 이중-가닥)를 검출하기 위한 본 개시내용의 방법은 높은 민감도로 표적 DNA를 검출할 수 있다. 일부 경우에, 본 개시내용의 방법은 복수의 DNA를 포함하는(표적 DNA 및 복수의 비표적 DNA를 포함하는) 샘플에 존재하는 표적 DNA를 검출하기 위해 사용될 수 있으며, 표적 DNA는 1개 이상의 복제물/10^7개의 비-표적 DNA(예를 들어, 1개 이상의 복제물/10^6개의 비표적 DNA, 1개 이상의 복제물/10^5개의 비-표적 DNA, 1개 이상의 복제물/10^4개의 비-표적 DNA, 1개 이상의 복제물/10^3개의 비-표적 DNA, 1개 이상의 복제물/10^2개의 비-표적 DNA, 1개 이상의 복제물/50개의 비-표적 DNA, 1개 이상의 복제물/20개의 비-표적 DNA, 1개 이상의 복제물/10개의 비-표적 DNA, 또는 1개 이상의 복제물/5개의 비-표적 DNA)로 존재한다. 일부 경우에, 본 개시내용의 방법은 복수의 DNA를 포함하는(표적 DNA 및 복수의 비-표적 DNA를 포함하는) 샘플에 존재하는 표적 DNA를 검출하는 데 사용될 수 있으며, 표적 DNA는 1개 이상의 복제물/10^{18}개의 비-표적 DNA(예를 들어, 1개 이상의 복제물/10^{15}개의 비-표적 DNA, 1개 이상의 복제물/10^{12}개의 비-표적 DNA, 1개 이상의 복제물/10^9개의 비-표적 DNA, 1개 이상의 복제물/10^6개의 비-표적 DNA, 1개 이상의 복제물/10^5개의 비-표적 DNA, 1개 이상의 복제물/10^4개의 비-표적 DNA, 1개 이상의 복제물/10^3개의 비-표적 DNA, 1개 이상의 복제물/10^2개의 비-표적 DNA, 1개 이상의 복제물/50개의 비-표적 DNA, 1개 이상의 복제물/20개의 비-표적 DNA, 1개 이상의 복제물/10개의 비-표적 DNA, 또는 1

개 이상의 복제물/5개의 비-표적 DNA)로 존재한다.

[0302] 일부 경우에, 본 개시내용의 방법은 샘플에 존재하는 표적 DNA를 검출할 수 있으며, 표적 DNA는 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10개의 비-표적 DNA(예를 들어, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^6개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 또는 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA)로 존재한다.

[0303] 일부 경우에, 본 개시내용의 방법은 샘플에 존재하는 표적 DNA를 검출할 수 있으며, 표적 DNA는 1개의 복제물/10^{18}개의 비-표적 DNA 내지 1개의 복제물/10개의 비-표적 DNA(예를 들어, 1개의 복제물/10^{18}개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^{15}개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^{12}개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^9개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^6개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 또는 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA)로 존재한다.

[0304] 일부 경우에, 본 개시내용의 방법은 샘플에 존재하는 표적 DNA를 검출할 수 있으며, 표적 DNA는 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/100 비-표적 DNA(예를 들어, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/10^7개의 비-표적 DNA 내지 1개의 복제물/10^6개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/100개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/100개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^2개의 비-표적 DNA, 1개의 복제

물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 또는 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA)로 존재한다.

[0305] 일부 경우에, 샘플에서 표적 DNA를 검출하는 대상 방법에 대해 검출 역치는 10nM 이하이다. 따라서, 예를 들어, 표적 DNA는 10nM 이하의 농도로 샘플에 존재할 수 있다. 용어 "검출 역치"는 검출이 일어나기 위해 샘플에 존재하여야 하는 표적 DNA의 최소량을 기재하기 위해 본 명세서에서 사용된다. 따라서, 예로서, 검출 역치가 10nM일 때, 10nM 이상의 농도에서 샘플에 표적 DNA가 존재하는 경우에 신호가 검출될 수 있다. 일부 경우에, 본 개시내용의 방법은 5nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 1nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.5nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.1nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.05nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.01nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.005nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.001nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.0005nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.0001nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.00005nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 0.00001nM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 10pM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은1pM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 500fM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 250fM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 100fM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 50fM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 500aM(아토몰) 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 250aM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 100aM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 50aM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 10aM 이하의 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 1aM 이하의 검출 역치를 가진다.

[0306] 일부 경우에, (대상 방법에서 표적 DNA를 검출하기 위한) 검출 역치는 500fM 내지 1nM(예를 들어, 500fM 내지 500pM, 500fM 내지 200pM, 500fM 내지 100pM, 500fM 내지 10pM, 500fM 내지 1pM, 800fM 내지 1nM, 800fM 내지 500pM, 800fM 내지 200pM, 800fM 내지 100pM, 800fM 내지 10pM, 800fM 내지 1pM, 1pM 내지 1nM, 1pM 내지 500pM, 1pM 내지 200pM, 1pM 내지 100pM, 또는 1pM 내지 10pM)의 범위이다(농도는 표적 DNA가 검출될 수 있는 표적 DNA의 역치 농도를 지칭한다). 일부 경우에, 본 개시내용의 방법은 800fM 내지 100pM 범위에서 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 1pM 내지 10pM 범위에서 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 10fM 내지 500fM, 예를 들어, 10fM 내지 50fM, 50fM 내지 100fM, 100fM 내지 250fM, 또는 250fM 내지 500fM 범위에서 검출 역치를 가진다.

[0307] 일부 경우에, 표적 DNA가 샘플에서 검출될수 있는 최소 농도는 500fM 내지 1nM(예를 들어, 500fM 내지 500pM, 500fM 내지 200pM, 500fM 내지 100pM, 500fM 내지 10pM, 500fM 내지 1pM, 800fM 내지 1nM, 800fM 내지 500pM, 800fM 내지 200pM, 800fM 내지 100pM, 800fM 내지 10pM, 800fM 내지 1pM, 1pM 내지 1nM, 1pM 내지 500pM, 1pM 내지 200pM, 1pM 내지 100pM, 또는 1pM 내지 10pM)의 범위에 있다. 일부 경우에, 표적 DNA가 샘플에서 검출될 수 있는 최소 농도는 800fM 내지 100pM의 범위에 있다. 일부 경우에, 표적 DNA가 샘플에서 검출될 수 있는 최소 농도는 1pM 내지 10pM의 범위에 있다.

[0308] 일부 경우에, (대상 방법에서 표적 DNA를 검출하기 위한) 검출 역치는 1aM 내지 1nM(예를 들어, 1aM 내지 500pM, 1aM 내지 200pM, 1aM 내지 100pM, 1aM 내지 10pM, 1aM 내지 1pM, 100aM 내지 1nM, 100aM 내지 500pM, 100aM 내지 200pM, 100aM 내지 100pM, 100aM 내지 10pM, 100aM 내지 1pM, 250aM 내지 1nM, 250aM 내지 500pM, 250aM 내지 200pM, 250aM 내지 100pM, 250aM 내지 10pM, 250aM 내지 1pM, 500aM 내지 1nM, 500aM 내지 500pM, 500aM 내지 200pM, 500aM 내지 100pM, 500aM 내지 10pM, 500aM 내지 1pM, 750aM 내지 1nM, 750aM 내지 500pM, 750aM 내지 200pM, 750aM 내지 100pM, 750aM 내지 10pM, 750aM 내지 1pM, 1fM 내지 1nM, 1fM 내지 500pM, 1fM 내지 200pM, 1fM 내지 100pM, 1fM 내지 10pM, 1fM 내지 1pM, 500fM 내지 500pM, 500fM 내지 200pM, 500fM 내지 100pM, 500fM 내지 10pM, 500fM 내지 1pM, 800fM 내지 1nM, 800fM 내지 500pM, 800fM 내지 200pM, 800fM 내지 100pM, 800fM 내지 10pM, 800fM 내지 1pM, 1pM 내지 1nM, 1pM 내지 500pM, 1pM 내지 200pM, 1pM 내지 100pM, 또는 1pM 내지 10pM)의 범위에 있다(농도는 표적 DNA가 검출될 수 있는 표적 DNA의 역치 농도를 지칭한다). 일부 경우에, 본 개시내용의 방법은 1aM 내지 800aM 범위에서 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 50aM 내지 1pM 범위에서 검출 역치를 가진다. 일부 경우에, 본 개시내용의 방법은 50aM 내지 500fM 범위

에서 검출 역치를 가진다.

[0309] 일부 경우에, 표적 DNA는 1aM 내지 1nM(예를 들어, 1aM 내지 500pM, 1aM 내지 200pM, 1aM 내지 100pM, 1aM 내지 10pM, 1aM 내지 1pM, 100aM 내지 1nM, 100aM 내지 500pM, 100aM 내지 200pM, 100aM 내지 100pM, 100aM 내지 10pM, 100aM 내지 1pM, 250aM 내지 1nM, 250aM 내지 500pM, 250aM 내지 200pM, 250aM 내지 100pM, 250aM 내지 10pM, 250aM 내지 1pM, 500aM 내지 1nM, 500aM 내지 500pM, 500aM 내지 200pM, 500aM 내지 100pM, 500aM 내지 10pM, 500aM 내지 1pM, 750aM 내지 1nM, 750aM 내지 500pM, 750aM 내지 200pM, 750aM 내지 100pM, 750aM 내지 10pM, 750aM 내지 1pM, 1fM 내지 1nM, 1fM 내지 500pM, 1fM 내지 200pM, 1fM 내지 100pM, 1fM 내지 10pM, 1fM 내지 1pM, 500fM 내지 500pM, 500fM 내지 200pM, 500fM 내지 100pM, 500fM 내지 10pM, 500fM 내지 1pM, 800fM 내지 1nM, 800fM 내지 500pM, 800fM 내지 200pM, 800fM 내지 100pM, 800fM 내지 10pM, 800fM 내지 1pM, 1pM 내지 1nM, 1pM 내지 500pM, 1pM 내지 200pM, 1pM 내지 100pM, 또는 1pM 내지 10pM) 범위에서 샘플에 존재한다. 일부 경우에, 표적 DNA는 1aM 내지 800aM 범위에서 샘플에 존재한다. 일부 경우에, 표적 DNA는 50aM 내지 1aM 범위에서 샘플에 존재한다. 일부 경우에, 표적 DNA는 50aM 내지 500fM 범위에서 샘플에 존재한다.

[0310] 일부 경우에, 표적 DNA가 샘플에서 검출될 수 있는 최소 농도는 1aM 내지 1nM(예를 들어, 1aM 내지 500pM, 1aM 내지 200pM, 1aM 내지 100pM, 1aM 내지 10pM, 1aM 내지 1pM, 100aM 내지 1nM, 100aM 내지 500pM, 100aM 내지 200pM, 100aM 내지 100pM, 100aM 내지 10pM, 100aM 내지 1pM, 250aM 내지 1nM, 250aM 내지 500pM, 250aM 내지 200pM, 250aM 내지 100pM, 250aM 내지 10pM, 250aM 내지 1pM, 500aM 내지 1nM, 500aM 내지 500pM, 500aM 내지 200pM, 500aM 내지 100pM, 500aM 내지 10pM, 500aM 내지 1pM, 750aM 내지 1nM, 750aM 내지 500pM, 750aM 내지 200pM, 750aM 내지 100pM, 750aM 내지 10pM, 750aM 내지 1pM, 1fM 내지 1nM, 1fM 내지 500pM, 1fM 내지 200pM, 1fM 내지 100pM, 1fM 내지 10pM, 1fM 내지 1pM, 500fM 내지 500pM, 500fM 내지 200pM, 500fM 내지 100pM, 500fM 내지 10pM, 500fM 내지 1pM, 800fM 내지 1nM, 800fM 내지 500pM, 800fM 내지 200pM, 800fM 내지 100pM, 800fM 내지 10pM, 800fM 내지 1pM, 1pM 내지 1nM, 1pM 내지 500pM, 1pM 내지 200pM, 1pM 내지 100pM, 또는 1pM 내지 10pM)의 범위에 있다. 일부 경우에, 표적 DNA가 샘플에서 검출될 수 있는 최소 농도는 1aM 내지 500pM의 범위에 있다. 일부 경우에, 표적 DNA가 샘플에서 검출될 수 있는 최소 농도는 100aM 내지 500pM의 범위에 있다.

[0311] 일부 경우에, 표적 DNA는 1aM 내지 1nM(예를 들어, 1aM 내지 500pM, 1aM 내지 200pM, 1aM 내지 100pM, 1aM 내지 10pM, 1aM 내지 1pM, 100aM 내지 1nM, 100aM 내지 500pM, 100aM 내지 200pM, 100aM 내지 100pM, 100aM 내지 10pM, 100aM 내지 1pM, 250aM 내지 1nM, 250aM 내지 500pM, 250aM 내지 200pM, 250aM 내지 100pM, 250aM 내지 10pM, 250aM 내지 1pM, 500aM 내지 1nM, 500aM 내지 500pM, 500aM 내지 200pM, 500aM 내지 100pM, 500aM 내지 10pM, 500aM 내지 1pM, 750aM 내지 1nM, 750aM 내지 500pM, 750aM 내지 200pM, 750aM 내지 100pM, 750aM 내지 10pM, 750aM 내지 1pM, 1fM 내지 1nM, 1fM 내지 500pM, 1fM 내지 200pM, 1fM 내지 100pM, 1fM 내지 10pM, 1fM 내지 1pM, 500fM 내지 500pM, 500fM 내지 200pM, 500fM 내지 100pM, 500fM 내지 10pM, 500fM 내지 1pM, 800fM 내지 1nM, 800fM 내지 500pM, 800fM 내지 200pM, 800fM 내지 100pM, 800fM 내지 10pM, 800fM 내지 1pM, 1pM 내지 1nM, 1pM 내지 500pM, 1pM 내지 200pM, 1pM 내지 100pM, 또는1pM 내지 10pM) 범위에서 샘플에 존재한다. 일부 경우에, 표적 DNA는 1aM 내지 500pM 범위에서 샘플에 존재한다. 일부 경우에, 표적 DNA는 100aM 내지 500pM 범위에서 샘플에 존재한다.

[0312] 일부 경우에, 대상 조성물 또는 방법은 아토몰(aM)의 검출 민감도를 나타낸다. 일부 경우에, 대상 조성물 또는 방법은 펨토몰(fM)의 검출 민감도를 나타낸다. 일부 경우에, 대상 조성물 또는 방법은 피코몰(pM)의 검출 민감도를 나타낸다. 일부 경우에, 대상 조성물 또는 방법은 나노몰(nM)의 검출 민감도를 나타낸다.

[0313] ***표적 DNA***

[0314] 표적 DNA는 단일 가닥(ssDNA) 또는 이중 가닥(dsDNA)일 수 있다. 표적 DNA가 단일 가닥일 때, 표적 DNA에서 PAM 서열에 대한 선호 또는 필요는 없다. 그러나, 표적 DNA가 dsDNA일 때, PAM은 보통 표적 DNA의 표적 서열에 인접하여 존재한다(예를 들어, 본 명세서의 다른 곳의 PAM의 논의를 참조). 표적 DNA의 공급원은, 예를 들어, 이하에 기재하는 바와 같은 샘플 공급원과 동일할 수 있다.

[0315] 표적 DNA의 공급원은 임의의 공급원일 수 있다. 일부 경우에, 표적 DNA는 바이러스 DNA(예를 들어, DNA 바이러스의 게놈 DNA)이다. 이렇게 해서, 대상 방법은 핵산 집단 중에서(예를 들어, 샘플에서) 바이러스 DNA의 존재를 검출하기 위한 것일 수 있다. 대상 방법은 또한 표적 DNA의 존재에서 비-표적 ssDNA의 절단을 위해 사용될 수 있다. 예를 들어, 방법이 세포에서 일어난다면, 대상 방법은 특정 표적 DNA가 세포에 존재할 때(예를 들어, 세포가 바이러스로 감염되고 바이러스 표적 DNA가 검출될 때) 세포 내 비-표적 ssDNA(가이드 RNA의 가이드 서열과 혼성화되지 않는 ssDNA)를 무차별적으로 절단하기 위해 사용될 수 있다.

[0316] 가능한 표적 DNA의 예는 바이러스 DNA, 예컨대 파포바바이러스(예를 들어, 인유두종 바이러스(HPV), 폴리오마바이러스); 헤파드나바이러스(예를 들어, B형 간염 바이러스(HBV)); 헤르페스 바이러스(예를 들어, 단순포진 바이러스(HSV), 수두 대상 포진 바이러스(VZV), 엡스타인-바르 바이러스(EBV), 거대세포 바이러스(CMV), 헤르페스 림프친화성 바이러스, 장미색 비강진, 카포시 육종-연관 헤르페스 바이러스); 아데노바이러스(예를 들어, 앳아데노바이러스(atadenovirus), 아비아데노바이러스, 이타아데노바이러스, 마스트아데노바이러스, 시아데노바이러스); 폭스바이러스(예를 들어, 두창, 백시니아 바이러스, 우두바이러스, 원두증바이러스, orf 바이러스, 가성우두, 소 구진성 구내염 바이러스; 타나폭스 바이러스, 야바 원숭이 종양 바이러스; 연속종 바이러스(MCV)); 파보바이러스(예를 들어, 아데노-연관 바이러스(AAV), 파보바이러스 B19, 인간 보카바이러스, 부파바이러스, 인간 parv4 G1); 제미니바이러스과(Geminiviridae); 나노바이러스과(Nanoviridae); 피코드나바이러스과 등을 포함하지만, 이들로 제한되지 않는다. 일부 경우에, 표적 DNA는 기생충 DNA이다. 일부 경우에, 표적 DNA는 박테리아 DNA, 예를 들어, 병원성 박테리아의 DNA이다.

[0317] ***샘플***

[0318] 대상 샘플은 핵산(예를 들어, 복수의 핵산)을 포함한다. 용어 "복수"는 본 명세서에서 2 이상을 의미하기 위해 사용된다. 따라서, 일부 경우에 샘플은 2개 이상의(예를 들어, 3개 이상, 5개 이상, 10개 이상, 20개 이상, 50개 이상, 100개 이상, 500개 이상, 1,000개 이상, 또는 5,000개 이상의) 핵산(예를 들어, DNA)을 포함한다. 대상 방법은 샘플에(예를 들어, 핵산, 예컨대 DNA의 복잡한 혼합물에) 존재하는 표적 DNA를 검출하기 위해 매우 민감한 방법으로서 사용될 수 있다. 일부 경우에, 샘플은 서열 내에서 서로 다른 5개 이상의 DNA(예를 들어, 10개 이상, 20개 이상, 50개 이상, 100개 이상, 500개 이상, 1,000개 이상, 또는 5,000개 이상의 DNA)를 포함한다. 일부 경우에, 샘플은 10개 이상, 20 이상, 50개 이상, 100개 이상, 500개 이상, 10^3개 이상, 5×10^3개 이상, 10^4개 이상, 5×10^4개 이상, 10^5개 이상, 5×10^5개 이상, 10^6개 이상 5×10^6개 이상, 또는 10^7개 이상의 DNA를 포함한다. 일부 경우에, 샘플은 10 내지 20, 20 내지 50, 50 내지 100, 100 내지 500, 500 내지 10^3, 10^3 내지 5×10^3, 5×10^3 내지 10^4, 10^4 내지 5×10^4, 5×10^4 내지 10^5, 내지 10^5 내지 5×10^5, 5×10^5 내지 10^6, 10^6 내지 5×10^6, 또는 5×10^6 내지 10^7, 또는 10^7개 초과의 DNA를 포함한다. 일부 경우에, 샘플은 5 내지 10^7개의 DNA(예를 들어, 서열 내에서 서로 다름)(예를 들어, 5 내지 10^6개, 5 내지 10^5, 5 내지 50,000, 5 내지 30,000, 10 내지 10^6, 10 내지 10^5, 10 내지 50,000, 10 내지 30,000, 20 내지 10^6, 20 내지 10^5, 20 내지 50,000, 또는 20 내지 30,000개의 DNA)를 포함한다. 일부 경우에, 샘플은 서열 내에서 서로 다른 20개 이상의 DNA를 포함한다. 일부 경우에, 샘플은 세포 용해물(예를 들어, 진핵 세포 용해물, 포유류 세포 용해물, 인간 세포 용해물, 원핵 세포 용해물, 식물 세포 용해물 등)로부터의 DNA를 포함한다. 예를 들어, 일부 경우에 샘플은 세포, 예컨대 진핵 세포, 예를 들어, 포유류 세포, 예컨대 인간 세포로부터의 DNA를 포함한다.

[0319] 용어 "샘플"은 DNA를 포함하는 임의의 샘플을 의미하기 위해 (예를 들어, 표적 DNA가 DNA의 집단 중에 존재하는지의 여부를 결정하기 위해) 본 명세서에서 사용된다. 샘플은 임의의 공급원으로부터 유래될 수 있고, 예를 들어, 샘플은 정제된 DNA의 합성 조합물일 수 있으며; 샘플은 세포 용해물, DNA-풍부 세포 용해물, 또는 세포 용해물로부터 단리 및/또는 정제된 DNA일 수 있다. 샘플은 (예를 들어, 진단 목적을 위해) 환자로부터 유래될 수 있다. 샘플은 투과된 세포로부터 유래될 수 있다. 샘플은 가교된 세포로부터 유래될 수 있다. 샘플은 조직 절편에 있을 수 있다. 샘플은 가교 다음에 탈지(delipidation) 및 균일한 굴절률을 만들기 위한 조절에 의해 제조된 조직으로부터 유래될 수 있다. 가교 다음에 탈지 및 균일한 굴절률을 만들기 위한 조절에 의한 조직 제조의 예는, 예를 들어, 문헌[Shah et al., Development (2016) 143, 2862-2867 doi:10.1242/dev.138560]에 기재되어 있다.

[0320] "샘플"은 표적 DNA 및 복수의 비-표적 DNA를 포함할 수 있다. 일부 경우에, 표적 DNA는 1개의 복제물/10개의 비-표적 DNA, 1개의 복제물/20개의 비-표적 DNA, 1개의 복제물/25개의 비-표적 DNA, 1개의 복제물/50개의 비-표적 DNA, 1개의 복제물/100개의 비-표적 DNA, 1개의 복제물/500개의 비-표적 DNA, 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/5×10^3개의 비-표적 DNA, 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/5×10^4개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/5×10^5개의 비-표적 DNA, 1개의 복제물/10^6개의 비-표적 DNA, 또는 1개 미만의 복제물/10^6개의 비-표적 DNA로 샘플 내에 존재한다. 일부 경우에, 표적 DNA는 1개의 복제물/10개의 비-표적 DNA 내지 1개의 복제물/20개의 비-표적 DNA, 1개의 복제물/20개의 비-표적 DNA 내지 1개

의 복제물/50개의 비-표적 DNA, 1개의 복제물/50개의 비-표적 DNA 내지 1개의 복제물/100개의 비-표적 DNA, 1개의 복제물/100개의 비-표적 DNA 내지 1개의 복제물/500개의 비-표적 DNA, 1개의 복제물/500개의 비-표적 DNA 내지 1개의 복제물/10^3개의 비-표적 DNA, 1개의 복제물/10^3개의 비-표적 DNA 내지 1개의 복제물/5×10^3개의 비-표적 DNA, 1개의 복제물/5×10^3개의 비-표적 DNA 내지 1개의 복제물/10^4개의 비-표적 DNA, 1개의 복제물/10^4개의 비-표적 DNA 내지 1개의 복제물/10^5개의 비-표적 DNA, 1개의 복제물/10^5개의 비-표적 DNA 내지 1개의 복제물/10^6개의 비-표적 DNA, 또는 1개의 복제물/10^6개의 비-표적 DNA 내지 1개의 복제물/10^7개의 비-표적 DNA로 샘플 내에 존재한다.

[0321] 적합한 샘플은 타액, 혈액, 혈청, 혈장, 소변, 흡입물 및 생검 샘플을 포함하지만, 이들로 제한되지 않는다. 따라서, 환자에 대해 용어 "샘플"은 혈액 및 생물학적 유래의 다른 액체 샘플, 고형 조직 샘플, 예컨대 생검 표본 또는 조직 배양물 또는 이들로부터 유래된 세포 및 이들의 자손을 포함한다. 상기 정의는 또한 시약으로 처리와 같이 조달 후에 임의의 방법으로 조작되거나; 세척되거나; 또는 특정 세포 집단, 예컨대 암 세포가 농축된 샘플을 포함한다. 상기 정의는 또한 특정 유형의 분자, 예를 들어, DNA를 농축한 샘플을 포함한다. 용어 "샘플"은 생물학적 샘플, 예컨대 임상 샘플, 예컨대 혈액, 혈장, 혈청, 흡입물, 뇌척수액(CSF)을 포함하고, 또한 수술적 절제에 의해 얻은 조직, 생검에 의해 얻은 조직, 배양물 내 세포, 세포 상청액, 세포 용해물, 조직 샘플, 기관, 골수 등에 의해 얻은 조직을 포함한다. "생물학적 샘플"은 이들로부터 유래된 생물학적 유체(예를 들어, 암성 세포, 감염 세포 등), 예를 들어, 이러한 세포로부터 얻은 DNA를 포함하는 샘플(예를 들어, 세포 용해물 또는 DNA를 포함하는 다른 세포 추출물)을 포함한다.

[0322] 샘플은 임의의 다양한 세포, 조직, 기관 또는 비세포 유체를 포함할 수 있거나 이들로부터 얻을 수 있다. 적합한 샘플 공급원은 진핵 세포, 박테리아 세포 및 고세균 세포를 포함한다. 적합한 샘플 공급원은 단세포 유기체 및 다세포 유기체를 포함한다. 적합한 샘플 공급원은 단세포 진핵생물 유기체; 식물 또는 식물 세포; 조류 세포, 예를 들어, 보트리오코커스 브라우니, 클라미도모나스 레인하르티, 나노클로롭시스 가디타나, 사르가섬 파텐스(*Sargassum patens*), 씨. 아가드(*C. agardh*) 등; 진균 세포(예를 들어, 효모 세포); 동물 세포, 조직 또는 기관; 무척추동물(예를 들어, 과실파리, 강장동물, 극피동물, 선충, 곤충, 거미류 등)로부터의 세포, 조직 또는 기관; 척추동물로부터의 세포, 조직, 유체 또는 기관(예를 들어, 어류, 양서류, 파충류, 조류, 포유류); 포유류로부터의 세포, 조직, 유체 또는 기관(예를 들어, 인간; 비-인간 영장류; 유제류; 고양이; 소; 양; 염소 등)을 포함한다. 적합한 샘플 공급원은 선충, 원생동물 등을 포함한다. 적합한 샘플 공급원은 기생충, 예컨대 연충, 말라리아 기생충 등을 포함한다.

[0323] 적합한 샘플 공급원은 6가지의 계(kingdom), 예를 들어, 박테리아(예를 들어, 진정세균); 원시세균; 원생생물; 진균; 식물계; 및 동물계 중 어느 하나의 세포, 조직 또는 유기체를 포함한다. 적합한 샘플 공급원은 조류(예를 들어, 녹조류, 홍조류, 회조류, 남세균); 원생생물의 진균-유사 구성원, 예를 들어, 점균류, 물 곰팡이 등; 원생생물의 동물-유사 구성원, 예를 들어, 편모충류(예를 들어, 유글레나), 육질충류(예를 들어, 아메바), 포자충류(예를 들어, 정단복합체충류, 점액포자충류, 미포자충류), 및 섬모충류(예를 들어, 짚신벌레)를 포함하지만, 이들로 제한되지 않는 원생생물계의 식물-유사 구성원을 포함한다. 적합한 샘플 공급원은 하기의 문(phyla) 중 어느 것의 구성원을 포함하지만, 이들로 제한되지 않는 진균계의 구성원을 포함한다: 담자균문(곤봉상균류; 예를 들어, 아가리쿠스버섯, 광대버섯, 그물버섯, 꾀꼬리버섯 등의 구성원); 자낭균문(자낭균류, 예를 들어, 사카로마에세스(Saccharomyces)); 마이코피코문(Mycophycophyta)(지의류); 접합균문(접합 진균); 및 불완전균문. 적합한 샘플 공급원은 하기의 분류 중 어느 것의 구성원을 포함하지만, 이들로 제한되지 않는 식물계의 구성원을 포함한다: 선태식물(예를 들어, 선류), 뿔이끼문(예를 들어, 뿔이끼류), 우산이끼문(예를 들어, 태류), 석송문(예를 들어, 석송), 속새식물문(예를 들어, 쇠뜨기), 송엽란문(예를 들어, 솔잎란), 고사리삼과, 양치류(예를 들어, 양치식물), 소철강, 은행나무문, 나자식물문, 마황문 및 피자식물문(예를 들어, 현화식물). 적합한 샘플 공급원은 하기의 문 중 어느 것의 구성원을 포함하지만, 이들로 제한되지 않는 동물계의 구성원을 포함한다: 해면동물(해면); 판형동물; 직유동물(해양 무척추동물의 기생충); 능형류; 자포동물(산호, 아네모네, 해파리, 바다조름속, 산호류, 맹독 해파리); 유즐동물(빗해파리); 편형동물(편형충); 유형동물(끈벌레); 악구동물(턱이 있는 동물(jawed worms)); 복모동물; 윤형동물; 새예동물; 동문동물; 동갑동물; 구두충류; 내항동물; 선형동물; 유선형동물; 구륜동물; 연체동물(연체동물); 성구동물(땅콩 벌레); 환형동물(분절된 벌레); 완보동물(곰벌레); 유조동물(우단벌레); 절지동물(아문을 포함: 협각류, 다지류, 곤충강 및 갑각류, 여기서 협각류는, 예를 들어, 거미류, 퇴구강 및 바다거미강을 포함하고, 다지류는, 예를 들어, 순각류(지네), 배각강(노래기), 소각강, 및 종지네류를 포함하며, 곤충강은 곤충을 포함하고, 갑각류는 새우, 크릴, 따개비 등을 포함함; 추형동물; 태형동

물(이끼벌레류); 완족동물; 극피동물(예를 들어, 불가사리, 바다데이지류, 깃별나리류, 성게, 해삼, 거미불가사리, 브리틀 바스켓(brittle basket) 등); 모악동물(화살벌레); 반삭동물(장새류); 및 척삭동물. 척삭동물의 적합한 구성원은 다음의 아문의 임의의 구성원을 포함한다: 미삭동물(멍게; 해초강, 탈리아강 및 유형류를 포함); 두삭동물아문(창고기); 먹장어강 (먹장어); 및 척추동물, 여기서 척추동물의 구성원은, 예를 들어, 칠성장어강(칠성장어), 연골어류(연골어류), 조기아강(조기류), 총기어아강(실러캔스), 폐어류(폐어), 파충류(파충류, 예를 들어, 뱀, 앨리게이터, 악어, 도마뱀 등), 조류(새); 및 포유류(포유류)의 구성원을 포함한다. 적합한 식물은 임의의 단자엽식물 및 임의의 쌍자엽식물을 포함한다.

[0324] 샘플의 적합한 공급원은 유기체로부터 취한 세포, 유체, 조직 또는 기관; 유기체로부터 단리시킨 특정 세포 또는 세포 그룹 등을 포함한다. 예를 들어, 유기체가 식물인 경우에, 적합한 공급원은 물관부, 체관부, 형성층, 잎, 뿌리 등을 포함한다. 유기체가 동물인 경우, 적합한 공급원은 특정 조직(예를 들어, 폐, 간, 심장, 신장, 뇌, 비장, 피부, 태아성 조직 등), 또는 특정 세포 유형(예를 들어, 뉴런 세포, 상피 세포, 내피 세포내, 성상세포, 대식세포, 신경교세포, 도세포, T 림프구, B 림프구 등)을 포함한다.

[0325] 일부 경우에, 샘플의 공급원은 질환이 있는 세포, 유체, 조직 또는 기관이거나 질환이 있는 것으로 의심된다. 일부 경우에, 샘플의 공급원은 정상(비-질환) 세포, 유체, 조직 또는 기관이다. 일부 경우에, 샘플의 공급원은 병원균-감염 세포, 조직 또는 기관이거나 질환이 있는 것으로 의심된다. 예를 들어, 샘플의 공급원은 감염될 수 있는 또는 감염되지 않을 수 있는 개체일 수 있고 - 샘플은 개체로부터 수집한 임의의 생물학적 샘플(예를 들어, 혈액, 타액, 생검, 혈장, 혈청, 기관지폐포세척, 객담, 대변 샘플, 뇌척수액, 미세 침 흡입물, 채취물 샘플(예를 들어, 협측 채취물, 자궁경부 채취물, 비강 채취물), 간질액, 활액, 비루, 눈물, 버피 코트, 점막 샘플, 상피 세포 샘플(예를 들어, 상피세포 스크레이핑) 등)일 수 있다. 일부 경우에, 샘플은 무세포 액체 샘플이다. 일부 경우에, 샘플은 세포를 포함할 수 있는 액체 샘플이다. 병원균은 바이러스, 진균, 연충, 원생동물, 말라리아 기생충, 말라리아원충 기생충, 톡소플라즈마 기생충, 주혈흡충 기생충 등을 포함한다. "연충"은 회충, 사상충, 및 초식성 선충(선충), 흡충(흡충류), 구두충류 및 촌충(조충)을 포함한다. 원생동물 감염은 지아디아종(Giardia spp.), 트리코모나스 종(Trichomonas spp.), 아프리카 트리파노소미아시스(African trypanosomiasis), 아메바성 이질, 바베스원충, 대장섬모충이질, 샤가스병, 콕시디아증, 말라리아 및 톡소플라스마증으로부터의 감염을 포함한다. 병원균, 예컨대 기생충/원생동물 병원균의 예는 하기를 포함하지만, 이들로 제한되지 않는다: 플라스모듐 팔시파룸(*Plasmodium falciparum*), 플라스모듐 비박스(*Plasmodium vivax*), 트리파노소마 크루지(*Trypanosoma cruzi*) 및 톡소플라스마 곤디(*Toxoplasma gondii*). 진균 병원균은 하기를 포함하지만, 이들로 제한되지 않는다: 크립토코커스 네오포만스(*Cryptococcus neoformans*), 히스토플라스마 캡슐라툼(*Histoplasma capsulatum*), 콕시디오이데스 임미티스(*Coccidioides immitis*), 블라스토마이세스 더마티티디스(*Blastomyces dermatitidis*), 클라미디아 트라코마티스(*Chlamydia trachomatis*) 및 칸디다 알비칸스(*Candida albicans*). 병원성 바이러스는, 예를 들어, 면역결핍 바이러스(예를 들어, HIV); 인플루엔자 바이러스; 뎅기열; 웨스트 나일 바이러스; 헤르페스 바이러스; 황열 바이러스; C형 간염 바이러스; A형 간염 바이러스; B형 간염 바이러스; 유두종바이러스 등을 포함한다. 병원성 바이러스는 DNA 바이러스, 예컨대 파포바바이러스(예를 들어, 인유두종 바이러스(HPV), 폴리오마바이러스); 헤파드나바이러스(예를 들어, B형 간염 바이러스(HBV)); 헤르페스 바이러스(예를 들어, 단순포진 바이러스(HSV), 수두 대상 포진 바이러스(VZV), 엡스타인-바르 바이러스(EBV), 거대세포 바이러스(CMV), 헤르페스 림프친화성 바이러스, 장미색 비강진, 카포시 육종-연관 헤르페스 바이러스); 아데노바이러스(예를 들어, 앳아데노바이러스, 아비아데노바이러스, 이타아데노바이러스, 마스트아데노바이러스, 시아데노바이러스); 폭스바이러스(예를 들어, 두창, 백시니아 바이러스, 우두바이러스, 원두증바이러스, orf 바이러스, 가성우두, 소 구진성 구내염 바이러스; 타나폭스 바이러스, 야바 원숭이 종양 바이러스; 연속종 바이러스(MCV)); 파보바이러스(예를 들어, 아데노-연관 바이러스(AAV), 파보바이러스 B19, 인간 보카바이러스, 부파바이러스, 인간 parv4 G1); 제미니바이러스과; 나노바이러스과; 피코드나바이러스과 등을 포함할 수 있다. 병원균은, 예를 들어, DNA 바이러스[예를 들어, 파포바바이러스(예를 들어, 인유두종 바이러스(HPV), 폴리오마바이러스); 헤파드나바이러스(예를 들어, B형 간염 바이러스(HBV)); 헤르페스 바이러스(예를 들어, 단순포진 바이러스(HSV), 수두 대상 포진 바이러스(VZV), 엡스타인-바르 바이러스(EBV), 거대세포 바이러스(CMV), 헤르페스 림프친화성 바이러스, 장미색 비강진, 카포시 육종-연관 헤르페스 바이러스); 아데노바이러스(예를 들어, 앳아데노바이러스, 아비아데노바이러스, 이타아데노바이러스, 마스트아데노바이러스, 시아데노바이러스); 폭스바이러스(예를 들어, 두창, 백시니아 바이러스, 우두바이러스, 원두증바이러스, orf 바이러스, 가성우두, 소 구진성 구내염 바이러스; 타나폭스 바이러스, 야바 원숭이 종양 바이러스; 연속종 바이러스(MCV)); 파보바이러스(예를 들어, 아데노-연관 바이러스(AAV), 파보바이러스 B19, 인간 보카바이러스, 부파바이러스, 인간 parv4 G1); 제미니바이러스과; 나노바이러스과; 피코드나바이러스과 등], 마이코박테륨 투베르쿨로시스

(*Mycobacterium tuberculosis*), 스트렙토코커스 아갈락티아(*Streptococcus agalactiae*), 메티실린-내성 스타필로코커스 아우레우스, 레지오넬라 뉴모필라(*Legionella pneumophila*), 스트렙토코커스 피오게네스(*Streptococcus pyogenes*), 에스케리키아 콜라이(*Escherichia coli*), 네이세리나 고노르호에아(*Neisseria gonorrhoeae*), 네이세리아 메닌기티디스(*Neisseria meningitidis*), 뉴모코커스(Pneumococcus), 크립토코커스 네오포만스(*Cryptococcus neoformans*), 히스토플라스마 캡슐라툼(*Histoplasma capsulatum*), 헤모필러스 인플루엔자 B, 트레포네마 팔리둠(*Treponema pallidum*), 라임병 스피로헤테스(spirochetes), 슈도모나스 아에루기노사, 마이코박테륨 레프래(*Mycobacterium leprae*), 브루셀라 아보투스(*Brucella abortus*), 광견병 바이러스, 인플루엔자 바이러스, 거대세포 바이러스, 단순포진 바이러스 I, 단순포진 바이러스 II, 인간 혈청 파보-유사 바이러스, 호흡기 세포 융합 바이러스, 수두대상포진 바이러스, B형 간염 바이러스, C형 간염 바이러스, 홍역 바이러스, 아데노바이러스, 인간 T-세포 백혈병 바이러스, 엡스타인-바르 바이러스, 뮤린 백혈병 바이러스, 멈프스 바이러스, 수포성 구내염바이러스, 신드비스 바이러스, 림프구성 맥락수막염 바이러스, 사마귀 바이러스, 부루텅 바이러스, 센다이 바이러스, 고양이 백혈병 바이러스, 레오바이러스, 폴리오 바이러스, 유인원 바이러스 40, 마우스 유선 종양 바이러스, 뎅기 바이러스, 풍진바이러스, 웨스트 나일 바이러스, 플라스모듐 팔시파룸(*Plasmodium falciparum*), 플라스모듐 비박스, 톡소플라스마 곤디(*Toxoplasma gondii*), 트리파노소마 란젤리(*Trypanosoma rangeli*), 트리파노소마 크루지(*Trypanosoma cruzi*), 트리파노소마 로데시엔스(*Trypanosoma rhodesiense*), 트리파노소마 브루세이(*Trypanosoma brucei*), 쉬스토소마 만소니(*Schistosoma mansoni*), 쉬스토소마 자포니쿰(*Schistosoma japonicum*), 바베시아 보비스(*Babesia bovis*), 에이메리아 테넬라(*Eimeria tenella*), 온코세라 보불러스(*Onchocerca volvulus*), 리슈마니아 트로피카(*Leishmania tropica*), 마이코박테륨 투베르쿨로시스(*Mycobacterium tuberculosis*), 트리키넬라 스피랄리스(*Trichinella spiralis*), 테일러리아 파바(*Theileria parva*), 타에니아 하이다티게나(*Taenia hydatigena*), 타에니아 오비스(*Taenia ovis*), 타에니아 사기나타(*Taenia saginata*), 에키노코커스 그라눌로서스(*Echinococcus granulosus*), 메소세스토이데스 코티(*Mesocestoides corti*), 마이코플라즈마 아르트리티디스(*Mycoplasma arthritidis*), 마이코플라즈마 효리니스(*M. hyorhinis*), 마이코플라즈마 오랄레(*M. orale*), 마이코플라즈마 아르기니니(*M. arginini*), 아콜레플라즈마라이드라위(*Acholeplasma laidlawii*), 마이코플라즈마 살리바륨(*M. salivarium*) 및 마이코플라즈마 뉴모니아(*M. pneumoniae*)를 포함할 수 있따.

[0326] ***검출 가능한 신호의 측정***

[0327] 일부 경우에, 대상 방법은 측정하는 단계(예를 들어, CasZ-매개된 ssDNA 절단에 의해 생성된 검출 가능한 신호를 측정하는 단계)를 포함한다. 일단 활성화되면 CasZ 폴리펩타이드는 가이드 RNA가 CasZ 폴리펩타이드(일부 경우에, 또한 tranc RNA를 포함)의 존재 하에 표적 DNA와 혼성화 할 때 생기는 비-표적화된 ssDNA를 절단하기 때문에, 검출 가능한 신호는 ssDNA가 절단될 때 생긴 임의의 신호일 수 있다. 예를 들어, 일부 경우에 측정하는 단계는, 금 나노입자 기반 검출(예를 들어, 문헌[Xu et al., Angew Chem Int Ed Engl. 2007;46(19):3468-70]; 및 [Xia et al., Proc Natl Acad Sci U S A. 2010 Jun 15;107(24):10837-41]), 형광 편광, 콜로이드 상 전이/분산(예를 들어, 문헌[Baksh et al., Nature. 2004 Jan 8;427(6970):139-41]), 전기화학적 검출, 반도체-기반 센싱(예를 들어, 문헌[Rothberg et al., Nature. 2011 Jul 20;475(7356):348-52]; 예를 들어, 2'-3' 환식 포스파타제를 개방함으로써, 무기 포스파타제를 용액 내로 방출함으로써 ssDNA 절단 반응 후에 pH 변화를 생성하기 위해 포스파타제를 사용할 수 있음), 및 표지된 검출기 ssDNA의 검출(더 상세한 설명을 위해 본 명세서의 다른 곳을 참조) 중 하나 이상을 포함할 수 있다. 이러한 검출 방법의 판독은 임의의 편리한 판독일 수 있다. 가능한 판독의 예는 측정 가능한 양의 검출 가능한 형광 신호; 겔 상에서 밴드의 시각적 분석(예를 들어, 비절단 기질에 대해 절단된 생성물을 나타내는 밴드), 색의 존재 또는 부재의 시각적 또는 센서 기반 검출(즉, 색 검출 방법), 및 전기 신호의 (또는 특정 양의) 존재 또는 부재를 포함하지만, 이들로 제한되지 않는다.

[0328] 측정하는 단계는 일부 경우에, 예를 들어, 검출되는 신호의 양이 샘플에 존재하는 표적 DNA의 양을 결정하는 데 사용될 수 있다는 의미에서 정량적일 수 있다. 측정하는 단계는 일부 경우에, 예를 들어, 검출 가능한 신호의 존재 또는 부재가 표적화된 DNA(예를 들어, 바이러스, SNP 등)의 존재 또는 부재를 나타낼 수 있다는 의미에서 정성적일 수 있다. 일부 경우에, 달리 표적화된 DNA(들)(예를 들어, 바이러스, SNP 등)가 특정 역치 농도 초과에서 존재하지 않는 한, 검출 가능한 신호는 (예를 들어, 주어진 역치 수준 초과에서) 존재하지 않을 것이다. 일부 경우에, 검출 역치는, CasZ 폴리펩타이드, 가이드 RNA, 샘플 용적 및/또는 검출기 ssDNA의 양을 (하나가 사용된다면) 변형시킴으로써 적정될 수 있다. 이렇게 해서, 예를 들어, 당업자에 의해 이해될 바와 같이, 한 가지 이상의 반응을 셋업하기 위해 요망된다면, 다수의 대조군이 사용될 수 있으며, 각각은 표적 DNA의 상이한 역치 수준을 검출하도록 셋업하며, 따라서 이러한 일련의 반응은 샘플에 존재하는 표적 DNA의 양을 결정하는 데 사용될 수 있었다(예를 들어, 표적 DNA가 '적어도 X의 농도에서' 샘플에 존재하는지를 결정하는 이러한 일련의

반응을 사용할 수 있었다). 본 개시내용의 조성물 및 방법에 대한 적용/용도의 비제한적 예는 단일-뉴클레오타이드 다형성(SNP) 검출, 암 검진, 박테리아 감염의 검출, 항생제 내성의 검출, 바이러스 감염의 검출 등을 포함한다. 본 개시내용의 조성물 및 방법은 임의의 DNA 표적을 검출하는 데 사용될 수 있다. 예를 들어, 대상 샘플은 세포내 게놈 DNA를 포함할 수 있기 때문에 핵산 물질을 게놈 내로 통합하는 임의의 바이러스가 검출될 수 있고 - 가이드 RNA는 통합된 뉴클레오타이드 서열을 검출하도록 설계될 수 있다. 본 개시내용의 방법은 일부 경우에 증폭 단계를 포함하지 않는다. 본 개시내용의 방법은 일부 경우에 증폭 단계를 포함한다.

[0329] 일부 경우에, 본 개시내용의 방법은 샘플(예를 들어, 표적 DNA 및 복수의 비-표적 DNA를 포함하는 샘플)의 양을 결정하는 데 사용될 수 있다. 샘플 내 표적 DNA 양을 결정하는 것은 시험 샘플로부터 생성된 검출 가능한 신호의 양을 기준 샘플로부터 생성된 검출 가능한 신호의 양과 비교하는 것을 포함할 수 있다. 샘플 내 표적 DNA 양을 결정하는 것은 검출 가능한 신호를 측정하여 시험 측정치를 생성하는 것; 기준 샘플에 의해 생성된 검출 가능한 신호를 측정하여 기준 측정치를 생성하는 것; 및 시험 측정치를 기준 측정치와 비교하여 샘플에 존재하는 표적 DNA의 양을 결정하는 것을 포함할 수 있다.

[0330] 예를 들어, 일부 경우에, 샘플 내 표적 DNA의 양을 결정하기 위한 본 개시내용의 방법은: a) 샘플(예를 들어, 표적 DNA 및 복수의 비-표적 DNA를 포함하는 샘플)을: (i) 표적 DNA와 혼성화하는 가이드 RNA, (ii) 샘플에 존재하는 DNA를 절단하는 CasZ 폴리펩타이드, 및 (iii) 검출기 ssDNA와 접촉시키는 단계; b) CasZ 폴리펩타이드-매개 ssDNA 절단(예를 들어, 검출기 ssDNA의 절단)에 의해 생성된 검출 가능한 신호를 측정하여, 시험 측정치를 생성하는 단계; c) 기준 샘플에 의해 생성된 검출 가능한 신호를 측정하여 기준 측정치를 생성하는 단계; 및 d) 시험 측정치를 기준 측정치와 비교하여 샘플에 존재하는 표적 DNA의 양을 결정하는 단계를 포함한다.

[0331] 다른 예로서, 일부 경우에, 샘플 내 표적 DNA의 양을 결정하기 위한 본 개시내용의 방법은: a) 샘플(예를 들어, 표적 DNA 및 복수의 비-표적 DNA를 포함하는 샘플)을: (i) 표적 DNA와 혼성화하는 가이드 RNA, (ii) 샘플에 존재하는 DNA를 절단하는 CasZ 폴리펩타이드, (iii) tranc RNA; (iv) 검출기 ssDNA와 접촉시키는 단계; b) CasZ 폴리펩타이드-매개 ssDNA 절단(예를 들어, 검출기 ssDNA의 절단)에 의해 생성된 검출 가능한 신호를 측정하여, 시험 측정치를 생성하는 단계; c) 기준 샘플에 의해 생성된 검출 가능한 신호를 측정하여 기준 측정치를 생성하는 단계; 및 d) 시험 측정치를 기준 측정치와 비교하여 샘플에 존재하는 표적 DNA의 양을 결정하는 단계를 포함한다.

[0332] ***샘플 내 핵산의 증폭***

[0333] 일부 실시형태에서, 대상 조성물 및/또는 방법의 민감도는 (예를 들어, 세포 게놈 DNA에서 표적 DNA, 예컨대 바이러스 DNA 또는 SNP의 존재를 검출하기 위해) 검출을 핵산 증폭과 결합함으로써 증가될 수 있다. 일부 경우에, 샘플 내 핵산은 ssDNA를 절단하는 CasZ 폴리펩타이드와 접촉 전에 증폭된다(예를 들어, 샘플 내 핵산의 증폭은 CasZ 폴리펩타이드와의 접촉 전에 시작될 수 있다). 일부 경우에, 샘플 내 핵산은 CasZ 폴리펩타이드와의 접촉과 동시에 증폭된다. 예를 들어, 일부 경우에 대상 방법은 증폭된 샘플을 CasZ 폴리펩타이드와 접촉시키기 전에 (예를 들어, 샘플을 증폭 성분과 접촉시킴으로써) 샘플의 핵산을 증폭시키는 단계를 포함한다. 일부 경우에, 대상 방법은 샘플이 CasZ 폴리펩타이드와 접촉되는 것과 동일한 시간에(동시에) 샘플을 증폭 성분과 접촉시키는 단계를 포함한다. 모든 성분이 동시에 첨가된다면(증폭 성분 및 검출 성분, 예컨대 CasZ 폴리펩타이드, 가이드 RNA, 및 검출기 DNA), CasZ 폴리펩타이드의 트랜스-절단 활성은 핵산이 증폭을 겪는 것과 동시에 샘플의 핵산을 분해하기 시작하는 것이 가능하다. 그러나, 이런 경우라고 하더라도, 증폭시키는 단계 및 검출하는 단계는 증폭 없이 방법을 수행하는 것에 비해 여전히 민감도를 동시에 증가시킬 수 있다.

[0334] 일부 경우에, 구체적 서열(예를 들어, 바이러스의 서열, 관심 대상의 SNP를 포함하는 서열)은, 예를 들어, 프라이머를 이용하여 샘플로부터 증폭된다. 이렇게 해서, 대상 검출 방법의 민감도를 증가시키기 위해 가이드 RNA에 혼성화되는 서열이 증폭될 수 있고 - 이는 샘플에 존재하는 다른 서열에 비해 샘플에 존재하는 관심 대상의 서열의 복제물의 수를 증가시키기 위해 목적하는 서열의 편향 증폭을 달성할 수 있었다. 하나의 예시적 예로서, 대상 방법이 주어진 샘플이 특정 바이러스(또는 특정 SNP)를 포함하는지의 여부를 결정하는 데 사용되고 있다면, 바이러스 서열(또는 비-바이러스 게놈 서열)의 목적하는 영역은 증폭될 수 있고, 바이러스 서열(또는 SNP)이 사실 샘플에 존재한다면 증폭된 영역은 가이드 RNA에 혼성화되는 서열을 포함할 것이다.

[0335] 주목한 바와 같이, 일부 경우에 핵산은 증폭된 핵산을 CasZ 폴리펩타이드와 접촉시키기 전에(예를 들어, 증폭 성분과의 접촉에 의해) 증폭된다. 일부 경우에, 효소적으로 활성인 CasZ 폴리펩타이드와의 접촉 전에 증폭은 10초 이상(예를 들어, 30초 이상, 45초 이상, 1분 이상, 2분 이상, 3분 이상, 4분 이상, 5분 이상, 7.5분 이상, 10분 이상 등) 일어난다. 일부 경우에, 활성 CasZ 폴리펩타이드와 접촉 전에 증폭은 2분 이상(예를 들어, 3분

이상, 4분 이상, 5분 이상, 7.5분 이상, 10분 이상 등) 동안 일어난다. 일부 경우에, 증폭은 10초 내지 60분(예를 들어, 10초 내지 40분, 10초 내지 30분, 10초 내지 20분, 10초 내지 15분, 10초 내지 10분, 10초 내지 5분, 30초 내지 40분, 30초 내지 30분, 30초 내지 20분, 30초 내지 15분, 30초 내지 10분, 30초 내지 5분, 1분 내지 40분, 1분 내지 30분, 1분 내지 20분, 1분 내지 15분, 1분 내지 10분, 1분 내지 5분, 2분 내지 40분, 2분 내지 30분, 2분 내지 20분, 2분 내지 15분, 2분 내지 10분, 2분 내지 5분, 5분 내지 40분, 5분 내지 30분, 5분 내지 20분, 5분 내지 15분, 또는 5분 내지 10분)의 범위의 일정 기간 동안 일어난다. 일부 경우에, 증폭은 5분 내지 15분 범위에서 일정 기간 동안 일어난다. 일부 경우에, 증폭은 7분 내지 12분 범위에서 일정 기간 동안 일어난다.

[0336] 일부 경우에, 샘플은 CasZ 폴리펩타이드와의 접촉과 동시에 증폭 성분과 접촉된다. 일부 이러한 경우에, CasZ 폴리펩타이드는 접촉 시 비활성이고, 일단 샘플 내 핵산이 증폭된다면 활성화된다.

[0337] 다양한 증폭 방법 및 성분은 당업자에게 공지되어 있고, 임의의 편리한 방법이 사용될 수 있다(예를 들어, 문헌 [Zanoli and Spoto, Biosensors (Basel). 2013 Mar; 3(1): 18-43; Gill and Ghaemi, Nucleosides, Nucleotides, and Nucleic Acids, 2008, 27: 224-243; Craw and Balachandrana, Lab Chip, 2012, 12, 2469-2486]을 참조하며; 이들은 본 명세서에 전문이 참고로 포함된다). 핵산 증폭은 중합효소 연쇄 반응(PCR), 역전사 PCR(RT-PCR), 정량적 PCR(qPCR), 역전사 qPCR(RT-qPCR), 네스티드 PCR(nested PCR), 다중복합 PCR, 비대칭 PCR, 터치다운 PCR, 무작위 프라이머 PCR, 헤미-네스티드 PCR, 중합효소 사이클링 어셈블리(PCA), 콜로니 PCR, 리가제 연쇄 반응(LCR), 디지털 PCR, 메틸화 특이적-PCR(MSP), 낮은 변성 온도에서의 공동-증폭-PCR(COLD-PCR), 대립유전자-특이적 PCR, 서열간-특이적 PCR(ISS-PCR), 전체 게놈 증폭(WGA), 역 PCR, 및 열 비대칭 엇갈림(thermal asymmetric interlaced PCR)(TAIL-PCR)을 포함할 수 있다.

[0338] 일부 경우에, 증폭은 등온 증폭이다. 용어 "등온 증폭"은 단일 온도 인큐베이션을 사용하여 열 순환기에 대한 필요를 제거할 수 있는 핵산(예를 들어, DNA) 증폭(예를 들어, 효소 연쇄 반응을 이용) 방법을 나타낸다. 등온 증폭은 증폭 반응 동안 표적 핵산의 열 변성에 의존하지 않으며 따라서 온도의 다중의 빠른 변화를 필요로 하지 않을 수 있는 핵산 증폭 형태이다. 따라서 등온 핵산 증폭 방법은 실험실 환경 내에서 또는 밖에서 수행될 수 있다. 역 전사 단계와 조합함으로써, 이들 증폭 방법은 RNA를 등온으로 증폭시키는 데 사용될 수 있다.

[0339] 등온 증폭 방법의 예는 루프-매개 등온 증폭(LAMP), 헬리카제-의존적 증폭(HDA), 재조합효소 중합효소 증폭(RPA), 가닥 변위(strand displacement amplification: SDA), 핵산 서열-기반 증폭(NASBA), 전사 매개 증폭(TMA), 틈내기 효소 증폭 반응(NEAR), 구름원 증폭(rolling circle amplification: RCA), 다중 변위 증폭(MDA), 분지(Ramification: RAM), 원형 헬리카제-의존적 증폭(cHDA), 단일 프라이머 등온 증폭(SPIA), RNA 기술의 신호 매개 증폭(SMART), 자기-지속 서열 복제(3SR), 게놈 기하급수적 증폭 반응(GEAR) 및 등온 다중 변위 증폭(IMDA)을 포함하지만, 이들로 제한되지 않는다.

[0340] 일부 경우에, 증폭은 재조합효소 중합효소 증폭(RPA)이다(예를 들어, 미국 특허 제8,030,000호; 제8,426,134호; 제8,945,845호; 제9,309,502호; 및 제9,663,820호, 이들의 전문은 본 명세서에 참고로 포함됨). 재조합효소 중합효소 증폭(RPA)은 2개의 마주보는 프라이머(PCR와 매우 유사)를 사용하고, 3종의 효소, 즉, 재조합효소, 단일-가닥 DNA-결합 단백질(SSB) 및 가닥-변위 중합효소를 사용한다. 재조합효소는 이중가닥 DNA에서 상동성 서열을 갖는 올리고뉴클레오타이드 프라이머와 쌍을 이루고, SSB는 변위된 DNA 가닥에 결합하여 프라이머가 변위되는 것을 방지하고, 가닥 변위 중합효소는 DNA 합성을 시작하며, 여기서 프라이머는 표적 DNA에 결합된다. RPA 반응에 역전사효소를 첨가하는 것은 cDNA를 생성하는 별개의 단계에 대한 필요 없이 RNA뿐만 아니라 DNA의 검출을 용이하게 할 수 있다. RPA 반응에 대한 성분의 일례는 다음과 같다(예를 들어, 미국 특허 제8,030,000호; 제8,426,134호; 제8,945,845호; 제9,309,502호; 제9,663,820호 참조): 50mM Tris pH 8.4, 80mM 아세트산칼륨, 10mM 아세트산마그네슘, 2mM DTT, 5% PEG 화합물(카보왁스-20M), 3mM ATP, 30mM 포스포크레아틴, 100ng/㎕ 크레아틴 키나제, 420ng/㎕ gp32, 140ng/㎕ UvsX, 35ng/㎕ UvsY, 2000M dNTPs, 300nM 각각의 올리고뉴클레오타이드, 35ng/㎕ Bsu 중합효소 및 핵산-함유 샘플).

[0341] 전사 매개 증폭(TMA)에서, RNA 중합효소는 프라이머 영역에서 조작된 프로모터로부터 RNA를 생성하는 데 사용되고, 이어서, 역전사효소는 프라이머로부터 cDNA를 합성한다. 이어서, 제3 효소, 예를 들어, Rnase H는 열-변성 단계 없이 cDNA로부터 RNA 표적을 분해하는 데 사용될 수 있다. 이 증폭 기법은 자기-지속 서열 복제(3SR) 및 핵산 서열 기반 증폭(NASBA)과 유사하지만, 사용된 효소가 다르다. 다른 예에 대해, 헬리카제-의존적 증폭(HDA)은 풀려 있는 dsDNA를 가열하기보다는 열안정성 헬리카제(Tte-UvrD)를 이용하여, 이어서, 중합효소에 의한 프라이머의 혼성화 및 연장에 이용 가능한 단일-가닥을 생성한다. 또 다른 예에 대해, 루프 매개 증폭(LAMP)은

가닥 변위 능력 및 4개 이상의 특이적 설계 프라이머 세트를 갖는 열안정성 중합효소를 사용한다. 각각의 프라이머는, 일단 변위되면, 헤어핀이 자기-프라이밍 및 추가적인 중합효소 연장을 용이하게 하는 것을 시작하는 헤어핀 말단을 갖도록 설계된다. LAMP 반응에서, 반응은 등온 조건 하에 진행되지만, 초기 열 변성 단계는 이중-가닥 표적을 필요로 한다. 추가로, 증폭은 다양한 길이 생성물의 사다리 패턴을 수득한다. 또 다른 예에 대해, 가닥 변위 증폭(SDA)은 표적 DNA의 비변형 가닥에 틈을 내는 제한 엔도뉴클레아제의 능력과 틈에서 3' 말단을 연장시키고 하류의 DNA 가닥을 변위시키는 엑소뉴클레아제-결핍 DNA 중합효소의 능력을 조합한다.

[0342] *검출기 DNA*

[0343] 일부 경우에, 대상 방법은 샘플(예를 들어, 표적 DNA 및 복수의 비-표적 ssDNA를 포함하는 샘플)을: i) CasZ 폴리펩타이드; ii) 가이드 RNA; 및 iii) 단일 가닥이며 가이드 RNA의 가이드 서열과 혼성화되지 않는 검출기 DNA와 접촉시키는 단계를 포함한다.

[0344] 적합한 단일-가닥 검출기 DNA는 길이가 7개의 뉴클레오타이드 내지 25개의 뉴클레오타이드이다. 예를 들어, 적합한 단일-가닥 검출기 DNA는 길이가 7개의 뉴클레오타이드 내지 10개의 뉴클레오타이드, 11개의 뉴클레오타이드 내지 15개의 뉴클레오타이드, 15개의 뉴클레오타이드 내지 20개의 뉴클레오타이드, 또는 20개의 뉴클레오타이드 내지 25개의 뉴클레오타이드이다. 일부 경우에, 적합한 단일-가닥 검출기 DNA는 길이가 10개의 뉴클레오타이드 내지 15개의 뉴클레오타이드이다. 일부 경우에, 적합한 단일-가닥 검출기 DNA는 길이가 10개의 뉴클레오타이드이다. 일부 경우에, 적합한 단일-가닥 검출기 DNA는 길이가 11개의 뉴클레오타이드이다. 일부 경우에, 적합한 단일-가닥 검출기 DNA는 길이가 12개의 뉴클레오타이드이다. 일부 경우에, 적합한 단일-가닥 검출기 DNA는 길이가 13개의 뉴클레오타이드이다. 일부 경우에, 적합한 단일-가닥 검출기 DNA는 길이가 14개의 뉴클레오타이드이다. 일부 경우에, 적합한 단일-가닥 검출기 DNA는 길이가 15개의 뉴클레오타이드이다.

[0345] 일부 경우에, 대상 방법은: a) 샘플을 형광-방출 염료 쌍을 포함하는 표지된 단일 가닥 검출기 DNA(검출기 ssDNA); (표적 DNA에 혼성화되는 가이드 RNA와 관련하여 가이드 RNA에 결합함으로써) 활성화된 후에 표지된 검출기 ssDNA를 절단하는 CasZ 폴리펩타이드와 접촉시키는 단계; 및 b) 형광-방출 염료 쌍에 의해 생성되는 검출 가능한 신호를 측정하는 단계를 포함한다. 예를 들어, 일부 경우에, 대상 방법은 형광 공명 에너지 전달(FRET) 쌍 또는 퀀처/플루오르(quencher/fluor) 쌍 또는 둘 다를 포함하는 표지된 검출기 ssDNA와 샘플을 접촉시키는 단계를 포함한다. 일부 경우에, 대상 방법은 FRET 쌍을 포함하는 표지된 검출기 ssDNA와 샘플을 접촉시키는 단계를 포함한다. 일부 경우에, 대상 방법은 플루오르/퀀처 쌍을 포함하는 표지된 검출기 ssDNA와 샘플을 접촉시키는 단계를 포함한다.

[0346] 형광-방출 염료쌍은 FRET 쌍 또는 퀀처/플루오르 쌍을 포함한다. FRET 쌍과 퀀처/플루오르 쌍의 경우 둘 다에서, 염료 중 하나의 방출 스펙트럼은 쌍에서 다른 염료의 흡수 스펙트럼 영역과 중복된다. 본 명세서에서 사용되는 용어 "형광-방출 염료 쌍"은 "형광 공명 에너지 전달(FRET) 쌍"과 "퀀처/플루오르 쌍"을 둘 다 포함하기 위해 사용되는 일반적 용어이며, 이들 용어 둘 다 이하에 더 상세하게 논의한다. 용어 "형광-방출 염료 쌍"은 "FRET 쌍 및/또는 퀀처/플루오르 쌍"이라는 어구와 상호 호환적으로 사용된다.

[0347] 일부 경우에 (예를 들어, 검출기 ssDNA가 FRET 쌍을 포함할 때) 표지된 검출기 ssDNA는 절단되기 전에 검출 가능한 신호의 양을 생성하고, 측정되는 검출 가능한 신호의 양은 표지된 검출기 ssDNA가 절단될 때 감소된다. 일부 경우에, 표지된 검출기 ssDNA는 (예를 들어, FRET 쌍으로부터) 절단되기 전에 제1 검출 가능한 신호를 생성하고, 표지된 검출기 ssDNA가 (예를 들어, 퀀처/플루오르 쌍으로부터) 절단될 때 제2 검출 가능한 신호를 생성한다. 이렇게 해서, 일부 경우에, 표지된 검출기 ssDNA는 FRET 쌍 및 퀀처/플루오르 쌍을 포함한다.

[0348] 일부 경우에, 표지된 검출기 ssDNA는 FRET 쌍을 포함한다. FRET는 여기 상태의 형광단으로부터 매우 근위의 제2 발색단까지 에너지의 비복사 전달이 일어나는 과정이다. 에너지 전달이 일어날 수 있는 범위는 대략 10 나노미터(100 옹스트롬)으로 제한되며, 전달 효율은 형광단 사이의 이격 거리에 극도로 민감하다. 따라서, 본 명세서에서 사용되는 용어 "FRET"("형광 공명 에너지 전달"; "푀스터(Forster) 공명 에너지 전달"로도 알려짐)는 에너지의 일부가 양자 결합 효과를 통해 공여자로부터 수용자로 통과함에 따라, 공여자의 방출 스펙트럼이 수용자의 여기 스펙트럼과 중복되도록 선택되고, 추가로 공여자 및 수용자가 서로 근위일 때(보통 10㎚ 이하), 공여자의 여기가 수용자로부터의 여기 및 방출을 야기하도록 선택된 공여자 형광단 및 매칭 수용자 형광단을 수반하는 물리적 현상을 지칭한다. 따라서, FRET 신호는 공여자 및 수용자의 근위 게이지로서 작용하며; 이들이 서로 매우 근위에 있을 때에만 신호가 생성된다. FRET 공여자 모이어티(예를 들어, 공여자 형광단) 및 FRET 수용자 모이어티(예를 들어, 수용자 형광단)는 본 명세서에서 총괄적으로 "FRET 쌍"으로 지칭된다.

[0349] 공여자-수용자 쌍(FRET 공여자 모이어티 및 FRET 수용자 모이어티)은 본 명세서에서 "FRET 쌍" 또는 "신호 FRET 쌍"으로서 지칭된다. 따라서, 일부 경우에, 하나의 신호 상대가 FRET 공여자 모이어티이고 다른 신호 상대가 FRET 수용자 모이어티일 때, 대상 표지된 검출기 ssDNA는 2개의 신호 상대(신호 쌍)를 포함한다. 따라서 이러한 FRET 쌍(FRET 공여자 모이어티 및 FRET 수용자 모이어티)를 포함하는 대상 표지된 검출기 ssDNA는 신호 상대가 매우 근위일 때(예를 들어, 동일한 RNA 분자 상에 있는 동안) 검출 가능한 신호(FRET 신호)를 나타낼 것이지만, 상대가 분리될 때(예를 들어, CasZ 폴리펩타이드의 RNA 분자의 절단 후에) 신호는 감소(또는 부재)될 것이다.

[0350] FRET 공여자 및 수용자 모이어티(FRET 쌍)는 당업자에게 공지될 것이며, 임의의 편리한 FRET 쌍(예를 들어, 임의의 편리한 공여자 및 수용자 모이어티 쌍)이 사용될 수 있다. 적합한 FRET 쌍의 예는 표 1에 제시하는 것을 포함하지만, 이들로 제한되지 않는다. 또한 문헌[Bajar et al. Sensors (Basel). 2016 Sep 14;16(9); 및 Abraham et al. PLoS One. 2015 Aug 3;10(8):e0134436.] 참조.

표 6

[0351]

FRET 쌍(공여자 및 수용자 FRET 모이어티)의 예	
공여자	**수용자**
트립토판	단실
IAEDANS (1)	DDPM (2)
BFP	DsRFP
단실	플루오레세인 아이소티오사이아네이트(FITC)
단실	옥타데실로다민
시안 형광 단백질 (CFP)	녹색 형광 단백질(GFP)
CF (3)	텍사스 레드
플루오레세인	테트라메틸로다민
Cy3	Cy5
GFP	황색 형광 단백질(YFP)
BODIPY FL (4)	BODIPY FL (4)
로다민 110	Cy3
로다민 6G	말라키트 그린(Malachite Green)
FITC	에오신 티오세미카바자이드
B-피코에리트린	Cy5
Cy5	Cy5.5
(1) 5-(2-아이오도아세틸아미노에틸)아미노나프탈렌-1-설폰산 (2) N-(4-다이메틸아미노-3,5-다이나이트로페닐)말레이미드 (3) 카복시플루오레세인 석신이미딜 에스터 (4) 4,4-다이플루오로-4-보라-3a,4a-다이아자-s-인다센	

[0352] 일부 경우에, 표지된 검출기 ssDNA가 절단될 때 검출 가능한 신호가 생성된다(예를 들어, 일부 경우에, 표지된 검출기 ssDNA는 퀀처/플루오르쌍을 포함한다). 신호 퀀칭 쌍의 하나의 신호 상대는 검출 가능한 신호를 생성하며, 다른 신호 상대는 제1 신호 상대의 검출 가능한 신호를 퀀칭하는 퀀처 모이어티이다(즉, 신호 상대가 서로 근위일 때, 예를 들어, 신호쌍의 신호 상대가 근위일 때 신호 모이어티로부터의 신호가 감소(퀀칭)되도록 퀀처 모이어티는 신호 모이어티의 신호를 퀀칭시킨다).

[0353] 예를 들어, 일부 경우에, 검출 가능한 신호의 양은 표지된 검출기 ssDNA가 절단될 때 증가된다. 일부 경우에, 예를 들어, CasZ 폴리펩타이드에 의한 절단되기 전에 동일한 ssDNA 분자 상에 신호가 둘 다 존재할 때, 예를 들어, 하나의 신호 상대(신호 모이어티)에 의해 나타나는 신호는 다른 신호 상대(퀀처 신호 모이어티)에 의해 퀀칭된다. 이러한 신호 쌍은 본 명세서에서 "퀀처/플루오르 쌍", "퀀칭 쌍" 또는 "신호 퀀칭 쌍"으로서 지칭된다. 예를 들어, 일부 경우에, 하나의 신호 상대(예를 들어, 제1 신호 상대)는 제2 신호 상대(예를 들어, 퀀처 모이어티)에 의해 퀀칭되는 검출 가능한 신호를 생성하는 신호 모이어티이다. 따라서 이러한 퀀처/플루오르쌍의 신호 상대는 상대가 분리될 때(예를 들어, CasZ 폴리펩타이드에 의한 검출기 ssDNA의 절단된 후) 검출 가능한 신호를 생성할 것이지만, 상대가 매우 근위일 때(예를 들어, CasZ 폴리펩타이드에 의한 검출기 ssDNA의 절단되기 전에) 신호는 퀀칭될 것이다.

[0354] 퀀처 모이어티는 신호 모이어티로부터(예를 들어, CasZ 폴리펩타이드에 의한 검출기 ssDNA의 절단되기 전에) 다양한 정도로 신호를 퀀칭시킬 수 있다. 일부 경우에, 퀀처 모이어티는 퀀처 모이어티의 존재 하에(신호 상대가

서로 근위일 때) 검출된 신호가 퀀처 모이어티의 부재 하에(신호 상대가 분리될 때) 검출된 신호의 95% 이하인 경우 신호 모이어티로부터 신호를 퀀칭시킨다. 예를 들어, 일부 경우에, 퀀처 모이어티의 존재에서 검출된 신호는 퀀처 모이어티의 부재 하에 검출된 신호의 90% 이하, 80% 이하, 70% 이하, 60% 이하, 50% 이하, 40% 이하, 30% 이하, 20% 이하, 15% 이하, 10% 이하, 또는 5% 이하일 수 있다. 일부 경우에, 신호 없음(예를 들어, 배경 초과)은 퀀처 모이어티의 존재 하에 검출된다.

[0355] 일부 경우에, 퀀처 모이어티의 부재 하에(신호 상대가 분리될 때) 검출된 신호는 퀀처 모이어티의 존재 하에(신호 상대가 서로 근위일 때) 검출된 신호보다 적어도 1.2배 초과(예를 들어, 적어도 1.3배, 적어도 1.5 배, 적어도 1.7 배, 적어도 2 배, 적어도 2.5 배, 적어도 3 배, 적어도 3.5 배, 적어도 4 배, 적어도 5 배, 적어도 7 배, 적어도 10 배, 적어도 20배, 또는 적어도 50배 초과)이다.

[0356] 일부 경우에, 신호 모이어티는 형광 표지이다. 일부 이러한 경우에, 퀀처 모이어티는 (예를 들어, 표지의 방출 스펙트럼에서 에너지를 흡수함으로써) 형광 표지로부터의 신호(광 신호)를 퀀칭시킨다. 따라서, 퀀처 모이어티가 신호 모이어티에 근접하지 않을 때, 형광 표지로부터의 방출(신호)는 퀀처 모이어티에 의해 신호가 흡수되지 않기 때문에 검출 가능하다. 임의의 편리한 공여자 수용자 쌍(신호 모이어티/퀀처 모이어티 쌍)이 사용될 수 있고, 다수의 적합한 쌍은 당업계에 공지되어 있다.

[0357] 일부 경우에, 퀀처 모이어티는 신호 모이어티(또한 본 명세서에서 "검출 가능한 표지"로서 지칭됨)로부터 에너지를 흡수하고, 이어서, 신호(예를 들어, 상이한 파장에서의 광)를 방출한다. 따라서, 일부 경우에, 퀀처 모이어티는 신호 모이어티 그 자체이고(예를 들어, 신호 모이어티는 6-카복시플루오레세인일 수 있는 한편, 퀀처 모이어티는 6-카복시-테트라메틸로다민일 수 있음), 일부 경우에, 쌍은 또한 FRET 쌍일 수 있다. 일부 경우에, 퀀처 모이어티는 다크 퀀처이다. 다크 퀀처는 여기 에너지를 흡수하며, 상이한 방법으로 (예를 들어, 열로서) 에너지를 소멸시킨다. 따라서, 다크 퀀처는 그 자체의 형광이 최소이거나 없다(형광을 방출하지 않는다). 다크 퀀처의 예는 미국 특허 제8,822,673호 및 제8,586,718호; 미국 특허 공개 제20140378330호, 제20140349295호 및 제20140194611호; 및 국제 특허 출원 WO200142505 및 WO200186001에 추가로 기재되어 있고, 이들 모두는 본 명세서에 전문이 참고로 포함된다.

[0358] 형광 표지의 예는 알렉사 플루오르(Alexa Fluor)(등록상표) 염료, ATTO 염료(예를 들어, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO 740), DyLight 염료, 사이아닌 염료(예를 들어, Cy2, Cy3, Cy3.5, Cy3b, Cy5, Cy5.5, Cy7, Cy7.5), 플루오프로브(FluoProbes) 염료, 설포 Cy 염료, 세타(Seta) 염료, IRIS 염료, SeTau 염료, SRfluor 염료, 스퀘어(Square) 염료, 플루오레세인 아이소티오사이아네이트(FITC), 테트라메틸로다민(TRITC), 텍사스 레드(Texas Red), 오리건 그린(Oregon Green), 퍼시픽 블루(Pacific Blue), 퍼시픽 그린(Pacific Green), 퍼시픽 오렌지(Pacific Orange), 양자점 및 테더드(tethered) 형광 단백질을 포함하지만, 이들로 제한되지 않는다.

[0359] 일부 경우에, 검출 가능한 표지는 알렉사 플루오르(등록상표) 염료, ATTO 염료(예를 들어, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO 740), DyLight 염료, 사이아닌 염료(예를 들어, Cy2, Cy3, Cy3.5, Cy3b, Cy5, Cy5.5, Cy7, Cy7.5), 플루오프로브 염료, 설포 Cy 염료, 세타 염료, IRIS 염료, SeTau 염료, SRfluor 염료, 스퀘어 염료, 플루오레세인(FITC), 테트라메틸로다민(TRITC), 텍사스 레드(Texas Red), 오리건 그린, 퍼시픽 블루, 퍼시픽 그린 및 퍼시픽 오렌지로부터 선택된 형광 표지이다.

[0360] 일부 경우에, 검출 가능한 표지는 알렉사 플루오르(등록상표) 염료, ATTO 염료(예를 들어, ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO 740), DyLight 염료, 사이아닌 염료(예를 들어, Cy2, Cy3, Cy3.5, Cy3b, Cy5, Cy5.5, Cy7, Cy7.5), 플루오프로브 염료, 설포 Cy 염료, 세타 염료, IRIS 염료, SeTau 염료, SRfluor 염료, 스퀘어 염료, 플루오레세인(FITC), 테트라메틸로다민(TRITC), 텍사스 레드(Texas Red), 오리건 그린, 퍼시픽

블루, 퍼시픽 그린, 퍼시픽 오렌지, 양자점 및 테더드 형광 단백질로부터 선택된 형광 표지이다.

[0361] ATTO 염료의 예는 하기를 포함하지만, 이들로 제한되지 않는다: ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725 및 ATTO 740.

[0362] 알렉사플루오르 염료의 예는 하기를 포함하지만, 이들로 제한되지 않는다: 알렉사 플루오르(등록상표) 350, 알렉사 플루오르(등록상표) 405, 알렉사 플루오르(등록상표) 430, 알렉사 플루오르(등록상표) 488, 알렉사 플루오르(등록상표) 500, 알렉사 플루오르(등록상표) 514, 알렉사 플루오르(등록상표) 532, 알렉사 플루오르(등록상표) 546, 알렉사 플루오르(등록상표) 555, 알렉사 플루오르(등록상표) 568, 알렉사 플루오르(등록상표) 594, 알렉사 플루오르(등록상표) 610, 알렉사 플루오르(등록상표) 633, 알렉사 플루오르(등록상표) 635, 알렉사 플루오르(등록상표) 647, 알렉사 플루오르(등록상표) 660, 알렉사 플루오르(등록상표) 680, 알렉사 플루오르(등록상표) 700, 알렉사 플루오르(등록상표) 750, 알렉사 플루오르(등록상표) 790 등.

[0363] 퀀처 모이어티의 예는 다크 퀀처(dark quencher), 블랙 홀 퀀처(Black Hole Quencher)(등록상표)(BHQ(등록상표))(예를 들어, BHQ-0, BHQ-1, BHQ-2, BHQ-3), Qxl 퀀처, ATTO 퀀처(예를 들어, ATTO 540Q, ATTO 580Q, 및 ATTO 612Q), 다이메틸아미노아조벤젠설폰산(답실(Dabsyl)), 아이오와 블랙(Iowa Black) RQ, 아이오와 블랙 FQ, IRDye QC-1, QSY 염료(예를 들어, QSY 7, QSY 9, QSY 21), 앱솔루트퀀처(AbsoluteQuencher), 이클립스(Eclipse) 및 금속 클러스터, 예컨대, 금 나노입자 등을 포함하지만, 이들로 제한되지 않는다.

[0364] 일부 경우에, 퀀처 모이어티는 다크 퀀처, 블랙 홀 퀀처(등록상표)(BHQ(등록상표))(예를 들어, BHQ-0, BHQ-1, BHQ-2, BHQ-3), Qxl 퀀처, ATTO 퀀처(예를 들어, ATTO 540Q, ATTO 580Q 및 ATTO 612Q), 다이메틸아미노아조벤젠설폰산(답실), 아이오와 블랙 RQ, 아이오와 블랙 FQ, IRDye QC-1, QSY 염료(예를 들어, QSY 7, QSY 9, QSY 21), 앱솔루트퀀처, 이클립스 및 금속 클러스터로부터 선택된다.

[0365] ATTO 퀀처의 예는 ATTO 540Q, ATTO 580Q 및 ATTO 612Q를 포함하지만, 이들로 제한되지 않는다. 블랙홀 퀀처(등록상표)(BHQ(등록상표))의 예는 BHQ-0(493nm), BHQ-1(534nm), BHQ-2(579nm) 및 BHQ-3(672nm)를 포함하지만, 이들로 제한되지 않는다.

[0366] 일부 검출 가능한 표지(예를 들어, 형광 염료) 및/또는 퀀처 모이어티의 예는 예를 들어, 문헌[Bao et al., Annu Rev Biomed Eng. 2009;11:25-47]뿐만 아니라 미국 특허 제8,822,673호 및 제8,586,718호; 미국 특허 공개 제20140378330호, 제20140349295호, 제20140194611호, 제20130323851호, 제20130224871호, 제20110223677호, 제20110190486호, 제20110172420호, 제20060179585호 및 제20030003486호; 및 국제 특허 출원: WO200142505 및 WO200186001를 참조하며, 이들 모두는 본 명세서에 전문이 참고로 포함된다.

[0367] 일부 경우에, 표지된 검출기 ssDNA의 절단은 비색 판독을 측정함으로써 검출될 수 있다. 예를 들어, 형광단의 유리(예를 들어, FRET 쌍으로부터의 유리, 퀀처/플루오르 쌍으로부터의 유리 등)는 검출 가능한 신호의 파장 이동(따라서 색 이동)을 초래할 수 있다. 따라서, 일부 경우에, 대상 표지된 검출기 ssDNA의 절단은 색-이동에 의해 검출될 수 있다. 이러한 이동은 한 가지 색 신호(파장)양의 상실, 다른 색의 양의 획득, 한 가지 색의 다른 색으로의 할당 변화 등으로서 표현될 수 있다.

[0368] **표적 DNA를 검출하기 위한 키트**

[0369] 본 개시내용은, 예를 들어, 복수의 DNA를 포함하는 샘플에서 표적 DNA를 검출하기 위한 키트를 제공한다. 일부 경우에, 키트는 (a) 표지된 검출기 ssDNA(예를 들어, 형광-방출 염료쌍, 예를 들어, FRET 쌍 및/또는 퀀처/플루오르 쌍을 포함하는 표지된 검출기 ssDNA); 및 (b)(i) 가이드 RNA, 및/또는 상기 가이드 RNA를 암호화하는 핵산; 및 ii) CasZ 폴리펩타이드, 및/또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산 중 하나 이상을 포함한다. 일부 경우에, 가이드 RNA를 암호화하는 핵산은 사용자에 의한 가이드 서열의 삽입을 위해 서열 삽입 부위를 포함한다.

[0370] 일부 경우에, 키트는 (a) 표지된 검출기 ssDNA(예를 들어, 형광-방출 염료쌍, 예를 들어, FRET 쌍 및/또는 퀀처/플루오르 쌍을 포함하는 표지된 검출기 ssDNA); 및 (b)(i) 가이드 RNA, 및/또는 상기 가이드 RNA를 암호화하는 핵산; ii) tranc RNA 및/또는 상기 가이드 RNA를 암호화하는 핵산; 및 iii) CasZ 폴리펩타이드, 및/또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산 중 하나 이상을 포함한다. 일부 경우에, 가이드 RNA를 암호화하는 핵산은

사용자에 의한 가이드 서열의 삽입을 위해 서열 삽입 부위를 포함한다.

[0371] 일부 경우에, 키트는 (a) 표지된 검출기 ssDNA(예를 들어, 형광-방출 염료쌍, 예를 들어, FRET 쌍 및/또는 퀀처/플루오르 쌍을 포함하는 표지된 검출기 ssDNA); 및 (b) (i) 가이드 RNA 및 tranc RNA를 포함하는 단일-분자 RNA, 및/또는 단일-분자 RNA를 암호화하는 핵산; 및 iii) CasZ 폴리펩타이드, 및/또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산 중 하나 이상을 포함한다. 일부 경우에, 단일-분자 RNA를 암호화하는 핵산은 사용자에 의한 가이드 서열의 삽입을 위해 서열 삽입 부위를 포함한다.

[0372] 일부 경우에, 대상 키트는: (a) 형광-방출 염료쌍, 예를 들어, FRET 쌍 및/또는 퀀처/플루오르 쌍을 포함하는 표지된 검출기 ssDNA; 및 (b) (i) 가이드 RNA, 및/또는 상기 가이드 RNA를 암호화하는 핵산; 및/또는 i) CasZ 폴리펩타이드 중 하나 이상을 포함한다.

[0373] ***양성 대조군***

[0374] 본 개시내용의 키트(예를 들어, 표지된 검출기 ssDNA 및 CasZ 폴리펩타이드를 포함하는 것)는 또한 양성 대조군 표적 DNA를 포함할 수 있다. 일부 경우에, 키트는 또한 대조군 표적 DNA에 혼성화되는 뉴클레오타이드 서열을 포함하는 양성 대조군 가이드 RNA를 포함한다. 일부 경우에, 양성 대조군 표적 DNA는 다양한 양으로, 별개의 용기에서 제공된다. 일부 경우에, 양성 대조군 표적 DNA는 다양한 공지된 농도로, 별개의 용기에서 대조군 비-표적 DNA와 함께 제공된다.

[0375] ***핵산***

[0376] 본 개시내용의 RNA(예를 들어, 가이드 RNA, tranc RNA, 가이드 RNA 및 tranc RNA를 포함하는 단일-분자 RNA)가 임의의 편리한 방법(예를 들어, 화학적 합성, RNA 중합효소, 예를 들어, T7 중합효소, T3 중합효소, SP6 중합효소 등을 이용하여 시험관내)을 이용하여 합성될 수 있지만, 이러한 RNA를 암호화하는 핵산이 또한 생각된다. 추가적으로, 본 개시내용의 CasZ 폴리펩타이드는 (예를 들어, 키트의 부분으로서) 제공될 수 있지만, CasZ 폴리펩타이드를 암호화하는 핵산(예컨대, mRNA 및/또는 DNA)가 또한 제공될 수 있다.

[0377] 일부 경우에, 본 개시내용의 키트는 i) 가이드 RNA; 및 ii) tranc RNA를 암호화하는 단일-분자 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산(예를 들어, DNA, 예를 들어, 재조합 발현 벡터)을 포함한다. 일부 경우에, 뉴클레오타이드 서열은 가이드 서열 없이 단일-분자 RNA의 가이드 RNA 부분을 암호화한다. 예를 들어, 일부 경우에, 핵산은 i) 가이드 RNA(가이드 서열 없이 가이드 RNA)의 불변 영역을 암호화하는 뉴클레오타이드 서열을 포함하고, 가이드 서열; 및 ii) tranc RNA를 암호화하는 핵산에 대한 삽입 부위를 포함한다.

[0378] 예를 들어, 일부 경우에, 본 개시내용의 키트는 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산(예를 들어, DNA, 예를 들어, 재조합 발현 벡터)를 포함한다. 일부 경우에, 뉴클레오타이드 서열은 가이드 서열 없이 가이드 RNA를 암호화한다. 예를 들어, 일부 경우에, 핵산은 가이드 RNA(가이드 서열 없이 가이드 RNA)의 불변 영역을 암호화하는 뉴클레오타이드 서열을 포함하고, 가이드 서열을 암호화하는 핵산에 대한 삽입 부위를 포함한다. 일부 경우에, 본 개시내용의 키트는 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산(예를 들어, mRNA, DNA, 예를 들어, 재조합 발현 벡터)를 포함한다.

[0379] 일부 경우에, 가이드 RNA-암호화 뉴클레오타이드 서열은 프로모터, 예를 들어, 원핵 세포에서 기능성인 프로모터, 진핵 세포에서 기능성인 프로모터, 포유류 세포에서 기능성인 프로모터, 인간 세포에서 기능성인 프로모터 등에 작동 가능하게 연결된다. 일부 경우에, CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 프로모터, 예를 들어, 원핵 세포에서 기능성인 프로모터, 진핵 세포에서 기능성인 프로모터, 포유류 세포에서 기능성인 프로모터, 인간 세포에서 기능성인 프로모터, 세포 유형-특이적 프로모터, 조절 가능한 프로모터, 조직-특이적 프로모터 등에 작동 가능하게 연결된다.

[0380] **효용**

[0381] CasZ 조성물(예를 들어, 발현 벡터, 키트, 조성물, 핵산 등)은 다양한 방법에서의 용도를 발견한다. 예를 들어, 본 개시내용의 CasZ 조성물은 (i) 표적 핵산(DNA 또는 RNA; 단일 가닥 또는 이중 가닥)을 변형(예를 들어, 절단, 예를 들어, 틈내기; 메틸화 등); (ii) 표적 핵산의 전사를 조절; (iii) 표적 핵산을 표지; (iv) (예를 들어, 단리, 표지, 영상화, 추적 등의 목적을 위해) 표적 핵산에 결합; (v) 표적 핵산과 회합된 폴리펩타이드(예를 들어, 히스톤)를 변형 등을 위해 사용될 수 있다. 따라서, 본 개시내용은 표적 핵산을 변형시키는 방법을 제공한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을: a) 본 개시내용의 CasZ 폴리펩타이드; 및 b) 하나 이상의(예를 들어, 2개의) CasZ 가이드 RNA와 접촉시키는 단계를 포함한다. 일부 경

우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을: a) CasZ 폴리펩타이드; 및 b) 하나 이상의(예를 들어, 2개의) CasZ 가이드 RNA, 및 c) CasZ trancRNA와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 a) 본 개시내용의 CasZ 폴리펩타이드; b) CasZ 가이드 RNA; 및 c) 공여자 핵산(예를 들어, 공여자 주형)과 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 a) CasZ 폴리펩타이드; b) CasZ 가이드 RNA; c) CasZ trancRNA, 및 d) 공여자 핵산(예를 들어, 공여자 주형)과 접촉시키는 단계를 포함한다. 일부 경우에, 접촉시키는 단계는 시험관내 세포에서 수행된다. 일부 경우에, 접촉시키는 단계는 생체내 세포에서 수행된다. 일부 경우에, 접촉시키는 단계는 생체외 세포에서 수행된다.

[0382] CasZ 폴리펩타이드를 사용하는 방법은 (회합된 CasZ 가이드 RNA에 의해 거기서 표적화한 덕분에) 표적 핵산 내 특정 영역에 대한 CasZ 폴리펩타이드의 결합을 포함하기 때문에, 상기 방법은 일반적으로 본 명세서에서 결합 방법(예를 들어, 표적 핵산의 결합 방법)으로서 지칭된다. 그러나, 일부 경우에, 결합 방법이 표적 핵산의 결합에 불과할 수 있지만, 다른 경우에, 상기 방법은 상이한 최종 결과를 가질 수 있다는 것이 이해되어야 한다(예를 들어, 상기 방법은 표적 핵산의 변형, 예를 들어, 절단/메틸화/등, 표적 핵산으로부터 전사의 조절; 표적 핵산의 번역의 조절; 게놈 편집; 표적 핵산과 회합된 단백질의 조절; 표적 핵산의 단리 등을 초래할 수 있다).

[0383] (예를 들어, CRISPR/Cas9 시스템과 함께 사용되는) 적합한 방법의 예에 대해, 예를 들어, 문헌[Jinek et al., Science. 2012 Aug 17;337(6096):816-21; Chylinski et al., RNA Biol. 2013 May;10(5):726-37; Ma et al., Biomed Res Int. 2013;2013:270805; Hou et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9; Jinek et al., Elife. 2013;2:e00471; Pattanayak et al., Nat Biotechnol. 2013 Sep;31(9):839-43; Qi et al, Cell. 2013 Feb 28;152(5):1173-83; Wang et al., Cell. 2013 May 9;153(4):910-8; Auer et al., Genome Res. 2013 Oct 31; Chen et al., Nucleic Acids Res. 2013 Nov 1;41(20):e19; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; Cho et al., Genetics. 2013 Nov;195(3):1177-80; DiCarlo et al., Nucleic Acids Res. 2013 Apr;41(7):4336-43; Dickinson et al., Nat Methods. 2013 Oct;10(10):1028-34; Ebina et al., Sci Rep. 2013;3:2510; Fujii et al, Nucleic Acids Res. 2013 Nov 1;41(20):e187; Hu et al., Cell Res. 2013 Nov;23(11):1322-5; Jiang et al., Nucleic Acids Res. 2013 Nov 1;41(20):e188; Larson et al., Nat Protoc. 2013 Nov;8(11):2180-96; Mali et. at., Nat Methods. 2013 Oct;10(10):957-63; Nakayama et al., Genesis. 2013 Dec;51(12):835-43; Ran et al., Nat Protoc. 2013 Nov;8(11):2281-308; Ran et al., Cell. 2013 Sep 12;154(6):1380-9; Upadhyay et al., G3 (Bethesda). 2013 Dec 9;3(12):2233-8; Walsh et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15514-5; Xie et al., Mol Plant. 2013 Oct 9; Yang et al., Cell. 2013 Sep 12;154(6):1370-9]; 및 미국 특허 및 특허 출원 제8,906,616호; 제8,895,308호; 제8,889,418호; 제8,889,356호; 제8,871,445호; 제8,865,406호; 제8,795,965호; 제8,771,945호; 제8,697,359호; 제20140068797호; 제20140170753호; 제20140179006호; 제20140179770호; 제20140186843호; 제20140186919호; 제20140186958호; 제20140189896호; 제20140227787호; 제20140234972호; 제20140242664호; 제20140242699호; 제20140242700호; 제20140242702호; 제20140248702호; 제20140256046호; 제20140273037호; 제20140273226호; 제20140273230호; 제20140273231호; 제20140273232호; 제20140273233호; 제20140273234호; 제20140273235호; 제20140287938호; 제20140295556호; 제20140295557호; 제20140298547호; 제20140304853호; 제20140309487호; 제20140310828호; 제20140310830호; 제20140315985호; 제20140335063호; 제20140335620호; 제20140342456호; 제20140342457호; 제20140342458호; 제20140349400호; 제20140349405호; 제20140356867호; 제20140356956호; 제20140356958호; 제20140356959호; 제20140357523호; 제20140357530호; 제20140364333; 및 20140377868호를 참조하며; 이들 각각은 본 명세서에 그들의 전문이 참고로 포함된다.

[0384] 예를 들어, 본 개시내용은 표적 핵산의 절단 방법; 표적 핵산의 편집 방법; 표적 핵산으로부터 전사를 조절하는 방법; 표적 핵산을 단리시키는 방법, 표적 핵산을 결합시키는 방법, 표적 핵산의 영상화 방법, 표적 핵산을 변형시키는 방법 등을 제공한다(그러나 이들로 제한되지 않는다).

[0385] 본 명세서에서 사용되는 바와 같은, 예를 들어, CasZ 폴리펩타이드와 또는 CasZ 융합 폴리펩타이드 등과 "표적 핵산을 접촉시키다" 및 "표적 핵산을 접촉시키는"이라는 용어/어구는 표적 핵산을 접촉시키기 위한 모든 방법을 포함한다. 예를 들어, CasZ 폴리펩타이드는 단백질, RNA(CasZ 폴리펩타이드를 암호화) 또는 DNA(CasZ 폴리펩타이드를 암호화)로서 세포에 제공될 수 있는 반면; CasZ 가이드 RNA는 가이드 RNA로서 또는 가이드 RNA를 암호화하는 핵산으로서 제공될 수 있고, CasZ trancRNA는 trancRNA로서 또는 trancRNA를 암호화하는 핵산으로서 제공될 수 있다. 이렇게 해서, 예를 들어, 세포에서(예를 들어, 시험관내 세포 내에서, 생체내 세포 내에서, 생체외 세포 내에서) 방법을 수행할 때, 표적 핵산을 접촉시키는 단계를 포함하는 방법은 그들의 활성/최종 상태에서

(예를 들어, CasZ 폴리펩타이드에 대한 단백질(들)의 형태로; CasZ 융합 폴리펩타이드에 대한 단백질의 형태로; 일부 경우에 가이드 RNA에 대한 RNA의 형태로) 임의의 또는 모든 성분의 세포 내로의 도입을 포함하고, 그리고 또한 하나 이상의 성분을 암호화하는 하나 이상의 핵산(예를 들어, CasZ 폴리펩타이드 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열(들)을 포함하는 핵산(들), 가이드 RNA(들)를 암호화하는 뉴클레오타이드 서열(들)을 포함하는 핵산(들), 공여자 주형을 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 등)의 세포 내로의 도입을 포함한다. 상기 방법은 또한 세포밖 시험관내에서 수행될 수 있기 때문에, 표적 핵산을 접촉시키는 단계를 포함하는 방법은, (달리 구체화되지 않는 한) 시험관내 세포 밖에서, 시험관내 세포 내에서, 생체내 세포 내에서, 생체외 세포 내 등에서 접촉시키는 단계를 포함한다.

[0386] 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 세포 내로 CasZ 좌위, 예를 들어, CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열뿐만 아니라 CasZ 좌위를 포함하는 세포로부터의 CasZ-암호화 뉴클레오타이드 서열을 둘러싸는 길이가 약 1 킬로베이스(kb) 내지 5kb인 뉴클레오타이드 서열을 포함하는 핵산을 도입하는 단계를 포함하되(예를 들어, 일부 경우에 천연 상태(천연에서 생기는 상태)에서 세포는 CasZ 좌위를 포함함), 여기서 표적 세포는 정상적으로는(그의 천연 상태에서는) CasZ 좌위를 포함하지 않는다(예를 들어, 일부 경우에 좌위는 CasZ trancRNA를 포함한다). 그러나, 암호화된 crRNA(들)에 대한 가이드 서열을 암호화하는 하나 이상의 스페이서 서열은 관심 대상의 하나 이상의 표적 서열이 표적화되도록 변형될 수 있다. 따라서, 예를 들어, 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 세포 내로 CasZ 좌위, 예를 들어, 공급원 세포로부터 얻은 핵산을 도입하는 단계를 포함하되(예를 들어, 일부 경우에 천연 상태(천연에서 생기는 상태)의 세포는 CasZ 좌위를 포함함), 여기서 핵산은 길이가 100개의 뉴클레오타이드(nt) 내지 5 kb(예를 들어, 길이가 100 nt 내지 500 nt, 500 nt 내지 1 kb, 1 kb 내지 1.5 kb, 1.5 kb 내지 2 kb, 2 kb 내지 2.5 kb, 2.5 kb 내지 3 kb, 3 kb 내지 3.5 kb, 3.5 kb 내지 4 kb, 또는 4 kb 내지 5 kb)의 길이를 가지며, CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함한다. 상기 언급한 바와 같이, 일부 이러한 경우에, 암호화된 crRNA(들)에 대한 가이드 서열을 암호화하는 하나 이상의 스페이서 서열은 관심 대상의 하나 이상의 표적 서열이 표적화되도록 변형될 수 있다. 일부 경우에, 상기 방법은 표적 세포 내로: i) CasZ 좌위; 및 ii) 공여자 DNA 주형을 도입하는 단계를 포함한다. 일부 경우에, 표적 핵산은 시험관내 무세포 조성물이다. 일부 경우에, 표적 핵산은 표적 세포에 존재한다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 원핵 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 진핵 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 포유류 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 식물 세포이다.

[0387] 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 본 개시내용의 CasZ 폴리펩타이드와, 또는 본 개시내용의 CasZ 융합 폴리펩타이드와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 CasZ 폴리펩타이드 및 CasZ 가이드 RNA와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 CasZ 폴리펩타이드, CasZ 가이드 RNA 및 CasZ trancRNA와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 CasZ 폴리펩타이드, 제1 CasZ 가이드 RNA, 및 제2 CasZ 가이드 RNA(및 일부 경우에 CasZ trancRNA)와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 본 개시내용의 CasZ 폴리펩타이드 및 CasZ 가이드 RNA 및 공여자 DNA 주형과 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 본 개시내용의 CasZ 폴리펩타이드 및 CasZ 가이드 RNA 및 CasZ trancRNA 및 공여자 DNA 주형과 접촉시키는 단계를 포함한다.

[0388] 일부 경우에, 표적 핵산은 시험관내 무세포 조성물이다. 일부 경우에, 표적 핵산은 표적 세포에 존재한다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 원핵 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 진핵 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 포유류 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 식물 세포이다.

[0389] ***관심 대상의 표적 핵산 및 표적 세포***

[0390] 표적 핵산은 임의의 핵산(예를 들어, DNA, RNA)일 수 있고, 이중 가닥 또는 단일 가닥일 수 있으며, 임의의 유형의 핵산(예를 들어, 염색체(게놈 DNA), 염색체로부터 유래, 염색체 DNA, 플라스미드, 바이러스, 세포외, 세포내, 미토콘드리아, 엽록체, 선형, 원형 등)일 수 있고, (예를 들어, 표적 핵산이 표적화될 수 있도록 CasZ 가이드 RNA가 표적 핵산 내 표적 서열에 혼성화되는 뉴클레오타이드 서열을 포함하는 한) 임의의 유기체로부터 유래

될 수 있다.

[0391] 표적 핵산은 DNA 또는 RNA일 수 있다. 표적 핵산은 이중 가닥(예를 들어, dsDNA, dsRNA) 또는 단일 가닥(예를 들어, ssRNA, ssDNA)일 수 있다. 일부 경우에, 표적 핵산은 단일 가닥이다. 일부 경우에, 표적 핵산은 단일 가닥 RNA(ssRNA)이다. 일부 경우에, 표적 ssRNA(예를 들어, 표적 세포 ssRNA, 바이러스 ssRNA 등)은 mRNA, rRNA, tRNA, 비-암호 RNA(ncRNA), 긴 비-암호 RNA(lncRNA) 및 마이크로RNA(miRNA)로부터 선택된다. 일부 경우에, 표적 핵산은 단일 가닥 DNA(ssDNA)(예를 들어, 바이러스 DNA)이다. 상기 언급한 바와 같이, 일부 경우에, 표적 핵산은 단일 가닥이다.

[0392] 표적 핵산은 어디에서나, 예를 들어, 시험관내 세포 밖에서, 시험관내 세포 내에서, 생체내 세포 내에서, 생체외 세포 내 등에 위치될 수 있다. 적합한 표적 세포(표적 핵산, 예컨대 게놈 DNA를 포함할 수 있음)는 박테리아 세포; 고세균 세포; 단세포 진핵 유기체의 세포; 식물 세포; 조류 세포, 예를 들어, 보트리오코쿠스 브라우니, 클라미도모나스 레인하르티, 난노클로롭시스 가디타나, 클로렐라 피레노이도사, 모자반속, 매생이 등; 진균 세포(예를 들어, 효모 세포); 동물 세포; 무척추동물(예를 들어, 과실파리, 강장동물, 극피동물, 선충 등)로부터의 세포; 곤충(예를 들어, 모기; 벌; 해충 등)의 세포; 거미류(예를 들어, 거미; 진드기 등)의 세포; 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포; 포유류(예를 들어, 설치류로부터의 세포; 인간으로부터의 세포; 비-인간 포유류의 세포; 설치류(예를 들어, 마우스, 래트)의 세포; 토끼목(예를 들어, 토끼)의 세포; 유제류(예를 들어, 소, 말, 낙타, 라마, 비쿠냐, 양, 염소 등)의 세포; 해양 포유류(예를 들어, 고래, 바다표범, 바다 코끼리, 돌고래, 바다사자 등) 등을 포함하지만, 이들로 제한되지 않는다. 임의의 유형의 세포에 관심이 있을 수 있다(예를 들어, 줄기 세포, 예를 들어, 배아 줄기(ES) 세포, 유도 만능 줄기(iPS) 세포, 생식세포(예를 들어, 난모세포, 정자, 난원세포, 정원세포 등), 성체 줄기 세포, 체세포, 예를 들어, 섬유아세포, 조혈모세포, 뉴런, 근육 세포, 골세포, 간세포, 췌장 세포; 임의의 단계에서 배아의 시험관내 또는 생체내 배아 세포, 예를 들어, 1-세포, 2-세포, 4-세포, 8-세포 등 단계 제브라피쉬 배아 등).

[0393] 세포는 확립된 세포주로부터 유래될 수 있거나 또는 그들은 1차 세포일 수 있고, 여기서 "1차 세포", "1차 세포주" 및 "1차 배양물"은 대상으로부터 유래되고 배양물의 제한된 수의 계대, 즉, 분할 동안 시험관내에서 성장이 허용되는 세포 및 세포 배양물을 지칭하기 위해 본 명세서에서 상호 호환적으로 사용된다. 예를 들어, 1차 배양물은 0회, 1회, 2회, 4회, 5회, 10회 또는 15회(그러나, 위기의 단계를 통과하기에 충분한 시간은 아님) 계대될 수 있는 배양물이다. 전형적으로, 1차 세포주는 시험관내에서 10회보다 더 적은 계대 동안 유지된다. 표적 세포는 단세포 유기체일 수 있고/있거나 배양물에서 성장될 수 있다. 세포가 1차 세포라면, 그들은 임의의 편리한 방법에 의해 개체로부터 채취될 수 있다. 예를 들어, 백혈구는 분리반출법, 백혈구분리반출법, 밀도 구배 분리 등에 의해 편리하게 채취될 수 있는 한편, 조직, 예컨대 피부, 근육, 골수, 비장, 간, 췌장, 폐, 장, 위 등으로부터의 세포는 생검에 의해 편리하게 채취될 수 있다.

[0394] 일부 상기 적용에서, 대상 방법은 (예를 들어, 표적화된 mRNA에 의해 암호화된 단백질의 생성을 파괴하기 위해, 표적 DNA를 절단하거나 또는 달리 변형시키기 위해, 표적 세포 등을 유전적으로 변형시키기 위해) 생체내 및/또는 생체외 및/또는 시험관내 유사분열 또는 유사분열후 세포에서 표적 핵산 절단, 표적 핵산 변형을 유도하기 위해, 그리고/또는 표적 핵산에 결합하기 위해(예를 들어, 수집 및/또는 분석 등을 위한 시각화를 위해) 사용될 수 있다. 가이드 RNA는 표적 핵산에 혼성화함으로써 특이성을 제공하기 때문에, 개시된 방법에서 관심 대상의 유사분열 및/또는 유사분열후 세포는 임의의 유기체로부터의 세포(예를 들어, 박테리아 세포, 고세균 세포, 단일 세포 진핵 유기체의 세포, 식물 세포, 조류 세포, 예를 들어, 보트리오코쿠스 브라우니, 칼라마이도모나스 레인하르티, 난노클로롭시스 가디타나, 클로렐라 피레노이도사, 모자반속 등, 진균 세포(예를 들어, 효모 세포), 동물 세포, 무척추동물(예를 들어, 과실파리, 강장동물, 극피동물, 선충 등)로부터의 세포, 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포, 포유류로부터의 세포, 설치류로부터의 세포, 인간으로부터의 세포 등)를 포함할 수 있다. 일부 경우에, 대상 CasZ 단백질(및/또는 단백질을 암호화하는 핵산, 예컨대 DNA 및/또는 RNA), 및/또는 CasZ 가이드 RNA(및/또는 가이드 RNA를 암호화하는 DNA), 및/또는 공여자 주형, 및/또는 RNP는 개체에게 도입될 수 있다(즉, 표적 세포는 생체내일 수 있다)(예를 들어, 포유류, 래트, 마우스, 돼지, 영장류, 비-인간 영장류, 인간 등). 일부 경우에, 이러한 투여는, 예를 들어, 표적화된 세포의 게놈을 편집함으로써 질환을 치료하고/하거나 예방하는 목적을 위한 것일 수 있다.

[0395] 식물 세포는 외떡잎식물의 세포, 및 쌍떡잎식물의 세포를 포함한다. 세포는 뿌리 세포, 잎 세포, 물관부의 세포, 체관부의 세포, 형성층의 세포, 정단분열조직 세포, 격벽 유세포, 후각조직, 후막조직 세포 등일 수 있다. 식물 세포는 농작물, 예컨대 밀, 옥수수, 벼, 수수, 조, 대두 등의 세포를 포함한다. 식물 세포는 농업

과일 및 견과 식물, 예를 들어, 살구, 오렌지, 레놈, 사과, 자두, 배, 아몬드 등을 생산하는 식물을 포함한다.

[0396] 표적 세포의 추가적인 예는 상기 표제 "변형 세포" 부문에서 열거된다. 세포(표적 세포)의 비제한적 예는 원핵 세포, 진핵 세포, 박테리아 세포, 고세균 세포, 단세포 진핵 유기체의 세포, 원생동물 세포, 식물로부터의 세포(예를 들어, 식물 작물, 과일, 채소, 곡물, 대두, 옥수수, 마이스, 밀, 종자, 토마토, 벼, 카사바, 사탕수수, 호박, 건초, 감자, 목화, 대마초, 담배, 하훼, 구과 식물, 겉씨식물, 속씨식물, 양치식물, 석송, 뿔이끼류, 우산이끼류, 선태식물, 쌍떡잎식물, 외떡잎식물 등), 조류 세포(예를 들어, 보트리오코쿠스 브라우니, 칼라마이도모나스 레인하르티, 난노클로롭시스 가디타나, 클로렐라 피레노이도사, 모자반속 등), 해초(예를 들어, 켈프) 진균 세포(예를 들어, 효모 세포, 버섯으로부터의 세포), 동물 세포, 무척추동물(예를 들어, 과실파리, 강장동물, 극피동물, 선충 등)으로부터의 세포, 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포, 포유류(예를 들어, 유제류(예를 들어, 돼지, 소, 염소, 양)로부터의 세포; 설치류(예를 들어, 래트, 마우스); 비-인간 영장류; 인간; 고양이과(예를 들어, 고양이); 갯과(예를 들어, 개); 등) 등을 포함한다. 일부 경우에, 세포는 천연 유기체로부터 유래되지 않은 세포이다(예를 들어, 세포는 합성으로 만들어진 세포일 수 있고; 또한 인공 세포로서 지칭된다).

[0397] 세포는 시험관내 세포(예를 들어, 확립된 배양 세포주)일 수 있다. 세포는 생체외 세포(개체로부터의 배양 세포)일 수 있다. 세포는 생체내 세포(예를 들어, 개체 내 세포)일 수 있다. 세포는 단리된 세포일 수 있다. 세포는 유기체 내부의 세포일 수 있다. 세포는 유기체일 수 있다. 세포는 세포 배양물 내 세포(예를 들어, 시험관내 세포 배양물)일 수 있다. 세포는 세포 수집물 중 하나일 수 있다. 세포는 원핵 세포이거나 또는 원핵 세포로부터 유래될 수 있다. 세포는 박테리아 세포일 수 있거나 또는 박테리아 세포로부터 유래될 수 있다. 세포는 고세균 세포일 수 있거나 또는 고세균 세포로부터 유래될 수 있다. 세포는 진핵 세포일 수 있거나 또는 진핵 세포로부터 유래될 수 있다. 세포는 식물 세포일 수 있거나 또는 식물 세포로부터 유래될 수 있다. 세포는 동물 세포일 수 있거나 또는 동물 세포로부터 유래될 수 있다. 세포는 무척추동물 세포일 수 있거나 또는 무척추동물 세포로부터 유래될 수 있다. 세포는 척추동물 세포일 수 있거나 또는 척추동물 세포로부터 유래될 수 있다. 세포는 포유류 세포일 수 있거나 또는 포유류 세포로부터 유래될 수 있다. 세포는 설치류 세포일 수 있거나 또는 설치류 세포로부터 유래될 수 있다. 세포는 인간 세포일 수 있거나 또는 인간 세포로부터 유래될 수 있다. 세포는 미생물 세포일 수 있거나 또는 미생물 세포로부터 유래될 수 있다. 세포는 진균 세포일 수 있거나 또는 진균 세포로부터 유래될 수 있다. 세포는 곤충 세포일 수 있다. 세포는 절지동물 세포일 수 있다. 세포는 원생동물 세포일 수 있다. 세포는 유충 세포일 수 있다.

[0398] 적합한 세포는 줄기 세포(예를 들어, 배아 줄기(ES) 세포, 유도 만능 줄기(iPS) 세포; 생식세포(예를 들어, 난모세포, 정자, 난원세포, 정원세포 등); 체세포, 예를 들어, 섬유아세포, 희소돌기신경교, 교세포, 조혈모세포, 뉴런, 근육 세포, 골세포, 간세포, 췌장 세포 등을 포함한다.

[0399] 적합한 세포는 인간 배아 줄기 세포, 태아 심장 근육 세포, 근육섬유아세포, 간충직 줄기 세포, 자가이식 확장된 심장 근육 세포, 지방세포, 전분화 세포, 만능 세포, 혈액 줄기 세포, 근아세포, 성체 줄기 세포, 골수 세포, 간충직 세포, 배아 줄기 세포, 실질 세포, 상피세포, 내피세포, 중피세포, 섬유아세포, 조골세포, 연골세포, 외인성 세포, 내인성 세포, 줄기 세포, 조혈모세포, 골수 유래 전구체 세포, 심근세포, 골격 세포, 태아 세포, 미분화 세포, 다능 전구체 세포, 단분화능 전구체 세포, 단핵구, 심장 근아세포, 골격 근아세포, 대식세포, 모세혈관 내피세포, 이종 세포, 동종 세포 및 출산후 줄기 세포를 포함한다.

[0400] 일부 경우에, 세포는 면역세포, 뉴런, 상피세포 및 내피세포 또는 줄기 세포이다. 일부 경우에, 면역세포는 T 세포, B 세포, 단핵구, 자연살해 세포, 수지상 세포 또는 대식세포이다. 일부 경우에, 면역세포는 세포독성 T 세포이다. 일부 경우에, 면역세포는 헬퍼 T 세포이다. 일부 경우에, 면역세포는 조절 T 세포(Treg)이다.

[0401] 일부 경우에, 세포는 줄기 세포이다. 줄기 세포는 성체 줄기 세포를 포함한다. 성체 줄기 세포는 또한 체세포 줄기 세포로서 지칭된다.

[0402] 성체 줄기 세포는 분화된 조직 내에 존재하지만, 자기 재생 특성 및 다중 세포 유형, 보통 줄기 세포가 발견된 전형적인 조직의 세포 유형이 생기게 하는 능력을 보유한다. 체세포 줄기 세포의 수많은 예는 당업자에게 공지되어 있으며, 근육 줄기 세포; 조혈모세포; 상피 줄기 세포; 신경 줄기 세포; 간충직 줄기 세포; 유방 줄기 세포; 장 줄기 세포; 중배엽 줄기 세포; 내피 줄기 세포; 후각 줄기 세포; 신경 능선 줄기 세포 등을 포함한다.

[0403] 관심 대상의 줄기 세포는 포유류 줄기 세포를 포함하며, 용어 "포유류"는 인간; 비-인간 영장류; 가축 및 농장 동물; 및 동물원, 실험실, 운동 또는 반려 동물, 예컨대 개, 말, 고양이, 소, 마우스, 래트, 토끼 등을 포함하

는 포유류로서 분류된 임의의 동물을 지칭한다. 일부 경우에, 줄기 세포는 인간 줄기 세포이다. 일부 경우에, 줄기 세포는 설치류(예를 들어, 마우스; 래트) 줄기 세포이다. 일부 경우에, 줄기 세포는 비-인간 영장류 줄기 세포이다.

[0404] 줄기 세포는 하나 이상의 줄기 세포 마커, 예를 들어, SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17 및 PPARGC1A를 발현시킬 수 있다.

[0405] 일부 경우에, 줄기 세포는 조혈모세포(HSC)이다. HSC는 골수, 혈액, 제대혈, 태아간 및 난황낭으로부터 단리될 수 있는 중배엽-유래 세포이다. HSC는 $CD34^+$ 및 $CD3^-$으로서 특성규명된다. HSC는 생체내 적혈구, 호중성-대식세포, 거핵구 및 림프계 조혈세포 계통을 재생시킬 수 있다. 시험관내에서, HSC는 적어도 일부의 자기-재생 세포분열을 겪도록 유도될 수 있고, 생체내에서 알 수 있는 바와 같이 동일한 계통으로 분화되도록 유도될 수 있다. 이렇게 해서, HSC는 적혈구 세포, 거핵구, 호중구, 대식세포 및 림프구 세포 중 하나 이상으로 분화되도록 유도될 수 있다.

[0406] 다른 경우에, 줄기 세포는 신경 줄기 세포(NSC)이다. 신경 줄기 세포(NSC)는 뉴런, 및 교세포(희소돌기신경교 및 성상교세포를 포함)로 분화할 수 있다. 신경 줄기 세포는 다회 분열할 수 있는 다능성 줄기 세포이고, 특정 조건 하에서 신경아세포 또는 아교모세포, 예를 들어, 각각 뉴런 및 교세포 중 하나 이상의 유형이 되도록 기여하는 세포일 수 있는 신경 줄기 세포 또는 신경 전구체 세포인 딸 세포를 생성할 수 있다. NSC를 얻는 방법은 당업계에 공지되어 있다.

[0407] 다른 경우에, 줄기 세포는 간충직 줄기 세포(MSC)이다. MSC는 본래 배아 중배엽으로부터 유래되고, 성체 골수로부터 단리되며, 근육, 뼈, 연골, 지방, 골수 기질 및 힘줄을 형성하도록 분화될 수 있다. MSC를 단리시키는 방법은 당업계에 공지되어 있으며; MSC를 얻기 위한 임의의 공지된 방법이 사용될 수 있다. 예를 들어, 인간 MSC의 단리를 기재하는 미국 특허 제5,736,396호 참조.

[0408] 세포는 일부 경우에 식물 세포이다. 식물 세포는 외떡잎식물의 세포일 수 있다. 세포는 쌍떡잎식물의 세포일 수 있다.

[0409] 일부 경우에, 세포는 식물 세포이다. 예를 들어, 세포는 주요 농작물, 예를 들어, 보리, 콩류(건조 식용), 카놀라, 옥수수, 목화(피마), 목화(업랜드), 아마씨, 건초(알팔파), 건초(비-알팔파), 귀리, 땅콩, 벼, 수수, 대두, 사탕무, 사탕수수, 해바라기(오일), 해바라기(비-오일), 고구마, 담배(벌리), 담배(연관처리), 토마토, 밀(듀럼), 밀(봄), 밀(겨울) 등의 세포일 수 있다. 다른 예로서, 세포는, 예를 들어, 알팔파, 알로에 잎, 애로루트, 벗풀, 아티초크, 아스파라거스, 죽순, 바나나꽃, 콩나물, 콩류, 비이트톱, 근대, 비타멜론, 청경채, 브로콜리, 브로콜리 라베(엽경채), 방울다다기 양배추, 양배추, 방울 양배추, 가재발 선인장(노팔), 호리병박 열매, 카르둔, 당근, 콜리플라워, 셀러리, 차요테, 중국 아티초크(두루미냉이), 중국 양배추, 중국 셀러리, 중국 차이브, 채심, 국화잎(통호), 콜라드 그린, 옥수수대, 단옥수수, 오이, 대컨, 단데리온 그린, 타로토란, 다우 무에(피 팁), 동과(겨울멜론), 가지, 엔다이브, 에스카롤, 선수 장식 양치식물, 갓류 식물, 치커리, 갓(중국 겨자), 가일론, 갈랑가(시암, 태국 생강), 마늘, 생강 뿌리, 우엉, 나물, 하노버 샐러드 나물, 후아우존틀, 예루살렘 아티초크, 히카마, 케일 나물, 콜라비, 명아주(꼘리떼), 상추(비브), 상추(보스턴), 상추(보스턴 레드), 상추(녹색잎), 상추(아이스버그), 상추(롤라로사), 상추(오크리프 - 녹색), 상추(오크리프 - 적색), 상추(가공처리), 상추(적색잎), 상추(로메인), 상추(루비 로메인), 상추(러시안 레드 겨자), 린쿡, 로 복, 긴콩류, 연근, 마하, 용설란(아가베) 잎, 말랑가, 메스꿀랭 믹스, 겨자과, 모압(수세미오이), 무, 모쿠아(동과), 버섯, 겨자, 참마, 오크라, 옹초이, 그린 어니언, 오포(롱 스쿼시), 장식용 옥수수, 관상용 호박, 파슬리, 파스닙, 완두콩, 후추(벨형), 후추, 호박, 라디키오, 무순, 무, 레이프 그린, 레이프 그린, 대황, 로메인(베이비 레드), 루타베가, 함초(씨빈), 신쿠아(앵글드/줄기가 선 수세미), 시금치, 스쿼시, 짚 더미, 사탕수수, 고구마, 근대, 타마린드, 타로, 타로잎, 타로 슈트, 탓소이, 테페구아헤(구아헤), 빔바, 토마틸로, 토마토, 토마토 (체리), 토마토(포도형), 토마토(자두형), 강황, 순무 어린잎 나물, 순무, 마름, 얌피, 얌스(네임스), 유초이, 유카(카사바) 등을 포함하지만, 이들로 제한되지 않는 채소 작물의 세포이다.

[0410] 세포는 일부 경우에 절지동물 세포이다. 예를 들어, 세포는, 예를 들어, 협각류, 다지류, 헥시포디아, 거미강, 곤충강, 돌좀목, 좀류, 구시아강, 하루살이목, 잠자리목, 잠자리하목, 실잠자리아목, 신시류, 외시류, 강도래목, 흰개미붙이목, 메뚜기, 민벌레목, 집게벌레목, 바퀴목, 귀뚜라미붙이목, 갈르와벌레과, 만토파스마토데과, 대벌레목, 바퀴목, 흰개미목, 사마귀목, 책좀, 다듬이벌레목, 총채벌레목, 이목, 반시목, 내시류 또는 완전변태류, 막시류, 초시류, 부채벌레목, 약대벌레목, 뱀잠자리목, 풀잠자리목, 밑드리목, 버룩목, 파리목, 날도

래류 또는 인시목의 아목, 과, 아과, 군, 하위군, 또는 종의 세포일 수 있다.

[0411] 세포는 일부 경우에 곤충 세포이다. 예를 들어, 일부 경우에, 세포는 모기, 메뚜기, 노린재류, 파리, 벼룩, 벌, 말벌, 개미, 이, 나방 또는 딱정벌레의 세포이다.

[0412] ***공여자 폴리뉴클레오타이드(공여자 주형)***

[0413] CasZ 가이드 RNA에 의해 가이드된, CasZ 단백질은 일부 경우에 비-상동성 말단 접합(NHEJ) 또는 상동성-관련 재조합(HDR) 중 하나에 의해 수선된 이중-가닥 DNA(dsDNA) 표적 핵산 내에서 부위-특이적 이중 가닥 파손(double strand break: DSB) 또는 단일 가닥 파손(SSB)(예를 들어, CasZ 단백질이 틈내기효소 변이체일 때)을 생성한다.

[0414] 일부 경우에, 표적 DNA를 (CasZ 단백질 및 CasZ 가이드 RNA와) 접촉시키는 것은 비상동성 말단 결합 또는 상동 직접 수선을 허용하는 조건 하에서 일어난다. 따라서, 일부 경우에, 대상 방법은 표적 DNA를 공여자 폴리뉴클레오타이드와 (예를 들어, 공여자 폴리뉴클레오타이드를 세포 내로 도입함으로써) 접촉시키는 단계를 포함하되, 공여자 폴리뉴클레오타이드, 공여자 폴리뉴클레오타이드의 일부, 공여자 폴리뉴클레오타이드의 복제물, 또는 공여자 폴리뉴클레오타이드 복제물의 일부가 표적 DNA 내로 통합된다. 일부 경우에, 상기 방법은 세포를 공여자 폴리뉴클레오타이드와 접촉시키는 단계를 포함하지 않으며, 표적 DNA 내의 뉴클레오타이드가 결실되도록 표적 DNA는 변형된다.

[0415] 일부 경우에, CasZ trancRNA(또는 이를 암호화하는 핵산), CasZ 가이드 RNA(또는 이를 암호화하는 핵산), 및/또는 CasZ 단백질(또는 이를 암호화하는 핵산, 예컨대 RNA 또는 DNA, 예를 들어, 하나 이상의 발현 벡터)는 적어도 표적 DNA 서열에 상동성인 세그먼트를 포함하는 공여자 폴리뉴클레오타이드 서열과 함께 공동투여되며(예를 들어, 표적 핵산과 접촉되고, 세포에 투여됨 등), 대상 방법은 핵산 물질을 표적 DNA 서열에 첨가하기 위해, 즉, 삽입 또는 대체하기 위해(예를 들어, 핵산, 예를 들어, 단백질을 암호화하는 것, siRNA, miRNA 등을 "넉인"하기 위해), 태그(예를 들어, 6xHis, 형광 단백질(예를 들어, 녹색 형광 단백질; 황색형광 단백질 등), 혈구응집소(HA), FLAG 등)를 첨가하기 위해, 조절 서열을 유전자(예를 들어, 프로모터, 폴리아데닐화 신호, 내부 리보솜 유입 서열(IRES), 2A 펩타이드, 개시 코돈, 정지 코돈, 스플라이스 신호, 국소화 신호 등)를 첨가하기 위해, 핵산 서열(예를 들어, 돌연변이를 도입하기 위해, 정확한 서열을 도입함으로써 돌연변이를 야기하는 질환을 제거하기 위해) 등을 변형시키기 위해 사용될 수 있다. 이렇게 해서, CasZ 가이드 RNA 및 CasZ 단백질(또는 CasZ 가이드 RNA 및 CasZ trancRNA 및 CasZ 단백질)을 포함하는 복합체는 부위-특이적, 즉, "표적화된" 방법으로 DNA를 변형시키는 것이 바람직한 임의의 시험관내 또는 생체내 적용, 예를 들어, 유전자 요법에서 사용되는 바와 같은, 예를 들어, 유전자 넉인, 유전자 편집, 유전자 태깅 등에서, 예를 들어, 질환을 치료하거나 또는 항바이러스, 항병원성 또는 항암 치료제로서, 농업에서 유전자 변형된 유기체의 생산, 치료적, 진단적 또는 연구적 목적을 위한 세포에 의한 단백질의 생성, iPS 세포의 유도, 생물학적 연구, 검출 또는 대체 등을 위한 병원균 유전자의 표적화에서 유용하다.

[0416] 폴리뉴클레오타이드 서열을 표적 서열이 절단되는 게놈 내로 삽입하는 것이 바람직한 적용에서, 공여자 폴리뉴클레오타이드(공여자 서열을 포함하는 핵산)는 또한 세포에 제공될 수 있다. "공여자 서열" 또는 "공여자 폴리뉴클레오타이드" 또는 "공여자 주형"은 (예를 들어, dsDNA 절단 후에, 표적 DNA의 틈내기 후에, 표적 DNA의 이중 틈내기 후에 등) CasZ 단백질에 의해 절단된 부위에 삽입될 핵산 서열을 의미한다. 공여자 폴리뉴클레오타이드는 그것과 상동성을 보유하는 게놈 서열 사이의 상동 직접 수선을 뒷받침하기 위해, 예를 들어, 표적 부위의 약 50개 이하의 염기 이내, 예를 들어, 약 30개의 염기 이내, 약 15개의 염기 이내, 약 10개의 염기 이내, 약 5개의 염기 이내, 또는 표적 부위에 바로 측접하여 표적 부위에서 게놈 서열에 대한 충분한 상동성, 예를 들어, 표적 부위에 측접하는 뉴클레오타이드 서열과 70%, 80%, 85%, 90%, 95% 또는 100% 상동성을 함유할 수 있다. 공여자와 게놈 서열 사이의 서열 상동성의 대략 25, 50, 100 또는 200개 이상의 뉴클레오타이드, 또는 200개 초과의 뉴클레오타이드(또는 10 내지 200개 이상의 뉴클레오타이드의 임의의 정수값)는 상동 직접 수선을 뒷받침할 수 있다. 공여자 폴리뉴클레오타이드는 임의의 길이, 예를 들어, 10개 이상의 뉴클레오타이드, 50개 이상의 뉴클레오타이드, 100개 이상의 뉴클레오타이드, 250개 이상의 뉴클레오타이드, 500개 이상의 뉴클레오타이드, 1000개 이상의 뉴클레오타이드, 5000개 이상의 뉴클레오타이드 등을 가질 수 있다.

[0417] 공여자 서열은 전형적으로 그것이 대체하는 게놈 서열과 동일하지 않다. 오히려, 상동 직접 수선을 뒷받침하기 위해(예를 들어, 유전자 보정을 위해, 예를 들어, 질환 원인 염기쌍을 비질환 원인 염기쌍으로 전환시키기 위해) 충분한 상동성이 존재한다면 공여자 서열은 게놈 서열에 대해 적어도 하나 이상의 단일 염기 변화, 삽입, 결실, 역위 또는 재배열을 함유할 수 있다. 일부 실시형태에서, 표적 DNA 영역과 두 측접 서열 사이의 상동 직접 수선이 표적 영역에서 비상동성 서열의 삽입을 초래하도록 공여자 서열은 상동체의 두 영역에 의해 측접되는

비상동성 서열을 포함한다. 공여자 서열은 또한 관심 대상의 DNA 영역에 상동성이 아니고 관심 대상의 DNA 영역 내로의 삽입에 대해 의도되지 않는 서열을 함유하는 벡터 골격을 포함할 수 있다. 일반적으로, 공여자 서열의 상동성 영역(들)은 재조합이 요망되는 게놈 서열에 대해 적어도 50%의 서열 동일성을 가질 것이다. 특정 실시형태에서, 60%, 70%, 80%, 90%, 95%, 98%, 99% 또는 99.9%의 서열 동일성이 존재한다. 공여자 폴리뉴클레오타이드의 길이에 따라서 1% 내지 100%의 서열 동일성의 임의의 값이 존재할 수 있다.

[0418] 공여자 서열은 게놈 서열, 예를 들어, 제한 부위, 뉴클레오타이드 다형성, 선택 가능한 마커(예를 들어, 약물 내성 유전자, 형광 단백질, 효소 등) 등에 비교하여 특정 서열 차이를 포함할 수 있는데, 이는 절단 부위에서 공여자 서열의 성공적인 삽입에 대한 평가를 위해 사용될 수 있거나 또는 일부 경우에 다른 목적을 위해(예를 들어, 표적화된 게놈 좌위에서 발현을 나타내기 위해) 사용될 수 있다. 일부 경우에, 암호 영역에 위치된다면, 이러한 뉴클레오타이드 서열 차이는 아미노산 서열을 변화시키지 않을 것이고, 또는 침묵 아미노산 변화(즉, 단백질의 구조 또는 기능에 영향을 미치지 않는 변화)를 만들 것이다. 대안적으로, 이들 서열 차이는 측접하는 재조합 서열, 예컨대 FLP, loxP 서열 등을 포함할 수 있는데, 이는 마커 서열의 제거를 위해 이후의 시간에 활성화될 수 있다.

[0419] 일부 경우에, 공여자 서열은 단일-가닥 DNA로서 세포에 제공된다. 일부 경우에, 공여자 서열은 이중-가닥 DNA로서 세포에 제공된다. 이는 선형 또는 원형 형태로 세포에 도입될 수 있다. 선형 형태로 도입된다면, 공여자 서열의 말단은 임의의 편리한 방법에 의해 (예를 들어, 핵산외부분해로부터) 보호될 수 있고, 이러한 방법은 당업자에게 공지되어 있다. 예를 들어, 하나 이상의 다이데옥시뉴클레오타이드 잔기가 선형 분자의 3' 말단에 첨가될 수 있고/있거나 자기-상보성 올리고뉴클레오타이드가 말단 중 하나 또는 둘 다에 결찰될 수 있다. 예를 들어, 문헌[Chang et al. (1987) Proc. Natl. Acad Sci USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889] 참조. 분해로부터 외인성 폴리뉴클레오타이드를 보호하기 위한 추가적인 방법은 말단 아미노기(들)의 첨가 및 변형된 뉴클레오타이드간 결합, 예를 들어, 포스포로티오에이트, 포스포르아미데이트 및 O-메틸 리보스 또는 데옥시리보스 잔기의 사용을 포함하지만, 이들로 제한되지 않는다. 선형 공여자 서열 말단 보호에 대한 대안으로서, 서열의 추가적인 길이가 재조합에 영향을 미치는 일 없이 분해될 수 있는 영역 외부에 포함될 수 있다. 공여자 서열은 추가적인 서열, 예를 들어, 복제 기점, 프로모터 및 항생제 내성을 암호화하는 유전자를 갖는 벡터의 부분으로서 세포 내로 도입될 수 있다. 게다가, 공여자 서열은 네이키드 핵산으로서, 리포좀 또는 폴록사머와 같은 제제와 복합체화된 핵산으로서 도입될 수 있거나, 또는 CasZ 가이드 RNA 및/또는 CasZ 융합 폴리펩타이드 및/또는 공여자 폴리뉴클레오타이드를 암호화하는 핵산에 대해 본 명세서의 다른 곳에 기재된 바와 같이, 바이러스(예를 들어, 아데노바이러스, AAV)에 의해 전달될 수 있다.

[0420] **유전자이식, 비인간 유기체**

[0421] 상기 기재한 바와 같이, 일부 경우에, 본 개시내용(예를 들어, CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산; CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 등)의 핵산(예를 들어, 재조합 발현 벡터)은 본 개시내용의 CasZ 폴리펩타이드 또는 CasZ 융합 폴리펩타이드를 생성하는 유전자이식 비-인간 유기체를 생성하기 위해 이식유전자로서 사용된다. 본 개시내용은 본 개시내용의 CasZ 폴리펩타이드, 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 유전자이식 비인간 유기체를 제공한다.

[0422] ***유전자이식, 비-인간 동물***

[0423] 본 개시내용은 유전자이식 비-인간 동물을 제공하는데, 이 동물은 CasZ 폴리펩타이드 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 포함하는 이식유전자를 포함한다. 일부 실시형태에서, 유전자이식 비-인간 동물의 게놈은 본 개시내용의 CasZ 폴리펩타이드 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 유전자이식 비-인간 동물은 유전자 변형에 대해 동형접합적이다. 일부 경우에, 유전자이식 비-인간 동물은 유전자 변형에 대해 이형접합적이다. 일부 실시형태에서, 유전자이식 비-인간 동물은 척추동물, 예를 들어, 어류(예를 들어, 연어, 송어, 제브라피쉬, 금붕어, 복어, 동굴어 등), 양서류(개구리, 영원, 도롱뇽 등), 조류(예를 들어, 닭, 칠면조 등), 파충류(예를 들어, 뱀, 도마뱀 등), 비-인간 포유류(예를 들어, 유제류, 예를 들어, 돼지, 소, 염소, 양 등; 토끼목(예를 들어, 토끼); 설치류(예를 들어, 래트, 마우스); 비-인간 영장류 등) 등이다. 일부 경우에, 유전자이식 비-인간 동물은 무척추동물이다. 일부 경우에, 유전자이식 비-인간 동물은 곤충(예를 들어, 모기; 해충 등)이다. 일부 경우에, 유전자이식 비-인간 동물은 거미류이다.

[0424] 본 개시내용의 CasZ 폴리펩타이드, 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 미공지

프로모터(예를 들어, 핵산이 숙주 세포 게놈 내로 무작위로 통합될 때) 제어 하에 있을 수 있거나(즉, 이에 작동 가능하게 연결됨) 또는 공지된 프로모터의 제어 하에 있을 수 있다(즉, 이에 작동 가능하게 연결됨). 적합한 공지된 프로모터는 임의의 공지된 프로모터일 수 있고, 구성적으로 활성인 프로모터(예를 들어, CMV 프로모터), 유도성 프로모터(예를 들어, 열 충격 프로모터, 테트라사이클린-조절된 프로모터, 스테로이드-조절된 프로모터, 금속-조절된 프로모터, 에스트로겐 수용체-조절 프로모터 등), 공간적으로 제한된 그리고/또는 일시적으로 제한된 프로모터(예를 들어, 조직 특이적 프로모터, 세포 유형 특이적 프로모터 등) 등을 포함한다.

[0425] ***유전자이식 식물***

[0426] 상기 기재한 바와 같이, 일부 경우에, 본 개시내용(예를 들어, 본 개시내용의 CasZ 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산; 본 개시내용의 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 등)의 핵산(예를 들어, 재조합 발현 벡터)은 본 개시내용의 CasZ 폴리펩타이드 또는 CasZ 융합 폴리펩타이드를 생성하는 유전자이식 식물을 생성하기 위해 이식유전자로서 사용된다. 본 개시내용은 본 개시내용의 CasZ 폴리펩타이드, 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 유전자이식 식물을 제공한다. 일부 실시형태에서, 유전자이식 식물의 게놈은 대상 핵산을 포함한다. 일부 실시형태에서, 유전자이식 식물은 유전자 변형에 대해 동형접합적이다. 일부 실시형태에서, 유전자이식 식물은 유전자 변형에 대해 이형접합적이다.

[0427] 식물 세포 내로 외인성 핵산을 도입하는 방법은 당업계에 잘 공지되어 있다. 이러한 식물 세포는 상기 정의한 바와 같이 "형질전환되는" 것으로 고려된다. 적합한 방법은 바이러스 감염(예컨대 이중 가닥 DNA 바이러스), 형질감염, 접합, 원형질체 융합, 전기천공법, 입자충법, 인산칼슘 침전, 직접 미량주사법, 탄화규소 위스커 기법, 아그로박테륨-매개 형질전환 등을 포함한다. 방법의 선택은 일반적으로 형질감염 중인 세포의 유형 및 형질감염이 일어나는 환경(즉, 시험관내, 생체외 또는 생체내)에 의존한다.

[0428] 토양 박테리아 아그로박테륨 투메파시엔스(*Agrobacterium tumefaciens*)에 기반한 형질전환 방법은 외인성 핵산 분자를 관다발 식물 내로 도입하는 데 특히 유용하다. 아그로박테륨의 야생형 형태는 숙주 식물 상에서 종양형성성 뿌리혹병 성장의 생산을 지시하는 Ti(종양-유도성) 플라스미드를 함유한다. 식물 게놈에 대한 Ti 플라스미드의 종양 유도성 T-DNA 영역의 전달은 Ti 플라스미드-암호화된 독력 유전자뿐만 아니라 T-DNA 경계를 필요로 하는데, 이는 전달될 영역을 설명하는 직접 DNA 반복부의 세트이다. 아그로벡테리움계 벡터는 Ti 플라스미드의 변형된 형태인데, 이때 종양 유도성 작용은 식물 숙주 내로 도입될 관심 대상의 핵산 서열에 의해 대체된다.

[0429] 아그로박테륨-매개 형질전환은 일반적으로 합체 벡터 또는 바이너리 벡터 시스템을 사용하는데, 이때 Ti 플라스미드의 성분은 아그로박테륨 숙주에서 영구적으로 체류하고 독력 유전자를 운반하는 헬퍼 벡터와 T-DNA 서열에 의해 결합되는 관심 대상의 유전자를 함유하는 셔틀 벡터 사이에서 분할된다. 다양한 바이너리 벡터는 당업계에 잘 공지되어 있고, 예를 들어, 클론테크(Clontech)(캘리포니아주 팔로알토에 소재)로부터 상업적으로 입수 가능하다. 배양된 식물 세포 또는 상처 조직, 예컨대 잎 조직, 뿌리 외식편, 배축, 줄기 조각 또는 덩이줄기와 함께 아그로박테륨을 배양시키는 방법은, 예를 들어, 또한 당업계에 잘 공지되어 있다. 예를 들어, 문헌[Glick and Thompson, (eds.), Methods in Plant Molecular Biology and Biotechnology, Boca Raton, Fla.: CRC Press (1993)] 참조.

[0430] 미세투사물-매개 형질전환이 또한 대상 유전자이식 식물을 생산하는 데 사용될 수 있다. 문헌[Klein et al. (Nature 327:70-73 (1987))]에 의해 처음 기재된 이 방법은 염화칼슘, 스퍼미딘 또는 폴리에틸렌 글리콜에 의한 침전에 의해 목적으로 하는 핵산 분자로 코팅된 금 또는 텅스텐과 같은 미세투사물에 의존한다. 미세투사물 입자는 BIOLISTIC PD-1000(바이오라드(Biorad); 캘리포니아주 허큘레스에 소재)와 같은 장치를 이용하여 속씨식물 내로 고속으로 가속화된다.

[0431] 핵산이, 예를 들어 생체내 또는 생체외 프로토콜을 통해 식물 세포(들) 내로 들어갈 수 있는 방식으로 본 개시내용의 핵산(예를 들어, 본 개시내용의 CasZ 폴리펩타이드 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산(예를 들어, 재조합 발현 벡터))은 식물 내로 도입될 수 있다. "생체내"는 핵산이 식물의 살아있는 몸체에 투여되는 것, 예를 들어, 침윤을 의미한다. "생체외"는 세포 또는 외식편이 식물 외부에서 변형되고, 이어서, 이러한 세포 또는 기관이 식물에 대해 재생된다는 것을 의미한다. 문헌[Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology Academic Press, 및 Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers]에 기재된 것을 포함하는 식물 세포의 안정한 형질전환 또는 유전자이식 식물의 확립에 적합한 다수의 벡터가 기재되었다. 구체적 예는 아그로박테륨 투메파시엔스의 Ti 플라스미드로부터 유래된 것뿐만 아니라 문헌[Herrera-Estrella et al. (1983) Nature 303: 209, Bevan

(1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technolo 3: 637-642]에 의해 개시된 것을 포함한다. 대안적으로, 비-Ti 벡터는 유리 DNA 전달 기법을 이용함으로써 식물 및 세포에 대한 DNA의 전달을 위해 사용될 수 있다. 이들 방법을 이용함으로써, 유전자이식 식물, 예컨대, 밀, 벼(Christou (1991) Bio/Technology 9:957-9 및 4462) 및 옥수수(Gordon-Kamm (1990) Plant Cell 2: 603-618)가 생산될 수 있다. 미성숙 배아는 또한 입자총(문헌[Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technolo 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48]을 이용함으로써 직접 DNA 전달 기법을 위한 그리고 아그로박테륨-매개 DNA 전달(Ishida et al. (1996) Nature Biotech 14: 745-750)을 위한 외떡잎식물에 대한 양호한 표적 조직일 수 있다. 엽록체 내로 DNA의 도입을 위한 예시적인 방법은 유전자충격, 프로토플라스트의 폴리에틸렌 글리콜 형질전환, 및 미량주사법이다(문헌[Danieli et al Nat. Biotechnol 16:345-348, 1998; Staub et al Nat. Biotechnol 18: 333-338, 2000; O'Neill et al Plant J. 3:729-738, 1993; Knoblauch et al Nat. Biotechnol 17: 906-909]; 미국 특허 제5,451,513호, 제5,545,817호, 제5,545,818호 및 제5,576,198호; 국제 특허 출원 WO 95/16783; 및 문헌[Boynton et al., Methods in Enzymology 217: 510-536 (1993), Svab et al., Proc. Natl. Acad. Sci. USA 90: 913-917 (1993), 및 McBride et al., Proc. Natl. Acad. Sci. USA 91: 7301-7305 (1994)]). 유전자충격, 프로토플라스트의 폴리에틸렌 글리콜 형질전환 및 미량주사법의 방법에 적합한 임의의 벡터는 엽록체 형질전환을 위한 표적화 벡터로서 적합할 것이다. 임의의 이중 가닥 DNA 벡터는 형질전환 벡터, 특히 도입 방법이 아그로박테륨을 이용하지 않을 때 사용될 수 있다.

[0432] 유전자 변형될 수 있는 식물은 곡물, 사료용 작물, 과일, 채소, 오일시드 작물, 야자나무, 삼림 및 덩굴식물을 포함한다. 변형될 수 있는 식물의 구체적 예는 다음에 따를 수 있다: 마이스, 바나나, 땅콩, 경협종완두, 해바라기, 토마토, 카놀라, 담배, 밀, 보리, 귀리, 감자, 대두, 목화, 카네이션, 수수, 루핀 및 벼.

[0433] 본 개시내용은 형질전환된 식물 세포, 조직, 식물 및 형질전환 식물 세포를 함유하는 생성물을 제공한다. 대상 형질전환 세포, 및 조직 및 이를 포함하는 생성물의 특징은 게놈 내로 통합되는 대상 핵산의 존재, 및 본 개시내용의 CasZ 폴리펩타이드, 또는 CasZ 융합 폴리펩타이드의 식물 세포에 의한 생성이다. 본 발명의 재조합 식물 세포는 재조합 세포의 집단으로서, 또는 조직, 종자, 전체 식물, 줄기, 열매, 잎, 뿌리, 꽃, 줄기, 덩이줄기, 곡물, 동물 사료, 식물의 밭 등으로서 유용하다.

[0434] 본 개시내용의 CasZ 폴리펩타이드, 또는 CasZ 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 미공지 프로모터(예를 들어, 핵산이 숙주 세포 게놈 내로 무작위로 통합될 때) 제어 하에 있을 수 있거나(즉, 이에 작동 가능하게 연결됨) 또는 공지된 프로모터의 제어 하에 있을 수 있다(즉, 이에 작동 가능하게 연결됨). 적합한 알려진 프로모터는 임의의 공지된 프로모터일 수 있고 구성적으로 활성인 프로모터, 유도성 프로모터, 공간적으로 제한된 그리고/또는 일시적으로 제한된 프로모터 등을 포함한다.

[0435] ***개시내용의 비제한적 양상의 예***

[0436] 상기 기재한 본 대상의 실시형태를 포함하는 양상은 단독으로 또는 하나 이상의 다른 양상 또는 실시형태와 조합하여 유리하게 될 수 있다. 앞서 언급한 설명을 제한하는 일 없이, 1 내지 36으로 넘버링된 개시내용의 특정 비제한적 양상을 이하에 제공한다. 본 개시내용을 읽을 때 당업자에게 명백할 바와 같이, 각각의 개별적으로 넘버링된 양상은 임의의 앞서 선행하는 또는 뒤에 오는 개개로 넘버링된 양상과 함께 사용되거나 또는 조합될 수 있다. 이는 모든 이러한 조합에 대한 근거를 제공하는 것으로 의도되며, 이하에 명확하게 제공되는 양상의 조합으로 제한되지 않는다:

[0437] **양상**

[0438] 양상 1. CasZ 폴리펩타이드를 표적 핵산의 표적 서열에 가이드하는 방법으로서, 표적 핵산을 (a) CasZ 폴리펩타이드; 및 (b) 표적 핵산의 표적 서열에 혼성화되는 가이드 서열을 포함하고, CasZ 폴리펩타이드에 결합하는 영역을 포함하는 CasZ 가이드 RNA를 포함하는 조작되고/되거나 비천연 유래 복합체와 접촉시키는 단계를 포함하는, 방법.

[0439] 양상 2. 양상 1에 있어서, 상기 방법은 표적 핵산의 변형, 표적 핵산으로부터 전사의 조절, 또는 표적 핵산과 연관된 폴리펩타이드의 변형을 초래하는, 방법.

[0440] 양상 3. 양상 2에 있어서, 상기 표적 핵산은 절단에 의해 변형되는, 방법.

[0441] 양상 4. 양상 1 내지 3 중 어느 하나에 있어서, 상기 표적 핵산은 이중 가닥 DNA, 단일 가닥 DNA, RNA, 게놈 DNA, 및 염색체외 DNA로부터 선택된, 방법.

[0442] 양상 5. 양상 1 내지 4 중 어느 하나에 있어서, CasZ 폴리펩타이드에 결합하는 상기 가이드 서열 및 영역은 서로에 대해 이종성인, 방법.

[0443] 양상 6. 양상 1 내지 5 중 어느 하나에 있어서, 상기 접촉시키는 단계는 게놈 편집을 초래하는, 방법.

[0444] 양상 7. 양상 1 내지 5 중 어느 하나에 있어서, 상기 접촉시키는 단계는 박테리아 세포 외에서 그리고 고세균 세포 외에서 일어나는, 방법.

[0445] 양상 8. 양상 1 내지 5 중 어느 하나에 있어서, 상기 접촉시키는 단계는 세포 외부의 시험관내에서 일어나는, 방법.

[0446] 양상 9. 양상 1 내지 7 중 어느 하나에 있어서, 상기 접촉시키는 단계는 표적 세포 내에서 일어나는, 방법.

[0447] 양상 10. 양상 9에 있어서, 상기 접촉시키는 단계는: (a) 상기 CasZ 폴리펩타이드 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산; 및 (b) 상기 CasZ 가이드 RNA 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산 중 적어도 하나를 상기 표적 세포에 도입하는 것을 포함하는, 방법.

[0448] 양상 11. 양상 10에 있어서, 상기 CasZ 폴리펩타이드를 암호화하는 핵산은 표적 세포에서의 발현을 위해 코돈 최적화된 비천연 서열인, 방법.

[0449] 양상 12. 양상 9 내지 11 중 어느 하나에 있어서, 상기 표적 세포는 진핵 세포인, 방법.

[0450] 양상 13. 양상 9 내지 12 중 어느 하나에 있어서, 상기 표적 세포는 시험관내 배양물에 있는, 방법.

[0451] 양상 14. 양상 9 내지 12 중 어느 하나에 있어서, 상기 표적 세포는 생체 내에 있는, 방법.

[0452] 양상 15. 양상 9 내지 12 중 어느 하나에 있어서, 상기 표적 세포는 생체 외에 있는, 방법.

[0453] 양상 16. 양상 12에 있어서, 상기 진핵 세포는 식물 세포, 진균 세포, 단세포 진핵 유기체, 포유류 세포, 파충류 세포, 곤충 세포, 조류 세포, 어류 세포, 기생충 세포, 절지동물 세포, 거미류 세포, 무척추동물의 세포, 척추동물의 세포, 설치류 세포, 마우스 세포, 래트 세포, 영장류 세포, 비-인간 영장류 세포, 및 인간 세포로 이루어진 군으로부터 선택되는, 방법.

[0454] 양상 17. 양상 9 내지 16 중 어느 하나에 있어서, 상기 접촉시키는 단계는 DNA 공여자 주형을 상기 표적 세포 내로 도입하는 것을 더 포함하는, 방법.

[0455] 양상 18. 양상 1 내지 17 중 어느 하나에 있어서, 상기 방법은 상기 표적 핵산을 CasZ 전사촉진 비암호화 RNA(trancRNA)와 접촉시키는 단계를 포함하는, 방법.

[0456] 양상 19. 양상 9 내지 17 중 어느 하나에 있어서, 상기 접촉시키는 단계는 CasZ 전사촉진 비암호화 RNA(trancRNA) 및/또는 상기 CasZ trancRNA를 암호화하는 핵산을 표적 세포 내로 도입하는 것을 포함하는, 방법.

[0457] 양상 20. 양상 18 또는 양상 19에 있어서, 상기 trancRNA는 표 2의 trancRNA 서열과 70% 이상(적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100%)의 뉴클레오타이드 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는, 방법.

[0458] 양상 21. 조작된 그리고/또는 비천연 유래 복합체를 포함하는 조성물로서, (a) CasZ 폴리펩타이드 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산; 및 (b) CasZ 가이드 RNA 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산을 포함하되, 상기 CasZ 가이드 RNA는 표적 핵산의 표적 서열에 상보성인 가이드 서열을 포함하고, 상기 CasZ 폴리펩타이드에 결합할 수 있는 영역을 포함하는, 조성물.

[0459] 양상 22. 양상 21에 있어서, CasZ 전사촉진 비암호화 RNA(trancRNA), 또는 상기 CasZ trancRNA를 암호화하는 핵산을 더 포함하는, 조성물.

[0460] 양상 23. 조작된 그리고/또는 비천연 유래 복합체를 포함하는 조성물로서, (a) CasZ 폴리펩타이드, 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산; (b) CasZ 가이드 RNA, 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산을 포함하되, 상기 CasZ 가이드 RNA는 표적 핵산의 표적 서열에 상보성인 가이드 서열을 포함하고, 상기 CasZ 폴리펩타이드에 결합할 수 있는 영역을 포함하는, 키트.

[0461] 양상 24. 양상 23에 있어서, CasZ 전사촉진 비암호화 RNA(trancRNA), 또는 상기 CasZ trancRNA를 암호화하는

핵산을 더 포함하는, 키트.

[0462] 양상 25. 유전자 변형된 진핵 세포로서, (a) CasZ 폴리펩타이드, 또는 상기 CasZ 폴리펩타이드를 암호화하는 핵산; (b) CasZ 가이드 RNA, 또는 상기 CasZ 가이드 RNA를 암호화하는 핵산; 및 (c) CasZ 전사촉진 비암호화 RNA(trancRNA), 또는 상기 CasZ trancRNA를 암호화하는 핵산을 포함하되, 상기 CasZ 가이드 RNA는 표적 핵산의 표적 서열에 상보성인 가이드 서열을 포함하고, 상기 CasZ 폴리펩타이드에 결합할 수 있는 영역을 포함하는, 키트.

[0463] 양상 26. 양상 1 내지 25 중 어느 하나에 있어서, (a) 상기 CasZ 폴리펩타이드를 암호화하는 핵산은, (i) 상기 CasZ 폴리펩타이드를 암호화하고, (ii) 이종성 프로모터에 작동 가능하게 연결된, 뉴클레오타이드 서열을 포함하는 것; (b) 상기 CasZ 가이드 RNA를 암호화하는 상기 핵산은, (i) 상기 CasZ 가이드 RNA를 암호화하고, (ii) 이종성 프로모터에 작동 가능하게 연결된, 뉴클레오타이드 서열을 포함하는 것; 그리고 (c) 상기 CasZ trancRNA를 암호화하는 상기 핵산은, (i) 상기 CasZ trancRNA를 암호화하고, (ii) 이종성 프로모터에 작동 가능하게 연결된, 뉴클레오타이드 서열을 포함하는 것 중 적어도 하나를 특징으로 하는, 조성물, 키트 또는 진핵 세포.

[0464] 양상 27. 양상 1 내지 26 중 어느 하나에 있어서, 환자의 치료 방법에서 사용하기 위한, 조성물, 키트 또는 진핵 세포.

[0465] 양상 28. 양상 1 내지 27 중 어느 하나에 있어서, 상기 CasZ 폴리펩타이드를 암호화하는 핵산, 상기 CasZ 가이드 RNA를 암호화하는 핵산, 및 상기 CasZ trancRNA를 암호화하는 핵산 중 적어도 하나는 재조합 발현 벡터인, 방법, 조성물, 키트 또는 진핵 세포.

[0466] 양상 29. 양상 1 내지 28 중 어느 하나에 있어서, 상기 CasZ 가이드 RNA 및/또는 상기 CasZ trancRNA는 변형된 핵염기, 변형된 골격 또는 비천연 뉴클레오타이드간 결합, 변형된 당 모이어티, 잠금 핵산, 펩타이드 핵산 및 데옥시리보뉴클레오타이드 중 하나 이상을 포함하는, 방법, 조성물, 키트 또는 진핵 세포.

[0467] 양상 30. 양상 1 내지 29 중 어느 하나에 있어서, 상기 CasZ 폴리펩타이드는 대응하는 야생형 CasZ 단백질에 비해 감소된 뉴클레아제 활성을 갖는 변이체 CasZ 폴리펩타이드인, 방법, 조성물, 키트 또는 진핵 세포.

[0468] 양상 31. 양상 1 내지 30 중 어느 하나에 있어서, 상기 CasZ 폴리펩타이드, 상기 CasZ 폴리펩타이드를 암호화하는 핵산, 상기 CasZ 가이드 RNA, 상기 CasZ 가이드 RNA를 암호화하는 상기 핵산, 상기 CasZ trancRNA, 및 상기 CasZ trancRNA를 암호화하는 핵산 중 적어도 하나는 이종성 모이어티에 접합된, 방법, 조성물, 키트 또는 진핵 세포.

[0469] 양상 32. 양상 31에 있어서, 상기 이종성 모이어티는 이종성 폴리펩타이드인, 방법, 조성물, 키트 또는 진핵 세포.

[0470] 양상 33. 양상 1 내지 32 중 어느 하나에 있어서, 상기 CasZ 폴리펩타이드는 대응하는 야생형 CasZ 단백질에 비해 감소된 뉴클레아제 활성을 갖고, 이종성 폴리펩타이드에 융합된, 방법, 조성물, 키트 또는 진핵 세포.

[0471] 양상 34. 양상 33에 있어서, 상기 이종성 폴리펩타이드는 (i) DNA 변형 활성을 갖고, (ii) 전사를 증가시키거나 감소시키는 능력을 나타내고/내거나 (iii) DNA와 연관된 폴리펩타이드를 변형시키는 효소 활성을 갖는, 방법, 조성물, 키트 또는 진핵 세포.

[0472] 양상 35. 양상 1 내지 34 중 어느 하나에 있어서, 상기 CasZ 폴리펩타이드는 도 1 또는 도 7의 CasZ 단백질과 70% 이상(적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100%)의 아미노산 서열을 갖는 아미노산 서열을 포함하는, 방법, 조성물, 키트 또는 진핵 세포.

[0473] 양상 36. 양상 1 내지 35 중 어느 하나에 있어서, CasZ 폴리펩타이드에 결합하는 상기 가이드 서열 및 영역은 서로에 대해 이종성인, 방법, 조성물, 키트 또는 진핵 세포.

[0474] 양상 37. 샘플 내 표적 DNA를 검출하는 방법으로서, (a) 샘플을 (i) CasZ 폴리펩타이드; (ii) CasZ 폴리펩타이드에 결합하는 영역, 및 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA; 및 (iii) 단일 가닥이며 가이드 RNA의 가이드 서열과 혼성화되지 않는 검출기 DNA와 접촉시키는 단계; 및 (b) CasZ 폴리펩타이드에 의한 단일 가닥 검출기 DNA의 절단에 의해 생성되는 검출 가능한 신호를 측정하여, 표적 DNA를 검출하는 단계를 포함하는, 방법.

[0475] 양상 38. 양상 37에 있어서, 상기 표적 DNA는 단일 가닥인, 방법.

[0476] 양상 39. 양상 37 또는 38에 있어서, 상기 표적 DNA는 이중 가닥인, 방법.

[0477] 양상 40. 양상 37 내지 39 중 어느 하나에 있어서, 상기 표적 DNA는 바이러스 DNA인, 방법.

[0478] 양상 41. 양상 37 내지 40 중 어느 하나에 있어서, 상기 표적 DNA는 파포바바이러스, 헤파드나바이러스, 헤르페스 바이러스, 아데노바이러스, 폭스바이러스 또는 파보바이러스 DNA인, 방법.

[0479] 양상 42. 양상 37 내지 41 중 어느 하나에 있어서, 상기 CasZ 폴리펩타이드는 도 1 내지 도 7에 중 어느 하나에 제시된 CasZ 아미노산 서열에 대해 적어도 85%의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, 방법.

[0480] 양상 43. 양상 37 내지 41 중 어느 하나에 있어서, 상기 CasZ 폴리펩타이드는 Cas14a 폴리펩타이드인, 방법.

[0481] 양상 44. 양상 37 내지 43 중 어느 하나에 있어서, 상기 샘플은 세포 용해물로부터의 DNA 분자를 포함하는, 방법.

[0482] 양상 45. 양상 37 내지 44 중 어느 하나에 있어서, 상기 샘플은 세포를 포함하는, 방법.

[0483] 양상 46. 양상 37 내지 45 중 어느 하나에 있어서, 상기 접촉시키는 단계는 시험관내, 생체외 또는 생체내의 세포 내에서 수행되는, 방법.

[0484] 양상 47. 양상 46에 있어서, 상기 세포는 진핵 세포인, 방법.

[0485] 양상 48. 양상 37 내지 47 중 어느 하나에 있어서, 상기 표적 DNA는 10aM만큼 낮은 농도에서 검출될 수 있는, 방법.

[0486] 양상 49. 양상 37 내지 48 중 어느 하나에 있어서, 상기 샘플에 존재하는 상기 표적 DNA의 양을 결정하는 단계를 포함하는, 방법.

[0487] 양상 50. 양상 49에 있어서, 상기 결정하는 단계는 검출 가능한 신호를 측정하여 시험 측정치를 생성하는 것; 기준 샘플에 의해 생성된 검출 가능한 신호를 측정하여 기준 측정치를 생성하는 것; 및 시험 측정치를 기준 측정치와 비교하여 샘플에 존재하는 표적 DNA의 양을 결정하는 것을 포함하는, 방법.

[0488] 양상 51. 양상 37 내지 50 중 어느 하나에 있어서, 검출 가능한 신호를 측정하는 단계는 금 나노입자 기반 검출, 형광 편광, 콜로이드 상 전이/분산, 전기 화학적 검출 및 반도체-기반 센싱 중 하나 이상을 포함하는, 방법.

[0489] 양상 52. 양상 37 내지 51 중 어느 하나에 있어서, 상기 단일 가닥 검출기 DNA는 형광-방출 염료쌍을 포함하는, 방법.

[0490] 양상 53. 양상 52에 있어서, 상기 형광-방출 염료쌍은 상기 단일 가닥 검출기 DNA의 절단되기 전에 검출 가능한 신호의 양을 생성하고, 검출 가능한 신호의 상기 양은 상기 단일 가닥 검출기 DNA의 절단 후에 감소되는, 방법.

[0491] 양상 54. 양상 52에 있어서, 상기 단일 가닥 검출기 DNA는 절단되기 전에 제1 검출 가능한 신호를 생성하고, 상기 단일 가닥 검출기 DNA의 절단 후에 제2 검출 가능한 신호를 생성하는, 방법.

[0492] 양상 55. 양상 52 내지 54 중 어느 하나에 있어서, 상기 형광-방출 염료쌍은 형광 공명 에너지 전달(FRET) 쌍인, 방법.

[0493] 양상 56. 양상 18에 있어서, 검출 가능한 신호의 양은 상기 단일 가닥 검출기 DNA의 절단 후에 증가되는, 방법.

[0494] 양상 57. 양상 52 또는 양상 56에 있어서, 상기 형광-방출 염료쌍은 퀀처/플루오르 쌍인, 방법.

[0495] 양상 58. 양상 52 내지 57 중 어느 하나에 있어서, 상기 단일 가닥 검출기 DNA는 둘 이상의 형광-방출 염료쌍을 포함하는, 방법.

[0496] 양상 59. 양상 58에 있어서, 상기 둘 이상의 형광-방출 염료쌍은 형광 공명 에너지 전달(FRET) 쌍 및 퀀처/플루오르 쌍을 포함하는, 방법.

[0497] 양상 60. 양상 37 내지 59 중 어느 하나에 있어서, 상기 단일 가닥 검출기 DNA는 변형된 핵염기, 변형된 당 모이어티, 및/또는 변형된 핵산 결합을 포함하는, 방법.

[0498] 양상 61. 양상 37 내지 60 중 어느 하나에 있어서, 상기 방법은 상기 샘플에서 핵산을 증폭시키는 단계를 포함

하는, 방법.

[0499] 양상 62. 양상 61에 있어서, 상기 증폭시키는 단계는 등온 증폭을 포함하는, 방법.

[0500] 양상 63. 양상 62에 있어서, 상기 등온 증폭은 재조합효소 중합효소 증폭(RPA)을 포함하는, 방법.

[0501] 양상 64. 양상 61 내지 63 중 어느 하나에 있어서, 상기 증폭시키는 단계는 단계 (a)의 접촉시키는 단계 전에 시작하는, 방법.

[0502] 양상 65. 양상 61 내지 63 중 어느 하나에 있어서, 상기 증폭시키는 단계는 단계 (a)의 접촉시키는 단계와 함께 시작하는, 방법.

[0503] 양상 66. 샘플 내 표적 DNA를 검출하기 위한 키트로서, 상기 키트는 (a) 가이드 RNA가 CasZ 폴리펩타이드에 결합하는 영역, 및 표적 DNA에 상보성인 가이드 서열을 포함하는, 가이드 RNA, 또는 상기 가이드 RNA를 암호화하는 핵산; 및 (b) 단일 가닥이고 상기 가이드 RNA의 가이드 서열과 혼성화되지 않는, 표지된 검출기 DNA를 포함하는, 키트.

[0504] 양상 67. 양상 66에 있어서, CasZ 폴리펩타이드를 더 포함하는, 키트.

[0505] 양상 68. 양상 67에 있어서, 상기 CasZ 폴리펩타이드는 도 1 내지 도 7에 중 어느 하나에 제시된 CasZ 아미노산 서열에 대해 적어도 85%의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, 키트.

[0506] 양상 69. 양상 67에 있어서, 상기 CasZ 폴리펩타이드는 Cas14a 폴리펩타이드인, 키트.

[0507] 양상 70. 양상 66 내지 69 중 어느 하나에 있어서, 상기 단일 가닥 검출기 DNA는 형광-방출 염료쌍을 포함하는, 키트.

[0508] 양상 71. 양상 70에 있어서, 상기 형광-방출 염료쌍은 FRET 쌍인, 키트.

[0509] 양상 72. 양상 70에 있어서, 상기 형광-방출 염료쌍은 퀀처/플루오르 쌍인, 키트.

[0510] 양상 73. 양상 70 내지 72 중 어느 하나에 있어서, 상기 단일 가닥 검출기 DNA는 둘 이상의 형광-방출 염료쌍을 포함하는, 키트.

[0511] 양상 74. 양상 73에 있어서, 상기 둘 이상의 형광-방출 염료쌍은 제1 검출 가능한 신호를 생성하는 제1 형광-방출 염료쌍 및 제2 검출 가능한 신호를 생성하는 제2 형광-방출 염료쌍을 포함하는, 키트.

[0512] 양상 75. 양상 66 내지 74 중 어느 하나에 있어서, 핵산 증폭 성분을 더 포함하는, 키트.

[0513] 양상 76. 양상 75에 있어서, 상기 핵산 증폭 성분은 재조합효소 중합효소 증폭(RPA)을 위한 성분인, 키트.

[0514] 양상 77. 단일 가닥 DNA(ssDNA)를 절단하는 방법으로서, 상기 방법은 핵산의 집단을 (i) CasZ 폴리펩타이드; 및 (ii) CasZ 폴리펩타이드에 결합하는 영역 및 표적 DNA와 혼성화하는 가이드 서열을 포함하는, 가이드 RNA와 접촉시키는 단계를 포함하되, 상기 집단은 표적 DNA 및 복수의 비표적 ssDNA를 포함하고, 상기 CasZ 폴리펩타이드는 상기 복수의 비-표적 ssDNA를 절단하는, 방법.

[0515] 양상 78. 양상 77에 있어서, 상기 접촉시키는 단계는 시험관내, 생체외 또는 생체내의 세포 내에서인, 방법.

[0516] 양상 79. 양상 78에 있어서, 상기 세포는 진핵 세포인, 방법.

[0517] 양상 80. 양상 79에 있어서, 상기 진핵 세포는 식물 세포인, 방법.

[0518] 양상 81. 양상 78 내지 80 중 어느 하나에 있어서, 상기 비-표적 ssDNA는 상기 세포에 대해 외래인, 방법.

[0519] 양상 82. 양상 81에 있어서, 상기 비표적 ssDNA는 바이러스 DNA인, 방법.

[0520] 양상 83. 양상 77 내지 82 중 어느 하나에 있어서, 상기 표적 DNA는 단일 가닥인, 방법.

[0521] 양상 84. 양상 77 내지 82 중 어느 하나에 있어서, 상기 표적 DNA는 이중 가닥인, 방법.

[0522] 양상 85. 양상 77 내지 84 중 어느 하나에 있어서, 상기 표적 DNA는 바이러스 DNA인, 방법.

[0523] 양상 86. 양상 77 내지 84 중 어느 하나에 있어서, 상기 표적 DNA는 파포바바이러스, 헤파드나바이러스, 헤르페스 바이러스, 아데노바이러스, 폭스바이러스 또는 파보바이러스 DNA인, 방법.

[0524] **실시예**

[0525] 다음의 실시예는 당업자에게 본 발명의 제조하고 사용하는 방법의 완전한 개시내용 및 설명을 제공하기 위해 제시하며, 본 발명자들이 본 발명으로서 간주하는 것의 범주를 제한하는 것으로 의도되지도 또는 이하의 실험이 모든 또는 수행되는 실험만을 나타내는 것으로 의도하지도 않는다. 사용한 숫자(예를 들어, 양, 온도 등)에 관한 정확성을 보장하기 위한 노력을 하였지만, 일부 실험적 오차 및 편차를 고려하여야 한다. 달리 나타내지 않는 한, 부분은 중량부이고, 분자량은 중량 평균 분자량이며, 온도는 섭씨이고, 압력은 대기압에서 또는 대기압 근처이다. 표준 약어, 예를 들어, bp, 염기쌍(들); kb, 킬로베이스(들); pl, 피코리터(들); s 또는 sec, 초(들); min, 분(들); h 또는 hr, 시간(들); aa, 아미노산(들); kb, 킬로베이스(들); bp, 염기쌍(들); nt, 뉴클레오타이드(들); i.m., 근육내(로); i.p., 복강내(로); s.c., 피하(로); 등.

[0526] ***물질 및 방법***

[0527] 다음의 물질 및 방법은 일반적으로 달리 언급되는 경우를 제외하고 본 명세서에 기재된 실시예에 제시하는 결과에 적용한다.

[0528] ***메타게놈 및 메타전사체분석***

[0529] 2개의 사이트: 콜로라도주 라이플 근처의 콜로라도 강에 인접한 라이플 통합 분야 연구소(Rifle Integrated Field Research: IFRC) 사이트 및 유타주의 콜로라도 고원의 크리스탈 가이저(Crystal Geyser), 차가운, CO_2-유도 간헐천(CO_2-driven geyser)으로부터 이전에 조립하고 저장한 메타게놈에 대해 초기 분석을 수행하였다. IFRC 사이트로부터의 메타전사체 데이터를 사용하여 천연에서의 비-암호화 요소의 전사를 검출하였다. 이어서, 추가로 CRISPR-Cas14 시스템의 마이닝(mining)을 IMG/M으로부터의 공공 메타게놈에 대해 수행하였다.

[0530] ***CRISPR-Cas 컴퓨터 계산 분석***

[0531] 은닉 마르코프 모델(HMM) 프로파일을 이용하여 공지된 Cas 단백질에 대해 허머(HMMer) 스위트를 이용하여 다양한 샘플로부터의 조립 콘틱을 스캐닝하였다. 800개 미만의 aa을 함유한 추정적 V형 효과기의 MAFFT 정렬에 기반하여 Cas14 단백질에 대해 추가적인 HMM을 작제하였고, 획득 cas 유전자 및 CRISPR 어레이에 인접하였다. 이들 HMM는 기존의 Cas14 HMM 모델을 이용하여 발견한 수동으로 선택한 신규한 추정적 Cas14 서열을 이용하여 이들을 증가시킴으로써 반복적으로 개선시켰다. Cas14 반복부 서열의 서열을 표 3에 제공한다. CrisprFinder 소프트웨어 및 CRISPRDetect의 로컬 버전을 이용하여 CRISPR 어레이를 확인하였다. 치환 모델로서 PROTGAMMALG를 이용하는 RAxML 및 100 부트스트랩(bootstrap) 샘플링을 이용하여 Cas1 및 V형 효과기 단백질의 계통수를 구성하였다. FigTree 1.4.1(http://tree.bio.ed.ac.uk/software/figtree/)을 이용하여 계통수를 시각화하였다. Bowtie2를 이용하여 조립된 콘틱에 메타전사체 판독을 맵핑하였다. RNase 존재 분석은 KEGG 데이터베이스로부터 다운로드한 KEGG 오솔로그 그룹(KO)의 정렬로부터 구성한 HMM에 기반하였다.

표 3

본 명세서에서 사용한 모든 Cas14 단백질(예를 들어, 도 7 참조)의 반복부 서열(Cas14 가이드 RNA의 비-가이드 부분)을 표 3에 나타냄.

Cas14 단백질	반복부 서열	서열 번호	스캐폴드 수탁번호
Cas14a.1 (CasZa.3)	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC	53	**NCBI:** MK005734
Cas14a.2 (Za.8)	CTTGCAGAACCCGGATAGACGAATGAAGGAATGCAAC	295	**NCBI:** MK005733
Cas14a.3 (CasZa.3)	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC	53	**NCBI:** MK005732
Cas14a.4 (CasZa.4)	CTATCATATTCAGAACAAAGGGATTAAGGAATGCAAC	54	**NCBI:** MK005735
Cas14a.5 (CasZa.5, CasZb.3)	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55	**NCBI:** MK005736
Cas14a.6 (CasZa.6)	GTCTACAACTCATTGATAGAAATCAATGAGTTAGACA	56	**IMG/M:** Ga0137385 _10000156
Cas14b.1 (Cas Zb.2)	GTTGCAGAAATAGAATAAAGGAATTAAGGAATGCAAC	59	**NCBI:** MK005737
Cas14b.2 (CasZa.5, CasZb.3)	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55	**NCBI:** MK005738
Cas14b.3 (CasZb.4)	ATTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	61	**NCBI:** MK005739
Cas14b.4 (CasZb.16)	GTTTCAGCGCACGAATTAACGAGATGAGAGATGCAAC	303	**NCBI:** MK005740
Cas14b.5 (CasZb.6)	CTTGCAGAAGCTGAATAGACGAATCAAGGAATGCAAC	63	**NCBI:** MK005741
Cas14b.6 (Za.12)	CTTGCAGGCCTTGAATAGAGGAGTTAAGGAATGCAAC	296	**NCBI:** MK005742
Cas14b.7 (CasZb.9)	GTTGCAGCGCCCGAACTGACGAGACGAGAGATGCAAC	66	**IMG/M:** Ga0172369 _10000737
Cas14b.8 (Cas Zb.10)	GTTGCGCGAATAGAATAAAGGAATTAAGGAATGCAAC	67	**IMG/M:** Ga0172369 _10010464
Cas14b.9 (CasZb.11)	AGTTGCATTCCTTAATCCCTCTGTTCAGTTTGTGCAAT	68	**IMG/M:** Ga0172365 _10004421
Cas14b.10 (Zb.13)	GTTGCACAGTGCTAATTAGAGAAACTAGGAATGCAAC	297	**NCBI:** MK005743
Cas14b.11 (CasZc.2)	GTTGCGGCGCGCGAATAAACGAGACTAGGAATGCAAC	70	**NCBI:** MK005744
Cas14b.12 (Zb.14)	CTAGCATATTCAGAACAAAGGGATTAAGGAATGCAAC	298	**NCBI:** MK005745
Cas14b.13	CTTTCATATTCAGAACAAAGGGATTAAGGAATGCAAC	72	**NCBI:**

[0532]

[0533]

Cas14 단백질	반복부 서열	서열 번호	스캐폴드 수탁번호
(CasZc.4)			MK005746
Cas14b.14 (Zb.15)	CTTTCATATTCAGAAACTAGGGGTTAAGGACTGCAAC	299	**NCBI:** MK005747
Cas14b.15 (CasZc.6)	GTTGCAGCCCCCGAACTAACGAGATGAGAGATGCAAC	74	**IMG/M:** Ga0116204 _1008574
Cas14b.16 (casZc.7)	CTTGCAGAACAATCATATATGACTAATCAGACTGCAAC	75	**IMG/M:** Ga0078972 _1001015a
Cas14c.1 (Zb.8)	GTTGCATCCCTACGTCGTGAGCACCGGTGAGTGCAAC	300	**NCBI:** MK005748
Cas14c.2 (CasZd.2)	GTCCCTACTCGCTAGGGAAACTAATTGAATGGAAAC	77	**IMG/M:** JGI12048 J13642_ 10201286
Cas14d.1 (CasZe.2)	CTTCCAAACTCGAGCCAGTGGGGAGAGAAGTGGCA	79	**NCBI:** MK005750
Cas14d.2 (CasZe.3)	CCTGTAGACCGGTCTCATTCTGAGAGGGGTATGCAACT	80	**NCBI:** MK005751
Cas14d.3 (CasZe.4)	GTCTCGAGACCCTACAGATTTTGGAGAGGGGTGGGAC	81	**NCBI:** MK005752
Cas14e.1 (CasZf.1)	GTAGCAGGACTCTCCTCGAGAGAAACAGGGGTATGCT	83	**NCBI:** MK005753
Cas14e.2 (CasZf.2)	GTACAATACCTCTCCTTTAAGAGAGGGAGGGGTACGCTAC	84	**NCBI:** MK005754
Cas14e.3 (Zc.5)	GGAAAGGAATCCCCTGAAGGAAACGAGGGGG	301	**NCBI:** MK005755
Cas14f.1 (CasZg.1)	GGTTCCCCCGGGCGCGGGTGGGGTGGCG	86	**NCBI:** MK005756
Cas14f.2 (CasZg.2)	GGCTGCTCCGGGTGCGCGTGGAGCGAGG	87	**IMG/M:** Ga0105042 _100140
Cas14g.1 (Ze.3)	GTGTCCATCAATCAGATTTGCGTTGGCCGGTGCAAT	302	**NCBI:** MK005758
Cas14g.2 (CasZi.2)	GCCGCAGCGGCCGACGCGGCCCTGATCGATGGACAC	90	**IMG/M:** Ga0123330 _1010394
Cas14h.1 (CasZk.1)	GGCTAGCCCGTGCGCGCAGGGACGAGTGG	92	**IMG/M:** Ga0070762 _10001740
Cas14h.2 (CasZk.2)	GCCCGTGCGCGCAGGGACGAGTGG	93	**IMG/M:** Ga0070766 _10011912
Cas14h.3 (CasZk.4)	CCATCGCCCCGCGCGCACGTGGATGAGCC	95	**IMG/M:** Ga0116216 _10000905
Cas14u.1 (CasZa.7)	GTTATAAAGGCGGGGATCGCGACCGAGCGATTGAAAG	57	**IMG/M:** Ga0066793

Cas14 단백질	반복부 서열	서열 번호	스캐폴드 수탁번호
			_10010091
Cas14u.2 (CasZb.8)	GTCTCCATGACTGAAAAGTCGTGGCCGAATTGAAAC	65	**IMG/M:** JGI24730 J26740 _1002785
Cas14u.3 (CasZe.1)	GTTGCATTCGGGTGCAAAACAGGGAGTAGAGTGTAAC	78	**NCBI:** MK005749
Cas14u.4 (CasZj.2)	CTTTTAGACAGTTTAAATTCTAAAGGGTATAAAAC	307	**NCBI:** MK005757
Cas14u.5 (CasZj.1)	GTCGAAATGCCCGCGCGGGGGCGTCGTACCCGCGAC	308	**IMG/M:** Ga0137373 _10000316
Cas14u.6 (CasZk.3)	GTTGCAGCGGCCGACGGAGCGCGAGCGTGGATGCCAC	309	**IMG/M:** Ga0070717 _10000077
Cas14u.7 (CasZl.1)	CTTTAGACTTCTCCGGAAGTCGAATTAATGGAAAC	310	**IMG/M:** JGI12210 **IMG/M:** J13797 _10004690
Cas14u.8 (CasZl.2)	GGGCGCCCCGCGCGAGCGGGGGTTGAAG	311	**IMG/M:** Ga0073904 _10021651

[0534]

[0535] ***발현 플라스미드, RNA 및 DNA 기질의 생성***

[0536] 추정적 시스템에 대한 최소 CRISPR 좌위를 획득 단백질을 제거하고 단일 스페이서를 갖는 최소 어레이를 생성함으로써 설계하였다. 이들 최소 좌위는 gBlocks(IDT)로서 정리하며, 좌위의 발현을 유도하는 테트라사이클린 유도성 프로모터를 이용하여 플라스미드 내로 조립한다. 플라스미드 맵은 애드진(Addgene) 상에서 그리고 도면에서 이용 가능하다. T7 중합효소 및 dsDNA 주형으로서 PCR 생성물을 이용하여 모든 RNA를 시험관내에서 전사시켰다. 얻어진 IVT를 겔 추출하고, 에탄올 침전시켰다. IDT로부터 DNA 기질을 얻었고, 이들의 서열을 표 4에서 이용 가능하다. 방사성표지된 절단 분석을 위해, 방사성표지 전에 PAGE 겔로부터 DNA 올리고를 겔 추출하였다. FQ 분석을 위해, 추가 정제 없이 DNA 기질을 사용하였다.

[0537] ***이콜라이 RNAseq***

[0538] 문헌[Harrington et al. 2017]에서의 변형에 의해 이전에 기재한 바와 같이 작은 RNA 서열분석을 수행하였다. Cas14a1 ORF 상류의 테트라사이클린 유도성 프로모터 또는 Cas14에 융합된 N-말단의 10x-히스티딘 태그를 갖는 동일한 플라스미드를 갖는 플라스미드 발현 Cas14a1 시스템을 이용하여 이콜라이 NEB Stable3을 형질전환시켰다. 스타터를 밤새 SOB에서 성장시키고 나서, 214nM 안하이드로테트라사이클린을 함유하는 5㎖의 신선한 SOB에서 1:100으로 희석시키고 25℃에서 밤새 성장시켰다. Cas14a로 풀다운시킨 RNA의 서열분석을 위해, Cas14a1에 융합된 N-말단의 His-태그를 함유하는 플라스미드를 최대 18℃에서 성장시킨 후에, "단백질 정제"에 기재한 바와 같이 용리 및 정제시키고, Ni-NTA 용리 후에 중단시켰다. 세포를 펠릿화하고, 앞서 기재한 바와 같이 뜨거운 페놀을 이용하여 RNA를 추출하였다. TURBO DNase를 이용하여 총 핵산을 처리하고 나서, 페놀 추출하였다. 얻어진 RNA를 T4 PNK의 첨가 전에 열 비활성화시킨 rSAP로 처리하였다. NEBnext 스몰 RNA키트를 이용하여 작은 RNA에 어댑터를 결찰시키고, 8% 천연 PAGE 겔 상에서 겔-추출하였다. RNA를 단일 말단 300bp 판독으로 MiSeq 상에서 서열분석하였다. 분석을 위해, 쿠타댑트(Cutadapt)를 이용하여 얻어진 판독을 다듬고, 8nt 미만의 서열을 버리고, Bowtie2를 이용하여 플라스미드 기준에 대해 맵핑하였다.

[0539] ***PAM 고갈 분석***

[0540] 문헌[Burstein et al. (2017)]에서 앞서 기재한 바와 같이 PAM 고갈 분석을 수행하였다. 표적 서열에 인접한 무작위화된 PAM 영역을 함유하는 프라이머를 이용하여 무작위화된 플라스미드 라이브러리를 생성하였다. 무작위화된 프라이머를 프라이머의 3' 말단에 상보성인 프라이머로 혼성화시키고, 클레노우 단편(Klenow Fragment: NEB)을 이용하여 이중가닥을 연장시켰다. 표적을 함유하는 dsDNA를 EcoRI 및 NcoI로 분해시키고 나서, pUC19 골격에 결찰시키고, 이콜라이 DH5α에 형질전환시키고, 107개 초과의 세포를 채취하였다. 다음에, 이콜라이 NEBstable을 CRISPR 플라스미드 또는 빈 벡터 대조군 중 하나로 형질전환시키고, 이들 형질전화된 이콜라이를 10% 글리세롤로 반복 세척함으로써 전기천공적격으로 만들었다. 이들 전기천공적격 세포를 200ng의 표적 라이브러리로 형질전환시키고, 표적(카베니실린, 100mg l-1) 및 CRISPR 플라스미드(클로람페니콜, 30mg l-1)에 대한

선택을 함유하는 생물검정 접시 상에 플레이팅하였다. 세포를 채취하고 나서, Illumina MiSeq 상에서 앰플리콘 서열분석을 위해 준비하였다. Cutadapt를 이용하여 PAM 영역을 추출하고 나서, 파이썬(python)에서 고갈 값을 계산하였다. WebLogo를 이용하여 PAM을 시각화하였다.

[0541] ***전사체 RNA 맵핑***

[0542] 인비트로젠(Invitrogen) TRIzol 시약을 이용하여 0.2㎜ 필터로부터 RNA를 추출한 후에, 게놈 DNA를 제거하고, 퀴아젠(Qiagen) RNase-Free DNase 세트 키트 및 퀴아젠 미니 RNeasy 키트를 이용하여 세정하였다. 애질런트(Agilent) 2100 바이오어날라이저(Bioanalyzer)(애질런트 테크놀로지(Agilent Technologies))를 사용하여 RNA 샘플의 완전함을 평가하였다. 어플라이드 바이오시스템즈(Applied Biosystems) 솔리드 토탈(SOLiD Total) RNA-Seq 키트를 사용하여 cDNA 주형 라이브러리를 생성하였다. 솔리드 EZ 비드 시스템(라이프 테크놀로지즈)를 사용하여 SOLiD 플랫폼 서열분석을 위한 비드 주형을 생성하기 위한 에멀션 클론 비드 증폭을 수행하였다. 5500XL SOLiD 플랫폼 상에서 퍼시픽 노스웨스트 국립연구소(Pacific Northwest National Laboratory)에서 샘플을 서열분석하였다. 문헌[Brown et al. 2015]에서와 같이 낫(Sickle)을 이용하여 50 bp 단일 판독을 잘라내었다.

[0543] ***단백질 정제***

[0544] 앞서 기재한 바와 같이, 변형하여 Cas14a1를 정제하였다. 이콜라이 BL21(DE3) RIL을 10xHis-MBP-Cas14a1 발현 플라스미드로 형질전환시키고, 테리픽 브로스(Terrific Broth: TB)에서 OD600 =0.5까지 성장시키고 나서, 0.5mM IPTG로 유도하였다. 세포를 18℃에서 밤새 성장시킨 후에, 원심분리에 의해 수집하고, 용해 완충제(50mM Tris-HCl, pH 7.5, 20mM 이미다졸, 0.5mM TCEP, 500mM NaCl)에서 재현탁시키고 나서, 음파처리에 의해 분해시켰다. 용해물을 Ni-NTA 수지 상에 배취 장입시키고 나서, 상기 완충제로 세척한 후 용리 완충제(50mM Tris-HCl, pH 7.5, 300mM 이미다졸, 0.5mM TCEP, 500mM NaCl)로 용리시켰다. 4℃에서 TEV와 함께 밤새 인큐베이션시킴으로써 MBP 및 His-태그를 제거하였다. 얻어진 단백질을 완충제 A(20mM HEPES, pH 7.5, 0.5mM TCEP, 150mM NaCl)로 교환시키고 나서, 탠덤 MBP, 헤어핀 칼럼(GE, Hi-Trap)에 장입시키고 완충제 A 내지 완충제 B(20mM HEPES, pH 7.5, 0.5mM TCEP, 1250mM NaCl)의 선형 구배로 용리시켰다. Cas14a1을 함유하는 얻어진 분획을 S200 겔 여과 칼럼 상에 장입시키고, 순간 냉동시키고 나서, 사용까지 -80℃에서 저장하였다.

[0545] **시험관내 절단 분석**

[0546] *방사선표지*

[0547] 방사선표지된 절단 분석을 1× 절단 완충제(25mM NaCl, 20mM HEPES, pH 7.5, 1mM DTT, 5% 글리세롤)에서 수행하였다. 100nM Cas14a1를 RT에서 10분 동안 125nM crRNA 및 125nM tracrRNA와 복합체화하였다. 대략 1nM의 방사성표지된 DNA 또는 RNA 기질을 첨가하고 나서, 37℃에서 30분 동안 반응시켰다. 2× Quench 완충제(90% 폼아마이드, 25mM EDTA 및 미량의 브로모페놀 블루)를 첨가함으로써 반응을 중단시키고, 95℃로 2분 동안 가열하고 나서, 7M 유레아 및 0.5×TBE를 함유하는 10% 폴리아크릴아마이드 겔 상에서 실행하였다. 인형상화기법(phosphorimaging)에 의해 생성물을 시각화하였다.

[0548] *M13 DNA 절단*

[0549] M13 DNA 절단 분석을 100mM NaCl, 20mM HEPES, pH 7.5, 1mM DTT, 5% 글리세롤에서 수행하였다. 250nM Cas14a1을 250nM crRNA 및 250nM tracrRNA 및 250nM ssDNA 활성체와 복합체화하였다. 5nM M13 ssDNA 플라스미드의 첨가에 의해 반응을 개시하였고, 10mM EDTA로 보충한 장입 완충제의 첨가에 의해 반응을 중단시켰다. SYBR 금(써모피셔(Thermofisher))로 사전염색한 1.5% 아가로스 TAE 겔 상에서 생성물을 분리시켰다.

[0550] *트랜스-절단의 FQ 검출*

[0551] 문헌[Chen et al. (2018)]에서 앞서 기재한 바와 같이, 변형하여 FQ 검출을 수행하였다. 100nM Cas14a1을 37℃에서 10분 동안 1×절단 완충제에서 125nM crRNA, 125nM tracrRNA, 50nM FQ 프로브 및 2nM ssDNA 활성체와 복합체화하였다. 이어서, 가변적 RNA 성분을 사용하여 개시한 RNA 최적화 실험을 제외하고 모든 반응에 대해 활성체 DNA의 첨가에 의해 반응을 개시하였다. 1분마다 형광측정을 취하면서(λex: 485nm; λem: 535nm) 최대 120분 동안 37℃에서 형광 플레이트 판독기에서 반응을 모니터링하였다. 얻어진 데이터를 활성체의 부재 하에 취한 판독을 이용하여 배경을 차감하고, 단일 지수 붕괴 곡선을 이용하여 적합화시켰다.

[0552] **데이터 이용 가능성**

[0553] 본 명세서에서 사용한 플라스미드는 애드진(Addgene)(플라스미드 번호 112500, 112501, 112502, 112503,

112504, 112505, 112506)에서 이용 가능하다. 본 명세서에서 사용한 올리고뉴클레오타이드는 표 4 및 표 5에서 제공한다. 본 명세서에서 사용한 플라스미드를 도 24에 제공한다. 본 명세서에서 사용하는 Cas14 단백질 서열을 도 7에 제공한다.

표 4

본 명세서에서 사용되는 올리고뉴클레오타이드 및 플라스미드

[0554]

DNA 명칭	서열	서열 번호
방사성표지된 DNA 활성체 1T 가닥	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGC AC	220
방사성표지된 DNA 활성체 1 NT 가닥	GTGCCCGcaTTTTTtaccttacaccgcttgcgaaagttgtcagaagataatCGGCGT Ag	221
방사성표지된 DNA 활성체 2T 가닥	tttatatgtttctcctggagataacgcaatcgtgacaactttcgcaagcggtgtaaggtaGCAGG CTTCcgaattccgcgtttttacggc	222
방사성표지된 DNA 활성체 2 NT 가닥	gccgtaaaaacgcggaattcgGAAGCCTGCtaccttacaccgcttgcgaaagttgtcacgat tgcgttatctccaggagaaacatataaa	223
방사성표지된 활성체 3	gatcttcagcTATACATTATTGCACCAACACTAAGGCAGAGTATGt ttacctggac	224
방사성표지된 활성체 4	gatcttcagcTTTGTATTACTGGAAGGATGCTTGCTTGAGGTGTAa aaacctggac	225
F-Q 5nt	/56-FAM/TTTTT/3IABkFQ/	226
F-Q 6nt	/56-FAM/TTTTTT/3IABkFQ/	227
F-Q 7nt	/56-FAM/TTTTTTT/3IABkFQ/	228
F-Q 8nt	/56-FAM/TTTTTTTT/3IABkFQ/	229
F-Q 9nt	/56-FAM/TTTTTTTTT/3IABkFQ/	230
F-Q 10nt	/56-FAM/TTTTTTTTTT/3IABkFQ/	231
F-Q 11nt	/56-FAM/TTTTTTTTTTT/3IABkFQ/	232
F-Q 12nt	/56-FAM/TTTTTTTTTTTT/3IABkFQ/	233
F-Q 10nt A/T	/56-FAM/TATATATATA/3IABkFQ/	371

[0555]

F-Q 6nt A/T	/56-FAM/TATATA/3IABkFQ/	372
F-Q 5nt A/T	/56-FAM/TATAT/3IABkFQ/	373
표적 2, 완전 AAAT 3'	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT AAACGGCCACAAGC	234
표적 2, 완전 TCGT 3'	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT GCTCGGCCACAAGC	235
표적 2, 1-2MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAGCT AAACGGCCACAAGC	236
표적 2, 3-4MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCCTCGT AAACGGCCACAAGC	237
표적 2, 5-6MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGCGGACGT AAACGGCCACAAGC	238
표적 2, 7-8MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGAGCGCGACGT AAACGGCCACAAGC	239
표적 2, 9-10MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGCTCGGCGACGT AAACGGCCACAAGC	240
표적 2, 11-12MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCACGACGGCGACGT AAACGGCCACAAGC	241
표적 2, 13-14MM	GCCGGGGTGGTGCCCATCCTGGTCGACGTGGACGGCGACGT AAACGGCCACAAGC	242
표적 2, 15-16MM	GCCGGGGTGGTGCCCATCCTGGTCCTGCTGGACGGCGACGT AAACGGCCACAAGC	243
표적 2, 17-18MM	GCCGGGGTGGTGCCCATCCTGGAGGAGCTGGACGGCGACGT AAACGGCCACAAGC	244
표적 2, 19-20MM	GCCGGGGTGGTGCCCATCCTCCTCGAGCTGGACGGCGACGT AAACGGCCACAAGC	245
HERC2 Amp Fwd	G*T*G*T*TAATACAAAGGTACAGGAACAAAGAATTTG	246
HERC2 Amp Rev	CAAAGAGAAGCCTCGGCC	247
표적 1, 완전 AAAT 3'	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCCAGGT AAACTAACACAACT	248
표적 1, 완전 TCGT 3'	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCCAGGT GCTCTAACACAACT	249
표적 1, 1-2MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCCACCTA AACTAACACAACT	250
표적 1, 3-4MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCGTGGT AAACTAACACAACT	251
표적 1, 5-6MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAACAGGCAGGT AAACTAACACAACT	252
표적 1, 7-8MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAAGTCCCAGGT AAACTAACACAACT	253
표적 1, 9-10MM	TTTATTCAAGGCAATCACTATCAGCTGTGGTTCACCCAGGTA AACTAACACAACT	254
표적 1, 11-12MM	TTTATTCAAGGCAATCACTATCAGCTGTCCAACACCCAGGTA AACTAACACAACT	255

[0556]

표적 1, 13-14MM	TTTATTCAAGGCAATCACTATCAGCTCAGGAACACCCAGGTAAACTAACACAACT	256
표적 1, 15-16MM	TTTATTCAAGGCAATCACTATCAGGAGTGGAACACCCAGGTAAACTAACACAACT	257
표적 1, 17-18MM	TTTATTCAAGGCAATCACTATCTCCTGTGGAACACCCAGGTAAACTAACACAACT	258
표적 1, 19-20MM	TTTATTCAAGGCAATCACTAAGAGCTGTGGAACACCCAGGTAAACTAACACAACT	259
표적 3, 완전	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	220
표적 3, 1-2MM	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggCGAAAAAtgCGGGCAC	260
표적 3, 3-4MM	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaAAtaAAAAAtgCGGGCAC	261
표적 3, 5-6MM	cTACGCCGattatcttctgacaactttcgcaagcggtgtGGggtaAAAAAtgCGGGCAC	262
표적 3, 7-8MM	cTACGCCGattatcttctgacaactttcgcaagcggtACaaggtaAAAAAtgCGGGCAC	263
표적 3, 9-10MM	cTACGCCGattatcttctgacaactttcgcaagcgACgtaaggtaAAAAAtgCGGGCAC	264
표적 3, 11-12MM	cTACGCCGattatcttctgacaactttcgcaagTAgtgtaaggtaAAAAAtgCGGGCAC	265
표적 3, 13-14MM	cTACGCCGattatcttctgacaactttcgcaGAcggtgtaaggtaAAAAAtgCGGGCAC	266
표적 3, 15-16MM	cTACGCCGattatcttctgacaactttcgTGagcggtgtaaggtaAAAAAtgCGGGCAC	267
표적 3, 17-18MM	cTACGCCGattatcttctgacaactttTAcaagcggtgtaaggtaAAAAAtgCGGGCAC	268
표적 3, 19-20MM	cTACGCCGattatcttctgacaactCCcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	269
표적 3, 21-22MM	cTACGCCGattatcttctgacaaTCttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	270
표적 3, 23-24MM	cTACGCCGattatcttctgacGGctttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	271
표적 3, 25-26MM	cTACGCCGattatcttctgGTaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	272
전장 활성체	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	220
-20 5′ 활성체 표적	tatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	273
-25 5′ 활성체 표적	tctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	274
-30 5′ 활성체 표적	caactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	275
-35 5′ 활성체 표적	ttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	276

-5 3′ 활성체 표적	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCG	277
-9 3′ 활성체 표적	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAA	278
-14 3′ 활성체 표적	cTACGCCGattatcttctgacaactttcgcaagcggtgta	279
-19 3′ 활성체 표적	cTACGCCGattatcttctgacaactttcgcaagcg	280
-24 3′ 활성체 표적	cTACGCCGattatcttctgacaactttcgc	281
-29 3′ 활성체 표적	cTACGCCGattatcttctgacaact	282
-34 3′ 활성체 표적	cTACGCCGattatcttctga	283
루프 없음	caactttcgcaagcggtgtaaggtaAAAAAtgCG	284
0nt SL	atggaatgtggcgaacgctttcaacGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	285
5nt SL	atggaatgtggcgaacgcttagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	286
10nt SL	atggaatgtggcgaagcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	287
15nt SL	atggaatgtgcgcttgcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	288
20nt SL	atggatacaccgcttgcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	289
25nt SL	taccttacaccgcttgcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	290
M13_1 올리고	GTTTTATCTTCTGCTGGTGGTTCGTTCGGTATTTTTAATG	291
M13_2 올리고	CATTAAAAATACCGAACGAACCACCAGCAGAAGATAAAAC	292
M13_3 올리고	GACCATTTGCGAAATGTATCTAATGGTCAAACTAAATCTACTC	293
M13_4 올리고	GAGTAGATTTAGTTTGACCATTAGATACATTTCGCAAATGGTC	294

[0557]

표 5

RNA 올리고뉴클레오타이드

RNA 명칭	서열	서열 번호
ssRNA 표적	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCACcc	332
crRNA	GGAATGCAACtaccttacaccgcttgcgaa	333
tracrRNA	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTT	334

[0558]

[0559]

crRNA 1	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaa	335
crRNA 2	GACGAATGAAGGAATGCAACccttacaccgcttgcgaaag	336
crRNA 3	GACGAATGAAGGAATGCAACttacaccgcttgcgaaagtt	337
crRNA 4	GACGAATGAAGGAATGCAACacaccgcttgcgaaagttgt	338
crRNA MM 표적 2	GACGAATGAAGGAATGCAACCGTCGCCGTCCAGCTCGACCA	339
crRNA MM 표적 3	GACGAATGAAGGAATGCAACGATCGTTACGCTAACTATGA	340
HERC2 A sgRNA	TTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCC CTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCAT CAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTgaaaGAATGAAGGAATGCAACacttgacac ttaatgctcaa	341
LbCas12a HERC2 crRNA	GGGTAATTTCTACTAAGTGTAGATacttgacacttaatgctcaa	342
25nt 스페이서 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaaagttg	343
20nt 스페이서 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaa	335
18nt 스페이서 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcg	344
16nt 스페이서 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacaccgcttg	345
14nt 스페이서 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacaccgct	346
12nt 스페이서 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacaccg	347
10nt 스페이서 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacac	348
완전 반복부 crRNA 표적 4	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAACtaccttaca ccgcttgcgaa	349
20nt 반복부 crRNA 표적 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaa	335
17nt 반복부 crRNA 표적 4	GAATGAAGGAATGCAACtaccttacaccgcttgcgaa	350
15nt 반복부 crRNA 표적 4	ATGAAGGAATGCAACtaccttacaccgcttgcgaa	351

10nt 반복부 crRNA 표적 4	GGAATGCAACtaccttacaccgcttgcgaa	333
tracrRNA +41nt	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTTTCCTCTCCAATTCTGCACAAAAAAA GGTGAGTCCTTAT	352
tracrRNA +3nt	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTT	334
tracrRNA -26	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTT	353
tracrRNA -65	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGG	354
tracrRNA +0	TTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCC CTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCAT CAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCA	355
tracrRNA +90nt	ttcacacTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCT GTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTT GCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAA CCCTCGAAACAAATTCAttttttcctctccaattctgcacaaaaaaaggtgagtccttat aaaccggcgtgcagaacgccggctcacctttttttcttcattcgatttta	356
sgRNA 1	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTgaaaGAATGAAGGAATGCAACtaccttaca ccgcttgcgaa	357
sgRNA 2	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTTTCCTCTCCAATTCTGCACAAgaaaGT TGCAGAACCCGAATAGACGAATGAAGGAATGCAACtaccttacac cgcttgcgaa	358
M13 표적 1 crRNA	GACGAATGAAGGAATGCAACTACCGAACGAACCACCAGCA GAAGA	359
M13 표적 2 crRNA	GACGAATGAAGGAATGCAACTCTTCTGCTGGTGGTTCGTTC GGTA	360
M13 표적 3 crRNA	GACGAATGAAGGAATGCAACGTTTGACCATTAGATACATTT CG	361
M13 표적 4 crRNA	GACGAATGAAGGAATGCAACCGAAATGTATCTAATGGTCAA AC	362

[0560]

[0561] 실시예 1

[0562] **도 1**은 천연 유래 CasZ 단백질 서열의 예를 도시한 도면.

[0563] **도 2**는 CasZ 단백질에 추가로 Cas1 좌위의 개략적 표현을 도시한 도면.

[0564] **도 3**은 다른 클래스 2 CRISPR/Cas 효과기 단백질 서열에 관해 CasZ 서열의 계통수를 도시한 도면.

[0565] **도 4**는 다른 클래스 2 CRISPR/Cas 좌위로부터의 Cas1 서열에 관해 CasZ 좌위로부터의 Cas1 서열의 계통수를 도시한 도면.

[0566] **도 5**는 CasZ 좌위로부터의 trancRNA의 발현을 입증하는 전사체 RNA 맵핑 데이터를 도시한 도면. trancRNA는 CasZ 반복부 어레이에 인접하지만, 반복부 서열을 포함하지 않고, 반복부 서열에 대해 상보성이 아니다. 다음의 좌위에 대한 RNA 맵핑 데이터를 나타낸다: CasZa3, CasZb4, CasZc5, CasZd1 및 CasZe3. 작은 반복 정렬 화살표는 CRISPR 어레이의 반복부를 나타내고(가이드 RNA-암호화 서열의 존재를 나타냄); 반복부 어레이 밖의 그리고 인접한 피크는 고도로 전사된 trancRNA를 나타낸다.

[0567] 이런 메타전사체 데이터는 16S 고갈되지 않았고, 따라서 데이터의 대부분은 16S에 대해 맵핑되며, mRNA는, 예를 들어, 이들 판독에서 전혀 제시하지 않는다. 그럼에도 불구하고, 예측된 trancRNA 영역에 대한 RNA 맵핑이 관찰되었다.

[0568] 실시예 2

[0569] CRISPR 게놈 좌위 내로 DNA를 통합하는 것을 초래하는 Cas1 및 Cas2 단백질을 포함하는, 단지 400 내지 700개의 아미노산의 예외적으로 조밀한 RNA-가이드 뉴클레아제의 패밀리인 Cas14를 함유한 비배양 고세균으로부터의 CRISPR-Cas 시스템의 세트를 본 명세서에 개시하였고, 이들의 CRISPR 어레이가 새로운 감염에 능동적으로 적합

하게 되는 증거를 나타내었다. 이들의 작은 크기에도 불구하고, Cas14 단백질은 제한 서열 필요 없이 RNA-가이드된 단일-가닥 DNA(ssDNA)를 절단할 수 있었다. 게다가, Cas14에 의한 표적 인식은 ssDNA 분자의 비-특이적 절단을 촉발하였다. 메타게놈 데이터는 다중 CRISPR-Cas14 시스템이 독립적으로 진화되었고 단일-효과기 CRISPR-기반 적응 면역의 잠재적 진화 기원을 시사한다는 것을 나타내었다.

[0570] 미생물과 바이러스 사이의 경쟁은 감염제에 대한 보호를 제공하기 위해 CRISPR-기반 적응 면역의 진화를 자극하였다. 클래스 2 CRISPR-Cas 시스템에서, 다중 기능성 도메인을 갖는 단일 100 내지 200 킬로달톤(kD) CRISPR-연관(Cas) 단백질은 RNA-가이드 결합 및 DNA 또는 RNA 기질의 절단을 수행하였다. 보다 단순하고, 더 작은 RNA-가이드 단백질이 천연에서 생기는지의 여부를 결정하기 위해, 테라베이스-규모의 메타게놈 데이터세트를 CRISPR 어레이와 cas1, 즉, 보편적 CRISPR 인테그라제를 암호화한 유전자 둘 다에 대해 근위인 비특성규명 유전자에 대해 질의하였다. 이 분석은 40 내지 70kD 폴리펩타이드를 암호화하는 본 명세서에 기재된 cas1, cas2, cas4 및 cas14를 함유하는 CRISPR-Cas 시스템의 다양한 패밀리를 확인하였다(도 8, 패널 A). 비교적 서열 분석에 기반하여 3개의 하위 그룹(Cas14a 내지 c)으로 클러스터링한 스물 넷(24)의 상이한 cas14 유전자 변이체를 확인하였다(도 8, 패널 A 내지 B, 도 9, 도 10). Cas14 단백질은 앞서 알려진 클래스 2 CRISPR RNA-가이드 효소 크기의 대략 절반인 대략 400 내지 700개의 아미노산(aa)이었다(도 8, 패널 C 내지 D). 확인한 Cas14 단백질은 상당한 서열 다양성을 나타내었지만, 모두 예측된 RuvC 뉴클레아제 도메인의 존재에 의해 통합되었고, 이의 조직화는 V형 CRISPR-Cas DNA-표적 효소의 특징이었다(도 8, 패널 D).

[0571] 확인한 Cas14 단백질은 DPANN 내에서 거의 배타적으로 생기며, 공생 고세균의 상문은 작은 세포 및 게놈 크기를 특징으로 한다. 계통발생 비교는 때때로 CRISPR 어레이 근처에서 발견되지만 결코 다른 *cas* 유전자와 함께 하지 않는 박테리아 RuvC-도메인-함유 단백질의 패밀리인 C2c10 및 C2c9에 대해 유사성을 갖는 Cas14 단백질이 매우 다양화되었다는 것을 발견하였다(도 8, 패널 B 및 도 9). 이런 관찰 및 *c2c10* 및 *cas14* 유전자의 작은 크기는 이들 시스템이 독립된 CRISPR 효과기로서 작용하는 것을 불가능하게 만들었다.

[0572] 도 8, 패널 A 내지 D는 CRISPR-Cas14 게놈 좌위의 구성 및 계통발생을 도시한다. 도 8, 패널 A는 V형 CRISPR 시스템의 계통수를 도시한다. 새로 확인한 미니어처 CRISPR 시스템은 오렌지색으로 강조한다. 도 8, 패널 B는 C2c10 및 CRISPR-Cas14 시스템의 대표적인 좌위 구조를 도시한다. 도 8, 패널 C는 Cas12a 내지 e 및 Cas9에 비교한 Cas14a-c 시스템의 길이 분포를 도시한다. 도 8, 패널 D는 Cas9 및 Cas12a에 비교한 Cas14a의 도메인 조직화를 도시한다. 단백질 길이를 일정한 비율로 도시한다.

[0573] 도 9는 공지된 V형 CRISPR 효과기 및 RuvC 도메인을 함유하는 클래스 2 후보에 대한 최대 가능성 계통수를 도시한다. 삽입도는 각각의 새로 확인한 아형에 대한 개개 오솔로그를 나타낸다. 도 10은 공지된 CRISPR 시스템으로부터의 Cas1에 대한 최대 가능도(Maximum likelihood tree)를 도시한다.

[0574] 실시예 3

[0575] 감염의 유전자 기억을 생성하는 것을 초래하는 보존된 유전자(*cas1*, *cas2*, *cas4*)에 대한 근접도에 기반하여(도 11, 패널 A), CRISPR-Cas14 시스템이 DNA 서열을 이들의 CRISPR 어레이 내로 적극적으로 획득하는지의 여부를 연구하였다. 다중 CRISPR-Cas14 좌위에 대한 조립된 메타게놈 인접 DNA 서열(콘틱)은 달리 동일한 CRISPR 시스템이 이들의 CRISPR 어레이에서 다양성을 나타내고, 새로운 감염에 대한 적극적인 적응을 시사한다는 것을 나타내었다(도 11, 패널 B 및 도 12, 패널 A). 임의의 특정 이론으로 구속되기를 의도하지는 않지만, 새로운 DNA 서열의 적극적 획득은 이들 CRISPR-Cas14 좌위가 이들의 작은 크기에도 불구하고 핵산 표적화 활성을 갖는 기능적 효소를 암호화하였다는 것을 나타낸다는 것을 제안한다. 이런 가능성을 시험하기 위해, RNA 성분이 CRISPR-Cas14 좌위로부터 생성되었는지의 여부를 연구하였다. 환경적 메타전사체 서열분석 데이터를 CRISPR-Cas14a를 함유하는 천연 고세균 숙주로부터의 RNA의 존재에 대해 분석하였다(도 12, 패널 B 및 도 13, 패널 a). CRISPR RNA(crRNA)에 추가로, 매우 흔한 비-암호화 RNA를 *cas14a*와 인접한 CRISPR 어레이 사이에 위치된 약 130-염기쌍 서열에 맵핑하였다. 이 전사체의 3' 말단에서 20개의 뉴클레오타이드(nt)는 Cas9, Cas12b 및 Cas12e CRISPR 시스템과 관련하여 발견한 전사촉진 활성 CRISPR RNA(tracrRNA)에 대해 관찰한 바와 같이, crRNA의 반복부 세그먼트에 대해 가장 상보성이었다(도 12, 패널 C 및 도 13, 패널 b). 이들 앞서 연구한 시스템에서, 이중-가닥-RNA-절단 효소 리보뉴클레아제 III(RNase III)는 성숙 tracrRNA 및 crRNA를 생성하였지만, RNase III를 암호화하는 유전자는 *cas14*-함유 재작제 게놈에 존재하지 않았다(도 14, 패널 A). 이 관찰은 CRISPR-연관 RNA 가공에 대한 대안의 메커니즘이 이들 숙주에 존재한다는 것을 나타내었다.

[0576] Cas14a 단백질 및 연관된 RNA 성분이 이종성 유기체에서 조립될 수 있는지의 여부를 시험하기 위해, 플라스미드를 추정적 tracrRNA를 함유하는 CRISPR 어레이 및 통합 영역인 Cas14 유전자를 포함한 최소 CRISPR-Cas14a 좌

위를 함유하는 이콜라이에 도입하였다. 세포 용해물로부터의 Cas14a 단백질의 친화도 정제 및 공동 정제하는 RNA의 서열분석은 고도로 흔한 성숙 crRNA뿐만 아니라 추정적 tracrRNA를 나타내었고, 이는 Cas14가 crRNA와 tracrRNA 둘 다와 연관된다는 것을 시사한다(도 14, 패널 B). 조립된 Cas14a 단백질-tracrRNA-crRNA 입자의 계산된 질량은 스트렙토코커스 피오게네스(*S. pyogenes*) Cas9(SpCas9)에 대해 단지 17%(도 12, 패널 D) 및 프란시셀라 노비시다(*F. novicida*) Cas12a(FnCas12a)에 대해 8%에 비교되는 48중량% RNA였으며, 이는 Cas14a 복합체의 구조에서 RNA의 중요한 역할을 암시한다. 공지된 클래스 2 CRISPR 시스템은 표적 이중-가닥 DNA(dsDNA)에 대한 프로토스페이서 인접 모티프(PAM)으로 불리는 짧은 서열을 필요로 한다. Cas14a가 PAM을 필요로 하고 dsDNA 간섭을 수행할 수 있는지의 여부를 시험하기 위해, Cas14 어레이에서 표적-암호화 서열(스페이서)에 매칭되는 서열 다음에 무작위화된 PAM 영역을 함유하는 dsDAN 플라스미드를 이용하여 최소 Cas14a 좌위를 발현시키는 이콜라이를 형질전환시켰다. 이콜라이 형질전환체에서 PAM 서열의 고갈은 검출되지 않았는데, 이는 CRISPR-Cas14a 시스템이 dsDNA를 표적화할 수 없었고, PAM를 필요로 하는 일 없이 가능하거나, 이런 이종성 숙주에서 비활성이라는 것을 시사한다(도 15, 패널 a 내지 d).

[0577] 도 11, 패널 A 내지 B는 CRISPR-Cas14 시스템에 의한 새로운 스페이서의 획득을 도시한 도면. 도 11, 패널 A는 Cas14 Cas1 오솔로그의 정렬을 도시한다. 확장은 적색 박스로 강조한 앞서 나타낸 활성 부위 잔기의 보존을 나타낸다. 도 11, 패널 B는 이들 CRISPR 시스템에 대한 스페이서 다양성을 나타내는 다양한 CRISPR-Cas14 시스템에 대해 조립한 다중 CRISPR 어레이를 도시한다. 오렌지색 화살표는 반복부를 나타내지만, 다양한 색상의 박스는 독특한 스페이서를 나타낸다.

[0578] 도 12, 패널 A 내지 D는 CRISPR-Cas14a가 적극적으로 적합하게 되고, tracrRNA를 암호화하는 것을 도시한다. 도 12, 패널 A는 오렌지색의 다이어그램으로 나타낸 CRISPR 반복부 및 다른 색으로 나타낸 독특한 스페이서에 의한 Cas14a 및 Cas14b에 대한 스페이서 다양성을 도시한다. 도 12, 패널 B는 Cas14a1 및 Cas14a3에 맵핑된 메타전사체 판독을 도시한다. 삽입도는 가장 흔한 반복부 및 스페이서 서열의 확장을 나타낸다. 도 12, 패널 C는 Cas14a1 crRNA 및 tracrRNA의 인실리코 예측 구조를 도시한다. RNase III 오솔로그는 숙주 게놈에서 확인되지 않았다(도 14, 패널 A). 도 12, 패널 D는 RNA 및 단백질에 의해 구성되는 다양한 CRISPR 복합체의 분획을 도시한다.

[0579] 도 13, 패널 a 내지 b는 CRISPR-Cas14 좌위에 대한 메타전사체를 도시한다. 도 13, 패널 a는 Cas14a 오솔로그에 대한 환경적 RNA 서열분석 판독을 도시한다. Cas14 및 CRISPR 어레이의 위치를 이하에 나타낸다. 우측의 RNA 구조는 메타전사체로부터 확인한 tracrRNA의 인실리코 예측 구조를 나타낸다. 도 13, 패널 b는 Cas14a1 crRNA:tracrRNA 이중가닥에 대한 예측된 혼성화를 도시한다.

[0580] 도 14, 패널 A 내지 B는 RNA 가공 및 CRISPR-Cas14에 의한 이종성 발현을 도시한다. 도 14, 패널 A는 Cas14 함유 게놈에서 통상적인 RNase 오솔로그의 존재를 도시한다. 연보라색은 주어진 RNase에 대해 예상되는 길이보다 상당히 더 짧은 히트를 나타낸다. RNase III는 모든 연구한 게놈에서 없다는 것을 주목한다. 도 14, 패널 B는 메타전사체 판독(도 14, 패널 B, 상부 그래프)에 비해 이콜라이 내 Cas14a1 좌위(도 14, 패널 B, 하부 2개의 그래프)의 이종성 발현으로부터의 작은 RNAseq 판독을 도시한다. 풀 다운은 Ni-NTA 친화도 정제된 Cas14a1과 함께 공동정제한 RNA를 지칭한다.

[0581] 도 15, 패널 a 내지 d는 Cas14a1 및 SpCas9에 의한 플라스미드 고갈을 도시한다. 도 15, 패널 a는 PAM 발견 실험의 윤곽을 나타내는 다이어그램을 도시한다. 관심대상의 CRISPR 시스템을 발현시키는 이콜라이를 표적에 측접하는 무작위화된PAM 서열을 함유하는 플라스미드로 시험감염시켰다. 살아있는 (형질전환된) 세포를 채취하고 나서, 빈 벡터를 보유하는 대조군과 함께 서열분석하였다. 이어서, 고갈된 서열을 서열분석하고 나서, PAM 고갈값 역치(PDVT) 초과로 고갈된 PAM을 사용하여 웹로고(Weblogo)를 생성하였다. 도 15, 패널 b 내지 d는 표적의 무작위화된 PAM 서열 5'(도 15, 패널 b) 또는 3'(도 15, 패널 c)을 함유하는 표적 플라스미드로 형질전환시킨 Cas14a1을 이종성으로 발현시킴으로써 고갈된 PAM 서열을 도시한다. "서열 없음"은 주어진 PDVT에서 또는 초과에서 고갈된 서열이 발견되지 않았다는 것을 나타내었다.

[0582] 실시예 4

[0583] 정제된 Cas14a-tracrRNA-crRNA 복합체가 시험관내에서 RNA-가이드된 핵산 절단을 할 수 있는지의 여부를 시험하였다. 모든 현재 재구성된 DNA-표적화 클래스 2 간섭 복합체는 dsDNA와 ssDNA 기질을 둘 다 인식할 수 있었다. 정제한 Cas14a-tracrRNA-crRNA 복합체를 crRNA 가이드 서열에 상보성인 20-뉴클레오타이드 서열을 보유하는 방사성표지된 표적 올리고뉴클레오타이드(ssDNA, dsDNA 및 ssRNA), 또는 비상보성 ssDNA와 함께 인큐베이션시키고, 이들 기질을 Cas14a-매개 절단에 대해 분석하였다. 상보성 ssDNA 기질의 존재에서만 임의의

절단 생성물이 검출되었고(도 16, 패널 a 및 도 17, 패널 A 내지 C), 절단은 tracrRNA와 crRNA 둘 다의 존재에 의존하며, 이는 또한 단일-가이드 RNA(sgRNA)에 조합될 수 있었다(도 16, 패널 b 및 도 18). 검출 가능한 dsDNA 절단의 결여는 Cas14a가 ssDNA를 선택적으로 표적화하였지만, 일부 다른 인자 및 서열 필요가 천연 숙주에서 dsDNA 인식을 가능하게 할 수 있다는 것을 시사한다. Cas14a RuvC 도메인에서 보존된 활성 부위 잔기의 돌연변이는 절단 활성을 제거하였는데(도 17, 패널 A), 이는 DNA 절단을 초래하는 도메인으로서 RuvC를 나타낸다. 게다가, Cas14a DNA 절단은 천연 생성 서열보다 더 짧은 길이로의 RNA 성분의 절단에 민감하였다(도 16, 패널 b 및 도 19, 패널 A 내지 D). 이들 결과는 가장 짧은 클래스 2 CRISPR 효과기로서 Cas14a가 프로그램 가능한 RNA-가이드 DNA 절단을 수행한다는 것을 입증하였다는 것을 확인하였다.

[0584] Cas14a가 ssDNA 기질에 걸쳐 Cas14a 가이드를 타일링(tiling)함으로써 시험관내 ssDNA 절단을 위해 PAM을 필요로 하는지의 여부를 시험하였다(도 16, 패널 C). 이들 상이한 가이드의 표적에 인접한 서열 변이에도 불구하고, 모두 4개의 서열에 대해 절단이 관찰되었다. 절단 부위는 ssDNA의 가이드-상보성 영역을 지나서 일어났고, 가이드 결합 위치에 반응하여 이동하였다(도 16, 패널 C). 이들 데이터는 Cas14a가 활성화를 위해 PAM을 필요로 하지 않는 ssDNA-표적화 CRISPR 엔도뉴클레아제였다는 것을 입증하였다.

[0585] Cas14a가 crRNA/DNA 표적화 이형이중가닥 외부에서 절단한다는 관찰에 기반하여, Cas14a는 RuvC-함유 효소 Cas12a와 유사하게 표적-활성화된 비-특이적 ssDNA 절단 활성을 가질 수 있다는 것을 제안하였다. 이 가능성을 시험하기 위해, Cas14a-tracrRNA-crRNA를 상보성 활성체 DNA, 및 Cas14a crRNA 또는 활성체에 상보성인 서열을 보유하지 않는 M13 박테리오파지 ssDNA의 분취액과 함께 인큐베이션시켰다(도 16, 패널 D). M13 ssDNA는 보존된 Cas14a RuvC 활성 부위의 돌연변이에 의해 제거된 활성인 작은 단편으로 빠르게 분해되었고, 이는 Cas14a의 활성화가 비특이적 ssDNA 분해를 초래하였다는 것을 시사한다.

[0586] Cas14a에 의한 표적-의존적 비-특이적 DNA 절단 활성의 특이성을 연구하기 위해, 형광단-퀀처(FQ) 분석을 적합화하였고, 이때 염료-표지된 ssDNA의 절단은 형광 신호를 생성한다(도 20, 패널 A). Cas14a를 다양한 가이드 RNA-표적 ssDNA 쌍과 함께 인큐베이션시켰을 때, 동족 표적의 존재 하에서만 형광 신호가 관찰되었고 더 긴 FQ-함유 기질에 대한 강한 선호도를 나타내었다(도 19, 패널 F 및 도 20, 패널 A). 다양한 ssDNA 기질에서 표적화된 영역에 걸쳐 2-nt 미스매치를 타일링함으로써 Cas14a 미스매치 내성(mismatch tolerance)을 시험하였다. ssDNA 표적 중간 근처의 미스매치는 Cas14a 활성을 강하게 저해하였는데, 이는 PAM-근위 종자 영역과 별개인 내부 종자 서열이 dsDNA-표적화 CRISPR-Cas 시스템에 대해 관찰되었다는 것을 나타낸다(도 20, 패널 B 및 도 21, 패널 A 내지 D). 게다가, 강한 2차 구조를 함유하는 DNA 기질은 Cas14a의 감소된 활성화를 초래하였다(도 21, 패널 E). ssDNA 기질의 절단은 또한 감소된 또는 검출 가능하지 않은 트랜스 절단을 초래하였다(도 21, 패널 F). 이들 결과는 dsDNA-표적화 클래스 2 CRISPR 시스템과 별개의 신뢰도 메커니즘을 시사하였는데, RNA-표적화 Cas13a 효소와 유사하게 절단 활성을 개폐하기 위해 crRNA의 사전 정돈된 영역을 이용할 가능성이 있다.

[0587] 메타게놈 데이터에서 조밀한 V형 시스템의 추가적인 연구는 Cas14a 내지 c와 유사하게, 획득-연관 *cas* 유전자 및 CRISPR 어레이에 인접한 짧은 RuvC-함유 단백질을 암호화하는 유전자를 포함한다는 시스템의 큰 다양성을 나타내었다. 5가지 주요 패밀리(Cas14d 내지 h)에 클러스터링된 스무 개(20)의 추가적인 이러한 시스템을 발견하였다. 이들 패밀리는 V형 CRISPR 효과기의 진화 모체로서 연루된 트랜스포사제-연관 단백질인 TnpB의 독립적 순화(domestication) 사건으로부터 진화되는 것으로 보인다. *cas*12b와 관련된 *cas*14g를 제외하고, *cas14*-유사 유전자는 V형 효과기 계통발생 상에서 별개의 계통군을 형성하였고(도 22, 패널 A 내지 B), 이들의 *cas1* 유전자는 상이한 기원을 가졌다(도 10, 패널 A). 8가지의 패밀리(Cas14a 내지 h)에 속하는 38개의 CRISPR-Cas14 시스템을 확인하였고, 분석에 의해 클러스터링될 수 없는 8가지의 추가적인 시스템을 확인하였다(Cas14u로 칭함, 표 3).

[0588] 본 명세서에 기재한 Cas14 단백질의 작은 크기 및 V형 효과기 단백질에 대한 이런 유사성은 RNA-가이드된 ssDNA 절단이 조상의 클래스 2 CRISPR 시스템으로서 존재할 수 있다는 것을 시사하였다. 이런 시나리오에서, 작은, 길들여진 TnpB-유사 ssDNA 간섭 복합체는 시간에 따라 추가적인 도메인을 획득하여, dsDNA 인식 및 절단을 점진적으로 개선시킬 수 있다. 더 작은 Cas9 오솔로그는 이들의 더 큰 상대보다 더 약한r dsDNA-표적화 활성을 나타내었지만, ssDNA를 강하게 절단하는 능력을 보유하였다. 진화 영향과 별개로, 표적 ssDNA를 특이적으로 표적화하는 Cas14의 능력은 ssDNA 바이러스 또는 ssDNA 중간체를 통해 증식되는 전이 유전 인자(mobile genetic element: MGE)에 대한 방어에서의 어떤 역할을 시사하였다. 임의의 특정 이론에 의해 구속되는 것을 의도하지는 않지만, ssDNA-표적화 CRISPR 시스템은 ssDNA 바이러스가 대다수의 바이러스 존재비에 포함되는 특정 해양 환경에서 특히 유리할 수 있다.

[0589] **도 16, 패널 a 내지 D**는 CRISPR-Cas14a가 RNA-가이드된 DNA-엔도뉴클레아제라는 것을 도시한다. 도 16, 패널 a

는 ssDNA, dsDNA, ssRNA 및 비-표적 ssDNA를 표적화하는 Cas14a1의 절단 역학을 도시한다. 도 16, 패널 b는 표적 ssDNA에 결합된 Cas14a RNP의 다이어그램 및 다양한 RNA 성분의 존재 하에 방사성표지된 ssDNA의 Cas14a1 절단 역학을 도시한다. 도 16, 패널 C는 Cas14a1 가이드 서열에 의한 ssDNA 기질의 타일링을 도시한다. 도 16, 패널 D는 활성화된 Cas14a1에 의한 ssDNA 바이러스 M13 게놈의 절단을 도시한다.

[0590] **도 17, 패널 A 내지 E**는 Cas14a1에 의한 ssDNA의 분해를 도시한다. 도 17, 패널 A는 정제된 Cas14a1 및 Cas14a1 점 돌연변이의 SDS-PAGE를 도시한다. 도 17, 패널 B는 ssDNA 표적의 Cas14a1 절단을 위한 염, 양이온 및 온도의 최적화를 도시한다. 도 17, 패널 C는 다양한 길이의 스페이서 서열을 갖는 Cas14a1에 의한 ssDNA의 방사성표지된 절단을 도시한다. 도 17, 패널 D는 RuvC 활성 부위 잔기를 확인하기 위해 이전에 연구한 Cas12 단백질에 의한 Cas14의 정렬을 도시하고, 도 17, 패널 e는 정제된 Cas14a1 RuvC 점 돌연변이에 의한 ssDNA의 절단을 도시한다.

[0591] **도 18**은 다양한 가이드 RNA 성분에 의한 ssDNA의 Cas14a1 절단 역학을 도시한다.

[0592] **도 19, 패널 A 내지 F**는 Cas14a1 가이드 RNA 성분의 최적화를 도시한다. 도 19, 패널 A는 Cas14a1 표적화 ssDNA의 다이어그램을 도시한다. 스페이서 길이(도 19, 패널 B), 반복부 길이(도 19, 패널 C), tracrRNA(도 19, 패널 D)를 달리하는 것에 의한, crRNA와 tracrRNA를 함께 융합(도 19, 패널 E) 하는 것에 의한 기질의 Cas14a1 절단에 대한 영향. 도 19, 패널 f는 다양한 가이드 및 표적 조합의 존재 하에 ssDNA FQ 리포터의 절단으로부터 초래되는 배경 차감된 형광을 나타내는 히트맵을 도시한다.

[0593] **도 20, 패널 a 내지 E**는 CRISPR-Cas14a에 의한 고충실도 ssDNA SNP 검출을 도시한 도면. 도 20, 패널 a는 Cas14a1에 의한 ssDNA의 검출 및 다양한 길이 FQ 기질에 대한 절단 역학에 대한 형광-퀀처(FQ) 분석을 도시한다. 도 20, 패널 B는 기질을 가로질러 타일링된 미스매치를 갖는 Cas14a1에 대한 절단 역학을 도시한다(개개 점은 반복검증 측정을 나타낸다). 도 20, 패널 c는 Cas14-DETECTR 전략 및 HERC2 눈 색 SNP의 다이어그램을 도시한다. 도 20, 패널 D는 T7 엑소뉴클레아제의 적정 및 Cas14a-DETECTR에 대한 영향을 도시한다. 도 20, 패널 E는 Cas12a를 이용하는 ssDNA 검출에 비교되는 청색눈 및 갈색눈 타액 샘플에 대한 청색눈 표적화 가이드와 함께 Cas14a-DETECTR를 이용하는 SNP 검출을 도시한다.

[0594] **도 21**, 패널 a 내지 F는 Cas14a1 절단율에 대한 다양한 활성체의 영향을 도시하고, 도 21, 패널 A는 패널 A 내지 D에서 사용되는 미스매치 위치 및 표적 1에 대한 미스매치(MM) 위치의 대표적인 복제물에 대한 원 비율(raw rate)을 이용하여 ssDNA의 Cas14a1 표적화의 다이어그램을 도시한다. 3가지의 상이한 기질에 걸쳐 타일링된 돌연변이를 갖는 Cas14a 표적화 기질에 대한 절단율(도 21, 패널 B 내지 D). 도 21, 패널 E는 2차 구조의 양이 증가된 기질에 대한 트랜스 절단율을 도시한다. 도 21, 패널 F는 절단된 기질을 갖는 트랜스 세척률을 도시한다. 점은 개개의 측정을 나타낸다.

[0595] **도 22, 패널 a 내지 b**는 CRISPR-Cas14 시스템의 다양성을 도시한다. 도 22, 패널 A는 표시된 Cas14 시스템에 대한 대표적인 좌위 구조를 나타낸다. 단백질 길이를 일정한 비율로 도시한다. 도 22, 패널 B는 Cas14의 모두 8가지의 확인된 아형을 포함하는 V형 효과기에 대한 최대 가능성 계통수를 도시한다.

[0596] **도 23, 패널 A 내지 C**는 이종성 숙주에서의 Cas14a1 매개 간섭 시험을 도시한다. 이콜라이(E. coli)에서의 간섭을 시험하기 위한 Cas14a1 및 LbCas12a 작제물의 다이어그램. (B) Cas12a를 나타내지만 Cas14a1-매개 간섭을 나타내지 않는 이콜라이에 대해 스팟팅한 ΦX174의 플라크. 각각의 스팟은 ΦX174 저장액의 10배 희석물을 나타낸다. (c) ΦX174로 감염시킨 Cas14a1 또는 LbCas12a를 발현시키는 이콜라이의 성장 곡선(T, 표적화; NT, 비-표적화). 도 19, 패널 F는 30분의 인큐베이션 후에 다양한 가이드 및 표적 조합의 존재 하에 ssDNA FQ 리포터의 절단으로부터 초래되는 배경-차감 형광을 나타내는 히트맵을 나타낸다.

[0597] 본 발명을 이의 구체적 실시형태를 참고로 하여 기재하였지만, 본 발명의 진정한 정신과 범주로부터 벗어나는 일 없이 다양한 변화가 이루어질 수 있고 동등물이 치환될 수 있다는 것이 이해되어야 한다. 추가로, 본 발명의 객관적 정신 및 범주에 대해 특정 상황, 물질, 대상 조성물, 공정, 공정 단계 또는 단계들을 적용하여 다수의 변형이 이루어질 수 있다. 모든 이러한 변형은 본 명세서에 첨부하는 청구범위의 범주 내인 것으로 의도된다.

참고문헌

1. R. Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**, 1709–12 (2007).
2. S. A. Jackson *et al.*, CRISPR-Cas: Adapting to change. *Science*. **356** (6333), pp. 1–9 (2017).
3. S. Shmakov *et al.*, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
4. J. S. Chen, J. A. Doudna, The chemistry of Cas9 and its CRISPR colleagues. *Nat. Rev. Chem.* **1**, 0078 (2017).
5. C. T. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*. **523**, 208–211 (2015).
6. K. Anantharaman *et al.*, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
7. V. M. Markowitz *et al.*, IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42**, 568–573 (2014).
8. V. M. Markowitz *et al.*, IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, 115–122 (2012).
9. A. J. Probst *et al.*, Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2concentrations. *Environ. Microbiol.* **19**, 459–474 (2017).
10. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
11. J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. a. Doudna, Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*. **519**, 193–198 (2015).
12. S. Shmakov *et al.*, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell*. **60**, 385–397 (2015).
13. D. Burstein *et al.*, New CRISPR-Cas systems from uncultivated microbes. *Nature*. **542**, 237–241 (2017).
14. C. Rinke *et al.*, Insights into the phylogeny and coding potential of microbial dark matter. *Nature*. **499**, 431–437 (2013).
15. C. J. Castelle *et al.*, Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015).
16. E. Deltcheva *et al.*, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. **471**, 602–607 (2011).
17. K. E. Savell, J. J. Day, Applications of CRISPR/CAS9 in the mammalian central nervous system. *Yale J. Biol. Med.* **90** (2017), pp. 567–581.
18. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*. **155**, 733–740 (2009).
19. Y. Zhang, R. Rajan, H. S. Seifert, A. Mondragón, E. J. Sontheimer, DNase H Activity of Neisseria meningitidis Cas9. *Mol. Cell*. **60**, 242–255 (2015).
20. E. Ma, L. B. Harrington, M. R. O'Connell, K. Zhou, J. A. Doudna, Single-Stranded DNA

[0598]

Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell*. **60**, 398–407 (2015).
21. J. S. Chen *et al.*, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*. **360**, 436–439 (2018).
22. B. Zetsche *et al.*, Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. **163**, 759–771 (2015).
23. A. East-Seletsky *et al.*, Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*. **538**, 270–273 (2016).
24. L. Liu *et al.*, The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell*. **170**, 714–726.e10 (2017).
25. O. O. Abudayyeh *et al.*, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (80-.).* **353**, 1–9 (2016).
26. G. J. Knott *et al.*, Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nat. Struct. Mol. Biol.* **24**, 825–833 (2017).
27. S. Y. Li *et al.*, CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res.* **28**, 491–493 (2018).
28. H. Eiberg *et al.*, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum. Genet.* **123**, 177–187 (2008).
29. S. Shmakov *et al.*, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
30. E. V Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
31. K. S. Makarova *et al.*, An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*, 1–15 (2015).
32. O. Barabas *et al.*, Mechanism of IS 200 / IS 605 Family DNA Transposases : Activation and Transposon- Directed Target Site Selection, 208–220 (2008).
33. M. Yoshida *et al.*, Quantitative viral community DNA analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku, Japan. *Front. Microbiol.* **9**, 1–10 (2018).
34. C. T. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*. **523**, 208–211 (2015).
35. K. Anantharaman *et al.*, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
36. A. J. Probst *et al.*, Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2concentrations. *Environ. Microbiol.* **19**, 459–474 (2017).
37. V. M. Markowitz *et al.*, IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42**, 568–573 (2014).
38. V. M. Markowitz *et al.*, IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, 115–122 (2012).
39. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
40. D. Burstein *et al.*, New CRISPR-Cas systems from uncultivated microbes. *Nature*. **542**, 237–241 (2017).
41. I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**, W52–W57 (2007).
42. A. Biswas, R. H. J. Staals, S. E. Morales, P. C. Fineran, C. M. Brown, CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics*. **17**, 1–14 (2016).
43. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
44. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods*. **9**, 357–359 (2012).

[0599]

45. H. Ogata *et al.*, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
46. L. B. Harrington *et al.*, A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1–7 (2017).
47. G. Crooks, G. Hon, J. Chandonia, S. Brenner, NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res*. **14**, 1188–1190 (2004).
48. L. B. Harrington *et al.*, A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell*. **170**, 1224–1233.e15 (2017).
49. J. S. Chen *et al.*, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*. **360**, 436–439 (2018).
50. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*. **523**, 208–211 (2015), doi:10.1038/nature14486.

[0600]

도면

도면1a

>CasZa.1|rifcsphigho2_02_scaffold_2167_39 \N id=61050970
partial=0 scaf=rifcsphigho2_02_scaffold_2167
bin=RIFCSPHIGHO2_02_FULL_Archaea_Woesearchaeota_57_22
sample=Archaea-RifleO2 proj=Rifle Groundwater Metagenome
mEVQKTVMKTLSLRILRPLYSQEIEKEIKEEKERRKQAGGTGELDGGFYKKLEKKHSEMFSFDR
LNLLLNQLQREIAKVYNHAISELYIATIAQGNKSNKHYISSIVYNRAYGYFYNAYIALGICSKV
EANFRSNELLTQQSALPTAKSDNFPIVLHKQKGAEGEDGGFRISTEGSDLIFEIPIPFYEYNGE
NRKEPYKWVKKGGQKPVLKLILSTFRRQRNKGWAKDEGTDAEIRKVTEGKYQVSQIEINRGKKL
GEHQKWFANFSIEQPIYERKPNRSIVGGLDVGIRSPLVCAINNSFSRYSVDSNDVFKFSKQVFA
FRRRLLSKNSLKRKGHGAAHKLEPITEMTEKNDKFRKKIIERWAKEVTNFFVKNQVGIVQIEDL
STMKDREDHFFNQYLRGFWPYYQMQTLIENKLKEYGIEVKRVQAKYTSQLCSNPNCRYWNNYFN
FEYRKVNKFPKFKCEKCNLEISADYNAARNLSTPDIEKFVAKATKGINLPEK
(서열번호 : 1)

>CasZa.2|gwa2_scaffold_18027_12 Putative transposase DNA-binding
domain family id=2625699 partial=0 scaf=gwa2_scaffold_18027
bin=GWA2 Unbinned sample=GWA2 proj=Rifle Groundwater Metagenome
mEEAKTVSKTLSLRILRPLYSAEIEKEIKEEKERRKQGGKSGELDSGFYKKLEKKHTQMFGWDK
LNLMLSQLQRQIARVFNQSISELYIETVIQGKKSNKHYTSKIVYNRAYSVFYNAYLALGITSKV
EANFRSTELLMQKSSLPTAKSDNFPILLHKQKGVEGEEGGFKISADGNDLIFEIPIPFYEYDSA
NKKEPFKWIKKGGQKPTIKLILSTFRRQRNKGWAKDEGTDAEIRKVIEGKYQVSHIEINRGKKL
GDHQKWFVNFTIEQPIYERKLDKNIIGGIDVGIKSPLVCAVNNSFARYSVDSNDVLKFSKQAFA
FRRRLLSKNSLKRSGHGSKNKLDPITRMTEKNDRFRKKIIERWAKEVTNFFIKNQVGTVQIEDL
STMKDRQDNFFNQYLRGFWPYYQMQNLIENKLKEYGIETKRIKARYTSQLCSNPSCRHWNSYFS
FDHRKTNNFPKFKCEKCALEISADYNAARNISTPDIEKFVAKATKGINLPDKNENVILE
(서열번호 : 2)

>CasZa.3|gwa1_scaffold_1795_31 Transposase, IS605 OrfB family
id=2784443 partial=0 scaf=gwa1_scaffold_1795
bin=GWA1_Archaeon_57_19 sample=GWA1 proj=Rifle Groundwater
Metagenome
mAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYCTTQV
ERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKG
KGIANASSVEHYLSDVCYTRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI
PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLST
QRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIGEKSAWMLNLSIDVPKIDKGVDPS
IIGGIDVGVKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKP
ITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQmENLESMKRKEDSYFNIRLRGFWPYAEM
QNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENADYN
AALNISNPKLKSTKEEP (서열번호 : 3)

도면1b

>CasZa.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_2 \N
id=132542232 partial=0
scaf=CG10_big_fil_rev_8_21_14_0.10_scaffold_20906
bin=CG10_big_fil_rev_8_21_14_0_10_UNK
sample=CG10_big_fil_rev_8_21_14_0_10 proj=Crystal Geyser
mERQKVPQIRKTVRVVPLRTLRPKYSDVTENALKKFKEKGDDTNTNDFWRAIRDRDTEFFRKEL
NFSEDEINQLERDTLFRVGLDNRVLFSYFDFLQEKLMKDYNKIISKLFINRQSKSSFENDLTDE
EVEELIEKDVTPFYGAYIGKGIKSVIKSNLGGKFIKSVKIDRETKKVTKLTAINIGLMGLPVAK
SDTFPIKIIKTNPDYITFQKSTKENLQKIEDYETGIEYGDLLVQITIPWFKNENKDFSLIKTKE
AIEYYKLNGVGKKDLLNINLVLTTYHIRKKKSWQIDGSSQSLVREMANGELEEKWKSFFDTFIK
KYGDEGKSALVKRRVNKKSRAKGEKGRELNLDERIKRLYDSIKAKSFPSEINLIPENYKWKLHF
SIEIPPMVNDIDSNLYGGIDFGEQNIATLCVKNIEKDDYDFLTIYGNDLLKHAQASYARRRIMR
VQDEYKARGHGKSRKTKAQEDYSERMQKLRQKITERLVKQISDFFLWRNKFHMAVCSLRYEDLN
TLYKGESVKAKRMRQFINKQQLFNGIERKLKDYNSEIYVNSRYPHYTSRLCSKCGKLNLYFDFL
KFRTKNIIIRKNPDGSEIKYMPFFICEFCGWKQAGDKNASANIADKDYQDKLNKEKEFCNIRKP
KSKKEDIGEENEEERDYSRRFNRNSFIYNSLKKDNKLNQEKLFDEWKNQLKRKIDGRNKFEPKE

YKDRFSYLFAYYQEIIKNESES (서열번호 : 4)

>CasZa.5|rifcsplowo2_01_scaffold_34461_9 \N id=58382281
partial=0 scaf=rifcsplowo2_01_scaffold_34461
bin=RifCSPlowO2_01_full_UNK sample=RIFCSPLOWO2_01_FULL
proj=RifleCSP2_LowO2
mVPTELITKTLQLRVIRPLYFEEIEKELAELKEQKEKEFEETNSLLLESKKIDAKSLKKLKRKA
RSSAAVEFWKIAKEKYPDILTKPEMEFIFSEMQKMMARFYNKSMTNIFIEMNNDEKVNPLSLIS
KASTEANQVIKCSSISSGLNRKIAGSINKTKFKQVRDGLISLPTARTETFPISFYKSTANKDEI
PISKINLPSEEEADLTITLPFPFFEIKKEKKGQKAYSYFNIIEKSGRSNNKIDLLLSTHRRQRR
KGWKEEGGTSAEIRRLMEGEFDKEWEIYLGEAEKSEKAKNDLIKNMTRGKLSKDIKEQLEDIQV
KYFSDNNVESWNDLSKEQKQELSKLRKKKVEELKDWKHVKEILKTRAKIGWVELKRGKRQRDRN
KWFVNITITRPPFINKELDDTKFGGIDLGVKVPFVCAVHGSPARLIIKENEILQFNKMVSARNR
QITKDSEQRKGRGKKNKFIKKEIFNERNELFRKKIIERWANQIVKFFEDQKCATVQIENLESFD

RTSYK (서열번호 : 5)

>CasZa.6|037|3300012359.a:Ga0137385_10000156_01776
Zr6.cluA_1a.mostc10like protein
MKSDTKDKKIIIHQTKTLSLRIVKPQSIPMEEFTDLVRYHQMIIFPVYNNGAIDLYKKLFKAKI
QKGNEARAIKYFMNKIVYAPIANTVKNSYIALGYSTKMQSSFSGKRLWDLRFGEATPPTIKADF
PLPFYNQSGFKVSSENGEFIIGIPFGQYTKKTVSDIEKKTSFAWDKFTLEDTTKKTLIELLLST
KTRKMNEGWKNNEGTEAEIKRVMDGTYQVTSLEILQRDDSWFVNFNIAYDSLKKQPDRDKIAGI
HMGITRPLTAVIYNNKYRALSIYPNTVMHLTQKQLARIKEQRTNSKYATGGHGRNAKVTGTDTL
SEAYRQRRKKIIEDWIASIVKFAINNEIGTIYLEDISNTNSFFAAREQKLIYLEDISNTNSFLS
TYKYPISAISDTLQHKLEEKAIQVIRKKAYYVNQICSLCGHYNKGFTYQFRRKNKFPKMKCQGC

LEATSTEFNAAANVANPDYEKLLIKHGLLQLKK (서열번호 : 6)

도면1c

>CasZa.7|127|3300009029.a:Ga0066793_10010091_01298
Zr6.cluA_1.c10.inclusive
MSTITRQVRLSPTPEQSRLLMAHCQQYISTVNVLVAAFDSEVLTGKVSTKDFRAALPSAVKNQA
LRDAQSVFKRSVELGCLPVLKKPHCQWNNQNWRVEGDQLILPICKDGKTQQERFRCAAVALEGK
AGILRIKKKRGKWIADLTVTQEDAPESSGSAIMGVDLGIKVPAVAHIGGKGTRFFGNGRSQRSM
RRRFYARRKTLQKAKKLRAVRKSKGKEARWMKTINHQLSRQIVNHAHALGVGTIKIEALQGIRK
GTTRKSRGAAARKNNRMTNTWSFSQLTLFITYKAQRQGITVEQVDPAYTSQDCPACRARNGAQD
RTYVCSECGWRGHRDTVGAINISRRAGLSGHRRGATGA (서열번호 : 7)

>CasZb.1|bjp_ig2599_sub10_scaffold_488_2 transposase id=60827220
partial=0 scaf=bjp_ig2599_sub10_scaffold_488
bin=BJP_IG2599_SUB10_Micrarchaeota_NOVEL_41_24
sample=BJP_IG2599_SUB10 proj=Borehole JP
mIAQKTIKIKLNPTKEQIIKLNSIIEEYIKVSNFTAKKIAEIQESFTDSGLTQGTCSECGKEKT
YRKYHLLKKDNKLFCITCYKRKYSQFTLQKVEFQNKTGLRNVAKLPKTYYTNAIRFASDTFSGF
DEIIKKKQNRLNSIQNRLNFWKELLYNPSNRNEIKIKVVKYAPKTDTREHPHYYSEAEIKGRIK
RLEKQLKKFKMPKYPEFTSETISLQRELYSWKNPDELKISSITDKNESmNYYGKEYLKRYIDLI
NSQTPQILLEKENNSFYLCFPITKNIEMPKIDDTFEPVGIDWGITRNIAVVSILDSKTKKPKFV
KFYSAGYILGKRKHYKSLRKHFGQKKRQDKINKLGTKEDRFIDSNIHKLAFLIVKEIRNHSNKP
IILMENITDNREEAEKSMRQNILLHSVKSRLQNYIAYKALWNNIPTNLVKPEHTSQICNRCGHQ
DRENRPKGSKLFKCVKCNYMSNADFNASINIARKFYIGEYEPFYKDNEKMKSGVNSISM (서
열번호 : 8)

>CasZb.2|rifcsplowo2_01_scaffold_239_52 IS605 OrfB family
transposase id=57672634 partial=0
scaf=rifcsplowo2_01_scaffold_239 bin=RifCSPlowO2_01_full_UNK
sample=RIFCSPLOWO2_01_FULL proj=RifleCSP2_LowO2
LKLSEQENITTGVKFKLKLDKETSEGLNDYFDEYGKAINFAIKVIQKELAEDRFAGKVRLDENK
KPLLNEDGKKIWDFPNEFCSCGKQVNRYVNGKSLCQECYKNKFTEYGIRKRMYSAKGRKAEQDI
NIKNSTNKISKTHFNYAIREAFILDKSIKKQRKERFRRLREMKKKLQEFIEIRDGNKILCPKIE
KQRVERYIHPSWINKEKKLEDFRGYSMSNVLGKIKILDRNIKREEKSLKEKGQINFKARRLMLD
KSVKFLNDNKISFTISKNLPKEYELDLPEKEKRLNWLKEKIKIIKNQKPKYAYLLRKDDNFYLQ
YTLETEFNLKEDYSGIVGIDRGVSHIAVYTFVHNNGKNERPLFLNSSEILRLKNLQKERDRFLR
RKHNKKRKKSNMRNIEKKIQLILHNYSKQIVDFAKNKNAFIVFEKLEKPKKNRSKMSKKSQYKL
SQFTFKKLSDLVDYKAKREGIKVLYISPEYTSKECSHCGEKVNTQRPFNGNSSLFKCNKCGVEL
NADYNASINIAKKGLNILNSTN
(서열번호 : 9)

도면1d

>CasZb.3|rifcsplowo2_01_scaffold_282_93 IS605 OrfB family transposase id=57678680 partial=0 scaf=rifcsplowo2_01_scaffold_282 bin=RifCSPlowO2_01_full_UNK sample=RIFCSPLOWO2_01_FULL proj=RifleCSP2_LowO2
mEESIITGVKFKLRIDKETTKKLNEYFDEYGKAINFAVKIIQKELADDRFAGKAKLDQNKNPIL
DENGKKIYEFPDEFCSCGKQVNKYVNNKPFCQECYKIRFTENGIRKRMYSAKGRKAEHKINILN
STNKISKTHFNYAIREAFILDKSIKKQRKKRNERLRESKKRLQQFIDMRDGKREICPTIKGQKV
DRFIHPSWITKDKKLEDFRGYTLSIINSKIKILDRNIKREEKSLKEKGQIIFKAKRLMLDKSIR
FVGDRKVLFTISKTLPKEYELDLPSKEKRLNWLKEKIEIIKNQKPKYAYLLRKNIESEKKPNYE
YYLQYTLEIKPELKDFYDGAIGIDRGINHIAVCTFISNDGKVTPPKFFSSGEILRLKNLQKERD
RFLLRKHNKNRKKGNMRVIENKINLILHRYSKQIVDmAKKLNASIVFEELGRIGKSRTKMKKSQ
RYKLSLFIFKKLSDLVDYKSRREGIRVTYVPPEYTSKECSHCGEKVNTQRPFNGNYSLFKCNKC
GIQLNSDYNASINIAKKGLKIPNST
(서열번호 : 10)

>CasZb.4|rifcsphigho2_01_scaffold_36781_5 transposase, IS605 OrfB family id=55842010 partial=0 scaf=rifcsphigho2_01_scaffold_36781 bin=RifCSPhighO2_01_full_UNK sample=RIFCSPHIGHO2_01_FULL proj=RifleCSP2_HighO2
LWTIVIGDFIEMPKQDLVTTGIKFKLDVDKETRKKLDDYFDEYGKAINFAVKIIQKNLKEDRFA
GKIALGEDKKPLLDKDGKKIYNYPNESCSCGNQVRRYVNAKPFCVDCYKLKFTENGIRKRMYSA
RGRKADSDINIKNSTNKISKTHFNYAIREGFILDKSLKKQRSKRIKKLLELKRKLQEFIDIRQG
QMVLCPKIKNQRVDKFIHPSWLKRDKKLEEFRGYSLSVVEGKIKIFNRNILREEDSLRQRGHVN
FKANRIMLDKSVRFLDGGKVNFNLNKGLPKEYLLDLPKKENKLSWLNEKISLIKLQKPKYAYLL
RREGSFFIQYTIENVPKTFSDYLGAIGIDRGISHIAVCTFVSKNGVNKAPVFFSSGEILKLKSL
QKQRDLFLRGKHNKIRKKSNMRNIDNKINLILHKYSRNIVNLAKSEKAFIVFEKLEKIKKSRFK
MSKSLQYKLSQFTFKKLSDLVEYKAKIEGIKVDYVPPEYTSKECSHCGEKVDTQRPFNGNSSLF
KCNKCRVQLNADYNASINIAKKSLNISNN (서열번호 : 11)

>CasZb.5|cg1_0.2_scaffold_785_c_37 transposase, IS605 OrfB family id=91826491 partial=0 scaf=cg1_0.2_scaffold_785_c bin=CG1_02_FULL_Micrarchaeota_47_40_curated sample=CG1_02_FULL proj=Crystal Geyser
mSKTTISVKLKIIDLSSEKKEFLDNYFNEYAKATTFCQLRIRRLLRNTHWLGKKEKSSKKWIFE
SGICDLCGENKELVNEDRNSGEPAKICKRCYNGRYGNQMIRKLFVSTKKREVQENMDIRRVAKL
NNTHYHRIPEEAFDMIKAADTAEKRRKKNVEYDKKRQMEFIEMFNDEKKRAARPKKPNERETRY
VHISKLESPSKGYTLNGIKRKIDGMGKKIERAEKGLSRKKIFGYQGNRIKLDSNWVRFDLAESE
ITIPSLFKEMKLRITGPTNVHSKSGQIYFAEWFERINKQPNNYCYLIRKTSSNGKYEYYLQYTY
EAEVEANKEYAGCLGVDIGCSKLAAAVYYDSKNKKAQKPIEIFTNPIKKIKMRREKLIKLLSRV
KVRHRRRKLMQLSKTEPIIDYTCHKTARKIVEMANTAKAFISMENLETGIKQKQQARETKKQKF
YRNMFLFRKLSKLIEYKALLKGIKIVYVKPDYTSQTCSSCGADKEKTERPSQAIFRCLNPTCRY
YQRDINADFNAAVNIAKKALNNTEVVTTLL (서열번호 : 12)

도면1e

>CasZb.6|rifcsphigho2_02_scaffold_55589_5 transposase, IS605
OrfB family id=61682131 partial=0
scaf=rifcsphigho2_02_scaffold_55589 bin=RifCSPhighO2_02_full_UNK
sample=RIFCSPHIGHO2_02_FULL proj=RifleCSP2_HighO2
mARAKNQPYQKLTTTTGIKFKLDLSEEEGKRFDEYFSEYAKAVNFCAKVIYQLRKNLKFAGKKE
LAAKEWKFEISNCDFCNKQKEIYYKNIANGQKVCKGCHRTNFSDNAIRKKMIPVKGRKVESKFN
IHNTTKKISGTHRHWAFEDAADIIESMDKQRKEKQKRLRREKRKLSYFFELFGDPAKRYELPKV
GKQRVPRYLHKIIDKDSLTKKRGYSLSYIKNKIKISERNIERDEKSLRKASPIAFGARKIKMSK
LDPKRAFDLENNVFKIPGKVIKGQYKFFGTNVANEHGKKFYKDRISKILAGKPKYFYLLRKKVA
ESDGNPIFEYYVQWSIDTETPAITSYDNILGIDAGITNLATTVLIPKNLSAEHCSHCGNNHVKP
IFTKFFSGKELKAIKIKSRKQKYFLRGKHNKLVKIKRIRPIEQKVDGYCHVVSKQIVEMAKERN
SCIALEKLEKPKKSKFRQRRREKYAVSMFVFKKLATFIKYKAAREGIEIIPVEPEGTSYTCSHC
KNAQNNQRPYFKPNSKKSWTSMFKCGKCGIELNSDYNAAFNIAQKALNMTSA
(서열번호 : 13)

>CasZb.7|CG03_land_8_20_14_0.80_scaffold_2214_9 transposase,
IS605 OrfB family id=135165146 partial=0
scaf=CG03_land_8_20_14_0.80_scaffold_2214
bin=CG03_land_8_20_14_0_80_cor_UNK
sample=CG03_land_8_20_14_0_80_cor proj=Crystal Geyser
mDEKHFFCSYCNKELKISKNLINKISKGSIREDEAVSKAISIHNKKEHSLILGIKFKLFIENKL
DKKKLNEYFDNYSKAVTFAARIFDKIRSPYKFIGLKDKNTKKWTFPKAKCVFCLEEKEVAYANE
KDNSKICTECYLKEFGENGIRKKIYSTRGRKVEPKYNIFNSTKELSSTHYNYAIRDAFQLLDAL
KKQRQKKLKSIFNQKLRLKEFEDIFSDPQKRIELSLKPHQREKRYIHLSKSGQESINRGYTLRF
VRGKIKSLTRNIEREEKSLRKKTPIHFKGNRLMIFPAGIKFDFASNKVKISISKNLPNEFNFSG
TNVKNEHGKSFFKSRIELIKTQKPKYAYVLRKIKREYSKLRNYEIEKIRLENPNADLCDFYLQY
TIETESRNNEEINGIIGIDRGITNLACLVLLKKGDKKPSGVKFYKGNKILGMKIAYRKHLYLLK
GKRNKLRKQRQIRAIEPKINLILHQISKDIVKIAKEKNFAIALEQLEKPKKARFAQRKKEKYKL
ALFTFKNLSTLIEYKSKREGIPVIYVPPEKTSQMCSHCAINGDEHVDTQRPYKKPNAQKPSYSL
FKCNKCGIELNADYNAAFNIAQKGLKTLMLNHSH (서열번호 : 14)

>CasZb.8|058|3300002172.a:JGI24730J26740_1002785_01697
Zr6.newTrpA
MLQTLLVKLDPSKEQYKMLYETMERFNEACNQIAETVFAIHSANKIEVQKTVYYPIREKFGLSA
QLTILAIRKVCEAYKRDKSIKPEFRLDGALVYDQRVLSWKGLDKVSLVTLQGRQIIPIKFGDYQ
KARMDRIRGQADLILVKGVFYLCVVVEVSEESPYDPKGVLGVDLGIKNLAVDSDGEVHSGEQTT
NTRERLDSLKARLQSKGTKSAKRHLKKLSGRMAKFSKDVNHCISKKLVAKAKGTLMSIALEDLQ
GIRDRVTVRKAQRRNLHTWNFGLLRMFVDYKAKIAGVPLVFVDPRNTSRTCPSCGHVAKANRPT
RDEFRCVSCGFAGAADHIAAMNIAFRAEVSQPIVTRFFVQSQAPSFRVG
(서열번호 : 15)

도면1f

>CasZb.9|114|3300013125.a:Ga0172369_10000737_00842
Zr6.cluA_2b.c10_like protein
MDEEPDSAEPNLAPISVKLKLVKLDGEKLAALNDYFNEYAKAVNFCELKMQKIRKNLVNIRGTY
LKEKKAWINQTGECCICKKIDELRCEDKNPDINGKICKKCYNGRYGNQMIRKLFVSTNKRAVPK
SLDIRKVARLHNTHYHRIPPEAADIIKAIETAERKRRNRILFDERRYNELKDALENEEKRVARP
KKPKEREVRYVPISKKDTPSKGYTMNALVRKVSGMAKKIERAKRNLNKRKKIEYLGRRILLDKN
WVRFDFDKSEISIPTMKEFFGEMRFEITGPSNVMSPNGREYFTKWFDRIKAQPDNYCYLLRKES
EDETDFYLQYTWRPDAHPKKDYTGCLGIDIGGSKLASAVYFDADKNRAKQPIQIFSNPIGKWKT
KRQKVIKVLSKAAVRHKTKKLESLRNIEPRIDVHCHRIARKIVGMALAANAFISMENLEGGIRE
KQKAKETKKQKFSRNMFVFRKLSKLIEYKALMEGVKVVYIVPDYTSQLCSSCGTNNTKRPKQAI

FMCQNTECRYFGKNINADFNAAINIAKKALNRKDIVRELS (서열번호 : 16)

>CasZb.10|115|3300013125.a:Ga0172369_10010464_01540
Zr6.cluA_2b.c10_like protein
MEKNNSEQTSITTGIKFKLKLDKETKEKLNNYFDEYGKAINFAVRIIQMQLNDDRLAGKYKRDE
KGKPILGEDGKKILEIPNDFCSCGNQVNHYVNGVSFCQECYKKRFSENGIRKRMYSAKGRKAEQ
DINIKNSTNKISKTHFNYAIREAFNLDKSIKKQREKRFKKLKDMKRKLQEFLEIRDGKRVICPK
IEKQKVERYIHPSWINKEKKLEEFRGYSLSIVNSKIKSFDRNIQREEKSLKEKGQINFKAQRLM
LDKSVKFLKDNKVSFTISKELPKTFELDLPKKEKKLNWLNEKLEIIKNQKPKYAYLLRKENNIF
LQYTLDSIPEIHSEYSGAVGIDRGVSHIAVYTFLDKDGKNERPFFLSSSGILRLKNLQKERDKF
LRKKHNKIRKKGNMRNIEQKINLILHEYSKQIVNFAKDKNAFIVFELLEKPKKSRERMSKKIQY
KLSQFTFKKLSDLVDYKAKREGIKVIYVEPAYTSKDCSHCGERVNTQRPFNGNFSLFKCNKCGI

VLNSDYNASLNIARKGLNISAN (서열번호 : 17)

>CasZb.11|134|3300013127.a:Ga0172365_10004421_00828
Zr6.cluA_2b.c10_like protein
MAEEKFFFCEKCNKDIKIPKNYINKQGAEEKARAKHEHRVHALILGIKFKIYPKKEDISKLNDY
FDEYAKAVTFTAKIVDKLKAPFLFAGKRDKDTSKKKWVFPVDKCSFCKEKTEINYRTKQGKNIC
NSCYLTEFGEQGLLEKIYATKGRKVSSSFNLFNSTKKLTGTHNNYVVKESLQLLDALKKQRSKR
LKKLSNTRRKLKQFEEMFEKEDKRFQLPLKEKQRELRFIHVSQKDRATEFKGYTMNKIKSKIKV
LRRNIEREQRSLNRKSPVFFRGTRIRLSPSVQFDDKDNKIKLTLSKELPKEYSFSGLNVANEHG
RKFFAEKLKLIKENKSKYAYLLRRQVNKNNKKPIYDYYLQYTVEFLPNIITNYNGILGIDRGIN
TLACIVLLENKKEKPSFVKFFSGKGILNLKNKRRKQLYFLKGVHNKYRKQQKIRPIEPRIDQIL
HDISKQIIDLAKEKRVAISLEQLEKPQKPKFRQSRKAKYKLSQFNFKTLSNYIDYKAKKEGIRV
IYIAPEMTSQNCSRCAMKNDLHVNTQRPYKNTSSLFKCNKCGVELNADYNAAFNIAQKGLKILN

S (서열번호 : 18)

도면1g

>CasZc.1|CG08_land_8_20_14_0.20_scaffold_1609_10 Tax=CG_Micra_03
id=133381002 partial=0 scaf=CG08_land_8_20_14_0.20_scaffold_1609
bin=CG08_land_8_20_14_0_20_UNK sample=CG08_land_8_20_14_0_20
proj=Crystal Geyser
mISLKLKLLPDEEQKKLLDEMFWKWASICTRVGFGRADKEDLKPPKDAEGVWFSLTQLNQANTD
INDLREAMKHQKHRLEYEKNRLEAQRDDTQDALKNPDRREISTKRKDLFRPKASVEKGFLKLKY
HQERYWVRRLKEINKLIERKTKTLIKIEKGRIKFKATRITLHQGSFKIRFGDKPAFLIKALSGK
NQIDAPFVVVPEQPICGSVVNSKKYLDEITTNFLAYSVNAMLFGLSRSEEMLLKAKRPEKIKKK
EEKLAKKQSAFENKKKELQKLLGRELTQQEEAIIEETRNQFFQDFEVKITKQYSELLSKIANEL
KQKNDFLKVNKYPILLRKPLKKAKSKKINNLSPSEWKYYLQFGVKPLLKQKSRRKSRNVLGIDR
GLKHLLAVTVLEPDKKTFVWNKLYPNPITGWKWRRRKLLRSLKRLKRRIKSQKHETIHENQTRK
KLKSLQGRIDDLLHNISRKIVETAKEYDAVIVVEDLQSMRQHGRSKGNRLKTLNYALSLFDYAN
VMQLIKYKAGIEGIQIYDVKPAGTSQNCAYCLLAQRDSHEYKRSQENSKIGVCLNPNCQNHKKQ

IDADLNAARVIASCYALKINDSQPFGTRKRFKKRTTN (서열번호 : 19)

>CasZc.2|CG_4_10_14_0.8_um_filter_scaffold_20762_4
Tax=CG_Micra_03 id=144037526 partial=0
scaf=CG_4_10_14_0.8_um_filter_scaffold_20762
bin=CG_4_10_14_0_8_um_filter_cor_UNK
sample=CG_4_10_14_0_8_um_filter_cor proj=Crystal Geyser
mETLSLKLKLNPSKEQLLVLDKMFWKWASICTRLGLKKAEMSDLEPPKDAEGVWFSKTQLNQAN
TDVNDLRKAMQHQGKRIEYELDKVENRRNEIQEMLEKPDRRDISPNRKDLFRPKAAVEKGYLKL
KYHKLGYWSKELKTANKLIERKRKTLAKIDAGKMKFKPTRISLHTNSFRIKFGEEPKIALSTTS
KHEKIELPLITSLQRPLKTSCAKKSKTYLDAAILNFLAYSTNAALFGLSRSEEMLLKAKKPEKI
EKRDRKLATKRESFDKKLKTLEKLLERKLSEKEKSVFKRKQTEFFDKFCITLDETYVEALHRIA
EELVSKNKYLEIKKYPVLLRKPESRLRSKKLKNLKPEDWTYYIQFGFQPLLDTPKPIKTKTVLG
IDRGVRHLLAVSIFDPRTKTFTFNRLYSNPIVDWKWRRRKLLRSIKRLKRRLKSEKHVHLHENQ
FKAKLRSLEGRIEDHFHNLSKEIVDLAKENNSVIVVENLGGMRQHGRGRGKWLKALNYALSHFD
YAKVMQLIKYKAELAGVFVYDVAPAGTSINCAYCLLNDKDASNYTRGKVINGKKNTKIGECKTC

KKEFDADLNAARVIALCYEKRLNDPQPFGTRKQFKPKKP (서열번호 : 20)

도면1h

>CasZc.3|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_2 IS605
OrfB family transposase id=130740989 partial=0
scaf=CG22_combo_CG10-13_8_21_14_all_scaffold_2003
bin=CG22_combo_CG10-13_8_21_14_all_UNK sample=CG22_combo_CG10-
13_8_21_14_all proj=Crystal Geyser
mKALKLQLIPTRKQYKILDEMFWKWASLANRVSQKGESKETLAPKKDIQKIQFNATQLNQIEKD
IKDLRGAMKEQQKQKERLLLQIQERRSTISEMLNDDNNKERDPHRPLNFRPKGWRKFHTSKHWV
GELSKILRQEDRVKKTIERIVAGKISFKPKRIGIWSSNYKINFFKRKISINPLNSKGFELTLMT
EPTQDLIGKNGGKSVLNNKRYLDDSIKSLLMFALHSRFFGLNNTDTYLLGGKINPSLVKYYKKN
QDMGEFGREIVEKFERKLKQEINEQQKKIIMSQIKEQYSNRDSAFNKDYLGLINEFSEVFNQRK
SERAEYLLDSFEDKIKQIKQEIGESLNISDWDFLIDEAKKAYGYEEGFTEYVYSKRYLEILNKI
VKAVLITDIYFDLRKYPILLRKPLDKIKKISNLKPDEWSYYIQFGYDSINPVQLMSTDKFLGID
RGLTHLLAYSVFDKEKKEFIINQLEPNPIMGWKWKLRKVKRSLQHLERRIRAQKMVKLPENQMK
KKLKSIEPKIEVHYHNISRKIVNLAKDYNASIVVESLEGGGLKQHGRKKNARNRSLNYALSLFD
YGKIASLIKYKADLEGVPMYEVLPAYTSQQCAKCVLEKGSFVDPEIIGYVEDIGIKGSLLDSLF
EGTELSSIQVLKKIKNKIELSARDNHNKEINLILKYNFKGLVIVRGQDKEEIAEHPIKEINGKF
AILDFVYKRGKEKVGKKGNQKVRYTGNKKVGYCSKHGQVDADLNASRVIALCKYLDINDPILFG

EQRKSFK (서열번호 : 21)

>CasZc.4|rifcsphigho2_01_scaffold_82367_2 Tax=CG_Micra_03
id=56177664 partial=0 scaf=rifcsphigho2_01_scaffold_82367
bin=RifCSPhighO2_01_full_UNK sample=RIFCSPHIGHO2_01_FULL
proj=RifleCSP2_HighO2
mVTRAIKLKLDPTKNQYKLLNEMFWKWASLANRFSQKGASKETLAPKDGTQKIQFNATQLNQIK
KDVDDLRGAMEKQGKQKERLLIQIQERLLTISEILRDDSKKEKDPHRPQNFRPFGWRRFHTSAY
WSSEASKLTRQVDRVRRTIERIKAGKINFKPKRIGLWSSTYKINFLKKKINISPLKSKSFELDL
ITEPQQKIIGKEGGKSVANSKKYLDDSIKSLLIFAIKSRLFGLNNKDKPLFENIITPNLVRYHK
KGQEQENFKKEVIKKFENKLKKEISQKQKEIIFSQIERQYENRDATFSEDYLRAISEFSEIFNQ
RKKERAKELLNSFNEKIRQLKKEVNGNISEEDLKILEVEAEKAYNYENGFIEWEYSEQFLGVLE
KIARAVLISDNYFDLKKYPILIRKPTNKSKKITNLKPEEWDYYIQFGYGLINSPMKIETKNFMG
IDRGLTHLLAYSIFDRDSEKFTINQLELNPIKGWKWKLRKVKRSLQHLERRMRAQKGVKLPENQ
MKKRLKSIEPKIESYYHNLSRKIVNLAKANNASIVVESLEGGGLKQHGRKKNSRHRALNYALSL
FDYGKIASLIKYKSDLEGVPMYEVLPAYTSQQCAKCVLKKGSFVEPEIIGYIEEIGFKENLLTL
LFEDTGLSSVQVLKKSKNKMTLSARDKEGKMVDLVLKYNFKGLVISQEKKKEEIVEFPIKEIDG
KFAVLDSAYKRGKERISKKGNQKLVYTGNKKVGYCSVHGQVDADLNASRVIALCKYLGINEPIV

FGEQRKSFK (서열번호 : 22)

도면1i

>CasZc.5|gwc1_scaffold_8732_6 Transposase, IS605 OrfB family
id=3520800 partial=0 scaf=gwc1_scaffold_8732 bin=GWC1 Unbinned
sample=GWC1 proj=Rifle Groundwater Metagenome
LDLITEPIQPHKSSSLRSKEFLEYQISDFLNFSLHSLFFGLASNEGPLVDFKIYDKIVIPKPEE
RFPKKESEEGKKLDSFDKRVEEYYSDKLEKKIERKLNTEEKNVIDREKTRIWGEVNKLEEIRSI
IDEINEIKKQKHISEKSKLLGEKWKKVNNIQETLLSQEYVSLISNLSDELTNKKKELLAKKYSK
FDDKIKKIKEDYGLEFDENTIKKEGEKAFLNPDKFSKYQFSSSYLKLIGEIARSLITYKGFLDL
NKYPIIFRKPINKVKKIHNLEPDEWKYYIQFGYEQINNPKLETENILGIDRGLTHILAYSVFEP
RSSKFILNKLEPNPIEGWKWKLRKLRRSIQNLERRWRAQDNVKLPENQMKKNLRSIEDKVENLY
HNLSRKIVDLAKEKNACIVFEKLEGQGMKQHGRKKSDRLRGLNYKLSLFDYGKIAKLIKYKAEI
EGIPIYRIDSAYTSQNCAKCVLESRRFAQPEEISCLDDFKEGDNLDKRILEGTGLVEAKIYKKL
LKEKKEDFEIEEDIAMFDTKKVIKENKEKTVILDYVYTRRKEIIGTNHKKNIKGIAKYTGNTKI
GYCMKHGQVDADLNASRTIALCKNFDINNPEIWK (서열번호 : 23)

>CasZc.6|109|3300010293.a:Ga0116204_1008574_00822
Zr6.cluA_2a.c10_like protein
MSDESLVSSEDKLAIKIKIVPNAEQAKMLDEMFKKWSSICNRISRGKEDIETLRPDEGKELQFN
STQLNSATMDVSDLKKAMARQGERLEAEVSKLRGRYETIDASLRDPSRRHTNPQKPSSFYPSDW
DISGRLTPRFHTARHYSTELRKLKAKEDKMLKTINKIKNGKIVFKPKRITLWPSSVNMAFKGSR
LLLKPFANGFEMELPIVISPQKTADGKSQKASAEYMRNALLGLAGYSINQLLFGMNRSQKMLAN
AKKPEKVEKFLEQMKNKDANFDKKIKALEGKWLLDRKLKESEKSSIAVVRTKFFKSGKVELNED
YLKLLKHMANEILERDGFVNLNKYPILSRKPMKRYKQKNIDNLKPNMWKYYIQFGYEPIFERKA
SGKPKNIMGIDRGLTHLLAVAVFSPDQQKFLFNHLESNPIMHWKWKLRKIRRSIQHMERRIRAE
KNKHIHEAQLKKRLGSIEEKTEQHYHIVSSKIINWAIEYEAAIVLESLSHMKQRGGKKSVRTRA
LNYALSLFDYEKVARLITYKARIRGIPVYDVLPGMTSKTCATCLLNGSQGAYVRGLETTKAAGK
ATKRKNMKIGKCMVCNSSENSMIDADLNAARVIAICKYKNLNDPQPAGSRKVFKRF
(서열번호 : 24)

>CasZc.7|126|3300005573.a:Ga0078972_1001015_00056
Zr6.cluA_2a.c10_like protein
MLALKLKIMPTEKQAEILDAMFWKWASICSRIAKMKKKVSVKENKKELSKKIPSNSDIWFSKTQ
LCQAEVDVGDHKKALKNFEKRQESLLDELKYKVKAINEVINDESKREIDPNNPSKFRIKDSTKK
GNLNSPKFFTLKKWQKILQENEKRIKKKESTIEKLKRGNIFFNPTKISLHEEEYSINFGSSKLL
LNCFYKYNKKSGINSDQLENKFNEFQNGLNIICSPLQPIRGSSKRSFEFIRNSIINFLMYSLYA
KLFGIPRSVKALMKSNKDENKLKLEEKLKKKKSSFNKTVKEFEKMIGRKLSDNESKILNDESKK
FFEIIKSNNKYIPSEEYLKLLKDISEEIYNSNIDFKPYKYSILIRKPLSKFKSKKLYNLKPTDY
KYYLQLSYEPFSKQLIATKTILGIDRGLKHLLAVSVFDPSQNKFVYNKLIKNPVFKWKKRYHDL
KRSIRNRERRIRALTGVHIHENQLIKKLKSMKNKINVLYHNVSKNIVDLAKKYESTIVLERLEN
LKQHGRSKGKRYKKLNYVLSNFDYKKIESLISYKAKKEGVPVSNINPKYTSKTCAKCLLEVNQL
SELKNEYNRDSKNSKIGICNIHGQIDADLNAARVIALCYSKNLNEPHFK
(서열번호 : 25)

도면1j

>CasZd.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_25 IS605
OrfB family transposase id=132383911 partial=0
scaf=CG10_big_fil_rev_8_21_14_0.10_scaffold_4477
bin=CG10_big_fil_rev_8_21_14_0_10_test2
sample=CG10_big_fil_rev_8_21_14_0_10 proj=Crystal Geyser
VINLFGYKFALYPNKTQEELLNKHLGECGWLYNKAIEQNEYYKADSNIEEAQKKFELLPDKNSD
EAKVLRGNISKDNYVYRTLVKKKKSEINVQIRKAVVLRPAETIRNLAKVKKKGLSVGRLKFIPI
REWDVLPFKQSDQIRLEENYLILEPYGRLKFKMHRPLLGKPKTFCIKRTATDRWTISFSTEYDD
SNMRKNDGGQVGIDVGLKTHLRLSNENPDEDPRYPNPKIWKRYDRRLTILQRRISKSKKLGKNR
TRLRLRLSRLWEKIRNSRADLIQNETYEILSENKLIAIEDLNVKGMQEKKDKKGRKGRTRAQEK
GLHRSISDAAFSEFRRVLEYKAKRFGSEVKPVSAIDSSKECHNCGNKKGMPLESRIYECPKCGL
KIDRDLNSAKVILARATGVRPGSNARADTKISATAGASVQTEGTVSEDFRQQMETSDQKPMQGE
GSKEPPMNPEHKSSGRGSKHVNIGCKNKVGLYNEDENSRSTEKQIMDENRSTTEDMVEIGALHS
PVLTT (서열번호 : 26)

>CasZd.2|053|3300001245.a:JGI12048J13642_10201286_01511
Zr6.c9.inclusive
MIASIDYEAVSQALIVFEFKAKGKDSQYQAIDEAIRSYRFIRNSCLRYWMDNKKVGKYDLNKYC
KVLAKQYPFANKLNSQARQSAAECSWSAISRFYDNCKRKVSGKKGFPKFKKHARSVEYKTSGWK
LSENRKAITFTDKNGIGKLKLKGTYDLHFSQLEDMKRVRLVRRADGYYVQFCISVDVKVETEPT
GKAIGLDVGIKYFLADSSGNTIENPQFYRKAEKKLNRANRRKSKKYIRGVKPQSKNYHKARCRY
ARKHLRVSRQRKEYCKRVAYCVIHSNDVVAYEDLNVKGMVKNRHLAKSISDVAWSTFRHWLEYF
AIKYGKLTIPVAPHNTSQNCSNCDKKVPKSLSTRTHICHHCGYSEDRDVNAAKNILKKALSTVG
QTGSLKLGEIEPLLVLEQSCTRKFDL (서열번호 : 27)

>CasZe.1|19ft_2_nophage_noknown_scaffold_0_545 \N id=8269792
partial=0 scaf=19ft_2_nophage_noknown_scaffold_0 bin=19ft_2_UNK
sample=19FT_2_THINNED proj=Rifle Sediment CSP1
LAEENTLHLTLAMSLPLNDLPENRTRSELWRRQWLPQKKLSLLLGVNQSVRKAAADCLRWFEPY
QELLWWEPTDPDGKKLLDKEGRPIKRTAGHMRVLRKLEEIAPFRGYQLGSAVKNGLRHKVADLL
LSYAKRKLDPQFTDKTSYPSIGDQFPIVWTGAFVCYEQSITGQLYLYLPLFPRGSHQEDITNNY
DPDRGPALQVFGEKEIARLSRSTSGLLLPLQFDKWGEATFIRGENNPPTWKATHRRSDKKWLSE
VLLREKDFQPKRVELLVRNGRIFVNVACEIPTKPLLEVENFMGVSFGLEHLVTVVVINRDGNVV
HQRQEPARRYEKTYFARLERLRRRGGPFSQELETFHYRQVAQIVEEALRFKSVPAVEQVGNIPK
GRYNPRLNLRLSYWPFGKLADLTSYKAVKEGLPKPYSVYSATAKMLCSTCGAANKEGDQPISLK
GPTVYCGNCGTRHNTGFNTALNLARRAQELFVKGVVAR (서열번호 : 28)

도면1k

>CasZe.2|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_49 Tax=RIFCSPHIGHO2_01_FULL_RIF_OD1_04_46_36_curated id=87353177 partial=0 scaf=RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646 bin=RIFCSPHIGHO2_01_FULL_OD1_Andersenbacteria_46_36_curated sample=2500-curated-genomes-non-redundant proj=2500 Genomes
mSQSLLKWHDMAGRDKDASRSLQKSAVEGVLLHLTASHRVALEMLEKSVSQTVAVTMEAAQQRL
VIVLEDDPTKATSRKRVISADLQFTREEFGSLPNWAQKLASTCPEIATKYADKHINSIRIAWGV
AKESTNGDAVEQKLQWQIRLLDVTMFLQQLVLQLADKALLEQIPSSIRGGIGQEVAQQVTSHIQ
LLDSGTVLKAELPTISDRNSELARKQWEDAIQTVCTYALPFSRERARILDPGKYAAEDPRGDRL
INIDPMWARVLKGPTVKSLPLLFVSGSSIRIVKLTLPRKHAAGHKHTFTATYLVLPVSREWINS
LPGTVQEKVQWWKKPDVLATQELLVGKGALKKSANTLVIPISAGKKRFFNHILPALQRGFPLQW
QRIVGRSYRRPATHRKWFAQLTIGYTNPSSLPEMALGIHFGMKDILWWALADKQGNILKDGSIP
GNSILDFSLQEKGKIERQQKAGKNVAGKKYGKSLLNATYRVVNGVLEFSKGISAEHASQPIGLG
LETIRFVDKASGSSPVNARHSNWNYGQLSGIFANKAGPAGFSVTEITLKKAQRDLSDAEQARVL
AIEATKRFASRIKRLATKRKDDTLFV (서열번호 : 29)

>CasZe.3|rifcsphigho2_01_scaffold_10981_6 Tax=RIFCSPHIGHO2_01_FULL_OD1_47_10b_curated id=55513311 partial=0 scaf=rifcsphigho2_01_scaffold_10981 bin=RIFCSPHIGHO2_01_FULL_OD1_47_10b sample=2500-genomes proj=2500 Genomes
VEPVEKERFYYRTYTFRLDGQPRTQNLTTQSGWGLLTKAVLDNTKHYWEIVHHARIANQPIVFE
NPVIDEQGNPKLNKLGQPRFWKRPISDIVNQLRALFENQNPYQLGSSLIQGTYWDVAENLASWY
ALNKEYLAGTATWGEPSFPEPHPLTEINQWMPLTFSSGKVVRLLKNASGRYFIGLPILGENNPC
YRMRTIEKLIPCDGKGRVTSGSLILFPLVGIYAQQHRRMTDICESIRTEKGKLAWAQVSIDYVR
EVDKRRRMRRTRKSQGWIQGPWQEVFILRLVLAHKAPKLYKPRCFAGISLGPKTLASCVILDQD
ERVVEKQQWSGSELLSLIHQGEERLRSLREQSKPTWNAAYRKQLKSLINTQVFTIVTFLRERGA
AVRLESIARVRKSTPAPPVNFLLSHWAYRQITERLKDLAIRNGMPLTHSNGSYGVRFTCSQCGA
TNQGIKDPTKYKVDIESETFLCSICSHREIAAVNTATNLAKQLLDE (서열번호 : 30)

>CasZe.4b|rifcsplowo2_02_scaffold_57876_2 hypothetical protein Tax=GWA2_OD1_56_11 id=66016044 partial=0 scaf=rifcsplowo2_02_scaffold_57876 bin=RifCSPlowO2_02_full_UNK sample=RIFCSPLOWO2_02_FULL proj=RifleCSP2_LowO2
mNDTETSETLTSHRTVCAHLHVVGETGSLPRLVEAALAELITLNGRATQALLSLAKNGLVLRRD
KEENLIAAELTLPCRKNKYADVAAKAGEPILATRINNKGKLVTKKWYGEGNSYHIVRFTPETGM
FTVRVFDRYAFDEELLHLHSEVVFGSDLPKGIKAKTDSLPANFLQAVFTSFLELPFQGFPDIVV
KPAMKQAAEQLLSYVQLEAGENQQAEYPDTNERDPELRLVEWQKSLHELSVRTEPFEFVRARDI
DYYAETDRRGNRFVNITPEWTKFAESPFARRLPLKIPPEFCILLRRKTEGHAKIPNRIYLGLQI
FDGVTPDSTLGVLATAEDGKLFWWHDHLDEFSNLEGKPEPKLKNKPQLLMVSLEYDREQRFEES
VGGDRKICLVTLKETRNFRRGWNGRILGIHFQHNPVITWALMDHDAEVLEKGFIEGNAFLGKAL
DKQALNEYLQKGGKWVGDRSFGNKLKGITHTLASLIVRLAREKDAWIALEEISWVQKQSADSVA
NHEIVEQPHHSLTR (서열번호 : 31)

도면11

>CasZe.4|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_33 hypothetical protein Tax=GWA2_OD1_56_11 id=88324948 partial=0
scaf=RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495
bin=RIFCSPLOWO2_01_FULL_OD1_Taylorbacteria_45_34b_curated
sample=2500-curated-genomes proj=2500 Genomes
mNDTETSETLTSHRTVCAHLHVVGETGSLPRLVEAALAELITLNGRATQALLSLAKNGLVLRRD
KEENLIAAELTLPCRKNKYADVAAKAGEPILATRINNKGKLVTKKWYGEGNSYHIVRFTPETGM
FTVRVFDRYAFDEELLHLHSEVVFGSDLPKGIKAKTDSLPANFLQAVFTSFLELPFQGFPDIVV
KPAMKQAAEQLLSYVQLEAGENQQAEYPDTNERDPELRLVEWQKSLHELSVRTEPFEFVRARDI
DYYAETDRRGNRFVNITPEWTKFAESPFARRLPLKIPPEFCILLRRKTEGHAKIPNRIYLGLQI
FDGVTPDSTLGVLATAEDGKLFWWHDHLDEFSNLEGKPEPKLKNKPQLLMVSLEYDREQRFEES
VGGDRKICLVTLKETRNFRRGRHGHTRTDRLPAGNTLWRADFATSAEVAAPKWNGRILGIHFQH
NPVITWALMDHDAEVLEKGFIEGNAFLGKALDKQALNEYLQKGGKWVGDRSFGNKLKGITHTLA
SLIVRLAREKDAWIALEEISWVQKQSADSVANRRFSMWNYSRLATLIEWLGTDIATRDCGTAAP
LAHKVSDYLTHFTCPECGACRKAGQKKEIADTVRAGDILTCRKCGFSGPIPDNFIAEFVAKKAL
ERMLKKKPV (서열번호 : 32)

>CasZf.1|rifcsphigho2_01_scaffold_566_121
Tax=RIFCSPHIGHO2_01_FULL_RIF_OD1_10_44_22b_curated id=55167017
partial=0 scaf=rifcsphigho2_01_scaffold_566
bin=RifCSPhighO2_01_full_UNK sample=RIFCSPHIGHO2_01_FULL
proj=RifleCSP2_HighO2
mAKRNFGEKSEALYRAVRFEVRPSKEELSILLAVSEVLRMLFNSALAERQQVFTEFIASLYAEL
KSASVPEEISEIRKKLREAYKEHSISLFDQINALTARRVEDEAFASVTRNWQEETLDALDGAYK
SFLSLRRKGDYDAHSPRSRDSGFFQKIPGRSGFKIGEGRIALSCGAGRKLSFPIPDYQQGRLAE
TTKLKKFELYRDQPNLAKSGRFWISVVYELPKPEATTCQSEQVAFVALGASSIGVVSQRGEEVI
ALWRSDKHWVPKIEAVEERMKRRVKGSRGWLRLLNSGKRRMHMISSRQHVQDEREIVDYLVRNH
GSHFVVTELVVRSKEGKLADSSKPERGGSLGLNWAAQNTGSLSRLVRQLEEKVKEHGGSVRKHK
LTLTEAPPARGAENKLWMARKLRESFLKEV (서열번호 : 33)

>CasZf.2|rifcsplowo2_01_scaffold_81231_1 transposase, IS605 OrfB family, central region id=58737995 partial=0
scaf=rifcsplowo2_01_scaffold_81231 bin=RifCSPlowO2_01_full_UNK
sample=RIFCSPLOWO2_01_FULL proj=RifleCSP2_LowO2
LAKNDEKELLYQSVKFEIYPDESKIRVLTRVSNILVLVWNSALGERRARFELYIAPLYEELKKF
PRKSAESNALRQKIREGYKEHIPTFFDQLKKLLTPMRKEDPALLGSVPRAYQEETLNTLNGSFV
SFMTLRRNNDMDAKPPKGRAEDRFHEISGRSGFKIDGSEFVLSTKEQKLRFPIPNYQLEKLKEA
KQIKKFTLYQSRDRRFWISIAYEIELPDQRPFNPEEVIYIAFGASSIGVISPEGEKVIDFWRPD
KHWKPKIKEVENRMRSCKKGSRAWKKRAAARRKMYAMTQRQQKLNHREIVASLLRLGFHFVVTE
YTVRSKPGKLADGSNPKRGGAPQGFNWSAQNTGSFGEFILWLKQKVKEQGGTVQTFRLVLGQSE
RPEKRGRDNKIEMVRLLREKYLESQTIVV (서열번호 : 34)

도면1m

>CasZf.3|rifcsphigho2_01_scaffold_4702_81 IS605 OrfB family
transposase id=55366684 partial=0
scaf=rifcsphigho2_01_scaffold_4702 bin=RifCSPhighO2_01_full_UNK
sample=RIFCSPHIGHO2_01_FULL proj=RifleCSP2_HighO2
mAKGKKKEGKPLYRAVRFEIFPTSDQITLFLRVSKNLQQVWNEAWQERQSCYEQFFGSIYERIG
QAKKRAQEAGFSEVWENEAKKGLNKKLRQQEISMQLVSEKESLLQELSIAFQEHGVTLYDQING
LTARRIIGEFALIPRNWQEETLDSLDGSFKSFLALRKNGDPDAKPPRQRVSENSFYKIPGRSGF
KVSNGQIYLSFGKIGQTLTSVIPEFQLKRLETAIKLKKFELCRDERDMAKPGRFWISVAYEIPK
PEKVPVVSKQITYLAIGASRLGVVSPKGEFCLNLPRSDYHWKPQINALQERLEGVVKGSRKWKK
RMAACTRMFAKLGHQQKQHGQYEVVKKLLRHGVHFVVTELKVRSKPGALADASKSDRKGSPTGP
NWSAQNTGNIARLIQKLTDKASEHGGTVIKRNPPLLSLEERQLPDAQRKIFIAKKLREEFLADQ
K (서열번호 : 35)

>CasZg.1|rifcsp13_1_sub10_scaffold_3_54 \N id=12787801 partial=0
scaf=rifcsp13_1_sub10_scaffold_3 bin=RifCSP13_1_UNK
sample=RIFCSP13_1 proj=Rifle Sediment CSP2
mAKREKKDDVVLRGTKMRIYPTDRQVTLMDMWRRRCISLWNLLLNLETAAYGAKNTRSKLGWRS
IWARVVEENHAKALIVYQHGKCKKDGSFVLKRDGTVKHPPRERFPGDRKILLGLFDALRHTLDK
GAKCKCNVNQPYALTRAWLDETGHGARTADIIAWLKDFKGECDCTAISTAAKYCPAPPTAELLT
KIKRAAPADDLPVDQAILLDLFGALRGGLKQKECDHTHARTVAYFEKHELAGRAEDILAWLIAH
GGTCDCKIVEEAANHCPGPRLFIWEHELAMIMARLKAEPRTEWIGDLPSHAAQTVVKDLVKALQ
TMLKERAKAAAGDESARKTGFPKFKKQAYAAGSVYFPNTTmFFDVAAGRVQLPNGCGSMRCEIP
RQLVAELLERNLKPGLVIGAQLGLLGGRIWRQGDRWYLSCQWERPQPTLLPKTGRTAGVKIAAS
IVFTTYDNRGQTKEYPMPPADKKLTAVHLVAGKQNSRALEAQKEKEKKLKARKERLRLGKLEKG
HDPNALKPLKRPRVRRSKLFYKSAARLAACEAIERDRRDGFLHRVTNEIVHKFDAVSVQKMSVA
PMMRRQKQKEKQIESKKNEAKKEDNGAAKKPRNLKPVRKLLRHVAMARGRQFLEYKYNDLRGPG
SVLIADRLEPEVQECSRCGTKNPQMKDGRRLLRCIGVLPDGTDCDAVLPRNRNAARNAEKRLRK
HREAHNA (서열번호 : 36)

>CasZg.2|033|3300009991.a:Ga0105042_100140_01533
Zr6.c9.inclusive
MNEVLPIPAVGEDAADTIMRGSKMRIYPSVRQAATMDLWRRRCIQLWNLLLELEQAAYSGENRR
TQIGWRSIWATVVEDSHAEAVRVAREGKKRKDGTFRKAPSGKEIPPLDPAMLAKIQRQMNGAVD
VDPKTGEVTPAQPRLFMWEHELQKIMARLKQAPRTHWIDDLPSHAAQSVVKDLIKALQAMLRER
KKRASGIGGRDTGFPKFKKNRYAAGSVYFANTQLRFEAKRGKAGDPDAVRGEFARVKLPNGVGW
MECRMPRHINAAHAYAQATLMGGRIWRQGENWYLSCQWKMPKPAPLPRAGRTAAIKIAAAIPIT
TVDNRGQTREYAMPPIDRERIAAHAAAGRAQSRALEARKRRAKKREAYAKKRHAKKLERGIAAK
PPGRARIKLSPGFYAAAAKLAKLEAEDANAREAWLHEITTQIVRNFDVIAVPRMEVAKLMKKPE
PPEEKEEQVKAPWQGKRRSLKAARVMMRRTAMALIQTTLKYKAVDLRGPQAYEEIAPLDVTAAA
CSGCGVLKPEWKMARAKGREIMRCQEPLPGGKTCNTVLTYTRNSARVIGRELAVRLAERQKA
(서열번호 : 37)

도면1n

>CasZh.1|rifcsp2_19_4_full_scaffold_168_120 transposase
id=102318404 partial=0 scaf=rifcsp2_19_4_full_scaffold_168
bin=RifCSP19_4_full_UNK sample=RIFCSP19_4_FULL proj=Rifle
Sediment CSP1
mTTQKTYNFCFYDQRFFELSKEAGEVYSRSLEEFWKIYDETGVWLSKFDLQKHMRNKLERKLLH
SDSFLGAMQQVHANLASWKQAKKVVPDACPPRKPKFLQAILFKKSQIKYKNGFLRLTLGTEKEF
LYLKWDINIPLPIYGSVTYSKTRGWKINLCLETEVEQKNLSENKYLSIDLGVKRVATIFDGENT
ITLSGKKFMGLMHYRNKLNGKTQSRLSHKKKGSNNYKKIQRAKRKTTDRLLNIQKEMLHKYSSF
IVNYAIRNDIGNIIIGDNSSTHDSPNMRGKTNQKISQNPEQKLKNYIKYKFESISGRVDIVPEP
YTSRKCPHCKNIKKSSPKGRTYKCKKCGFIFDRDGVGAINIYNENVSFGQIISPGRIRSLTEPI
GMKFHNEIYFKSYVAA (서열번호 : 38)

>CasZi.1|RBG_13_scaffold_1401_19
Tax=RBG_13_Planctomycetes_46_10_curated id=16386779 partial=0
scaf=RBG_13_scaffold_1401 bin=RBG_13_Planctomycetes_46_10
sample=2500-genomes proj=2500 Genomes
mSVRSFQARVECDKQTMEHLWRTHKVFNERLPEIIKILFKMKRGECGQNDKQKSLYKSISQSIL
EANAQNADYLLNSVSIKGWKPGTAKKYRNASFTWADDAAKLSSQGIHVYDKKQVLGDLPGMMSQ
MVCRQSVEAISGHIELTKKWEKEHNEWLKEKEKWESEDEHKKYLDLREKFEQFEQSIGGKITKR
RGRWHLYLKWLSDNPDFAAWRGNKAVINPLSEKAQIRINKAKPNKKNSVERDEFFKANPEMKAL
DNLHGYYERNFVRRRKTKKNPDGFDHKPTFTLPHPTIHPRWFVFNKPKTNPEGYRKLILPKKAG
DLGSLEMRLLTGEKNKGNYPDDWISVKFKADPRLSLIRPVKGRRVVRKGKEQGQTKETDSYEFF
DKHLKKWRPAKLSGVKLIFPDKTPKAAYLYFTCDIPDEPLTETAKKIQWLETGDVTKKGKKRKK
KVLPHGLVSCAVDLSMRRGTTGFATLCRYENGKIHILRSRNLWVGYKEGKGCHPYRWTEGPDLG
HIAKHKREIRILRSKRGKPVKGEESHIDLQKHIDYMGEDRFKKAARTIVNFALNTENAASKNGF
YPRADVLLLENLEGLIPDAEKERGINRALAGWNRRHLVERVIEMAKDAGFKRRVFEIPPYGTSQ
VCSKCGALGRRYSIIRENNRREIRFGYVEKLFACPNCGYCANADHNASVNLNRRFLIEDSFKSY
YDWKRLSEKKQKEEIETIESKLMDKLCAMHKISRGSISK (서열번호 : 39)

도면10

>CasZi.2|026|3300009652.a:Ga0123330_1010394_01528
Zr6.c2c1.inclusive
MHLWRTHCVFNQRLPALLKRLFAMRRGEVGGNEAQRQVYQRVAQFVLARDAKDSVDLLNAVSLR
KRSANSAFKKKATISCNGQAREVTGEEVFAEAVALASKGVFAYDKDDMRAGLPDSLFQPLTRDA
VACMRSHEELVATWKKEYREWRDRKSEWEAEPEHALYLNLRPKFEEGEAARGGRFRKRAERDHA
YLDWLEANPQLAAWRRKAPPAVVPIDEAGKRRIARAKAWKQASVRAEEFWKRNPELHALHKIHV
QYLREFVRPRRTRRNKRREGFKQRPTFTMPDPVRHPRWCLFNAPQTSPQGYRLLRLPQSRRTVG
SVELRLLTGPSDGAGFPDAWVNVRFKADPRLAQLRPVKVPRTVTRGKNKGAKVEADGFRYYDDQ
LLIERDAQVSGVKLLFRDIRMAPFADKPIEDRLLSATPYLVFAVEIKDEARTERAKAIRFDETS
ELTKSGKKRKTLPAGLVSVAVDLDTRGVGFLTRAVIGVPEIQQTHHGVRLLQSRYVAVGQVEAR
ASGEAEWSPGPDLAHIARHKREIRRLRQLRGKPVKGERSHVRLQAHIDRMGEDRFKKAARKIVN
EALRGSNPAAGDPYTRADVLLYESLETLLPDAERERGINRALLRWNRAKLIEHLKRMCDDAGIR
HFPVSPFGTSQVCSKCGALGRRYSLARENGRAVIRFGWVERLFACPNPECPGRRPDRPDRPFTC
NSDHNASVNLHRVFALGDQAVAAFRALAPRDSPARTLAVKRVEDTLRPQLMRVHKLADAGVDSP
F (서열번호 : 40)

>CasZj.1|130|3300012532.a:Ga0137373_10000316_00700 c2c4
MATLVYRYGVRAHGSARQQDAVVSDPAMLEQLRLGHELRNALVGVQHRYEDGKRAVWSGFASVA
AADHRVTTGETAVAELEKQARAEHSADRTAATRQGTAESLKAARAAVKQARADRKAAMAAVAEQ
AKPKIQALGDDRDAEIKDLYRRFCQDGVLLPRCGRCAGDLRSDGDCTDCGAAHEPRKLYWATYN
AIREDHQTAVKLVEAKRKAGQPARLRFRRWTGDGTLTVQLQRMHGPACRCVTCAEKLTRRARKT
DPQAPAVAADPAYPPTDPPRDPALLASGQGKWRNVLQLGTWIPPGEWSAMSRAERRRVGRSHIG
WQLGGGRQLTLPVQLHRQMPADADVAMAQLTRVRVGGRHRMSVALTAKLPDPPQVQGLPPVALH
LGWRQRPDGSLRVATWACPQPLDLPPAVADVVVSHGGRWGEVIMPARWLADAEVPPRLLGRRDK
AMEPVLEALADWLEAHTEACTARMTPALVRRWRSQGRLAGLTNRWRGQPPTGSAEILTYLEAWR
IQDKLLWERESHLRRRLAARRDDAWRRVASWLARHAGVLVVDDADIAELRRRDDPADTDPTMPA
SAAQAARARAALAAPGRLRHLATITATRDGLGVHTVASAGLTRLHRKCGHQAQPDPRYAASAVV
TCPGCGNGYDQDYNAAMLMLDRQQQP (서열번호 : 41)

>CasZk.1|012|3300005602.a:Ga0070762_10001740_01506
Zr6.c9.inclusive
MSRVELHRAYKFRLYPTPAQVAELAEWERQLRRLYNLAHSQRLAAMQRHVRPKSPGVLKSECLS
CGAVAVAEIGTDGKAKKTVKHAVGCSVLECRSCGGSPDAEGRTAHTAACSFVDYYRQGREMTQL
LEEDDQLARVVCSARQETLRDLEKAWQRWHKMPGFGKPHFKKRIDSCRIYFSTPKSWAVDLGYL
SFTGVASSVGRIKIRQDRVWPGDAKFSSCHVVRDVDEWYAVFPLTFTKEIEKPKGGAVGINRGA
VHAIADSTGRVVDSPKFYARSLGVIRHRARLLDRKVPFGRAVKPSPTKYHGLPKADIDAAAARV
NASPGRLVYEARARGSIAAAEAHLAALVLPAPRQTSQLPSEGRNREARRFLALAHQRVRRQRE
WFLHNESAHYAQSYTKIAIEDWSTKEMTSSEPRDAEEMKRVTRARNRSILDVGWYELGRQIAYK
SEATGAEFAKVDPGLRETETHVPEAIVRERDVDVSGMLRGEAGISGTCSRCGGLLRASASGHAD
AECEVCLHVEVGDVNAAVNVLKRAMFPGAAPPSKEKAKVTIGIKGRKKKRAA
(서열번호 : 42)

도면1p

>CasZk.2|016|3300005921.a:Ga0070766_10011912_01491
Zr6.c9.inclusive
MSRVELHRAYKFRLYPTPVQVAELSEWERQLRRLYNLGHEQRLLTLTRHLRPKSPGVLKGECLS
CDSTQVQEVGADGRPKTTVRHAEQCPTLACRSCGALRDAEGRTAHTVACAFVDYYRQGREMTEL
LAADDQLARVVCSARQEVLRDLDKAWQRWRKMPGFGKPRFKRRTDSCRIYFSTPKAWKLEGGHL
SFTGAATTVGAIKMRQDRNWPASVQFSSCHVVRDVDEWYAVFPLTFVAEVARPKGGAVGINRGA
VHAIADSTGRVVDSPRYYARALGVIRHRARLFDRKVPSGHAVKPSPTKYRGLSAIEVDRVARAT
GFTPGRVVTEALNRGGVAYAECALAAIAVLGHGPERPLTSDGRNREKARKFLALAHQRVRRQRE
WFLHNESAHYARTYSKIAIEDWSTKEMTASEPQGEETRRVTRSRNRSILDVGWYELGRQLAYKT
EATGAEFAQVDPGLKETETNVPKAIADARDVDVSGMLRGEAGISGTCSKCGGLLRAPASGHADA
ECEICLNVEVGDVNAAVNVLKRAMFPGDAPPASGEKPKVSIGIKGRQKKKKAA
(서열번호 : 43)

>CasZk.3|106|3300009698.a:Ga0116216_10000905_01565 c2c9
MEAIATGMSPERRVELGILPGSVELKRAYKFRLYPMKVQQAELSEWERQLRRLYNLAHEQRLAA
LLRYRDWDFQKGACPSCRVAVPGVHTAACDHVDYFRQAREMTQLLEVDAQLSRVICCARQEVLR
DLDKAWQRWRKKLGGRPRFKRRTDSCRIYLSTPKHWEIAGRYLRLSGLASSVGEIRIEQDRAFP
EGALLSSCSIVRDVDEWYACLPLTFTQPIERAPHRSVGLNRGVVHALADSDGRVVDSPKFFERA
LATVQKRSRDLARKVSGSRNAHKARIKLAKAHQRVRRQRAAFLHQESAYYSKGFDLVALEDMSV
RKMTATAGEAPEMGRGAQRDLNRGILDVGWYELARQIDYKRLAHGGELLRVDPGQTTPLACVTE
EQPARGISSACAVCGIPLARPASGNARMRCTACGSSQVGDVNAAENVLTRALSSAPSGPKSPKA
SIKIKGRQKRLGTPANRAGEASGGDPPVRGPVEGGTLAYVVEPVSESQSDT
(서열번호 : 44)

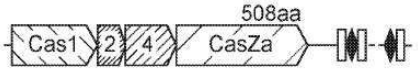
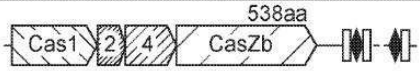
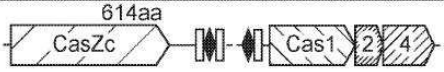
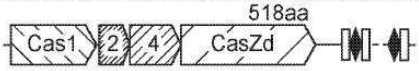
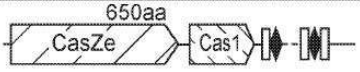
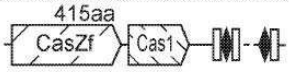
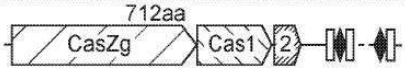
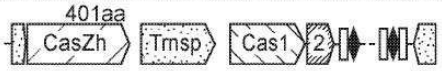
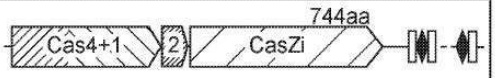
>CasZk.4|072|3300006028.a:Ga0070717_10000077_00263
Zr6.c9.inclusive
MTVRTYKYRAYPTPEQAEALTSWLRFASQLYNAALEHRKNAWGRHDAHGRGFRFWDGDAAPRKK
SDPPGRWVYRGGGGAHISKNDQGKLLTEFRREHAELLPPGMPALVQHEVLARLERSMAAFFQRA
TKGQKAGYPRWRSEHRYDSLTFGLTSPSKERFDPETGESLGRGKTVGAGTYHNGDLRLTGLGEL
RILEHRRIPMGAIPKSVIVRRSGKRWFVSIAMEMPSVEPAASGRPAVGLDMGVVTWGTAFTADT
SAAAALVADLRRMATDPSDCRRLEELEREAAQLSEVLAHCRARGLDPARPRRCPKELTKLYRRS
LHRLGELDRACARIRRRLQAAHDIAEPVPDEAGSAVLIEGSNAGMRHARRVARTQRRVARRTRA
GHAHSNRRKKAVQAYARAKERERSARGDHRHKVSRALVRQFEEISVEALDIKQLTVAPEHNPDP
QPDLPAHVQRRRNRGELDAAWGAFFAALDYKAADAGGRVARKPAPHTTQECARCGTLVPKPISL
RVHRCPACGYTAPRTVNSARNVLQRPLEEPGRAGPSGANGRGVPHAVA (서열번호 : 45)

도면1q

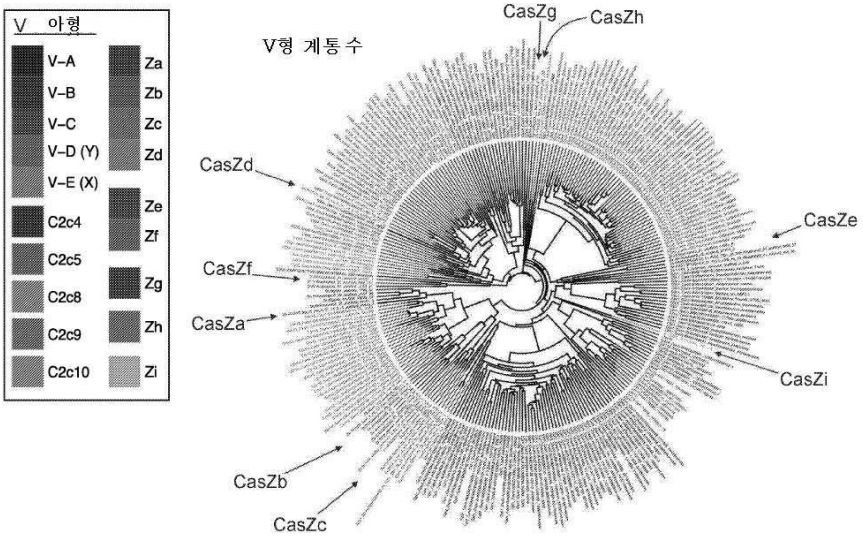
>CasZl.1|004|3300001256.a:JGI12210J13797_10004690_01983 c2c9
MNCRYRYRIYPTPGQRQSLARLFGCVRVVWNDALFLCRQSEKLPKNSELQKLCITQAKKTEARG
WLGQVSAIPLQQSVADLGVAFKNFFQSRSGKRKGKKVNPPRVKRRNNRQGARFTRGGFKVKTSK
VYLARIGDIKIKWSRPLPSEPSSVTVIKDCAGQYFLSFVVEVKPEIKPPKNPSIGIDLGLKTFA
SCSNGEKIDSPDYSRLYRKLKRCQRRLAKRQRGSKRRERMRVKVAKLNAQIRDKRKDFLHKLST
KVVNENQVIALEDLNVGGMLKNRKLSRAISQAGWYEFRSLCEGKAEKHNRDFRVISRWEPTSQV
CSECGYRWGKIDLSVRSIVCINCGVEHDRDDNASVNIEQAGLKVGVGHTHDSKRTGSACKTSNG
AVCVEPSTHREYVQLTLFDW (서열번호 : 46)

>CasZl.2|014|3300005660.a:Ga0073904_10021651_01988 c2c9
MKSRWTFRCYPTPEQEQHLARTFGCVRFVWNWALRARTDAFRAGERIGYPATDKALTLLKQQPE
TVWLNEVSSVCLQQALRDLQVAFSNFFDKRAAHPSFKRKEARQSANYTERGFSFDHERRILKLA
KIGAIKVKWSRKAIPHPSSIRLIRTASGKYFVSLVVETQPAPMPETGESVGVDFGVARLATLSN
GERISNPKHGAKWQRRLAFYQKRLARATKGSKRRMRIKRHVARIHEKIGNSRSDTLHKLSTDLV
TRFDLICVEDLNLRGMVKNHSLARSLHDASIGSAIRMIEEKAERYGKNVVKIDRWFPSSKTCSD
CGHIVEQLPLNVREWTCPECGTTHDRDANAAANILAVGQTVSAHGGTVRRSRAKASERKSQRSA
NRQGVNRA (서열번호 : 47)

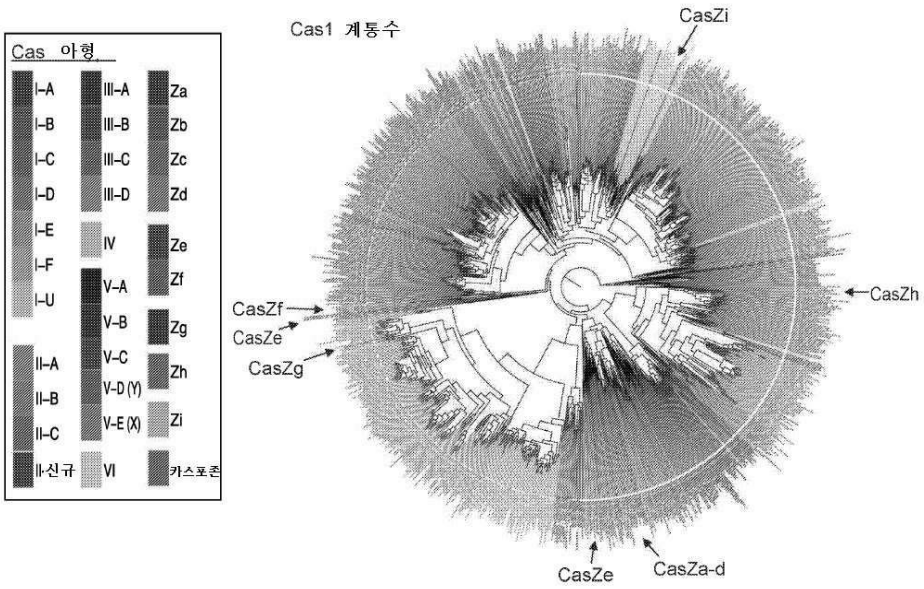
도면2

	#	Cas1	유기체	좌위 예시
Za	5	I-B	DPANN	
Zb	7	I-B	DPANN	
Zc	5	I-B	DPANN	
Zd	1	I-B	DPANN	
Ze	4	III-D*	CPR	
Zf	3	III-D	CPR	
Zg	1	I-E	파지 ?	
Zh	1	II-C	MGE	
Zi	1	I-U	부유균류	

도면3

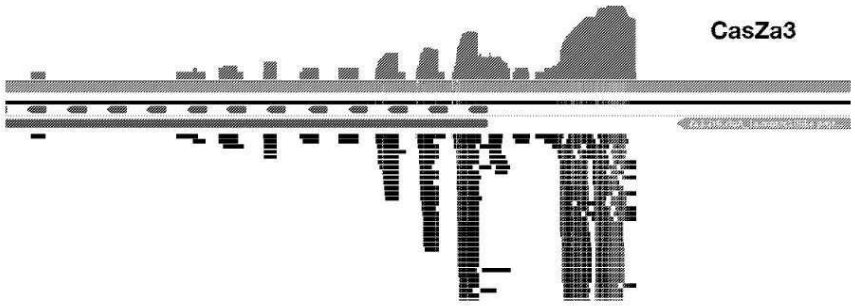


도면4



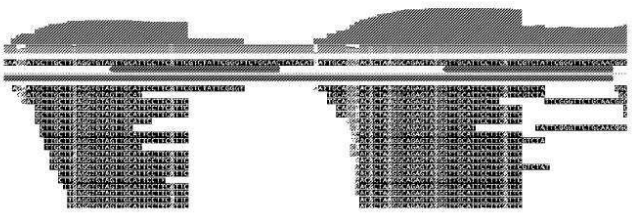
도면5a

전사체학

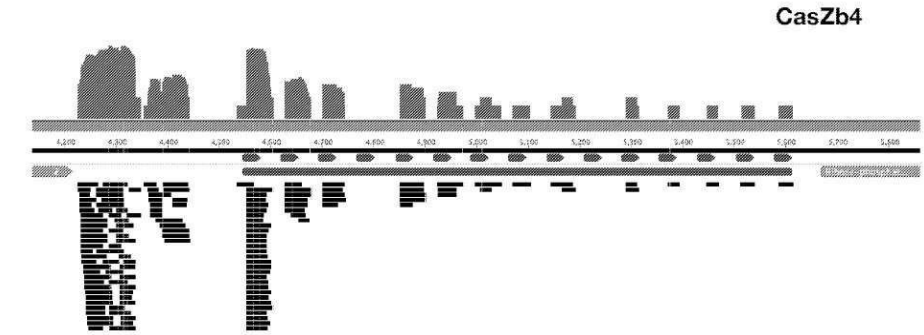


도면5b

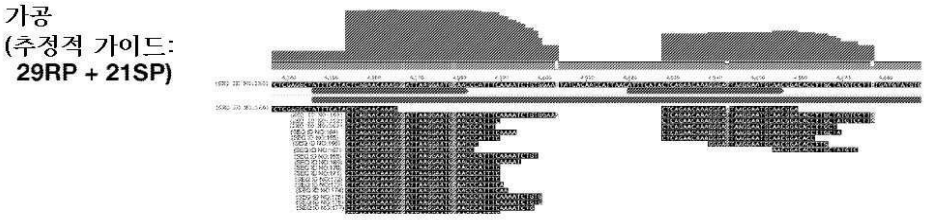
가공
(추정적 가이드:
17RP + 18SP)



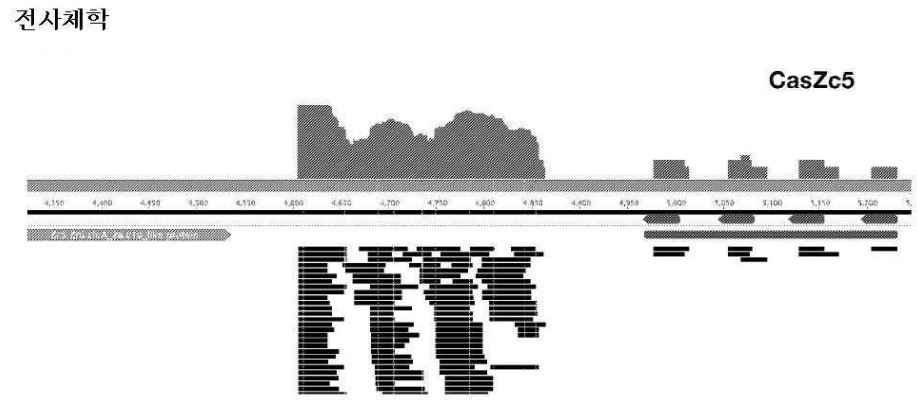
도면5c



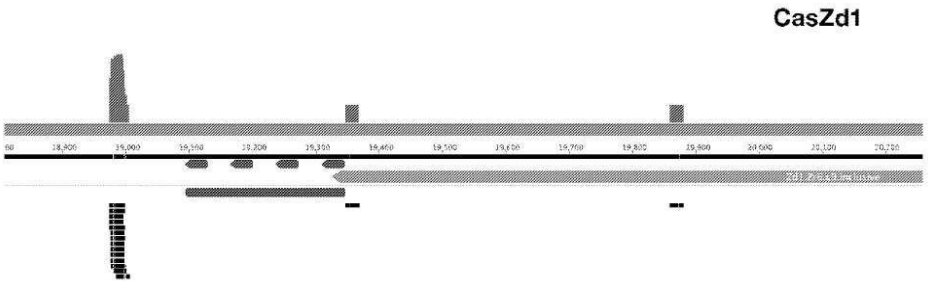
도면5d



도면5e

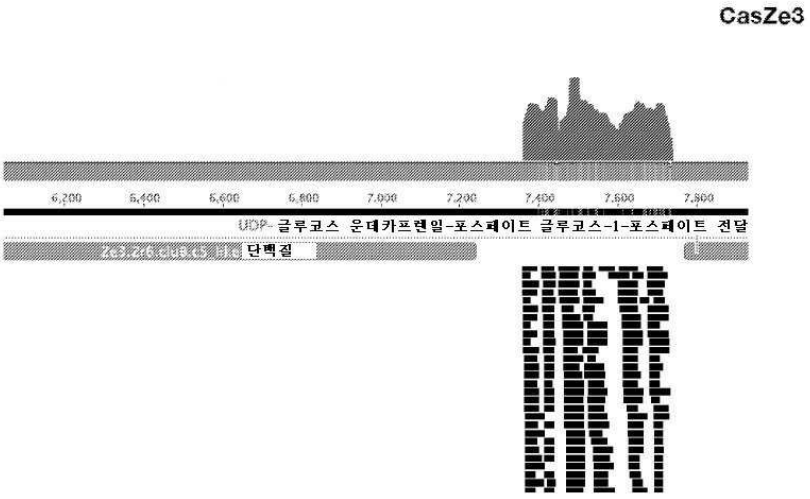


도면5f

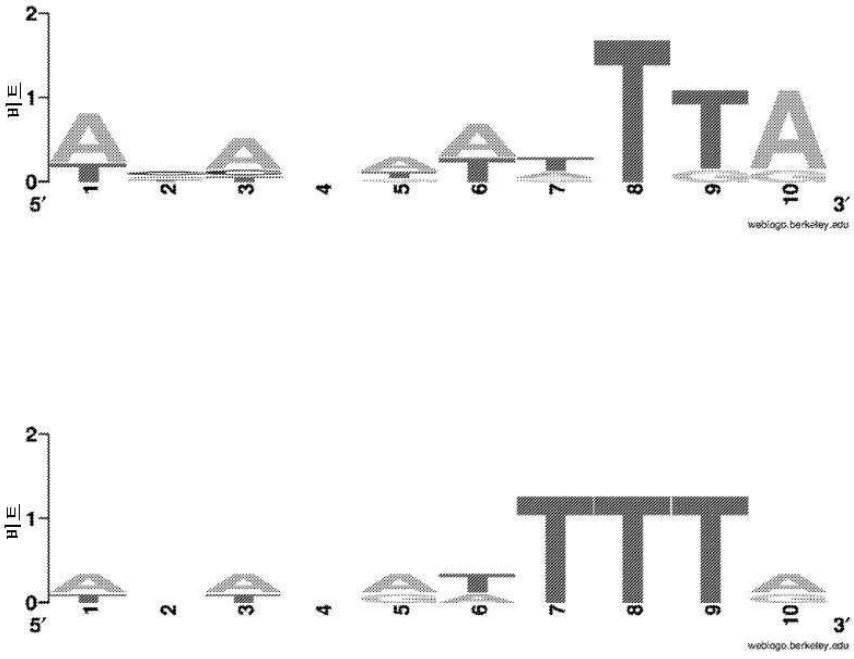


도면5g

전사체학



도면6



도면7a

>Cas14a.1|gwa1_scaffold_1795_curated|25635..27224|revcom
PEEKTSKLKPNSINLAANYDANEKFNCKECKFHPFKNKKRYEFNFYNNLHGCKSCTKST
NNPAVKRIEIGYQKLKFEIKNQMEAYPWFGRLRINFYSDEKRKMSELNEMQVTGVKNKI
FFDAIECAWREILKKRFRESKETLITIPKLKNKAGHGARKHRNKKLLIRRRAFMKKNFH
FLDNDSISYRSFANNIACVLPSKVGVDIGGIISPDVGKDIKPVDISLNLMWASKEGIKS
GRKVEIYSTQYDGNMVKKIEAETGEDKSWGKNRKRRQTSLLLSIPKPSKQVQEFDFKEW
PRYKDIEKKVQWRGFPIKIIFDSNHNSIEFGTYQGGKQKVLPIPFNDSKTTPLGSKMNK
LEKLRFNSKIKSRLGSAIAANKFLEAARTYCVDSLYHEVSSANAIGKGKIFIEYYLEIL
SQNYIEAAQKQLQRFIESIEQWFVADPFQGRLKQYFKDDLKRAKCFLCANREVQTTCYA
AVKLHKSCAEKVKDKNKELAIKERNNKEDAVIKEVEASNYPRVIRLKLTKTITNKAM
(서열번호 :306)

>Cas14a.2|gwa2_scaffold_18027_curated|7105..8628
ELIVNENKDPLNIGKTAKAVFKEIDPTSINRAANYDASIELACKECKFKPFNNTKRHDF
SFYSNWHRCSPNSCLQSTYRAKIRKTEIGYEKLKNEILNQMQYYPWFGRLYQNFFNDQR
DKMTSLDEIQVTGVQNKIFFNTVEKAWREIIKKRFRDNKETMRTIPDLKNKSGHGSRKL
SNKSLLRRRFAFAQKSFKLVDNSDVSYRAFSNNVACVLPSKIGVDIGGIINKDLKREYI
PQEITFNVFWKQHDGLKKGRNIEIHSVQYKGEIVKRIEADTGEDKAWGKNRQRRFTSLI
LKITPKQGGKKIWKFPEKKNASDYEYFPIPIEFILDNGDASIKFGGEEGEVGKQKHLLI
PFNDSKATPLSSKQMLLETSRFNAEVKSTIGLALYANYFVSYARNYVIKSTYHKNSKKG
QIVTEIYLESISQNFVRAIQRQLQSLMLNLKDWGFMQTHKKELKKYFGSDLEGSKGGQK
RREKEEKIEKEIEASYLPRLIRLSLTKSVTKAEEM (서열번호 :305)

>Cas14a.3|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom
KEPLNIGKTAKAVFKEIDPTSLNRAANYDASIELNCKECKFKPFKNVKRYEFNFYNNWY
RCNPNSCLQSTYKAQVRKVEIGYEKLKNEILTQMQYYPWFGRLYQNFFHDERDKMTSLD
EIQVIGVQNKVFFNTVEKAWREIIKKRFKDNKETMETIPELKHAAGHGKRKLSNKSLLR
RRFAFVQKSFKFVDNSDVSYRSFSNNIACVLPSRIGVDLGGVISRNPKREYIPQEISFN
AFWKQHEGLKKGRNIEIQSVQYKGETVKRIEADTGEDKAWGKNRQRRFTSLILKLVPKQ
GGKKVWKYPEKRNEGNYEYFPIPIEFILDSGETSIRFGGDEGEAGKQKHLVIPFNDSKA
TPLASQQTLLENSRFNAEVKSCIGLAIYANYFYGYARNYVISSIYHKNSKNGQAITAIY
LESIAHNYVKAIERQLQNLLLNLRDFSFMESHKKELKKYFGGDLEGTGGAQKRREKEEK
IEKEIEQSYLPRLIRLSLTKMVTKQVEM (서열번호 :304)

도면7b

>Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829
SESENKIIEQYYAFLYSFRDKYEKPEFKNRGDIKRKLQNKWEDFLKEQNLKNDKKLSNY
IFSNRNFRRSYDREEENEEGIDEKKSKPKRINCFEKEKNLKDQYDKDAINASANKDGAQ
KWGCFECIFFPMYKIESGDPNKRIIINKTRFKLFDFYLNLKGCKSCLRSTYHPYRSNVY
IESNYDKLKREIGNFLQQKNIFQRMRKAKVSEGKYLTNLDEYRLSCVAMHFKNRWLFFD
SIQKVLRETIKQRLKQMRESYDEQAKTKRSKGHGRAKYEDQVRMIRRRAYSAQAHKLLD
NGYITLFDYDDKEINKVCLTAINQEGFDIGGYLNSDIDNVMPPIEISFHLKWKYNEPIL
NIESPFSKAKISDYLRKIREDLNLERGKEGKARSKKNVRRKVLASKGEDGYKKIFTDFF
SKWKEELEGNAMERVLSQSSGDIQWSKKKRIHYTTLVLNINLLDKKGVGNLKYYEIAEK
TKILSFDKNENKFWPITIQVLLDGYEIGTEYDEIKQLNEKTSKQFTIYDPNTKIIKIPF
TDSKAVPLGMLGINIATLKTVKKTERDIKVSKIFKGGLNSKIVSKIGKGIYAGYFPTVD
KEILEEVEEDTLDNEFSSKSQRNIFLKSIIKNYDKMLKEQLFDFYSFLVRNDLGVRFLT
DRELQNIEDESFNLEKRFFETDRDRIARWFDNTNTDDGKEKFKKLANEIVDSYKPRLIR
LPVVRVIKRIQPVKQREM (서열번호 :178)

>Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521
KYSTRDFSELNEIQVTACKQDEFFKVIQNAWREIIKKRFLENRENFIEKKIFKNKKGRG
KRQESDKTIQRNRASVMKNFQLIENEKIILRAPSGHVACVFPVKVGLDIGGFKTDDLEK
NIFPPRTITINVFWKNRDRQRKGRKLEVWGIKARTKLIEKVHKWDKLEEVKKKRLKSLE
QKQEKSLDNWSEVNNDSFYKVQIDELQEKIDKSLKGRTMNKILDNKAKESKEAEGLYIE
WEKDFEGEMLRRIEASTGGEEKWGKRRQRRHTSLLLDIKNNSRGSKEIINFYSYAKQGK
KEKKIEFFPFPLTITLDAEEESPLNIKSIPIEDKNATSKYFSIPFTETRATPLSILGDR
VQKFKTKNISGAIKRNLGSSISSCKIVQNAETSAKSILSLPNVKEDNNMEIFINTMSKN
YFRAMMKQMESFIFEMEPKTLIDPYKEKAIKWFEVAASSRAKRKLKKLSKADIKKSELL
LSNTEEFEKEKQEKLEALEKEIEEFYLPRIVRLQLTKTILETPVM (서열번호 :179)

>Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734
KKLQLLGHKILLKEYDPNAVNAAANFETSTAELCGQCKMKPFKNKRRFQYTFGKNYHGC
LSCIQNVYYAKKRIVQIAKEELKHQLTDSIASIPYKYTSLFSNTNSIDELYILKQERAA
FFSNTNSIDELYITGIENNIAFKVISAIWDEIIKKRRQRYAESLTDTGTVKANRGHGGT
AYKSNTRQEKIRALQKQTLHMVTNPYISLARYKNNYIVATLPRTIGMHIGAIKDRDPQK
KLSDYAINFNVFWSDDRQLIELSTVQYTGDMVRKIEAETGENNKWGENMKRTKTSLLLE
ILTKKTTDELTFKDWAFSTKKEIDSVTKKTYQGFPIGIIFEGNESSVKFGSQNYFPLPF
DAKITPPTAEGFRLDWLRKGSFSSQMKTSYGLAIYSNKVTNAIPAYVIKNMFYKIARAE
NGKQIKAKFLKKYLDIAGNNYVPFIIMQHYRVLDTFEEMPISQPKVIRLSLTKTQHIII
KKDKTDSKM (서열번호 :180)

도면7c

>Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257
NTSNLINLGKKAINISANYDANLEVGCKNCKFLSSNGNFPRQTNVKEGCHSCEKSTYEP
SIYLVKIGERKAKYDVLDSLKKFTFQSLKYQSKKSMKSRNKKPKELKEFVIFANKNKAF
DVIQKSYNHLILQIKKEINRMNSKKRKKNHKRRLFRDREKQLNKLRLIESSNLFLPREN
KGNNHVFTYVAIHSVGRDIGVIGSYDEKLNFETELTYQLYFNDDKRLLYAYKPKQNKII
KIKEKLWNLRKEKEPLDLEYEKPLNKSITFSIKNDNLFKVSKDLMLRRAKFNIQGKEKL
SKEERKINRDLIKIKGLVNSMSYGRFDELKKEKNIWSPHIYREVRQKEIKPCLIKNGDR
IEIFEQLKKKMERLRRFREKRQKKISKDLIFAERIAYNFHTKSIKNTSNKINIDQEAKR
GKASYMRKRIGYETFKNKYCEQCLSKGNVYRNVQKGCSCFENPFDWIKKGDENLLPKKN
EDLRVKGAFRDEALEKQIVKIAFNIAKGYEDFYDNLGESTEKDLKLKFKVGTTINEQES
LKL (서열번호 :181)

>Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983
TSNPIKLGKKAINISANYDSNLQIGCKNCKFLSYNGNFPRQTNVKEGCHSCEKSTYEPP
VYTVRIGERRSKYDVLDSLKKFIFLSLKYRQSKKMKTRSKGIRGLEEFVISANLKKAMD
VIQKSYRHLILNIKNEIVRMNGKKRNKNHKRLLFRDREKQLNKLRLIEGSSFFKPPTVK
GDNSIFTCVAIHNIGRDIGIAGDYFDKLEPKIELTYQLYYEYNPKKESEINKRLLYAYK
PKQNKIIEIKEKLWNLRKEKSPLDLEYEKPLTKSITFLVKRDGVFRISKDLMLRKAKFI
IQGKEKLSKEERKINRDLIKIKSNIISLTYGRFDELKKDKTIWSPHIFRDVKQGKITPC
IERKGDRMDIFQQLRKKSERLRENRKKRQKKISKDLIFAERIAYNFHTKSIKNTSNLIN
IKHEAKRGKASYMRKRIGNETFRIKYCEQCFPKNNVYKNVQKGCSCFEDPFEYIKKGNE
DLIPNKNQDLKAKGAFRDDALEKQIIKVAFNIAKGYEDFYENLKKTTEKDIRLKFKVGT
IISEEM (서열번호 :182)

>Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217
NNSINLSKKAINISANYDANLQVRCKNCKFLSSNGNFPRQTDVKEGCHSCEKSTYEPPV
YDVKIGEIKAKYEVLDSLKKFTFQSLKYQLSKSMKFRSKKIKELKEFVIFAKESKALNV
INRSYKHLILNIKNDINRMNSKKRIKNHKGRLFLDRQKQLSKLKLIEGSSFFVPAKNVG
NKSVFTCVAIHSIGRDIGIAGLYDSFTKPVNEITYQIFFSGERRLLYAYKPKQLKILSI
KENLWSLKNEKKPLDLLYEKPLGKNLNFNVKGGDLFRVSKDLMIRNAKFNVHGRQRLSD
EERLINRNFIKIKGEVVSLSYGRFEELKKDRKLWSPHIFKDVRQNKIKPCLVMQGQRID
IFEQLKRKLELLKKIRKSRQKKLSKDLIFGERIAYNFHTKSIKNTSNKINIDSDAKRGR
ASYMRKRIGNETFKLKYCDVCFPKANVYRRVQNGCSCSENPYNYIKKGDKDLLPKKDEG
LAIKGAFRDEKLNKQIIKVAFNIAKGYEDFYDDLKKRTEKDVDLKFKIGTTVLDQKPME
IFDGIVITWL (서열번호 :183)

도면7d

>Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155
LLTTVVETNNLAKKAINVAANFDANIDRQYYRCTPNLCRFIAQSPRETKEKDAGCSSCT
QSTYDPKVYVIKIGKLLAKYEILKSLKRFLFMNRYFKQKKTERAQQKQKIGTELNEMSI
FAKATNAMEVIKRATKHCTYDIIPETKSLQMLKRRRHRVKVRSLLKILKERRMKIKKIP
NTFIEIPKQAKKNKSDYYVAAALKSCGIDVGLCGAYEKNAEVEAEYTYQLYYEYKGNSS
TKRILYCYNNPQKNIREFWEAFYIQGSKSHVNTPGTIRLKMEKFLSPITIESEALDFRV
WNSDLKIRNGQYGFIKKRSLGKEAREIKKGMGDIKRKIGNLTYGKSPSELKSIHVYRTE
RENPKKPRAARKKEDNFMEIFEMQRKKDYEVNKKRRKEATDAAKIMDFAEEPIRHYHTN
NLKAVRRIDMNEQVERKKTSVFLKRIMQNGYRGNYCRKCIKAPEGSNRDENVLEKNEGC
LDCIGSEFIWKKSSKEKKGLWHTNRLLRRIRLQCFTTAKAYENFYNDLFEKKESSLDII
KLKVSITTKSM (서열번호 :184)

>Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598
ASTMNLAKQAINFAANYDSNLEIGCKGCKFMSTWSKKSNPKFYPRQNNQANKCHSCTYS
TGEPEVPIIEIGERAAKYKIFTALKKFVFMSVAYKERRRQRFKSKKPKELKELAICSNR
EKAMEVIQKSVVHCYGDVKQEIPRIRKIKVLKNHKGRLFYKQKRSKIKIAKLEKGSFFK
TFIPKVHNNGCHSCHEASLNKPILVTTALNTIGADIGLINDYSTIAPTETDISWQVYYE
FIPNGDSEAVKKRLLYFYKPKGALIKSIRDKYFKKGHENAVNTGFFKYQGKIVKGPIKF
VNNELDFARKPDLKSMKIKRAGFAIPSAKRLSKEDREINRESIKIKNKIYSLSYGRKKT
LSDKDIIKHLYRPVRQKGVKPLEYRKAPDGFLEFFYSLKRKERRLRKQKEKRQKDMSEI
IDAADEFAWHRHTGSIKKTTNHINFKSEVKRGKVPIMKKRIANDSFNTRHCGKCVKQGN
AINKYYIEKQKNCFDCNSIEFKWEKAALEKKGAFKLNKRLQYIVKACFNVAKAYESFYE
DFRKGEEESLDLKFKIGTTTTLKQYPQNKARAM (서열번호 :185)

>Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634
..8466|revcom
HSHNLMLTKLGKQAINFAANYDANLEIGCKNCKFLSYSPKQANPKKYPRQTDVHEDGNI
ACHSCMQSTKEPPVYIVPIGERKSKYEILTSLNKFTFLALKYKEKKRQAFRAKKPKELQ
ELAIAFNKEKAIKVIDKSIQHLILNIKPEIARIQRQKRLKNRKGKLLYLHKRYAIKMGL
IKNGKYFKVGSPKKDGKKLLVLCALNTIGRDIGIIGNIEENNRSETEITYQLYFDCLDA
NPNELRIKEIEYNRLKSYERKIKRLVYAYKPKQTKILEIRSKFFSKGHENKVNTGSFNF
ENPLNKSISIKVKNSAFDFKIGAPFIMLRNGKFHIPTKKRLSKEEREINRTLSKIKGRV
FRLTYGRNISEQGSKSLHIYRKERQHPKLSLEIRKQPDSFIDEFEKLRLKQNFISKLKK
QRQKKLADLLQFADRIAYNYHTSSLEKTSNFINYKPEVKRGRTSYIKKRIGNEGFEKLY
CETCIKSNDKENAYAVEKEELCFVCKAKPFTWKKTNKDKLGIFKYPSRIKDFIRAAFTV
AKSYNDFYENLKKKDLKNEIFLKFKIGLILSHEKKNHISIAKSVAEDERISGKSIKNIL
NKSIKLEKNCYSCFFHKEDM (서열번호 :186)

도면7e

>Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom
SLERVIDKRNLAKKAINIAANFDANINKGFYRCETNQCMFIAQKPRKTNNTGCSSCLQS
TYDPVIYVVKVGEMLAKYEILKSLKRFVFMNRSFKQKKTEKAKQKERIGGELNEMSIFA
NAALAMGVIKRAIRHCHVDIRPEINRLSELKKTKHRVAAKSLVKIVKQRKTKWKGIPNS
FIQIPQKARNKDADFYVASALKSGGIDIGLCGTYDKKPHADPRWTYQLYFDTEDESEKR
LLYCYNDPQAKIRDFWKTFYERGNPSMVNSPGTIEFRMEGFFEKMTPISIESKDFDFRV
WNKDLLIRRGLYEIKKRKNLNRKAREIKKAMGSVKRVLANMTYGKSPTDKKSIPVYRVE
REKPKKPRAVRKEENELADKLENYRREDFLIRNRRKREATEIAKIIDAAEPPIRHYHTN
HLRAVKRIDLSKPVARKNTSVFLKRIMQNGYRGNYCKKCIKGNIDPNKDECRLEDIKKC
ICCEGTQNIWAKKEKLYTGRINVLNKRIKQMKLECFNVAKAYENFYDNLAALKEGDLKV
LKLKVSIPALNPEASDPEEDM (**서열번호** :187)

>Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom
NASINLGKRAINLSANYDSNLVIGCKNCKFLSFNGNFPRQTNVREGCHSCDKSTYAPEV
YIVKIGERKAKYDVLDSLKKFTFQSLKYQIKKSMRERSKKPKELLEFVIFANKDKAFNV
IQKSYEHLILNIKQEINRMNGKKRIKNHKKRLFKDREKQLNKLRLIGSSSLFFPRENKG
DKDLFTYVAIHSVGRDIGVAGSYESHIEPISDLTYQLFINNEKRLLYAYKPKQNKIIEL
KENLWNLKKEKKPLDLEFTKPLEKSITFSVKNDKLFKVSKDLMLRQAKFNIQGKEKLSK
EERQINRDFSKIKSNVISLSYGRFEELKKEKNIWSPHIYREVKQKEIKPCIVRKGDRIE
LFEQLKRKMDKLKKFRKERQKKISKDLNFAERIAYNFHTKSIKNTSNKINIDQEAKRGK
ASYMRKRIGNESFRKKYCEQCFSVGNVYHNVQNGCSCFDNPIELIKKGDEGLIPKGKED
RKYKGALRDDNLQMQIIRVAFNIAKGYEDFYNNLKEKTEKDLKLKFKIGTTISTQESNN
KEM (**서열번호** :188)

>Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom
SNLIKLGKQAINFAANYDANLEVGCKNCKFLSSTNKYPRQTNVHLDNKMACRSCNQSTM
EPAIYIVRIGEKKAKYDIYNSLTKFNFQSLKYKAKRSQRFKPKQPKELQELSIAVRKEK
ALDIIQKSIDHLIQDIRPEIPRIKQQKRYKNHVGKLFYLQKRRKNKLNLIGKGSFFKVF
SPKEKKNELLVICALTNIGRDIGLIGNYNTIINPLFEVTYQLYYDYIPKKNNKNVQRRL
LYAYKSKNEKILKLKEAFFKRGHENAVNLGSFSYEKPLEKSLTLKIKNDKDDFQVSPSL
RIRTGRFFVPSKRNLSRQEREINRRLVKIKSKIKNMTYGKFETARDKQSVHIFRLERQK
EKLPLQFRKDEKEFMEEFQKLKRRTNSLKKLRKSRQKKLADLLQLSEKVVYNNHTGTLK
KTSNFLNFSSSVKRGKTAYIKELLGQEGFETLYCSNCINKGQKTRYNIETKEKCFSCKD
VPFVWKKKSTDKDRKGAFLFPAKLKDVIKATFTVAKAYEDFYDNLKSIDEKKPYIKFKI
GLILAHVRHEHKARAKEEAGQKNIYNKPIKIDKNCKECFFFKEEAM (**서열번호** :189
)

도면7f

>Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975
NTTRKKFRKRTGFPQSDNIKLAYCSAIVRAANLDADIQKKHNQCNPNLCVGIKSNEQSR
KYEHSDRQALLCYACNQSTGAPKVDYIQIGEIGAKYKILQMVNAYDFLSLAYNLTKLRN
GKSRGHQRMSQLDEVVIVADYEKATEVIKRSINHLLDDIRGQLSKLKKRTQNEHITEHK
QSKIRRKLRKLSRLLKRRRWKWGTIPNPYLKNWVFTKKDPELVTVALLHKLGRDIGLVN
RSKRRSKQKLLPKVGFQLYYKWESPSLNNIKKSKAKKLPKRLLIPYKNVKLFDNKQKLE
NAIKSLLESYQKTIKVEFDQFFQNRTEEIIAEEQQTLERGLLKQLEKKKNEFASQKKAL
KEEKKKIKEPRKAKLLMEESRSLGFLMANVSYALFNTTIEDLYKKSNVVSGCIPQEPVV
VFPADIQNKGSLAKILFAPKDGFRIKFSGQHLTIRTAKFKIRGKEIKILTKTKREILKN
IEKLRRVWYREQHYKLKLFGKEVSAKPRFLDKRKTSIERRDPNKLADQTDDRQAELRNK
EYELRHKQHKMAERLDNIDTNAQNLQTLSFWVGEADKPPKLDEKDARGFGVRTCISAWK
WFMEDLLKKQEEDPLLKLKLSIM (서열번호 :190)

>Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219
PKKPKFQKRTGFPQPDNLRKEYCLAIVRAANLDADFEKKCTKCEGIKTNKKGNIVKGRT
YNSADKDNLLCYACNISTGAPAVDYVFVGALEAKYKILQMVKAYDFHSLAYNLAKLWKG
RGRGHQRMGGLNEVVIVSNNEKALDVIEKSLNHFHDEIRGELSRLKAKFQNEHLHVHKE
SKLRRKLRKISRLLKRRRWKWDVIPNSYLRNFTFTKTRPDFISVALLHRVGRDIGLVTK
TKIPKPTDLLPQFGFQIYYTWDEPKLNKLKKSRLRSEPKRLLVPYKKIELYKNKSVLEE
AIRHLAEVYTEDLTICFKDFFETQKRKFVSKEKESLKRELLKELTKLKKDFSERKTALK
RDRKEIKEPKKAKLLMEESRSLGFLAANTSYALFNLIAADLYTKSKKACSTKLPRQLST
ILPLEIKEHKSTTSLAIKPEEGFKIRFSNTHLSIRTPKFKMKGADIKALTKRKREILKN
ATKLEKSWYGLKHYKLKLYGKEVAAKPRFLDKRNPSIDRRDPKELMEQIENRRNEVKDL
EYEIRKGQHQMAKRLDNVDTNAQNLQTKSFWVGEADKPPELDSMEAKKLGLRTCISAWK
WFMKDLVLLQEKSPNLKLKLSLTEM (서열번호 :191)

>Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom
KFSKRQEGFLIPDNIDLYKCLAIVRSANLDADVQGHKSCYGVKKNGTYRVKQNGKKGVK
EKGRKYVFDLIAFKGNIEKIPHEAIEEKDQGRVIVLGKFNYKLILNIEKNHNDRASLEI
KNKIKKLVQISSLETGEFLSDLLSGKIGIDEVYGIIEPDVFSGKELVCKACQQSTYAPL
VEYMPVGELDAKYKILSAIKGYDFLSLAYNLSRNRANKKRGHQKLGGGELSEVVISANY
DKALNVIKRSINHYHVEIKPEISKLKKKMQNEPLKVMKQARIRRELHQLSRKVKRLKWK
WGMIPNPELQNIIFEKKEKDFVSYALLHTLGRDIGLFKDTSMLQVPNISDYGFQIYYSW
EDPKLNSIKKIKDLPKRLLIPYKRLDFYIDTILVAKVIKNLIELYRKSYVYETFGEEYG
YAKKAEDILFDWDSINLSEGIEQKIQKIKDEFSDLLYEARESKRQNFVESFENILGLYD
KNFASDRNSYQEKIQSMIIKKQQENIEQKLKREFKEVIERGFEGMDQNKKYYKVLSPNI
KGGLLYTDTNNLGFFRSHLAFMLLSKISDDLYRKNNLVSKGGNKGILDQTPETMLTLEF
GKSNLPNISIKRKFFNIKYNSSWIGIRKPKFSIKGAVIREITKKVRDEQRLIKSLEGVW
HKSTHFKRWGKPRFNLPRHPDREKNNDDNLMESITSRREQIQLLLREKQKQQEKMAGRL
DKIDKEIQNLQTANFQIKQIDKKPALTEKSEGKQSVRNALSAWKWFMEDLIKYQKRTPI
LQLKLAKM (서열번호 : 192)

도면7g

>Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom
KFSKRQEGFVIPENIGLYKCLAIVRSANLDADVQGHVSCYGVKKNGTYVLKQNGKKSIR
EKGRKYASDLVAFKGDIEKIPFEVIEEKKKEQSIVLGKFNYKLVLDVMKGEKDRASLTM
KNKSKKLVQVSSLGTDEFLLTLLNEKFGIEEIYGIIEPEVFSGKKLVCKACQQSTYAPL
VEYMPVGELDSKYKILSAIKGYDFLSLAYNLARHRSNKKRGHQKLGGGELSEVVISANN
AKALNVIKRSLNHYYSEIKPEISKLRKKMQNEPLKVGKQARMRRELHQLSRKVKRLKWK
WGKIPNLELQNITFKESDRDFISYALLHTLGRDIGMFNKTEIKMPSNILGYGFQIYYDW
EEPKLNTIKKSKNTPKRILIPYKKLDFYNDSILVARAIKELVGLFQESYEWEIFGNEYN
YAKEAEVELIKLDEESINGNVEKKLQRIKENFSNLLEKAREKKRQNFIESFESIARLYD
ESFTADRNEYQREIQSFIIEKQKQSIEKKLKNEFKKIVEKKFNEQEQGKKHYRVLNPTI
INEFLPKDKNNLGFLRSKIAFILLSKISDDLYKKSNAVSKGGEKGIIKQQPETILDLEF
SKSKLPSINIKKKLFNIKYTSSWLGIRKPKFNIKGAKIREITRRVRDVQRTLKSAESSW
YASTHFRRWGFPRFNQPRHPDKEKKSDDRLIESITLLREQIQILLREKQKGQKEMAGRL
DDVDKKIQNLQTANFQIKQTGDKPALTEKSAGKQSFRNALSAWKWFMENLLKYQNKTPD
LKLKIARTVM (서열번호 : 193)

>Cas14b.14|gwc1_scaffold_8732_curated|2705..4537
KWIEPNNIDFNKCLAITRSANLDADVQGHKMCYGIKTNGTYKAIGKINKKHNTGIIEKR
RTYVYDLIVTKEKNEKIVKKTDFMAIDEEIEFDEKKEKLLKKYIKAEVLGTGELIRKDL
NDGEKFDDLCSIEEPQAFRRSELVCKACNQSTYASDIRYIPIGEIEAKYKILKAIKGYD
FLSLKYNLGRLRDSKKRGHQKMGQGELKEFVICANKEKALDVIKRSLNHYLNEVKDEIS
RLNKKMQNEPLKVNDQARWRRELNQISRRLKRLKWKWGEIPNPELKNLIFKSSRPEFVS
YALIHTLGRDIGLINETELKPNNIQEYGFQIYYKWEDPELNHIKKVKNIPKRFIIPYKN
LDLFGKYTILSRAIEGILKLYSSSFQYKSFKDPNLFAKEGEKKITNEDFELGYDEKIKK
IKDDFKSYKKALLEKKKNTLEDSLNSILSVYEQSLLTEQINNVKKWKEGLLKSKESIHK
QKKIENIEDIISRIEELKNVEGWIRTKERDIVNKEETNLKREIKKELKDSYYEEVRKDF
SDLKKGEESEKKPFREEPKPIVIKDYIKFDVLPGENSALGFFLSHLSFNLFDSIQYELF
EKSRLSSSKHPQIPETILDL (서열번호 : 194)

>Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032
FRKFVKRSGAPQPDNLNKYKCIAIVRAANLDADIMSNESSNCVMCKGIKMNKRKTAKGA
AKTTELGRVYAGQSGNLLCTACTKSTMGPLVDYVPIGRIRAKYTILRAVKEYDFLSLAY
NLARTRVSKKGGRQKMHSLSELVIAAEYEIAWNIIKSSVIHYHQETKEEISGLRKKLQA
EHIHKNKEARIRREMHQISRRIKRLKWKWHMIPNSELHNFLFKQQDPSFVAVALLHTLG
RDIGMINKPKGSAKREFIPEYGFQIYYKWMNPKLNDINKQKYRKMPKRSLIPYKNLNVF
GDRELIENAMHKLLKLYDENLEVKGSKFFKTRVVAISSKESEKLKRDLLWKGELAKIKK
DFNADKNKMQELFKEVKEPKKANALMKQSRNMGFLLQNISYGALGLLANRMYEASAKQS
KGDATKQPSIVIPLEMEFGNAFPKLLLRSGKFAMNVSSPWLTIRKPKFVIKGNKIKNIT
KLMKDEKAKLKRLETSYHRATHFRPTLRGSIDWDSPYFSSPKQPNTHRRSPDRLSADIT
EYRGRLKSVEAELREGQRAMAKKLDSVDMTASNLQTSNFQLEKGEDPRLTEIDEKGRSI
RNCISSWKKFMEDLMKAQEANPVIKIKIALKDESSVLSEDSM (서열번호 : 195)

도면7h

>Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627
KFHPENLNKSYCLAIVRAANLDADIQGHINCIGIKSNKSDRNYENKLESLQNVELLCKA
CTKSTYKPNINSVPVGEKKAKYSILSEIKKYDFNSLVYNLKKYRKGKSRGHQKLNELRE
LVITSEYKKALDVINKSVNHYLVNIKNKMSKLKKILQNEHIHVGTLARIRRERNRISRK
LDHYRKKWKFVPNKILKNYVFKNQSPDFVSVALLHKLGRDIGLITKTAILQKSFPEYSL
QLYYKYDTPKLNYLKKSKFKSLPKRILISYKYPKFDINSNYIEESIDKLLKLYEESPIY
KNNSKIIEFFKKSEDNLIKSENDSLKRGIMKEFEKVTKNFSSKKKKLKEELKLKNEDKN
SKMLAKVSRPIGFLKAYLSYMLFNIISNRIFEFSRKSSGRIPQLPSCIINLGNQFENFK
NELQDSNIGSKKNYKYFCNLLLKSSGFNISYEEEHLSIKTPNFFINGRKLKEITSEKKK
IRKENEQLIKQWKKLTFFKPSNLNGKKTSDKIRFKSPNNPDIERKSEDNIVENIAKVKY
KLEDLLSEQRKEFNKLAKKHDGVDVEAQCLQTKSFWIDSNSPIKKSLEKKNEKVSVKKK
MKAIRSCISAWKWFMADLIEAQKETPMIKLKLALM (서열번호 : 196)

>Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curat
ed|19327..20880|revcom
TTLVPSHLAGIEVMDETTSRNEDMIQKETSRSNEDENYLGVKNKCGINVHKSGRGSSKH
EPNMPPEKSGEGQMPKQDSTEMQQRFDESVTGETQVSAGATASIKTDARANSGPRVGTA
RALIVKASNLDRDIKLGCKPCEYIRSELPMGKKNGCNHCEKSSDIASVPKVESGFRKAK
YELVRRFESFAADSISRHLGKEQARTRGKRGKKDKKEQMGKVNLDEIAILKNESLIEYT
ENQILDARSNRIKEWLRSLRLRLRTRNKGLKKSKSIRRQLITLRRDYRKWIKPNPYRPD
EDPNENSLRLHTKLGVDIGVQGGDNKRMNSDDYETSFSITWRDTATRKICFTKPKGLLP
RHMKFKLRGYPELILYNEELRIQDSQKFPLVDWERIPIFKLRGVSLGKKKVKALNRITE
APRLVVAKRIQVNIESKKKKVLTRYVYNDKSINGRLVKAEDSNKDPLLEFKKQAEEINS
DAKYYENQEIAKNYLWGCEGLHKNLLEEQTKNPYLAFKYGFLNIV (서열번호 :197)

>Cas14c.2|3300001245.a|JGI12048J13642_10201286|4257..5489|r
evcom
LDFKRTCSQELVLLPEIEGLKLSGTQGVTSLAKKLINKAANVDRDESYGCHHCIHTRTS
LSKPVKKDCNSCNQSTNHPAVPITLKGYKIAFYELWHRFTSWAVDSISKALHRNKVMGK
VNLDEYAVVDNSHIVCYAVRKCYEKRQRSVRLHKRAYRCRAKHYNKSQPKVGRIYKKSK
RRNARNLKKEAKRYFQPNEITNGSSDALFYKIGVDLGIAKGTPETEVKVDVSICFQVYY
GDARRVLRVRKMDELQSFHLDYTGKLKLKGIGNKDTFTIAKRNESLKWGSTKYEVSRAH
KKFKPFGKKGSVKRKCNDYFRSIASWSCEAASQRAQSNLKNAFPYQKALVKCYKNLDYK
GVKKNDMWYRLCSNRIFRYSRIAEDIAQYQSDKGKAKFEFVILAQSVAEYDISAIM
(서열번호 : 198)

도면7i

>Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom
VFLTDDKRKTALRKIRSAFRKTAEIALVRAQEADSLDRQAKKLTIETVSFGAPGAKNAF
IGSLQGYNWNSHRANVPSSGSAKDVFRITELGLGIPQSAHEASIGKSFELVGNVVRYTA
NLLSKGYKKGAVNKGAKQQREIKGKEQLSFDLISNGPISGDKLINGQKDALAWWLIDKM
GFHIGLAMEPLSSPNTYGITLQAFWKRHTAPRRYSRGVIRQWQLPFGRQLAPLIHNFFR
KKGASIPIVLTNASKKLAGKGVLLEQTALVDPKKWWQVKEQVTGPLSNIWERSVPLVLY
TATFTHKHGAAHKRPLTLKVIRISSGSVFLLPLSKVTPGKLVRAWMPDINILRDGRPDE
AAYKGPDLIRARERSFPLAYTCVTQIADEWQKRALESNRDSITPLEAKLVTGSDLLQIH
STVQQAVEQGIGGRISSPIQELLAKDALQLVLQQLFMTVDLLRIQWQLKQEVADGNTSE
KAVGWAIRISNIHKDAYKTAIEPCTSALKQAWNPLSGFEERTFQLDASIVRKRSTAKTP
DDELVIVLRQQAAEMTVAVTQSVSKELMELAVRHSATLHLLVGEVASKQLSRSADKDRG
AMDHWKLLSQSM (**서열번호 : 199**)

>Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom
EDLLQKALNTATNVAAIERHSCISCLFTESEIDVKYKTPDKIGQNTAGCQSCTFRVGYS
GNSHTLPMGNRIALDKLRETIQRYAWHSLLFNVPPAPTSKRVRAISELRVAAGRERLFT
VITFVQTNILSKLQKRYAANWTPKSQERLSRLREEGQHILSLLESGSWQQKEVVREDQD
LIVCSALTKPGLSIGAFCRPKYLKPAKHALVLRLIFVEQWPGQIWGQSKRTRRMRRRKD
VERVYDISVQAWALKGKETRISECIDTMRRHQQAYIGVLPFLILSGSTVRGKGDCPILK
EITRMRYCPNNEGLIPLGIFYRGSANKLLRVVKGSSFTLPMWQNIETLPHPEPFSPEGW
TATGALYEKNLAYWSALNEAVDWYTGQILSSGLQYPNQNEFLARLQNVIDSIPRKWFRP
QGLKNLKPNGQEDIVPNEFVIPQNAIRAHHVIEWYHKTNDLVAKTLLGWGSQTTLNQTR
PQGDLRFTYTRYYFREKEVPEV (**서열번호 : 200**)

>Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom
VPKKKLMRELAKKAVFEAIFNDPIPGSFGCKRCTLIDGARVTDAIEKKQGAKRCAGCEP
CTFHTLYDSVKHALPAATGCDRTAIDTGLWEILTALRSYNWMSFRRNAVSDASQKQVWS
IEELAIWADKERALRVILSALTHTIGKLKNGFSRDGVWKGGKQLYENLAQKDLAKGLFA
NGEIFGKELVEADHDMLAWTIVPNHQFHIGLIRGNWKPAAVEASTAFDARWLTNGAPLR
DTRTHGHRGRRFNRTEKLTVLCIKRDGGVSEEFRQERDYELSVMLLQPKNKLKPEPKGE
LNSFEDLHDHWWFLKGDEATALVGLTSDPTVGDFIQLGLYIRNPIKAHGETKRRLLICF
EPPIKLPLRRAFPSEAFKTWEPTINVFRNGRRDTEAYYDIDRARVFEFPETRVSLEHLS
KQWEVLRLEPDRENTDPYEAQQNEGAELQVYSLLQEAAQKMAPKVVIDPFGQFPLELFS
TFVAQLFNAPLSDTKAKIGKPLDSGFVVESHLHLLEEDFAYRDFVRVTFMGTEPTFRVI
HYSNGEGYWKKTVLKGKNNIRTALIPEGAKAAVDAYKNKRCPLTLEAAILNEEKDRRLV
LGNKALSLLAQTARGNLTILEALAAEVLRPLSGTEGVVHLHACVTRHSTLTESTETDNM
(**서열번호 : 201**)

도면7j

>Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313
VEKLFSERLKRAMWLKNEAGRAPPAETLTLKHKRVSGGHEKVKEELQRVLRSLSGTNQA
AWNLGLSGGREPKSSDALKGEKSRVVLETVVFHSGHNRVLYDVIEREDQVHQRSSIMHM
RRKGSNLLRLWGRSGKVRRKMREEVAEIKPVWHKDSRWLAIVEEGRQSVVGISSAGLAV
FAVQESQCTTAEPKPLEYVVSIWFRGSKALNPQDRYLEFKKLKTTEALRGQQYDPIPFS
LKRGAGCSLAIRGEGIKFGSRGPIKQFFGSDRSRPSHADYDGKRRLSLFSKYAGDLADL
TEEQWNRTVSAFAEDEVRRATLANIQDFLSISHEKYAERLKKRIESIEEPVSASKLEAY
LSAIFETFVQQREALASNFLMRLVESVALLISLEEKSPRVEFRVARYLAESKEGFNRKA
M (서열번호 : 202)

>Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217
VVITQSELYKERLLRVMEIKNDRGRKEPRESQGLVLRFTQVTGGQEKVKQKLWLIFEGF
SGTNQASWNFGQPAGGRKPNSGDALKGPKSRVTYETVVFHFGLRLLSAVIERHNLKQQR
QTMAYMKRRAAARKKWARSGKKCSRMRNEVEKIKPKWHKDPRWFDIVKEGEPSIVGISS
AGFAIYIVEEPNFPRQDPLEIEYAISIWFRRDRSQYLTFKKIQKAEKLKELQYNPIPFR
LKQEKTSLVFESGDIKFGSRGSIEHFRDEARGKPPKADMDNNRRLTMFSVFSGNLTNLT
EEQYARPVSGLLAPDEKRMPTLLKKLQDFFTPIHEKYGERIKQRLANSEASKRPFKKLE
EYLPAIYLEFRARREGLASNWVLVLINSVRTLVRIKSEDPYIEFKVSQYLLEKEDNKAL
(서열번호 : 203)

>Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom
KQDALFEERLKKAIFIKRQADPLQREELSLLPPNRKIVTGGHESAKDTLKQILRAINGT
NQASWNPGTPSGKRDSKSADALAGPKSRVKLETVVFHVGHRLLKKVVEYQGHQKQQHGL
KAFMRTCAAMRKKWKRSGKVVGELREQLANIQPKWHYDSRPLNLCFEGKPSVVGLRSAG
IALYTIQKSVVPVKEPKPIEYAVSIWFRGPKAMDREDRCLEFKKLKIATELRKLQFEPI
VSTLTQGIKGFSLYIQGNSVKFGSRGPIKYFSNESVRQRPPKADPDGNKRLALFSKFSG
DLSDLTEEQWNRPILAFEGIIRRATLGNIQDYLTVGHEQFAISLEQLLSEKESVLQMSI
EQQRLKKNLGKKAENEWVESFGAEQARKKAQGIREYISGFFQEYCSQREQWAENWVQQL
NKSVRLFLTIQDSTPFIEFRVARYLPKGEKKKGKAM (서열번호 : 204)

>Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041
ANHAERHKRLRKEANRAANRNRPLVADCDTGDPLVGICRLLRRGDKMQPNKTGCRSCEQ
VEPELRDAILVSGPGRLDNYKYELFQRGRAMAVHRLLKRVPKLNRPKKAAGNDEKKAEN
KKSEIQKEKQKQRRMMPAVSMKQVSVADFKHVIENTVRHLFGDRRDREIAECAALRAAS
KYFLKSRRVRPRKLPKLANPDHGKELKGLRLREKRAKLKKEKEKQAELARSNQKGAVLH
VATLKKDAPPMPYEKTQGRNDYTTFVISAAIKVGATRGTKPLLTPQPREWQCSLYWRDG
QRWIRGGLLGLQAGIVLGPKLNRELLEAVLQRPIECRMSGCGNPLQVRGAAVDFFMTTN
PFYVSGAAYAQKKFKPFGTKRASEDGAAAKAREKLMTQLAKVLDKVVTQAAHSPLDGIW
ETRPEAKLRAMIMALEHEWIFLRPGPCHNAAEEVIKCDCTGGHAILWALIDEARGALEH
KEFYAVTRAHTHDCEKQKLGGRLAGFLDLLIAQDVPLDDAPAARKIKTLLEATPPAPCY
KAATSIATCDCEGKFDKLWAIIDATRAGHGTEDLWARTLAYPQNVNCKCKAGKDLTHRL
ADFLGLLIKRDGPFRERPPHKVTGDRKLVFSGDKKCKGHQYVILAKAHNEEVVRAWISR
WGLKSRTNKAGYAATELNLLLNWLSICRRRWMDMLTVQRDTPYIRMKTGRLVVDDKKER
KAM (서열번호 : 205)

도면7k

>Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348
AKQREALRVALERGIVRASNRTYTLVTNCTKGGPLPEQCRMIERGKARAMKWEPKLVGC
GSCAAATVDLPAIEEYAQPGRLDVAKYKLTTQILAMATRRMMVRAAKLSRRKGQWPAKV
QEEKEEPPEPKKMLKAVEMRPVAIVDFNRVIQTTIEHLWAERANADEAELKALKAAAAY
FGPSLKIRARGPPKAAIGRELKKAHRKKAYAERKKARRKRAELARSQARGAAAHAAIRE
RDIPPMAYERTQGRNDVTTIPIAAAIKIAATRGARPLPAPKPMKWQCSLYWNEGQRWIR
GGMLTAQAYAHAANIHRPMRCEMWGVGNPLKVRAFEGRVADPDGAKGRKAEFRLQTNAF
YVSGAAYRNKKFKPFGTDRGGIGSARKKRERLMAQLAKILDKVVSQAAHSPLDDIWHTR
PAQKLRAMIKQLEHEWMFLRPQAPTVEGTKPDVDVAGNMQRQIKALMAPDLPPIEKGSP
AKRFTGDKRKKGERAVRVAEAHSDEVVTAWISRWGIQTRRNEGSYAAQELELLLNWLQI
CRRRWLDMTAAQRVSPYIRMKSGRMITDAADEGVAPIPLVENM (서열번호 : 206
)

>Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180
KSISGRSIKHMACLKDMLKSEITEIEEKQKKESLRKWDYYSKFSDEILFRRNLNVSANH
DANACYGCNPCAFLKEVYGFRIERRNNERIISYRRGLAGCKSCVQSTGYPPIEFVRRKF
GADKAMEIVREVLHRRNWGALARNIGREKEADPILGELNELLLVDARPYFGNKSAANET
NLAFNVITRAAKKFRDEGMYDIHKQLDIHSEEGKVPKGRKSRLIRIERKHKAIHGLDPG
ETWRYPHCGKGEKYGVWLNRSRLIHIKGNEYRCLTAFGTTGRRMSLDVACSVLGHPLVK
KKRKKGKKTVDGTELWQIKKATETLPEDPIDCTFYLYAAKPTKDPFILKVGSLKAPRWK
KLHKDFFEYSDTEKTQGQEKGKRVVRRGKVPRILSLRPDAKFKVSIWDDPYNGKNKEGT
LLRMELSGLDGAKKPLILKRYGEPNTKPKNFVFWRPHITPHPLTFTPKHDFGDPNKKTK
RRRVFNREYYGHLNDLAKMEPNAKFFEDREVSNKKNPKAKNIRIQAKESLPNIVAKNGR
WAAFDPNDSLWKLYLHWRGRRKTIKGGISQEFQEFKERLDLYKKHEDESEWKEKEKLWE
NHEKEWKKTLEIHGSIAEVSQRCVMQSMMGPLDGLVQKKDYVHIGQSSLKAADDAWTFS
ANRYKKATGPKWGKISVSNLLYDANQANAELISQSISKYLSKQKDNQGCEGRKMKFLIK
IIEPLRENFVKHTRWLHEMTQKDCEVRAQFSRVSM (서열번호 :207)

>Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123
FPSDVGADALKHVRMLQPRLTDEVRKVALTRAPSDRPALARFAAVAQDGLAFVRHLNVS
ANHDSNCTFPRDPRDPRRGPCEPNPCAFLREVWGFRIVARGNERALSYRRGLAGCKSCV
QSTGFPSVPFHRIGADDCMRKLHEILKARNWRLLARNIGREREADPLLTELSEYLLVDA
RTYPDGAAPNSGRLAENVIKRAAKKFRDEGMRDIHAQLRVHSREGKVPKGRLQRLRRIE
RKHRAIHALDPGPSWEAEGSARAEVQGVAVYRSQLLRVGHHTQQIEPVGIVARTLFGVG
RTDLDVAVSVLGAPLTKRKKGSKTLESTEDFRIAKARETRAEDKIEVAFVLYPTASLLR
DEIPKDAFPAMRIDRFLLKVGSVQADREILLQDDYYRFGDAEVKAGKNKGRTVTRPVKV
PRLQALRPDAKFRVNVWADPFGAGDSPGTLLRLEVSGVTRRSQPLRLLRYGQPSTQPAN
FLCWRPHRVPDPMTFTPRQKFGERRKNRRTRRPRVFERLYQVHIKHLAHLEPNRKWFEE
ARVSAQKWAKARAIRRKGAEDIPVVAPPAKRRWAALQPNAELWDLYAHDREARKRFRGG
RAAEGEEFKPRLNLYLAHEPEAEWESKRDRWERYEKKWTAVLEEHSRMCAVADRTLPQF
LSDPLGARMDDKDYAFVGKSALAVAEAFVEEGTVERAQGNCSITAKKKFASNASRKRLS
VANLLDVSDKADRALVFQAVRQYVQRQAENGGVEGRRMAFLRKLLAPLRQNFVCHTRWL
HM (서열번호 : 208)

도면71

>Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom
AARKKKRGKIGITVKAKEKSPPAAGPFMARKLVNVAANVDGVEVHLCVECEADAHGSAS
ARLLGGCRSCTGSIGAEGRLMGSVDVDRERVIAEPVHTETERLGPDVKAFEAGTAESKY
AIQRGLEYWGVDLISRNRARTVRKMEEADRPESSTMEKTSWDEIAIKTYSQAYHASENH
LFWERQRRVRQHALALFRRARERNRGESPLQSTQRPAPLVLAALHAEAAAISGRARAEY
VLRGPSANVRAAAADIDAKPLGHYKTPSPKVARGFPVKRDLLRARHRIVGLSRAYFKPS
DVVRGTSDAIAHVAGRNIGVAGGKPKEIEKTFTLPFVAYWEDVDRVVHCSSFKADGPWV
RDQRIKIRGVSSAVGTFSLYGLDVAWSKPTSFYIRCSDIRKKFHPKGFGPMKHWRQWAK
ELDRLTEQRASCVVRALQDDEELLQTMERGQRYYDVFSCAATHATRGEADPSGGCSRCE
LVSCGVAHKVTKKAKGDTGIEAVAVAGCSLCESKLVGPSKPRVHRQMAALRQSHALNYL
RRLQREWEALEAVQAPTPYLRFKYARHLEVRSM (**서열번호** : 209)

>Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081
AAKKKKQRGKIGISVKPKEGSAPPADGPFMARKLVNVAANVDGVEVNLCIECEADAHGS
APARLLGGCKSCTGSIGAEGRLMGSVDVDRADAIAKPVNTETEKLGPDVQAFEAGTAET
KYALQRGLEYWGVDLISRNRSRTVRRTEEGQPESATMEKTSWDEIAIKSYTRAYHASEN
HLFWERQRRVRQHALALFKRAKERNRGDSTLPREPGHGLVAIAALACEAYAVGGRNLAE
TVVRGPTFGTARAVRDVEIASLGRYKTPSPKVAHGSPVKRDFLRARHRIVGLARAYYRP
SDVVRGTSDAIAHVAGRNIGVAGGKPRAVEAVFTLPFVAYWEDVDRVVHCSSFQVSAPW
NRDQRMKIAGVTTAAGTFSLHGGELKWAKPTSFYIRCSDTRRKFRPKGFGPMKRWRQWA
KDLDRLVEQRASCVVRALQDDAALLETMERGQRYYDVFACAVTHATRGEADRLAGCSRC
ALTPCQEAHRVTTKPRGDAGVEQVQTSDCSLCEGKLVGPSKPRLHRTLTLLRQEHGLNY
LRRLQREWESLEAVQVPTPYLRFKYARHLEVRSM (**서열번호** : 210)

>Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504
TDSQSESVPEVVYALTGGEVPGRVPPDGGSAEGARNAPTGLRKQRGKIKISAKPSKPGS
PASSLARTLVNEAANVDGVQSSGCATCRMRANGSAPRALPIGCVACASSIGRAPQEETV
CALPTTQGPDVRLLEGGHALRKYDIQRALEYWGVDLIGRNLDRQAGRGMEPAEGATATM
KRVSMDELAVLDFGKSYYASEQHLFAARQRRVRQHAKALKIRAKHANRSGSVKRALDRS
RKQVTALAREFFKPSDVVRGDSDALAHVVGRNLGVSRHPAREIPQTFTLPLCAYWEDVD
RVISCSSLLAGEPFARDQEIRIEGVSSALGSLRLYRGAIEWHKPTSLYIRCSDTRRKFR
PRGGLKKRWRQWAKDLDRLVEQRACCIVRSLQADVELLQTMERAQRFYDVHDCAATHVG
PVAVRCSPCAGKQFDWDRYRLLAALRQEHALNYLRRLQREWESLEAQQVKMPYLRFKYA
RKLEVSGPLIGLEVRREPSMGTAIAEM (**서열번호** : 211)

>Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom
AGTAGRRHGSLGARRSINIAGVTDRHGRWGCESCVYTRDQAGNRARCAPCDQSTYAPDV
QEVTIGQRQAKYTIFLTLQSFSWTNTMRNNKRAAAGRSKRTTGKRIGQLAEIKITGVGL
AHAHNVIQRSLQHNITKMWRAEKGKSKRVARLKKAKQLTKRRAYFRRRMSRQSRGNGFF
RTGKGGIHAVAPVKIGLDVGMIASGSSEPADEQTVTLDAIWKGRKKKIRLIGAKGELAV
AACRFREQQTKGDKCIPLILQDGEVRWNQNNWQCHPKKLVPLCGLEVSRKFVSQADRLA
QNKVASPLAARFDKTSVKGTLVESDFAAVLVNVTSIYQQCHAMLLRSQEPTPSLRVQRT
ITSM (**서열번호** : 212)

도면7m

>Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1605|rev
com
GVRFSPAQSQVFFRTVIPQSVEARFAINMAAIHDAAGAFGCSVCRFEDRTPRNAKAVHG
CSPCTRSTNRPDVFVLPVGAIKAKYDVFMRLLGFNWTHLNRRQAKRVTVRDRIGQLDEL
AISMLTGKAKAVLKKSICHNVDKSFKAMRGSLKKLHRKASKTGKSQLRAKLSDLRERTN
TTQEGSHVEGDSDVALNKIGLDVGLVGKPDYPSEESVEVVVCLYFVGKVLILDAQGRIR
DMRAKQYDGFKIPIIQRGQLTVLSVKDLGKWSLVRQDYVLAGDLRFEPKISKDRKYAEC
VKRIALITLQASLGFKERIPYYVTKQVEIKNASHIAFVTEAIQNCAENFREMTEYLMKY
QEKSPDLKVLLTQLM (**서열번호** : 213)

>Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188.
.509648
RAVVGKVFLEQARRALNLATNFGTNHRTGCNGCYVTPGKLSIPQDGEKNAAGCTSCLMK
ATASYVSYPKPLGEKVAKYSTLDALKGFPWYSLRLNLRPNYRGKPINGVQEVAPVSKFR
LAEEVIQAVQRYHFTELEQSFPGGRRRLRELRAFYTKEYRRAPEQRQHVVNGDRNIVVV
TVLHELGFSVGMFNEVELLPKTPIECAVNVFIRGNRVLLEVRKPQFDKERLLVESLWKK
DSRRHTAKWTPPNNEGRIFTAEGWKDFQLPLLLGSTSRSLRAIEKEGFVQLAPGRDPDY
NNTIDEQHSGRPFLPLYLYLQGTISQEYCVFAGTWVIPFQDGISPYSTKDTFQPDLKRK
AYSLLLDAVKHRLGNKVASGLQYGRFPAIEELKRLVRMHGATRKIPRGEKDLLKKGDPD
TPEWWLLEQYPEFWRLCDAAAKRVSQNVGLLLSLKKQPLWQRRWLESRTRNEPLDNLPL
SMALTLHLTNEEAL (**서열번호** : 214)

>Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..856
57
AAVYSKFYIENHFKMGIPETLSRIRGPSIIQGFSVNENYINIAGVGDRDFIFGCKKCKY
TRGKPSSKKINKCHPCKRSTYPEPVIDVRGSISEFKYKIYNKLKQEPNQSIKQNTKGRM
NPSDHTSSNDGIIINGIDNRIAYNVIFSSYKHLMEKQINLLRDTTKRKARQIKKYNNSG
KKKHSLRSQTKGNLKNRYHMLGMFKKGSLTITNEGDFITAVRKVGLDISLYKNESLNKQ
EVETELCLNIKWGRTKSYTVSGYIPLPINIDWKLYLFEKETGLTLRLFGNKYKIQSKKF
LIAQLFKPKRPPCADPVVKKAQKWSALNAHVQQMAGLFSDSHLLKRELKNRMHKQLDFK
SLWVGTEDYIKWFEELSRSYVEGAEKSLEFFRQDYFCFNYTKQTTM (**서열번호** : 215
)

>Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286
PQQQRDLMLMAANYDQDYGNGCGPCTVVASAAYRPDPQAQHGCKRHLRTLGASAVTHVG
LGDRTATITALHRLRGPAALAARARAAQAASAPMTPDTDAPDDRRRLEAIDADDVVLVG
AHRALWSAVRRWADDRRAALRRRLHSEREWLLKDQIRWAELYTLIEASGTPPQGRWRNT
LGALRGQSRWRRVLAPTMRATCAETHAELWDALAELVPEMAKDRRGLLRPPVEADALWR
APMIVEGWRGGHSVVVDAVAPPLDLPQPCAWTAVRLSGDPRQRWGLHLAVPPLGQVQPP
DPLKATLAVSMRHRGGVRVRTLQAMAVDADAPMQRHLQVPLTLQRGGGLQWGIHSRGVR
RREARSMASWEGPPIWTGLQLVNRWKGQGSALLAPDRPPDTPPYAPDAAVAPAQPDTKR
ARRTLKEACTVCRCAPGHMRQLQVTLTGDGTWRRFRLRAPQGAKRKAEVLKVATQHDER
IANYTAWYLKRPEHAAGCDTCDGDSRLDGACRGCRPLLVGDQCFRRYLDKIEADRDDGL
AQIKPKAQEAVAAMAAKRDARAQKVAARAAKLSEATGQRTAATRDASHEARAQKELEAV
ATEGTTVRHDAAAVSAFGSWVARKGDEYRHQVGVLANRLEHGLRLQELMAPDSVVADQQ
RASGHARVGYRYVLTAM (**서열번호** :216)

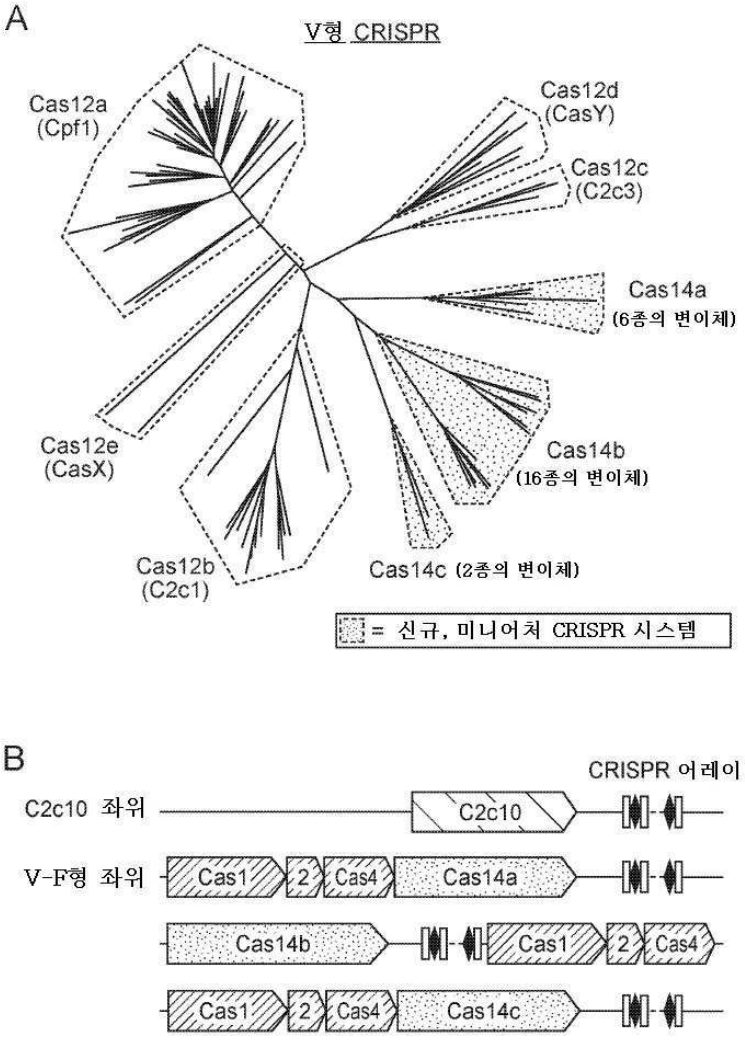
도면7n

>Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revc
om
AVAHPVGRGNAGSPGARGPEELPRQLVNRASNVTRPATYGCAPCRHVRLSIPKPVLTGC
RACEQTTHPAPKRAVRGGADAAKYDLAAFFAGWAADLEGRNRRRQVHAPLDPQPDPNHE
PAVTLQKIDLAEVSIEEFQRVLARSVKHRHDGRASREREKARAYAQVAKKRRNSHAHGA
RTRRAVRRQTRAVRRAHRMGANSGEILVASGAEDPVPEAIDHAAQLRRRIRACARDLEG
LRHLSRRYLKTLEKPCRRPRAPDLGRARCHALVESLQAAERELEELRRCDSPDTAMRRL
DAVLAAAAASTDATFATGWTVVGMDLGVAPRGSAAPEVSPMEMAISVFWRKGSRRVIVSK
PIAGMPIRRHELIRLEGLGTLRLDGNHYTGAGVTKGRGLSEGTEPDFREKSPSTLGFTL
SDYRHESRWRPYGAKQGKTARQFFAAMSRELRALVEHQVLAPMGPPLLEAHERRFETLL
KGQDNKSIHAGGGGRYVWRGPPDSKKRPAADGDWFRFGRGHADHRGWANKRHELAANYL
QSAFRLWSTLAEAQEPTPYARYKYTRVTM (**서열번호** :217)

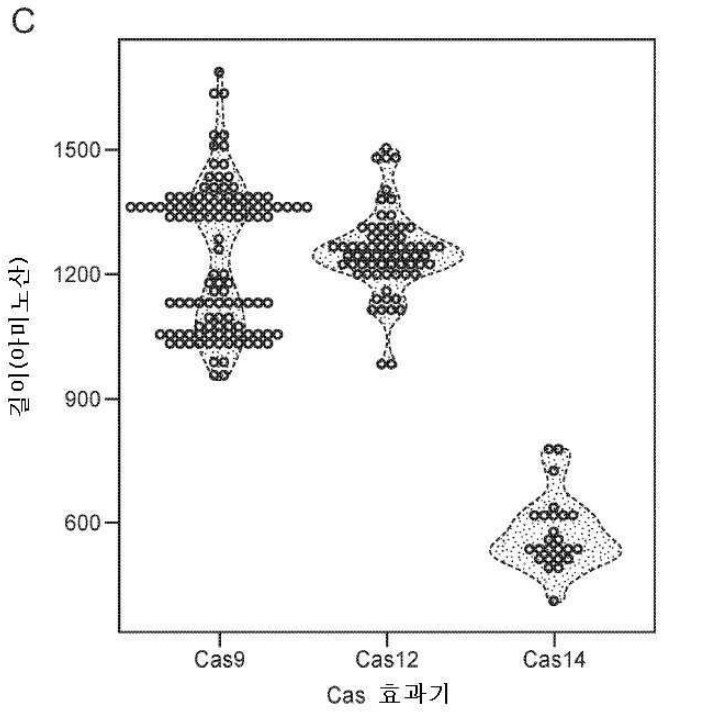
>Cas14u.7|3300001256.a|JGI12210J13797_10004690|5792..7006
WDFLTLQVYERHTSPEVCVAGNSTKCASGTRKSDHTHGVGVKLGAQEINVSANDDRDHE
VGCNICVISRVSLDIKGWRYGCESCVQSTPEWRSIVRFDRNHKEAKGECLSRFEYWGAQ
SIARSLKRNKLMGGVNLDELAIVQNENVVKTSLKHLFDKRKDRIQANLKAVKVRMRERR
KSGRQRKALRRQCRKLKRYLRSYDPSDIKEGNSCSAFTKLGLDIGISPNKPPKIEPKVE
VVFSLFYQGACDKIVTVSSPESPLPRSWKIKIDGIRALYVKSTKVKFGGRTFRAGQRNN
RRKVRPPNVKKGKRKGSRSQFFNKFAVGLDAVSQQLPIASVQGLWGRAETKKAQTICLK
QLESNKPLKESQRCLFLADNWVVRVCGFLRALSQRQGPTPYIRYRYRCNM (**서열번호**
218)

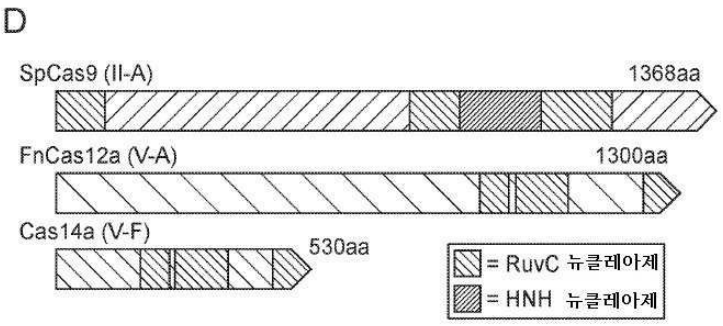
>Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943
ARNVGQRNASRQSKRESAKARSRRVTGGHASVTQGVALINAAANADRDHTTGCEPCTWE
RVNLPLQEVIHGCDSCTKSSPFWRDIKVVNKGYREAKEEIMRIASGISADHLSRALSHN
KVMGRLNLDEVCILDFRTVLDTSLKHLTDSRSNGIKEHIRAVHRKIRMRRKSGKTARAL
RKQYFALRRQWKAGHKPNSIREGNSLTALRAVGFDVGVSEGTEPMPAPQTEVVLSVFYK
GSATRILRISSPHPIAKRSWKVKIAGIKALKLIRREHDFSFGRETYNASQRAEKRKFSP
HAARKDFFNSFAVQLDRLAQQLCVSSVENLWVTEPQQKLLTLAKDTAPYGIREGARFAD
TRARLAWNWVFRVCGFTRALHQEQEPTPYCRFTWRSKM (**서열번호** :219)

도면8ab

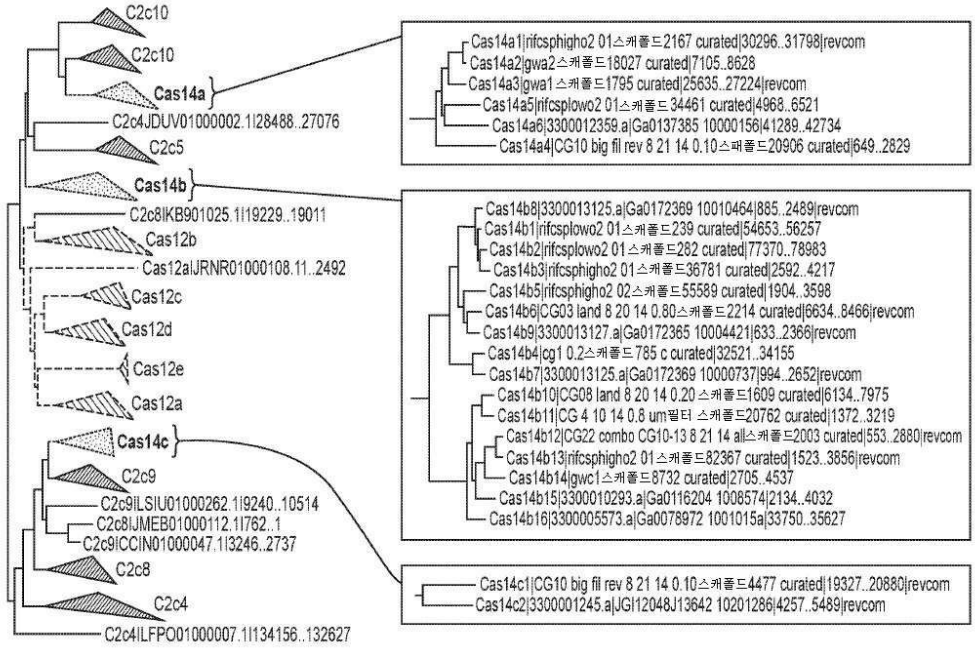

A
V형 CRISPR
Cas12a
(Cpf1)
Cas12d
(CasY)
Cas12c
(C2c3)
Cas14a
(6종의 변이체)
Cas14b
(16종의 변이체)
Cas12e
(CasX)
Cas12b
(C2c1)
Cas14c (2종의 변이체)
= 신규, 미니어처 CRISPR 시스템
B
CRISPR 어레이
C2c10 좌위
C2c10
V-F형 좌위
Cas1
2
Cas4
Cas14a
Cas14b
Cas1
2
Cas4
Cas1
2
Cas4
Cas14c

도면8cd

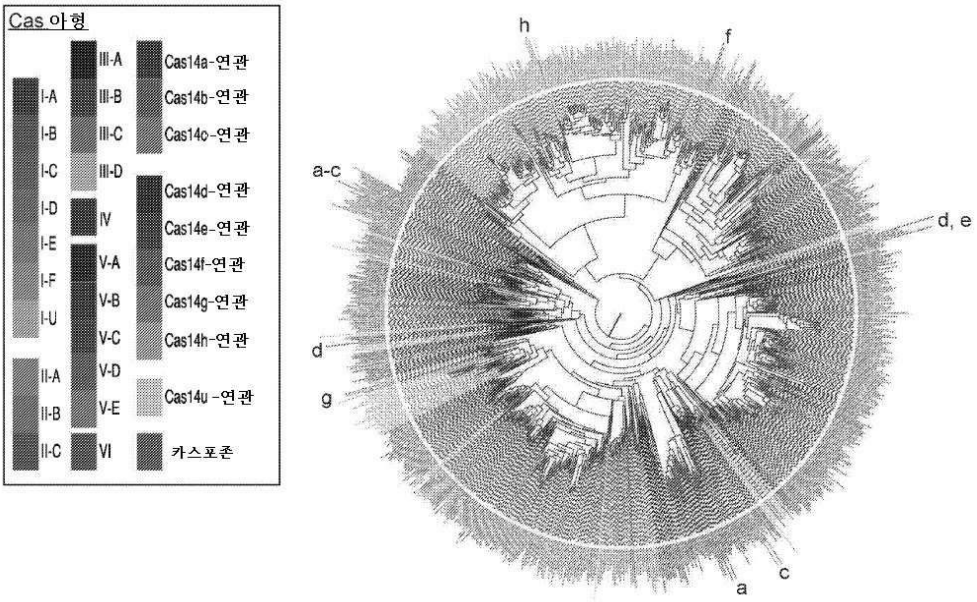

C
1500
1200
900
600
길이(아미노산)
Cas9
Cas12
Cas14
Cas 효과기


D
SpCas9 (II-A)
1368aa
FnCas12a (V-A)
1300aa
Cas14a (V-F)
530aa
= RuvC 뉴클레아제
= HNH 뉴클레아제

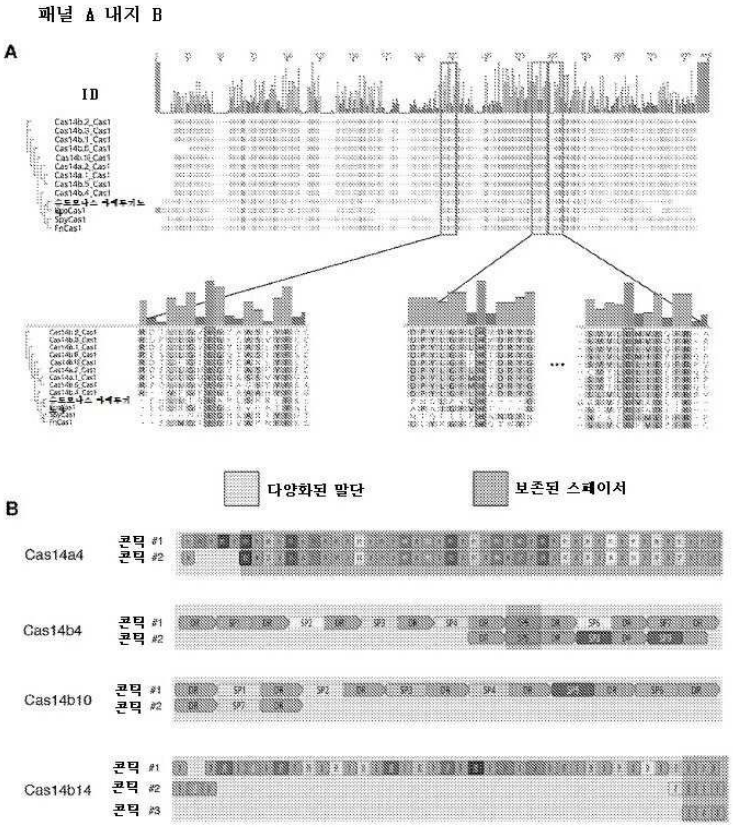
도면9


C2c10
C2c10
Cas14a
C2c4JDUV01000002.1|28488..27076
C2c5
Cas14b
C2c8|KB901025.1|19229..19011
Cas12b
Cas12a|JRNR01000108.11..2492
Cas12c
Cas12d
Cas12e
Cas12a
Cas14c
C2c9
C2c9|LSIU01000262.1|9240..10514
C2c8|JMEB01000112.1|762..1
C2c9|CCIN01000047.1|3246..2737
C2c8
C2c4
C2c4|LFPO01000007.1|134156..132627
Cas14a1|rifcsphigho2 01스캐폴드2167 curated|30296..31798|revcom
Cas14a2|gwa2스캐폴드18027 curated|7105..8628
Cas14a3|gwa1스캐폴드1795 curated|25635..27224|revcom
Cas14a5|rifcsplowo2 01스캐폴드34461 curated|4968..6521
Cas14a6|3300012359.a|Ga0137385 10000156|41289..42734
Cas14a4|CG10 big fil rev 8 21 14 0.10스캐폴드20906 curated|649..2829
Cas14b8|3300013125.a|Ga0172369 10010464|885..2489|revcom
Cas14b1|rifcsplowo2 01스캐폴드239 curated|54653..56257
Cas14b2|rifcsplowo2 01스캐폴드282 curated|77370..78983
Cas14b3|rifcsphigho2 01스캐폴드36781 curated|2592..4217
Cas14b5|rifcsphigho2 02스캐폴드55589 curated|1904..3598
Cas14b6|CG03 land 8 20 14 0.80스캐폴드2214 curated|6634..8466|revcom
Cas14b9|3300013127.a|Ga0172365 10004421|633..2366|revcom
Cas14b4|cg1 0.2스캐폴드785 c curated|32521..34155
Cas14b7|3300013125.a|Ga0172369 10000737|994..2652|revcom
Cas14b10|CG08 land 8 20 14 0.20 스캐폴드1609 curated|6134..7975
Cas14b11|CG 4 10 14 0.8 um필터 스캐폴드20762 curated|1372..3219
Cas14b12|CG22 combo CG10-13 8 21 14 all스캐폴드2003 curated|553..2880|revcom
Cas14b13|rifcsphigho2 01스캐폴드82367 curated|1523..3856|revcom
Cas14b14|gwc1스캐폴드8732 curated|2705..4537
Cas14b15|3300010293.a|Ga0116204 1008574|2134..4032
Cas14b16|3300005573.a|Ga0078972 1001015a|33750..35627
Cas14c1|CG10 big fil rev 8 21 14 0.10스캐폴드4477 curated|19327..20880|revcom
Cas14c2|3300001245.a|JGI12048J13642 10201286|4257..5489|revcom

도면10

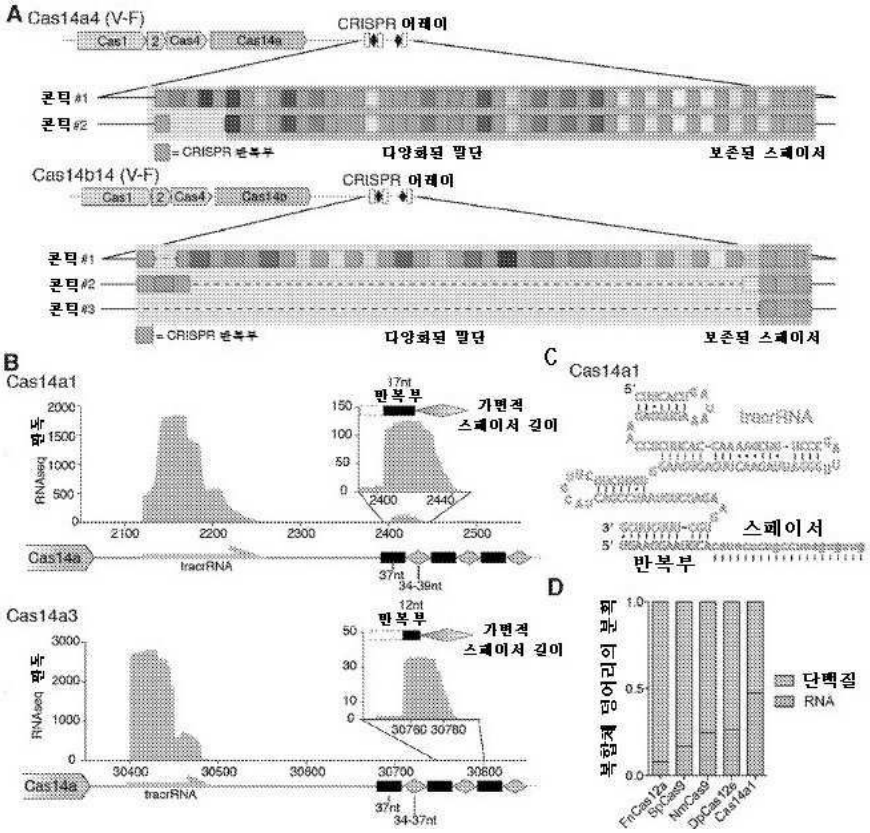

Cas 아형
I-A
I-B
I-C
I-D
I-E
I-F
I-U
II-A
II-B
II-C
III-A
III-B
III-C
III-D
IV
V-A
V-B
V-C
V-D
V-E
VI
Cas14a-연관
Cas14b-연관
Cas14c-연관
Cas14d-연관
Cas14e-연관
Cas14f-연관
Cas14g-연관
Cas14h-연관
Cas14u-연관
카스포존
h
f
a-c
d, e
d
g
a
c

도면11

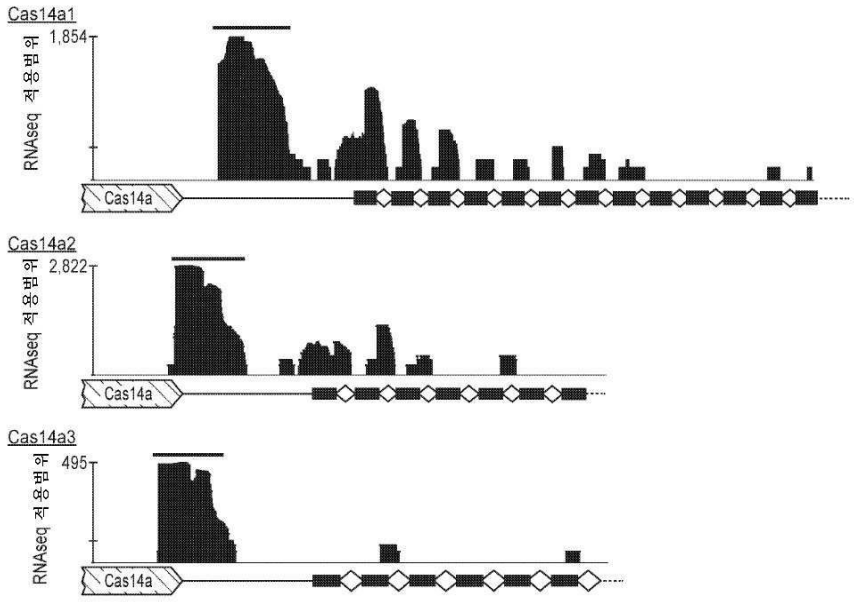

패널 A 내지 B
A
ID
다양화된 말단
보존된 스페이서
B
Cas14a4
콘틱 #1
콘틱 #2
Cas14b4
콘틱 #1
콘틱 #2
Cas14b10
콘틱 #1
콘틱 #2
Cas14b14
콘틱 #1
콘틱 #2
콘틱 #3

도면12

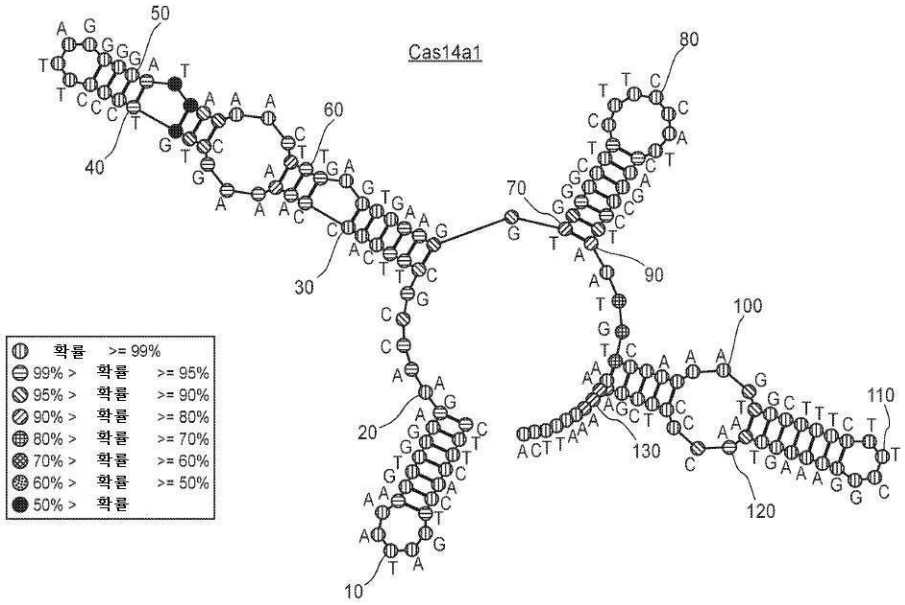
패널 A 내지 D



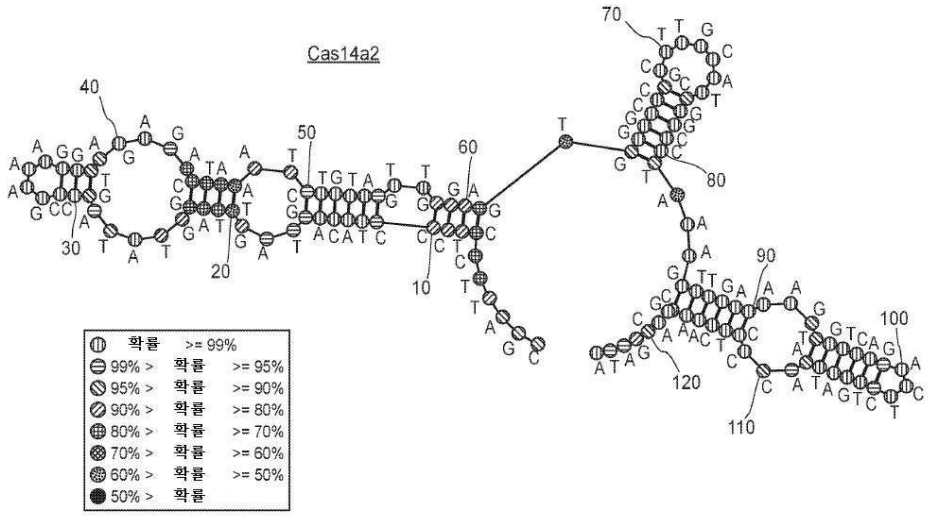
도면13aa



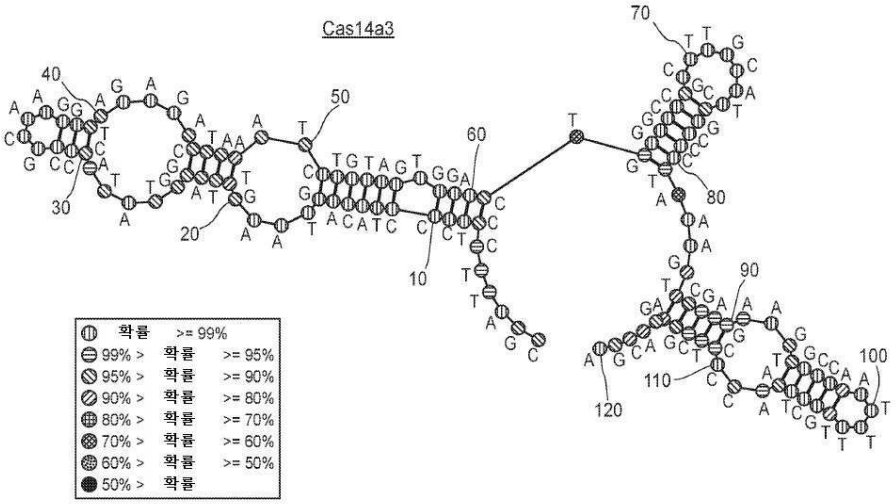
도면13ab



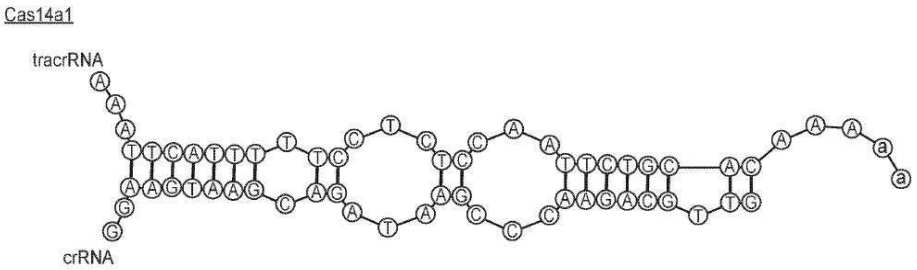
도면13ac



도면13ad

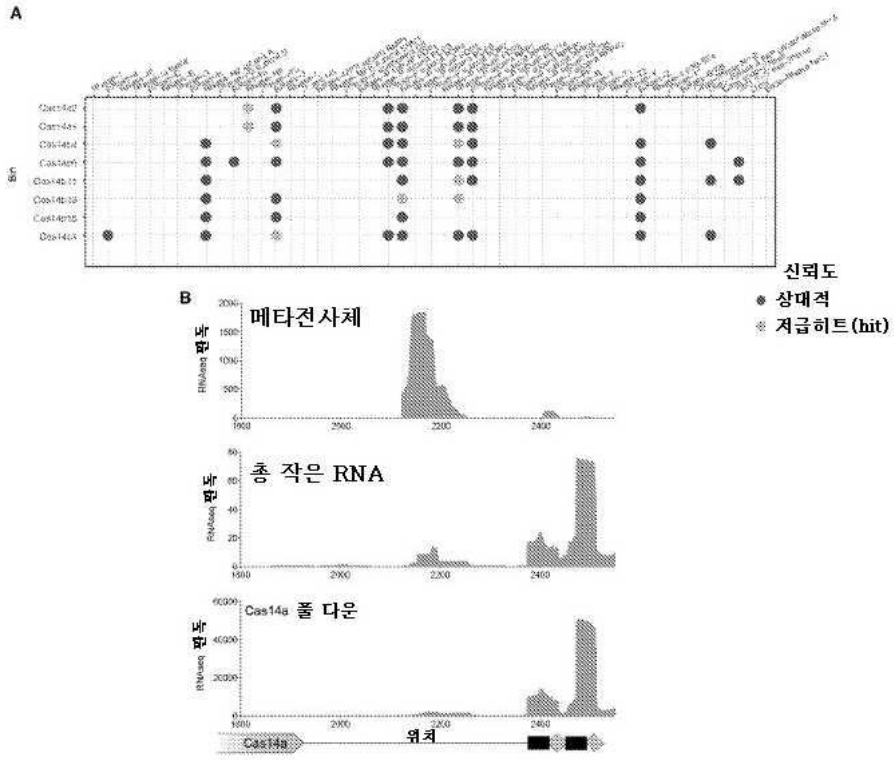


도면13b

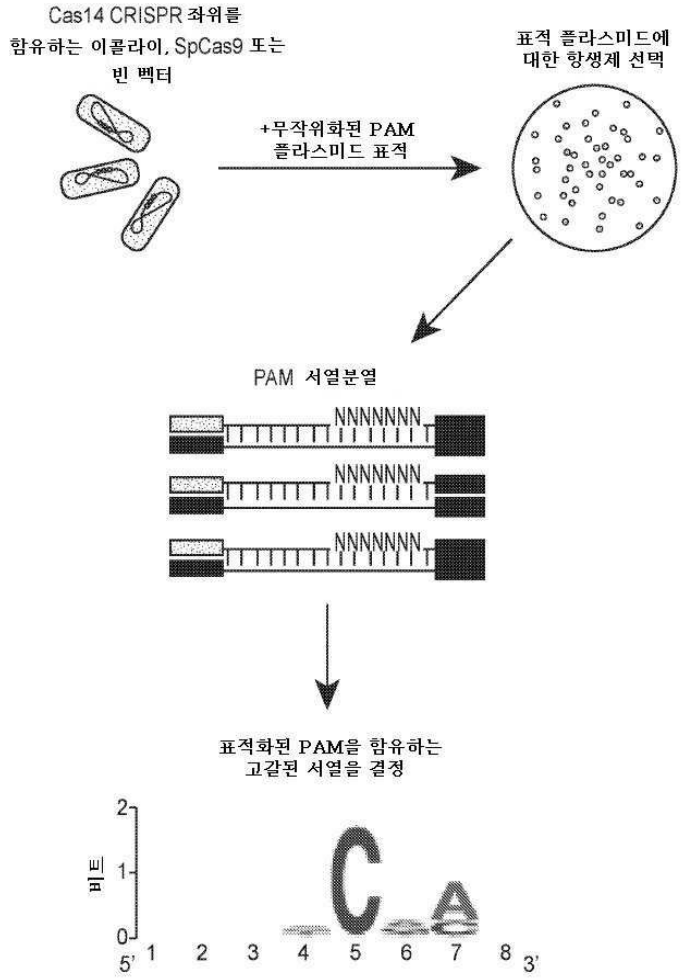


도면14

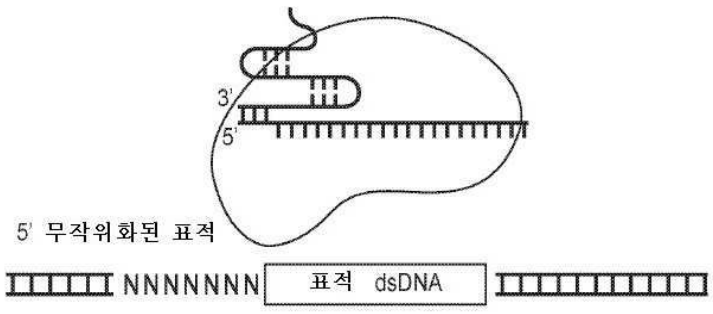
패널 A 내지 B



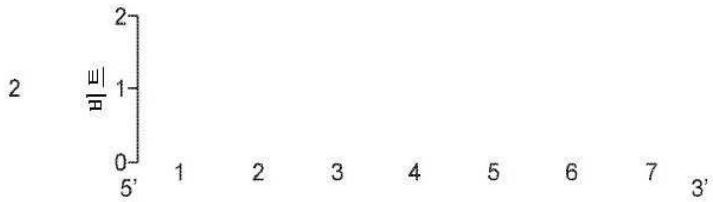
도면15a

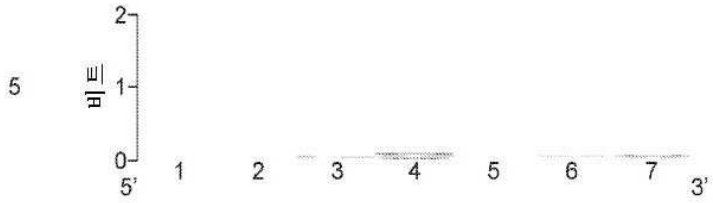

Cas14 CRISPR 좌위를
함유하는 이콜라이, SpCas9 또는
빈 벡터
+무작위화된 PAM
플라스미드 표적
표적 플라스미드에
대한 항생제 선택
PAM 서열분열
NNNNNNN
NNNNNNN
NNNNNNN
표적화된 PAM을 함유하는
고갈된 서열을 결정
비트
2
1
0
5'
1
2
3
4
5
6
7
8
3'

도면15b



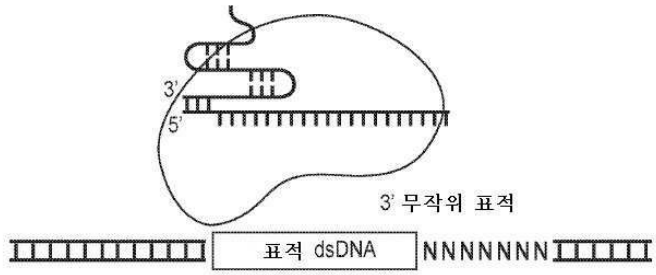
PDVT



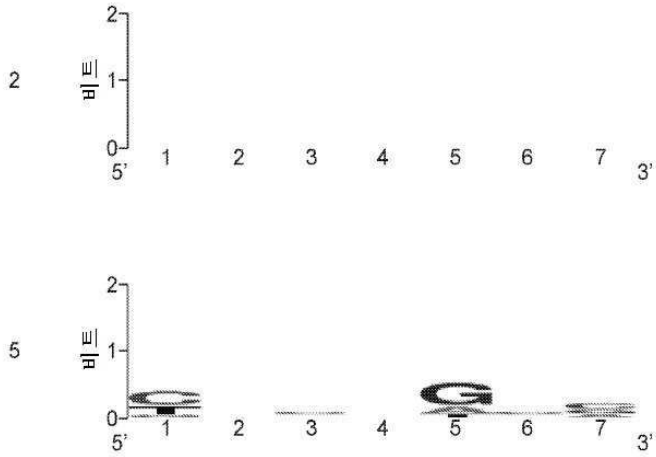


20 서열없음

도면15c

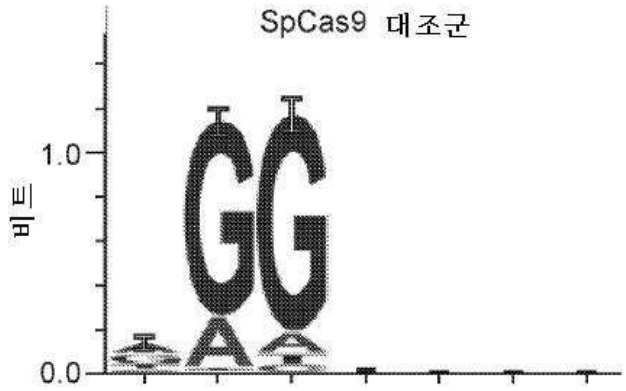


PDVT

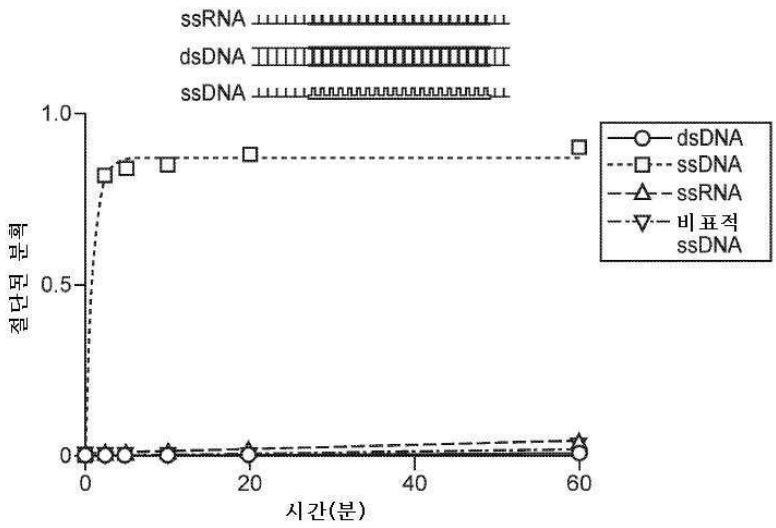


20 서열 없음

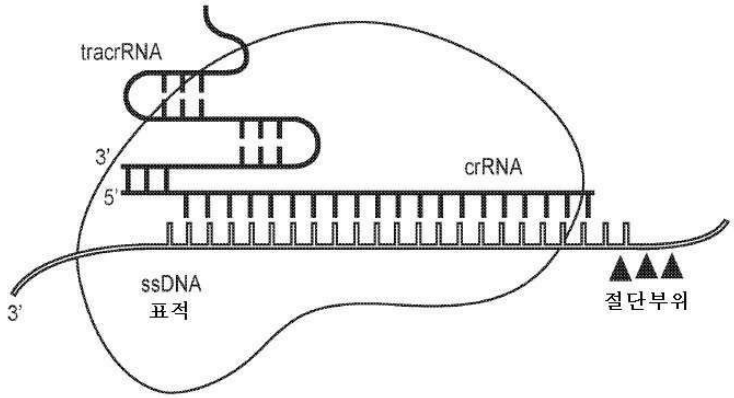
도면15d

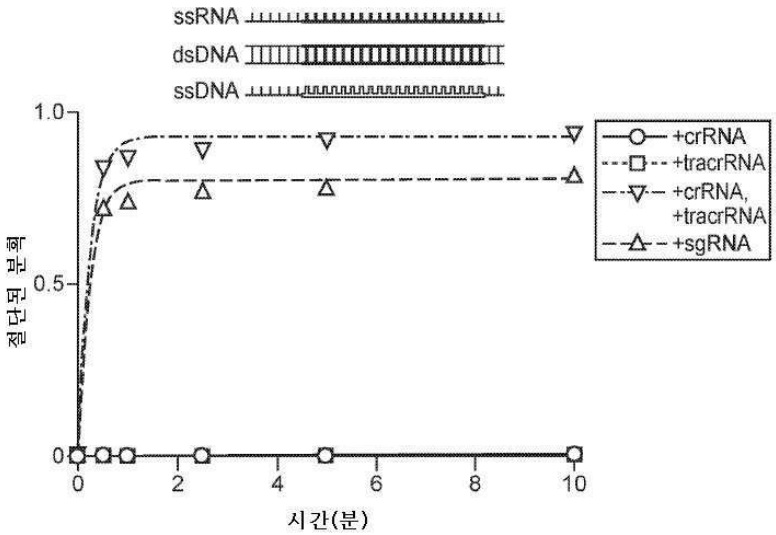


도면16a

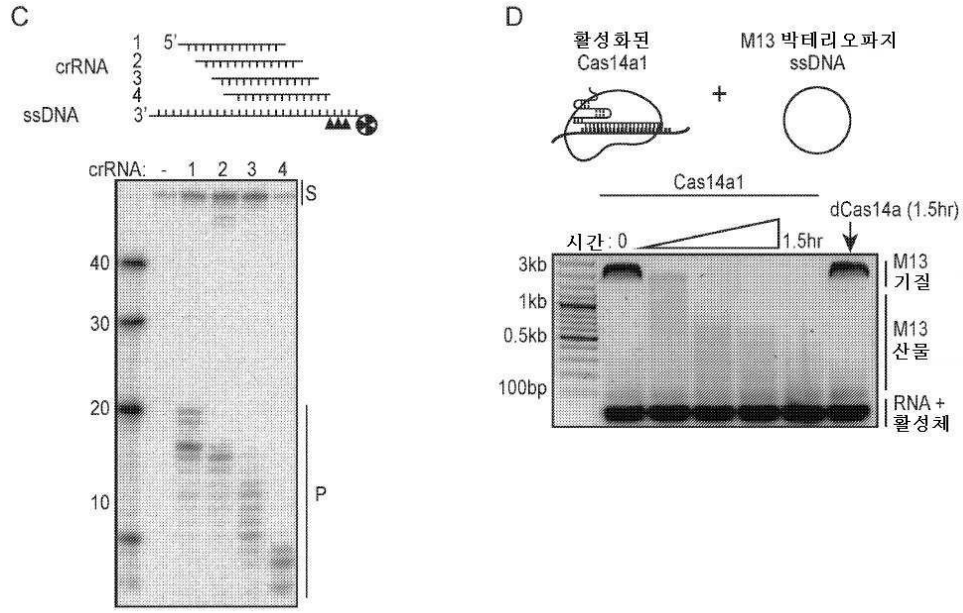

ssRNA
dsDNA
ssDNA
1.0
0.5
0
절단된 분획
0
20
40
60
시간(분)
dsDNA
ssDNA
ssRNA
비 표적 ssDNA

도면16b

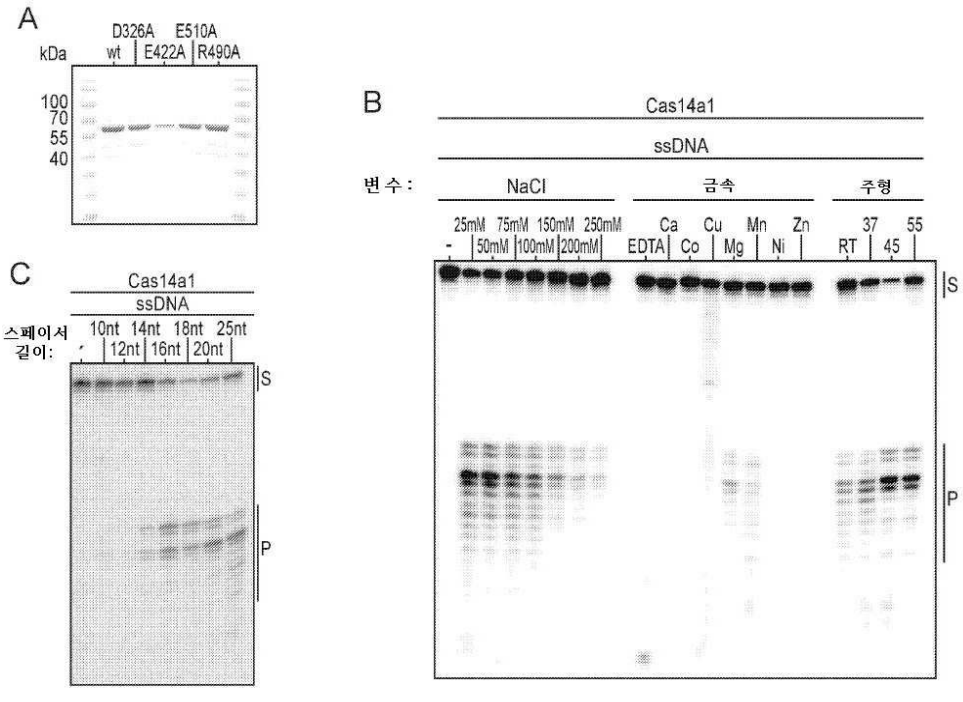

tracrRNA
3'
5'
crRNA
ssDNA
표적
3'
절단부위


ssRNA
dsDNA
ssDNA
1.0
0.5
0
절단된 분획
0
2
4
6
8
10
시간(분)
+crRNA
+tracrRNA
+crRNA, +tracrRNA
+sgRNA

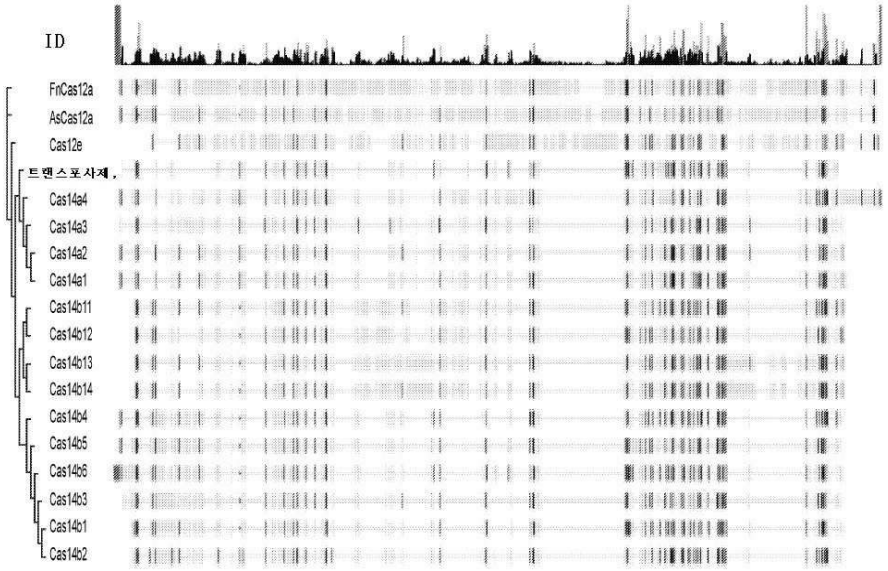
도면16cd



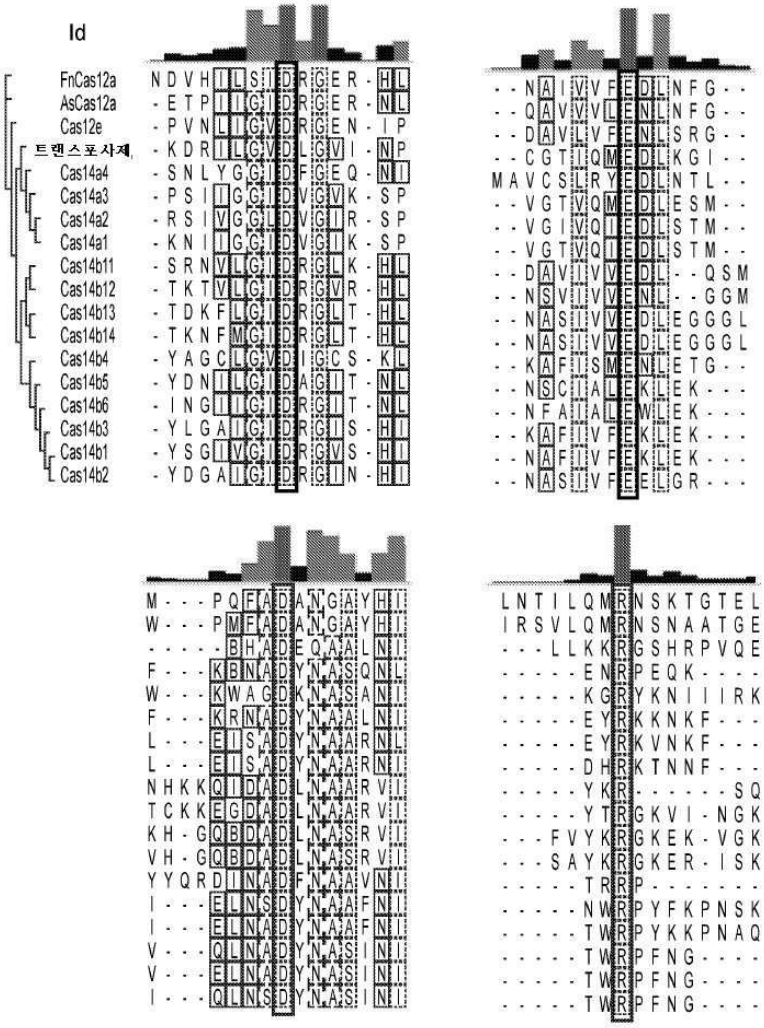
도면17ac



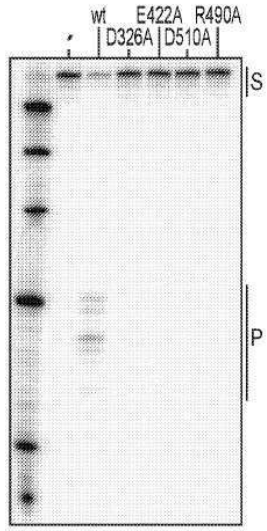
도면17da



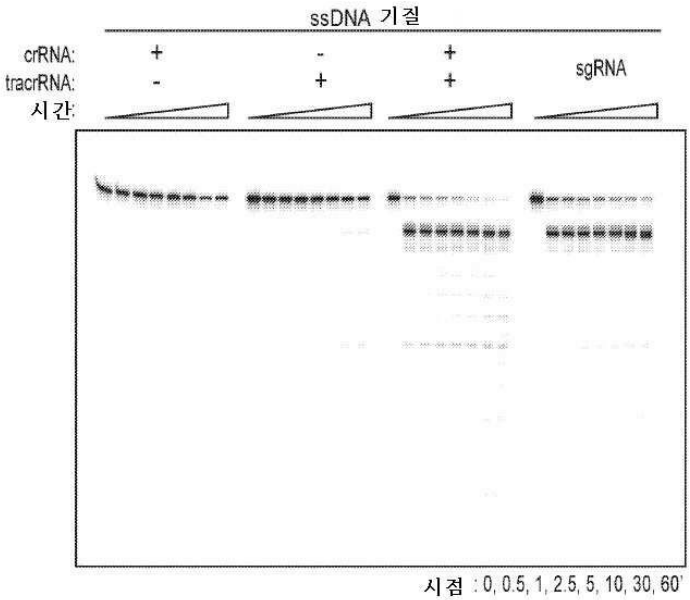
도면17db



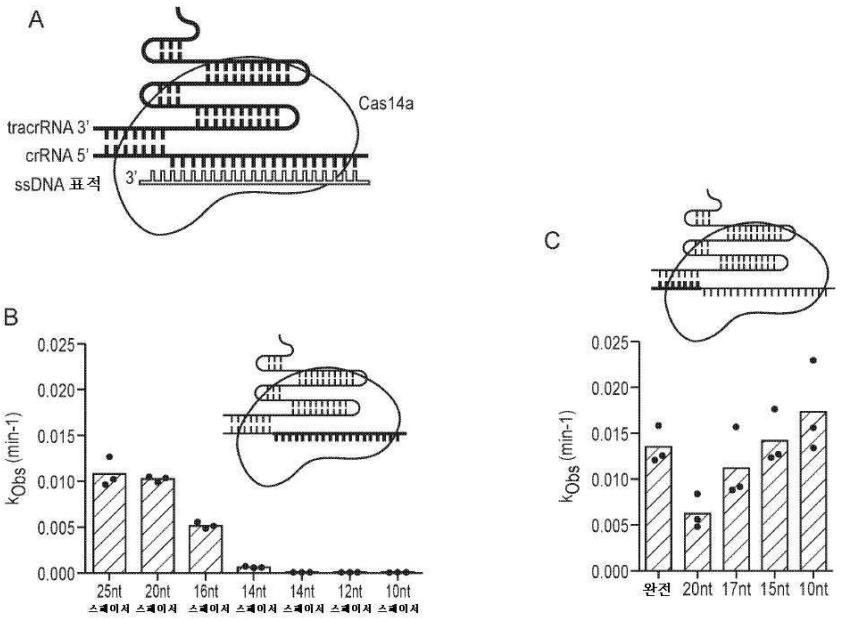
도면17e


wt
E422A
R490A
D326A
D510A
S
P

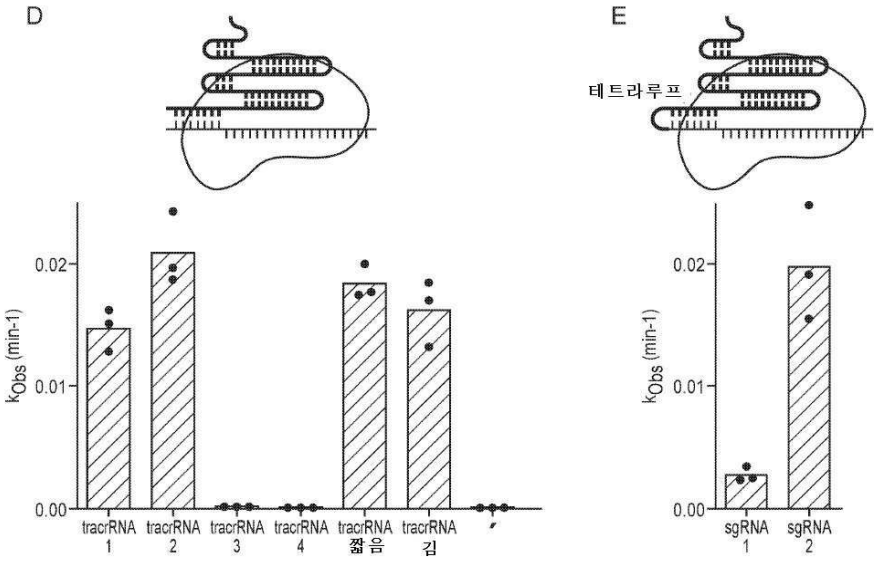
도면18


ssDNA 기질
crRNA:
tracrRNA:
시간:
+
-
-
+
+
+
sgRNA
시점 : 0, 0.5, 1, 2.5, 5, 10, 30, 60'

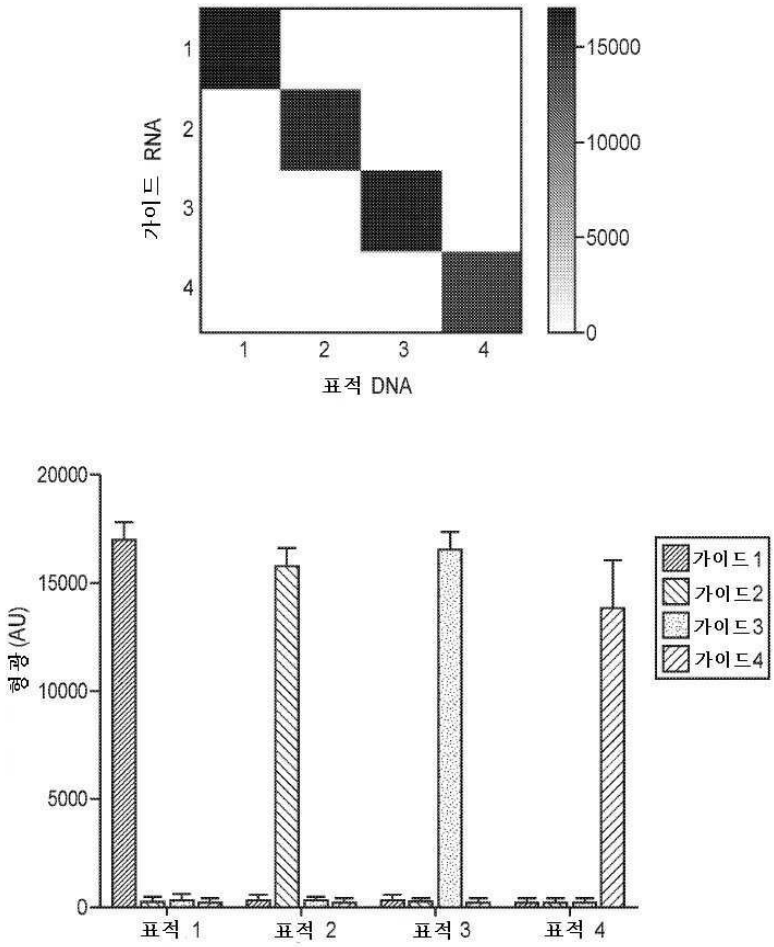
도면19ac



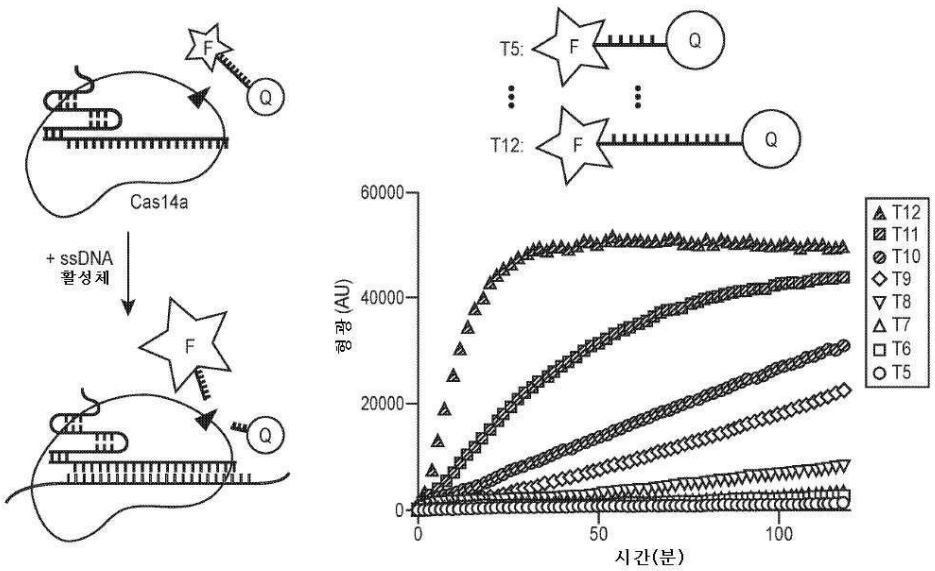
도면19de



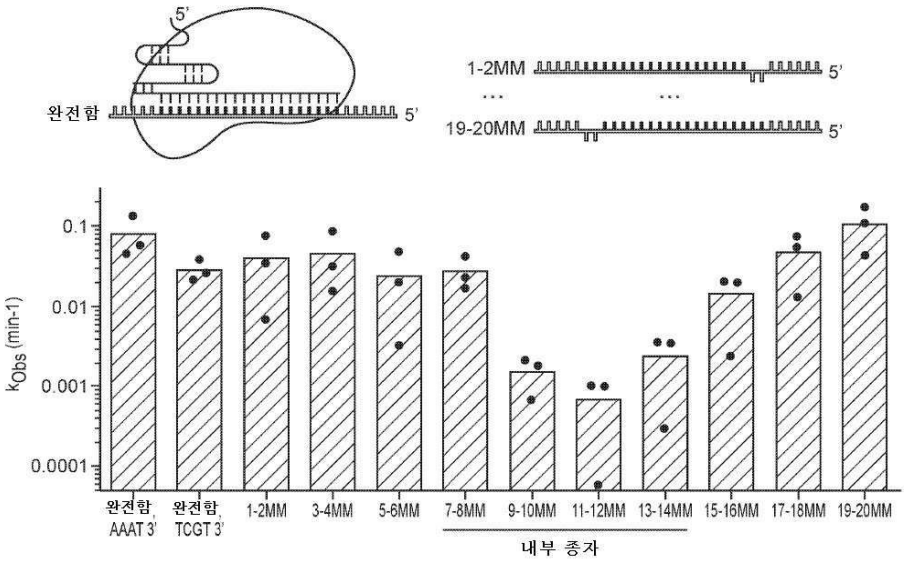
도면19f


가이드 RNA
1
2
3
4
표적 DNA
15000
10000
5000
0
형광 (AU)
20000
15000
10000
5000
0
표적 1
표적 2
표적 3
표적 4
가이드1
가이드2
가이드3
가이드4

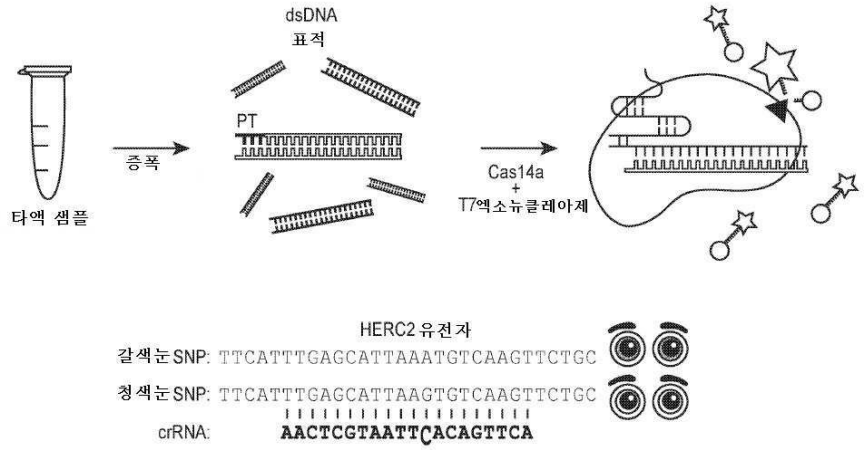
도면20a


F
Q
Cas14a
+ ssDNA
활성체
T5:
T12:
형광 (AU)
60000
40000
20000
0
50
100
시간(분)
T12
T11
T10
T9
T8
T7
T6
T5

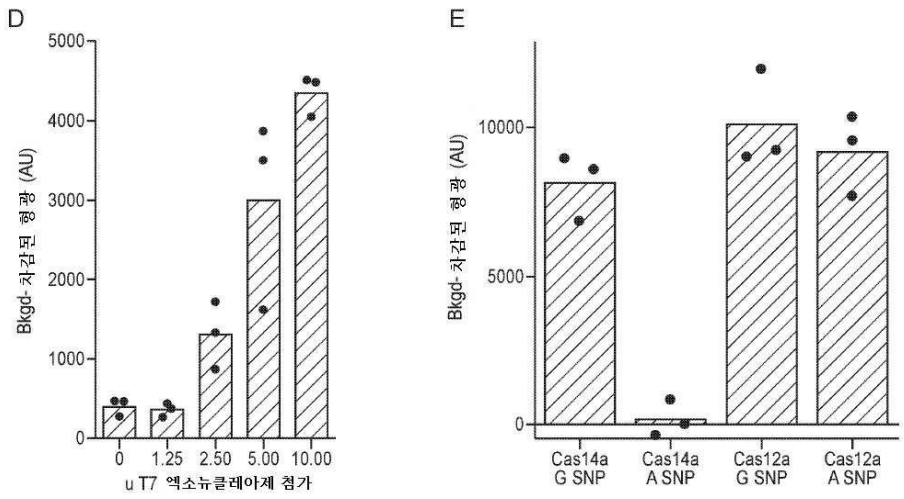
도면20b



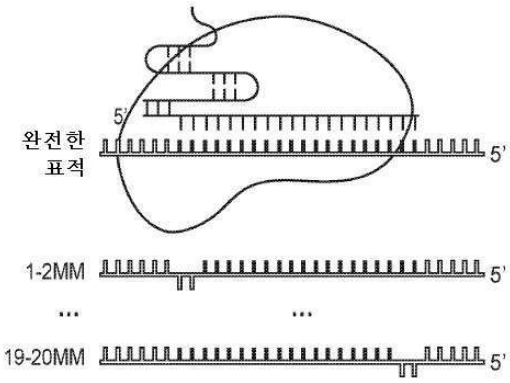
도면20c

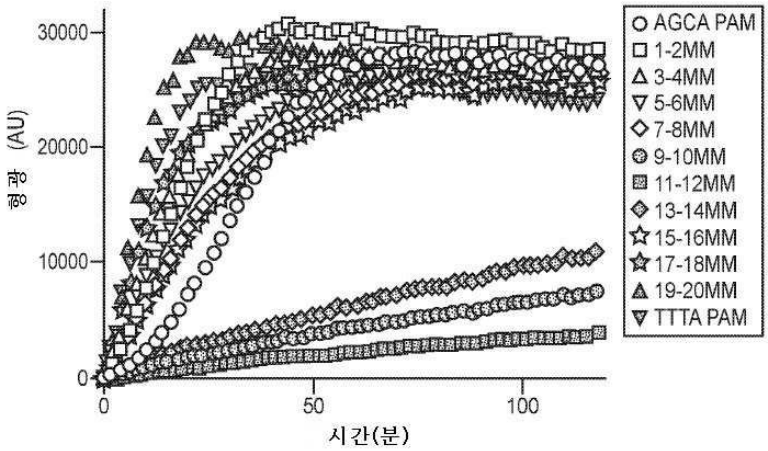


도면20de

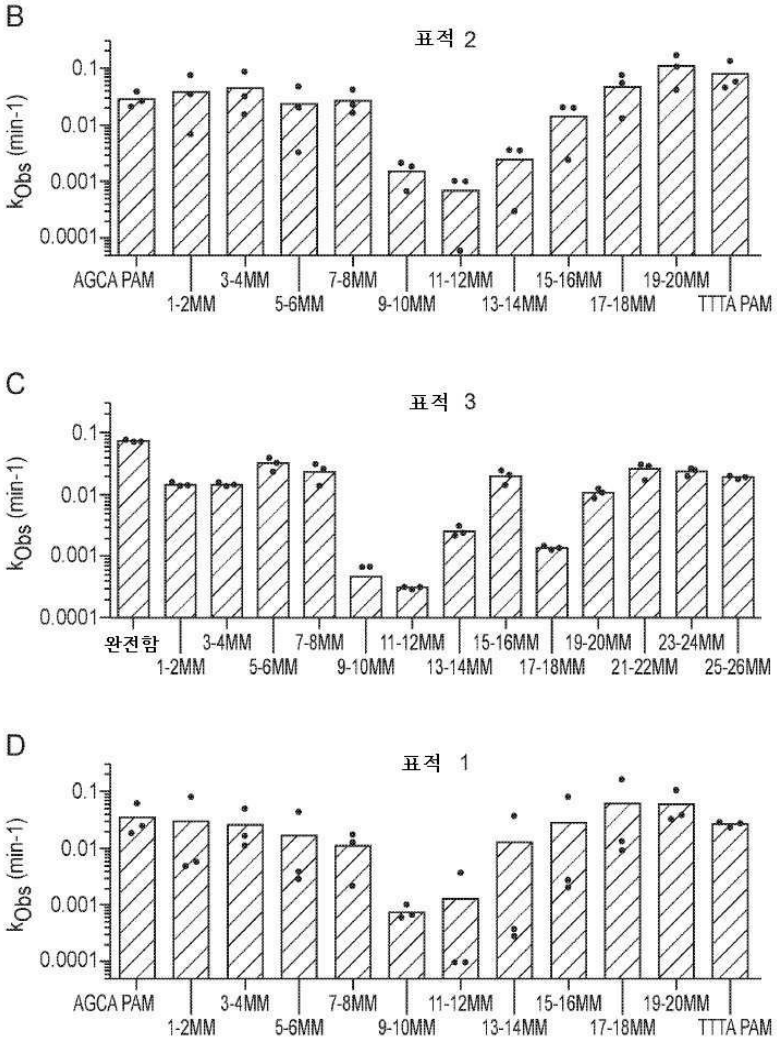


도면21a

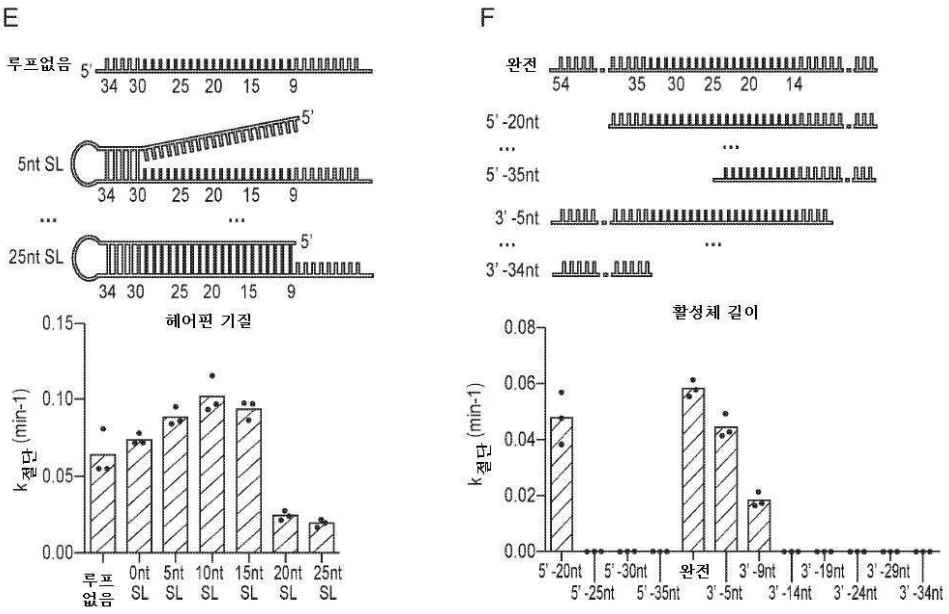

5'
완전한
표적
5'
1-2MM
5'
...
...
19-20MM
5'


30000
20000
10000
0
형광 (AU)
0
50
100
시간(분)
AGCA PAM
1-2MM
3-4MM
5-6MM
7-8MM
9-10MM
11-12MM
13-14MM
15-16MM
17-18MM
19-20MM
TTTA PAM

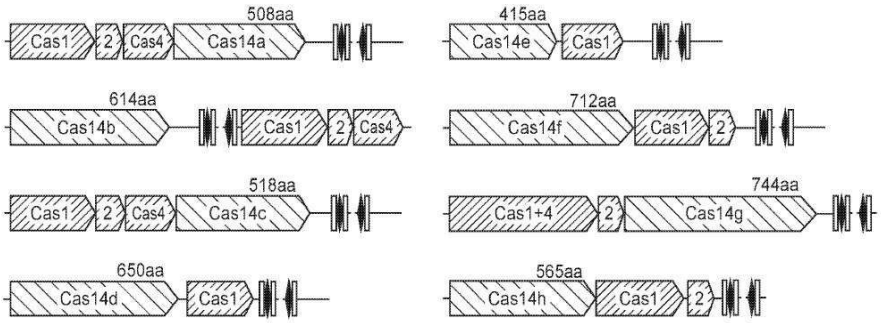
도면21bd



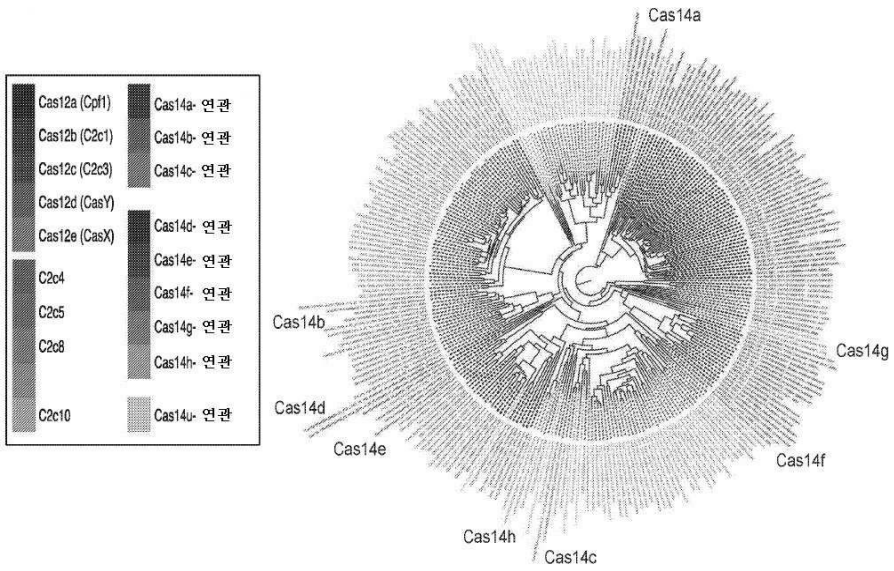
도면21ef



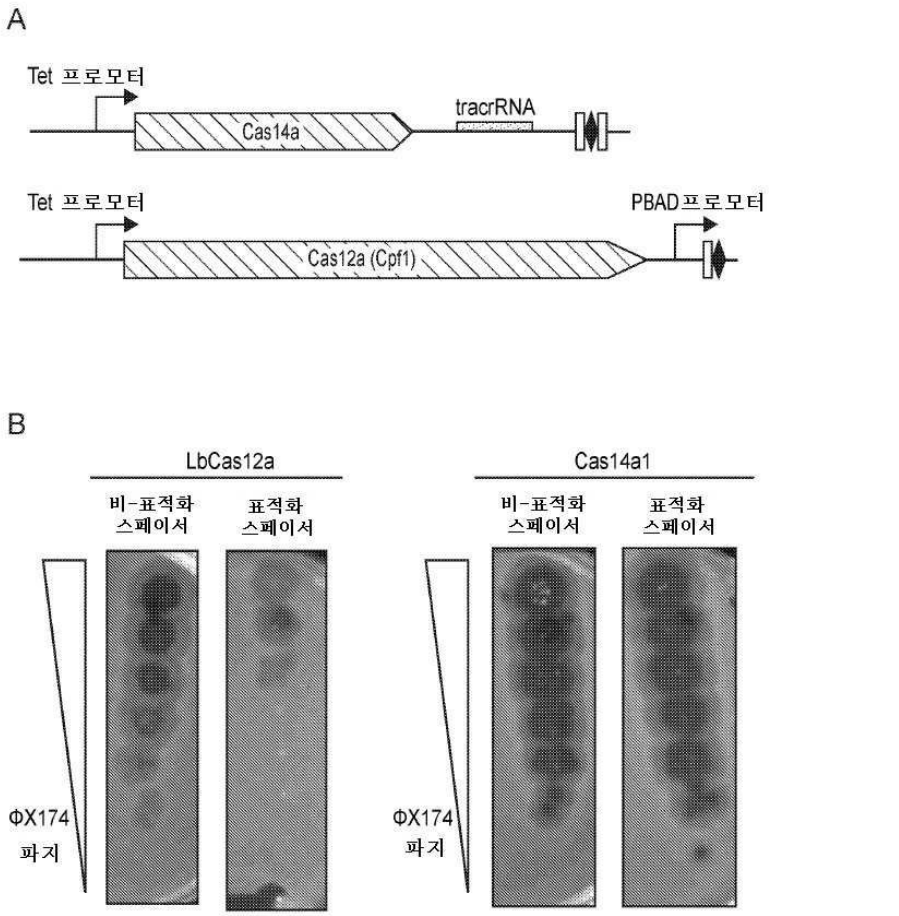
도면22a


508aa
Cas1
2
Cas4
Cas14a
415aa
Cas14e
Cas1
614aa
Cas14b
Cas1
2
Cas4
712aa
Cas14f
Cas1
2
518aa
Cas1
2
Cas4
Cas14c
744aa
Cas1+4
2
Cas14g
650aa
Cas14d
Cas1
565aa
Cas14h
Cas1
2

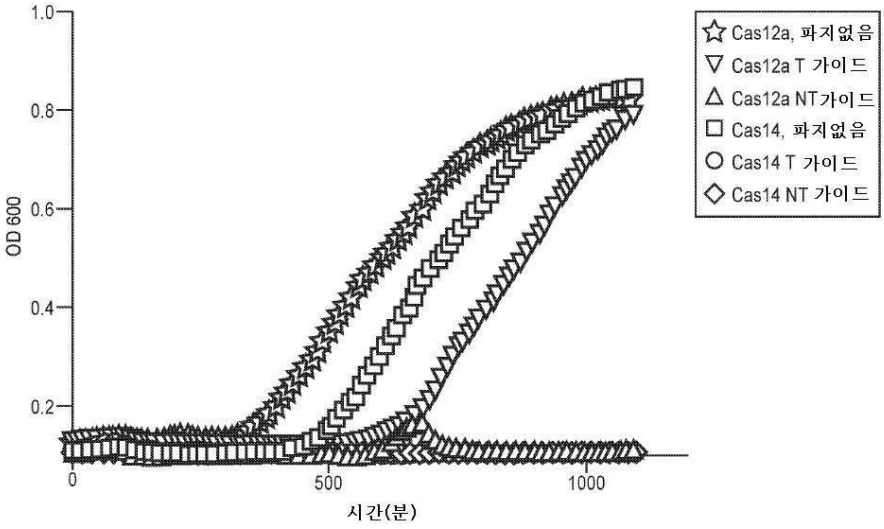
도면22b


Cas12a (Cpf1)
Cas12b (C2c1)
Cas12c (C2c3)
Cas12d (CasY)
Cas12e (CasX)
C2c4
C2c5
C2c8
C2c10
Cas14a- 연관
Cas14b- 연관
Cas14c- 연관
Cas14d- 연관
Cas14e- 연관
Cas14f- 연관
Cas14g- 연관
Cas14h- 연관
Cas14u- 연관
Cas14a
Cas14b
Cas14g
Cas14d
Cas14e
Cas14f
Cas14h
Cas14c

도면23ab



도면23c



도면24a

플라스미드 pLBH533_MBP-Cas14b2 발현 (애드진 플라스미드 번호 112506)

GCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCC
CGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGTGCCGGCTCCGGAGAGCTC
TTTAATTAAGCGGCCGCCCTGCAGGACTCGAGTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATA
TACATATGAAATCTTCTCACCATCACCATCACCATCACCATCACCATGGTTCTTCTATGAAAATCGAAGA
AGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTC
GAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTG
CGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGG
CCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTA
CGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATC
TGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAA
GAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTAT
GCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGG
GTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGA
AGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACC
AGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCG
TGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCT
GCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTAC
GAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATGC
CGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCG
TCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAAT
AACAACAACCTCGGGATCGAGGAAAACCTGTACTTCCAATCCAATGCAATGGAAGAAAGCATTATTACCG
GTGTGAAATTCAAACTGCGCATCGATAAAGAAACCACCAAAAAACTGAACGAGTACTTCGATGAATATGG
CAAAGCAATTAACTTCGCCGTGAAGATCATTCAGAAAGAACTGGCAGATGATCGTTTTGCAGGTAAAGCA
AAACTGGACCAGAATAAAAACCCGATCCTGGATGAAAACGGCAAAAAAATCTATGAATTCCCGGATGAAT
TTTGCAGCTGTGGTAAACAGGTTAACAAGTACGTTAACAACAAACCGTTTTGCCAAGAGTGCTATAAAAT
CCGCTTTACCGAAAATGGTATTCGCAAACGTATGTATAGCGCCAAAGGTCGTAAAGCCGAACATAAAATC
AATATCCTGAACAGCACCAACAAGATCAGCAAAACCCATTTTAACTATGCCATTCGCGAAGCCTTCATTC
TGGATAAAAGCATCAAAAAGCAGCGCAAAAAACGTAATGAACGTCTGCGTGAAAGTAAAAAACGTCTGCA
GCAGTTTATCGATATGCGTGATGGTAAACGTGAAATTTGCCCGACCATTAAAGGTCAGAAAGTGGATCGT
TTTATTCATCCGAGCTGGATCACCAAAGATAAAAAGCTGGAAGATTTTCGCGGTTATACCCTGAGCATTA

도면24b

TCAACAGCAAAATTAAGATTCTGGATCGCAACATCAAACGCGAAGAAAAAAGCCTGAAAGAAAAAGGCCA
GATCATCTTTAAAGCCAAACGTCTGATGCTGGATAAATCCATTCGTTTTGTTGGTGATCGCAAAGTGCTG
TTTACAATTAGTAAAACCCTGCCGAAAGAGTATGAACTGGATCTGCCGAGCAAAGAAAAACGGCTGAATT
GGCTGAAAGAGAAGATCGAGATTATCAAGAACCAGAAAACCGAAATATGCCTATCTGCTGCGCAAAAACAT
TGAGAGCGAAAAAAAACCGAACTATGAGTACTATCTGCAGTACACCCTGGAAATTAAACCGGAACTGAAA
GATTTTTATGATGGTGCCATTGGTATTGACCGTGGCATTAATCATATTGCCGTTTGCACCTTTATTAGCA
ACGATGGTAAAGTTACCCCTCCGAAATTTTTCAGCAGCGGTGAAATTCTGCGTCTGAAAAATCTGCAGAA
AGAGCGTGATCGCTTTCTGCTGCGTAAACACAACAAAAATCGCAAAAAAGGCAACATGCGCGTGATCGAA
AACAAAATCAATCTGATCCTGCACCGTTATAGCAAGCAGATTGTTGATATGGCCAAAAAGCTGAATGCCA
GCATTGTTTTTGAAGAACTGGGTCGTATTGGTAAAAGCCGCACCAAAATGAAAAAAAGCCAGCGTTATAA
ACTGAGCCTGTTCATCTTCAAGAAACTGAGCGATCTGGTTGATTACAAAAGCCGTCGTGAAGGTATTCGT
GTTACCTATGTTCCGCCTGAATATACCAGCAAAGAATGTAGCCATTGCGGTGAAAAAGTTAATACCCAGC
GTCCGTTTAATGGCAACTATAGCCTGTTTAAATGCAACAAATGTGGCATCCAGCTGAACAGCGATTATAA
TGCAAGCATCAACATTGCGAAAAAGGGCCTGAAAATTCCGAATAGCACCTAATAACATTGGAAGTGGATA
ACGGATCCGCGATCGCGGCGCGCCACCTGGTGGCCGGCCGGTACCACGCGTGCGCGCTGATCCGGCTGCT
AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG
CCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTG
TGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAA
GCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACG
CCATAGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCA
AGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACA
CGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGT
CAAACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATA
ATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTT
TCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAACATTGAAA
AAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCT
GTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGAT
GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGT
CGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATG
GCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCT
GACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT

도면24c

GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAA
TGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGA
CTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCT
GATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCT
CCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGA
GATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGAT
TTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCC
CTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG
GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCC
TTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCT
AATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAG
TTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGA
CCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGC
GGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC
TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTT
TGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCT
GATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGA
TGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAATGGTGCACTCTCAGTACAATCTGC
TCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCG
ACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCT
GTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGC
GGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTT
GAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGT
TTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGA
GGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACT
GGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGAT
GTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTG
ACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGA
CGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAA
CCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGC

도면24d

TGCCCGAGATGCGCCGCGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGT
TTGCGCATTCACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGG
TGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCGACGCAACGCGGGGAGGCAGAC
AAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCC
GTGACGATCAGCGGTCCAATGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCT
GATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAG
AAGAATCATAATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCG
GCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTT
GAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCG
GTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTC
ATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGG
GCATCGGTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
GAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCT
GCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGA
TGTCGGCGATATAGGCGCCA (서열번호 : 493)

도면24e

플라스미드pLBH547_Tet-Cas14b2 좌위 (애드진 플라스미드 번호 : 112504)

ATCGTTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAAAGAATTCAAAAGATCTAAAGAGGA
GAAAGGATCTATGGAAGAAAGCATTATTACCGGTGTGAAATTCAAACTGCGCATCGATAAAGAAACCACC
AAAAAACTGAACGAGTACTTCGATGAATATGGCAAAGCAATTAACTTCGCCGTGAAGATCATTCAGAAAG
AACTGGCAGATGATCGTTTTGCAGGTAAAGCAAAACTGGACCAGAATAAAAACCCGATCCTGGATGAAAA
CGGCAAAAAAATCTATGAATTCCCGGATGAATTTTGCAGCTGTGGTAAACAGGTTAACAAGTACGTTAAC
AACAAACCGTTTTGCCAAGAGTGCTATAAAATCCGCTTTACCGAAAATGGTATTCGCAAACGTATGTATA
GCGCCAAAGGTCGTAAAGCCGAACATAAAATCAATATCCTGAACAGCACCAACAAGATCAGCAAAACCCA
TTTTAACTATGCCATTCGCGAAGCCTTCATTCTGGATAAAAGCATCAAAAAGCAGCGCAAAAAACGTAAT
GAACGTCTGCGTGAAAGTAAAAAACGTCTGCAGCAGTTTATCGATATGCGTGATGGTAAACGTGAAATTT
GCCCGACCATTAAAGGTCAGAAAGTGGATCGTTTTATTCATCCGAGCTGGATCACCAAAGATAAAAAGCT
GGAAGATTTTCGCGGTTATACCCTGAGCATTATCAACAGCAAAATTAAGATTCTGGATCGCAACATCAAA
CGCGAAGAAAAAAGCCTGAAAGAAAAAGGCCAGATCATCTTTAAAGCCAAACGTCTGATGCTGGATAAAT
CCATTCGTTTTGTTGGTGATCGCAAAGTGCTGTTTACAATTAGTAAAACCCTGCCGAAAGAGTATGAACT
GGATCTGCCGAGCAAAGAAAACGGCTGAATTGGCTGAAAGAGAAGATCGAGATTATCAAGAACCAGAAA
CCGAAATATGCCTATCTGCTGCGCAAAAACATTGAGAGCGAAAAAAAACCGAACTATGAGTACTATCTGC
AGTACACCCTGGAAATTAAACCGGAACTGAAAGATTTTTATGATGGTGCCATTGGTATTGACCGTGGCAT
TAATCATATTGCCGTTTGCACCTTTATTAGCAACGATGGTAAAGTTACCCCTCCGAAATTTTTCAGCAGC
GGTGAAATTCTGCGTCTGAAAAATCTGCAGAAAGAGCGTGATCGCTTTCTGCTGCGTAAACACAACAAAA
ATCGCAAAAAAGGCAACATGCGCGTGATCGAAAACAAAATCAATCTGATCCTGCACCGTTATAGCAAGCA
GATTGTTGATATGGCCAAAAAGCTGAATGCCAGCATTGTTTTTGAAGAACTGGGTCGTATTGGTAAAAGC
CGCACCAAAATGAAAAAAAGCCAGCGTTATAAACTGAGCCTGTTCATCTTCAAGAAACTGAGCGATCTGG
TTGATTACAAAAGCCGTCGTGAAGGTATTCGTGTTACCTATGTTCCGCCTGAATATACCAGCAAAGAATG
TAGCCATTGCGGTGAAAAAGTTAATACCCAGCGTCCGTTTAATGGCAACTATAGCCTGTTTAAATGCAAC
AAATGTGGCATCCAGCTGAACAGCGATTATAATGCAAGCATCAACATTGCGAAAAAGGGCCTGAAAATTC
CGAATAGCACCTAATAATGTTGGTTAAGCCACAATATGGAATATTGTTCTTATGGACAGTATTGACTTAA
ATTAATAATCTTCGAAGGCTATATGCGGAAGATTTGGCGTTGTTGTAACGCAATAAGGGGTAACCCTGAA
AAGGTTTGAAATCATATAAACCTAGTTTTATTTGAGTTTAGGCTCAGATAAAATGAACAGACCAATCTTT
AATTCCGTTCTGATTTAAAAAATCAGAATCTCTTTATAAATAGTATTACAAAAAGTGTACATTCCAAAAT

도면24f

CCGAAAGCAGAATTGACCTTTTTAAGCCTAAAAAAGCCAAATTTCAAGGCTCTTTCATACTCAGAACAAA
GGGATTAAGGAATGCAACTACCTTACACCGCTTGCGAAAGTTGTCAGAAGATAATCTTTCATACTCAGAA
CAAAGGGATTAAGGAATGCAACTATCTTATCCATTTCTTGACATCAAATTTTCTTGCAGCATCTGAATTG
CTTAATTGCTTTCCTTGCTTCAGCAGGAAATAGCCAAGATTTTCCAGTTCTGCTGGCGTTATTGCAAGCA
CGGGGATTTTGTGCTTGCTGTAGATTTTGTTCTTCAGGACCTTCTTTTTCCATGCTAGAGTCACACTGGC
TCACCTTCGGGTGGGCCTTTCTGCGTTTATACCTAGGGATATATTCCGCTTCCTCGCTCACTGACTCGCT
ACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCC
AGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGA
CAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
TTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGT
TATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACT
GTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTC
ATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGG
TTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGC
AAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGAT
TTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTTACTAGTG
CTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTG
AGGTCATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCGATATCAAATTACGCCCCGCCCTGCCACT
CATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAAC
CTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCG
AAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGA
AAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGA
ATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGC
TCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATAC
GAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATT
TTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACT
GACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTT
TTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCT
TATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCGACG
TCTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTCAAGGCCGAATAAGA
AGGCTGGCTCTGCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAATGGCGGCATACTATCAGT

도면24g

AGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCTAAAGTAAAA
TGCCCCACAGCGCTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCATAAAAAGGCTAATT
GATTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGCTCCATCGCGATG
ACTTAGTAAAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCTTCTAAA
GGGCAAAAGTGAGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCCGCTTATTTT
TTACATGCCAATACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGGGTTGTTAAACCTTC
GATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACTTTACTTTTATCTAATCTAGACATCATT
AATTCCTAATTTTTGTTGACACTCT (서열번호 :494)

도면24h

플라스미드pLBH559_Tet-HisCas14a1 좌위 (애드진 플라스미드 번호: 112502)

TAATTCCTAATTTTTGTTGACACTCTATCGTTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGA
AAAGAATTCAAAAGATCTAAAGAGGAGAAAGGATCTATGAAATCTCACCATCACCATCACCATGAAAACC
TGTACTTCCAATCCAATATTGGAAGTGGAATGGCCAAAAACACCATTACCAAAACACTGAAACTGCGTAT
TGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATTGTTGCCGACGAAAAAAACAACCGCGAAAAAATC
GCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCTGCAGCAAACATCTGAAAGTTGCAGCATATTGTA
CCACACAGGTTGAACGTAATGCATGCCTGTTTTGTAAAGCACGTAAACTGGATGACAAATTCTACCAAAA
ACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAGAAATCAGCGAAATTTTTCGCCAGCTGCAGAAA
CAGGCAGCAGAAATCTATAATCAGAGCCTGATCGAACTGTACTACGAGATTTTTATCAAAGGCAAAGGTA
TTGCAAATGCCAGCAGCGTTGAACATTATCTGAGTGATGTTTGTTATACCCGTGCAGCAGAACTGTTTAA
AAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAAATCAAAAGCAATTTTCGTCTGAAAGAACTGAAAAAC
ATGAAAAGTGGTCTGCCGACCACCAAAAGCGATAATTTTCCGATTCCGCTGGTTAAACAGAAAGGTGGTC
AGTATACCGGTTTTGAAATTAGCAATCATAATAGCGACTTCATCATCAAGATTCCGTTTGGTCGTTGGCA
GGTCAAAAAAGAGATTGATAAATATCGTCCGTGGGAGAAATTTGACTTTGAACAGGTTCAGAAAAGCCCG
AAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTAAACGTAATAAAGGTTGGAGCAAAGATGAAGGCA
CCGAAGCCGAAATCAAAAAAGTTATGAATGGCGATTATCAGACCAGCTACATTGAAGTTAAACGTGGCAG
CAAAATCGGTGAAAAAAGCGCATGGATGCTGAATCTGAGCATTGATGTTCCGAAAATTGATAAAGGTGTG
GATCCGAGCATTATTGGTGGTATTGATGTTGGTGTTAAATCACCGCTGGTTTGCGCAATTAACAATGCAT
TTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCACTTCAACAAGAAAATGTTTGCACGTCGTCGTAT
CCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCATGGTGCAAAAAACAAACTGAAACCGATCACCATT
CTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGATTGAACGTTGGGCATGTGAAATCGCGGATTTCT
TCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAATCTGGAAAGCATGAAACGTAAAGAGGACAGCTA
TTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCAGAAATGCAGAACAAAATCGAATTCAAACTGAAG
CAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATAATACCAGCAAAACCTGTAGCAAATGTGGCCATC
TGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAACAAATTCCCGCACTTTAAATGCGAAAAATGCAA
CTTCAAAGAAAACGCCGATTATAATGCAGCCCTGAATATTTCAAACCCGAAACTGAAAAGCACCAAAGAG
GAACCGTAAATATTTATACTTTATTATCCTTCATTGACAAAAATGAGAATGTTATCCCAGATAACATTTG
ATGTACACAGATTCACACTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATT
AGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCC
TCGAAACAAATTCATTTTTCCTCTCCAATTCTGCACAAAAAAAGGTGAGTCCTTATAAACCGGCGTGCAG

도면24i

AACGCCGGCTCACCTTTTTTCTTCATTCGATTTTATGCTTAAAAGCCGTAAAAACGCGGAATTCGGCGCC
GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAACTACCTTACACCGCTTGCGAAAGTTGTCAGAAGA
TAATCTTGCAGAACCCGAATAGACGAATGAAGGAATGCAATCTTGACAGAGCCCGATTGCGTTATCTCCA
GGAGAAACATATAAAAGCATCAACCGCTGATCGGACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTT
TCTGCGTTTATACCTAGGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTG
CGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGG
AAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGA
CGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCC
TCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCA
TTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGT
TCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCA
CCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAA
ACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTC
AGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACC
AAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAATTTATCTCT
TCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTTACTAGTGCTTGGATTCTCACCAATAA
AAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTAT
CAACAGGAGTCCAAGCGAGCTCGATATCAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAA
TTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATC
AGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGG
CCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAA
CCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGC
CGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAAC
AAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATT
CATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAA
AAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAAT
GTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTC
CTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAG
TTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCGACGTCTTAAGACCCACTTTCAC
ATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTCAAGGCCGAATAAGAAGGCTGGCTCTGCACCTTG
GTGATCAAATAATTCGATAGCTTGTCGTAATAATGGCGGCATACTATCAGTAGTAGGTGTTTCCCTTTCT
TCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCTAAAGTAAAATGCCCCACAGCGCTGAGTG

도면24j

CATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCATAAAAAGGCTAATTGATTTTCGAGAGTTTCATA
CTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGATCCATCGCGATGACTTAGTAAAGCACATCTA
AAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCTTCTAAAGGGCAAAAGTGAGTATGGT
GCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCCGCTTATTTTTTACATGCCAATACAATGT
AGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGGGTTGTTAAACCTTCGATTCCGACCTCATTAAGC
AGCTCTAATGCGCTGTTAATCACTTTACTTTTATCTAATCTAGACATCAT (서열번호 : 495)

도면24k

플라스미드pLBH545_Tet-Cas14a1_좌위 (애드진 플라스미드 번호: 112501)

TTACTTTTATCTAATCTAGACATCATTAATTCCTAATTTTTGTTGACACTCTATCGTTGATAGAGTTATT
TTACCACTCCCTATCAGTGATAGAGAAAAGAATTCAAAAGATCTAAAGAGGAGAAAGGATCTATGGCCAA
AAACACCATTACCAAAACACTGAAACTGCGTATTGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATT
GTTGCCGACGAAAAAAACAACCGCGAAAAAATCGCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCT
GCAGCAAACATCTGAAAGTTGCAGCATATTGTACCACACAGGTTGAACGTAATGCATGCCTGTTTTGTAA
AGCACGTAAACTGGATGACAAATTCTACCAAAAACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAA
GAAATCAGCGAAATTTTCGCCAGCTGCAGAAACAGGCAGCAGAAATCTATAATCAGAGCCTGATCGAAC
TGTACTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCAGCAGCGTTGAACATTATCTGAGTGA
TGTTTGTTATACCCGTGCAGCAGAACTGTTTAAAAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAAATC
AAAAGCAATTTTCGTCTGAAAGAACTGAAAAACATGAAAAGTGGTCTGCCGACCACCAAAAGCGATAATT
TTCCGATTCCGCTGGTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAGCAATCATAATAGCGA
CTTCATCATCAAGATTCCGTTTGGTCGTTGGCAGGTCAAAAAAGAGATTGATAAATATCGTCCGTGGGAG
AAATTTGACTTTGAACAGGTTCAGAAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTA
AACGTAATAAAGGTTGGAGCAAAGATGAAGGCACCGAAGCCGAAATCAAAAAAGTTATGAATGGCGATTA
TCAGACCAGCTACATTGAAGTTAAACGTGGCAGCAAAATCGGTGAAAAAAGCGCATGGATGCTGAATCTG
AGCATTGATGTTCCGAAAATTGATAAAGGTGTGGATCCGAGCATTATTGGTGGTATTGATGTTGGTGTTA
AATCACCGCTGGTTTGCGCAATTAACAATGCATTTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCA
CTTCAACAAGAAAATGTTTGCACGTCGTCGTATCCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCAT
GGTGCAAAAAACAAACTGAAACCGATCACCATTCTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGA
TTGAACGTTGGGCATGTGAAATCGCGGATTTCTTCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAA
TCTGGAAAGCATGAAACGTAAAGAGGACAGCTATTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCA
GAAATGCAGAACAAAATCGAATTCAAACTGAAGCAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATA
ATACCAGCAAAACCTGTAGCAAATGTGGCCATCTGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAA
CAAATTCCCGCACTTTAAATGCGAAAAATGCAACTTCAAAGAAAACGCCGATTATAATGCAGCCCTGAAT
ATTTCAAACCCGAAACTGAAAAGCACCAAAGAGGAACCGTAAATATTTATACTTTATTATCCTTCATTGA
CAAAAATGAGAATGTTATCCCAGATAACATTTGATGTACACAGATTCACACTTCACTGATAAAGTGGAGA
ACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTA
ATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTTTCCTCTCCAATTCTGCACA

도면241

AAAAAAGGTGAGTCCTTATAAACCGGCGTGCAGAACGCCGGCTCACCTTTTTTCTTCATTCGATTTTATG
CTTAAAAGCCGTAAAAACGCGGAATTCGGCGCCGTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC
TACCTTACACCGCTTGCGAAAGTTGTCAGAAGATAATCTTGCAGAACCCGAATAGACGAATGAAGGAATG
CAATCTTGACAGAGCCCGATTGCGTTATCTCCAGGAGAAACATATAAAAGCATCAACCGCTGATCGGACT
AGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATACCTAGGGATATATTCCGCTTCCTCG
CTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATT
TCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGG
CTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTAT
AAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGG
TGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCG
CTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGT
CTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAG
TTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAA
GCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTT
TTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATA
AAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATA
AGTTGTTACTAGTGCTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATC
CAGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCGATATCAAATTACGC
CCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAA
ACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGG
TGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG
ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACAC
GCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAA
ACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTC
TTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAA
AACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG
TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGT
ATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACG
CCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTC
GCCAGATATCGACGTCTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTC
AAGGCCGAATAAGAAGGCTGGCTCTGCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAATGGC
GGCATACTATCAGTAGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGC

도면24m

AACCTAAAGTAAAATGCCCCACAGCGCTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCA
TAAAAAGGCTAATTGATTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTT
GCTCCATCGCGATGACTTAGTAAAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGC
TTTCCCCTTCTAAAGGGCAAAAGTGAGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAA
AGCCCGCTTATTTTTTACATGCCAATACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGG
GTTGTTAAACCTTCGATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACT (서열번호 :496)

도면24n

플라스미드pLBH531_MBP-Cas14a1 발현 (애드진 플라스미드 번호: 112500)

ACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCG
ATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCG
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGTGCCGGC
TCCGGAGAGCTCTTTAATTAAGCGGCCGCCCTGCAGGACTCGAGTTCTAGAAATAATTTTGTTTAACTTT
AAGAAGGAGATATACATATGAAATCTTCTCACCATCACCATCACCATCACCATCACCATGGTTCTTCTAT
GAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTC
GGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAAT
TCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTA
CGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACC
TGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTT
ATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAA
AGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCT
GACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTG
GCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTA
CTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCC
AACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAAC
CGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCT
CGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCG
CTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAG
GTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGC
CGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAAC
AACAATAACAATACAACAACCTCGGGATCGAGGAAAACCTGTACTTCCAATCCAATGCAATGGCCAAAA
ACACCATTACCAAAACACTGAAACTGCGTATTGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATTGT
TGCCGACGAAAAAAACAACCGCGAAAAAATCGCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCTGC
AGCAAACATCTGAAAGTTGCAGCATATTGTACCACACAGGTTGAACGTAATGCATGCCTGTTTTGTAAAG
CACGTAAACTGGATGACAAATTCTACCAAAAACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAGA
AATCAGCGAAATTTTTCGCCAGCTGCAGAAACAGGCAGCAGAAATCTATAATCAGAGCCTGATCGAACTG

도면24o

TACTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCAGCAGCGTTGAACATTATCTGAGTGATG
TTTGTTATACCCGTGCAGCAGAACTGTTTAAAAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAAATCAA
AAGCAATTTTCGTCTGAAAGAACTGAAAAACATGAAAAGTGGTCTGCCGACCACCAAAAGCGATAATTTT
CCGATTCCGCTGGTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAGCAATCATAATAGCGACT
TCATCATCAAGATTCCGTTTGGTCGTTGGCAGGTCAAAAAAGAGATTGATAAATATCGTCCGTGGGAGAA
ATTTGACTTTGAACAGGTTCAGAAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTAAA
CGTAATAAAGGTTGGAGCAAAGATGAAGGCACCGAAGCCGAAATCAAAAAAGTTATGAATGGCGATTATC
AGACCAGCTACATTGAAGTTAAACGTGGCAGCAAAATCTGTGAAAAAAGCGCATGGATGCTGAATCTGAG
CATTGATGTTCCGAAAATTGATAAAGGTGTGGATCCGAGCATTATTGGTGGTATTGATGTTGGTGTTAAA
TCACCGCTGGTTTGCGCAATTAACAATGCATTTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCACT
TCAACAAGAAAATGTTTGCACGTCGTCGTATCCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCATGG
TGCAAAAAACAAACTGAAACCGATCACCATTCTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGATT
GAACGTTGGGCATGTGAAATCGCGGATTTCTTCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAATC
TGGAAAGCATGAAACGTAAAGAGGACAGCTATTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCAGA
AATGCAGAACAAAATCGAATTCAAACTGAAGCAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATAAT
ACCAGCAAAACCTGTAGCAAATGTGGCCATCTGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAACA
AATTCCCGCACTTTAAATGCGAAAAATGCAACTTCAAAGAAAACGCCGATTATAATGCAGCCCTGAATAT
TTCAAACCCGAAACTGAAAAGCACCAAAGAGGAACCGTAATAACATTGGAAGTGGATAACGGATCCGCGA
TCGCGGCGCCCACCTGGTGGCCGGCCGGTACCACGCGTGCGCGCTGATCCGGCTGCTAACAAAGCCCGA
AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG
TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGA
TCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAG
TGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTG
GCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTA
CAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACT
GCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA
TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT
TAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATT
CAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAACATTGAAAAAGGAAGAGTAT
GAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGG

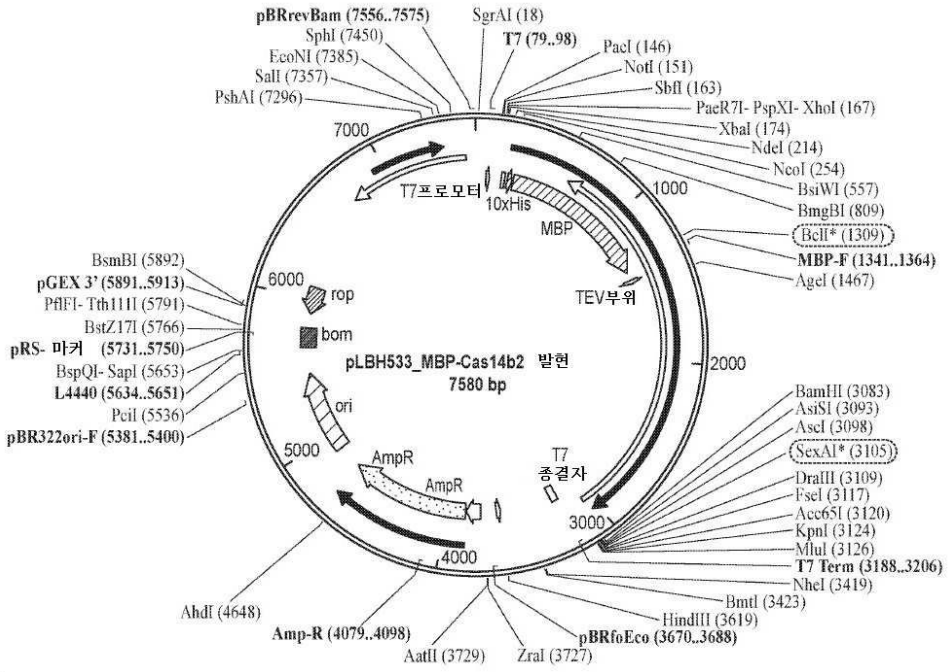
도면24p

ATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAA
AGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACAC
TATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA
GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGG
AGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAA
CCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGT
TGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGC
GGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA
GCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAG
TTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTC
ACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCAT
TTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGT
TTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTA
CCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACC
AGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAG
GCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTAT
AGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCC
TATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT
CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGC
CGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCAATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCA
TAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAAC
ACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCC
GGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCAT
CAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAG
AAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGAT
GCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGA
TACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGAT

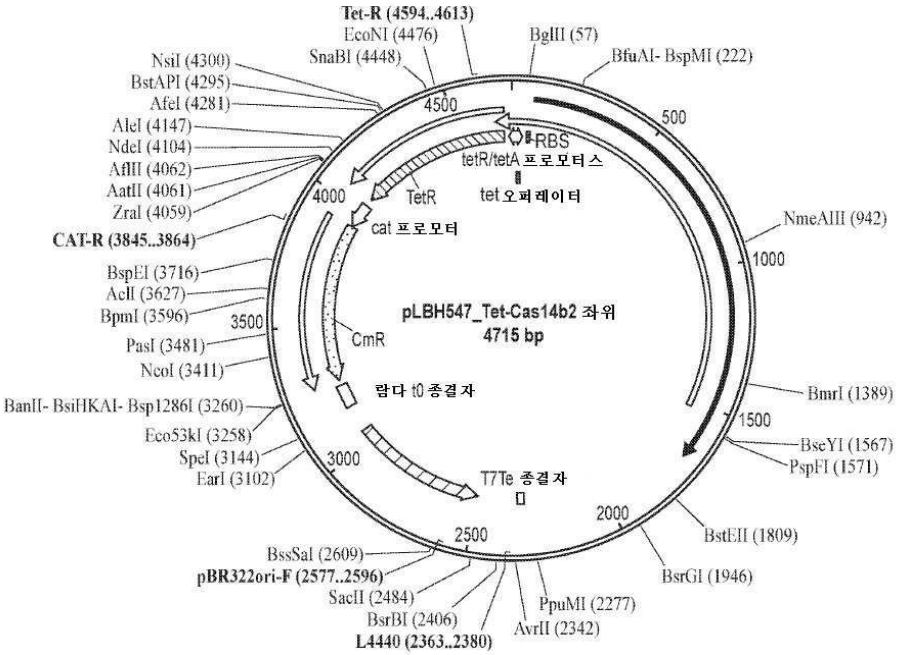
도면24q

GCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCA
CAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTT
CCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCA
GCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCT
AGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATGC
GCCGCGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTCAC
AGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGGTGCCGCCGGCTT
CCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCGACGCAACGCGGGGAGGCAGACAAGGTATAGGGC
GGCGCCTACAATCCATGCCAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGC
GGTCCAATGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCGTCAT
CTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAAT
GGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCG
GCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGT
GCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAA
AATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCG
ACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAC
GCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGC
CGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCAT
(서열번호): 497)

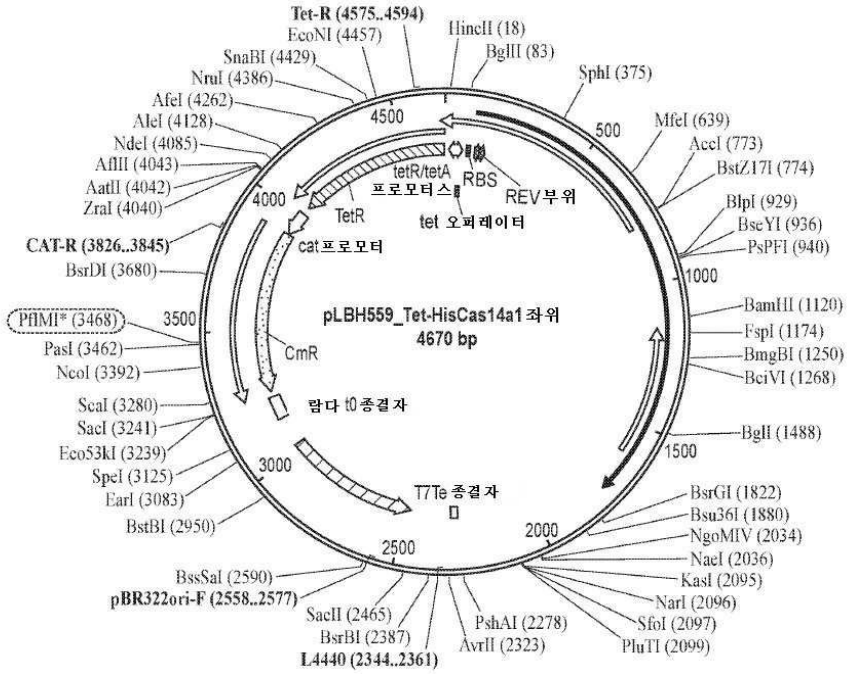
도면25a



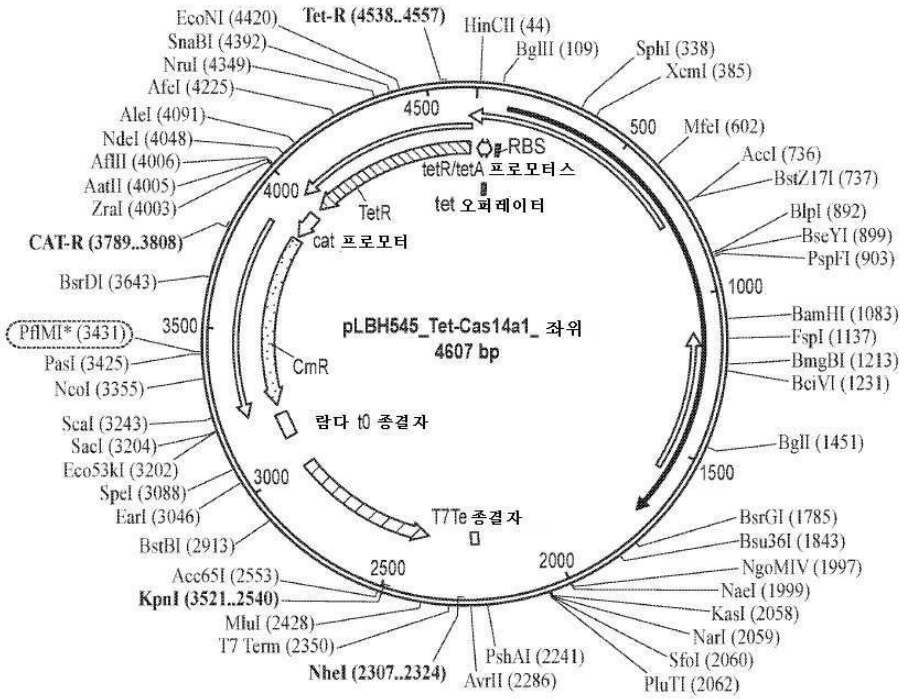
도면25b


Tet-R (4594..4613)
EcoNI (4476)
SnaBI (4448)
BglII (57)
BfuAI- BspMI (222)
NsiI (4300)
BstAPI (4295)
AfeI (4281)
AleI (4147)
NdeI (4104)
AflII (4062)
AatII (4061)
ZraI (4059)
CAT-R (3845..3864)
BspEI (3716)
AclI (3627)
BpmI (3596)
PasI (3481)
NcoI (3411)
BanII- BsiHKAI- Bsp1286I (3260)
Eco53kI (3258)
SpeI (3144)
EarI (3102)
BssSaI (2609)
pBR322ori-F (2577..2596)
SacII (2484)
BsrBI (2406)
L4440 (2363..2380)
AvrII (2342)
PpuMI (2277)
BsrGI (1946)
BstEII (1809)
BseYI (1567)
PspFI (1571)
BmrI (1389)
NmeAIII (942)
RBS
tetR/tetA 프로모터스
tet 오퍼레이터
TetR
cat 프로모터
CmR
pLBH547_Tet-Cas14b2 좌위
4715 bp
람다 t0 종결자
T7Te 종결자
4500
500
4000
1000
3500
1500
3000
2000
2500

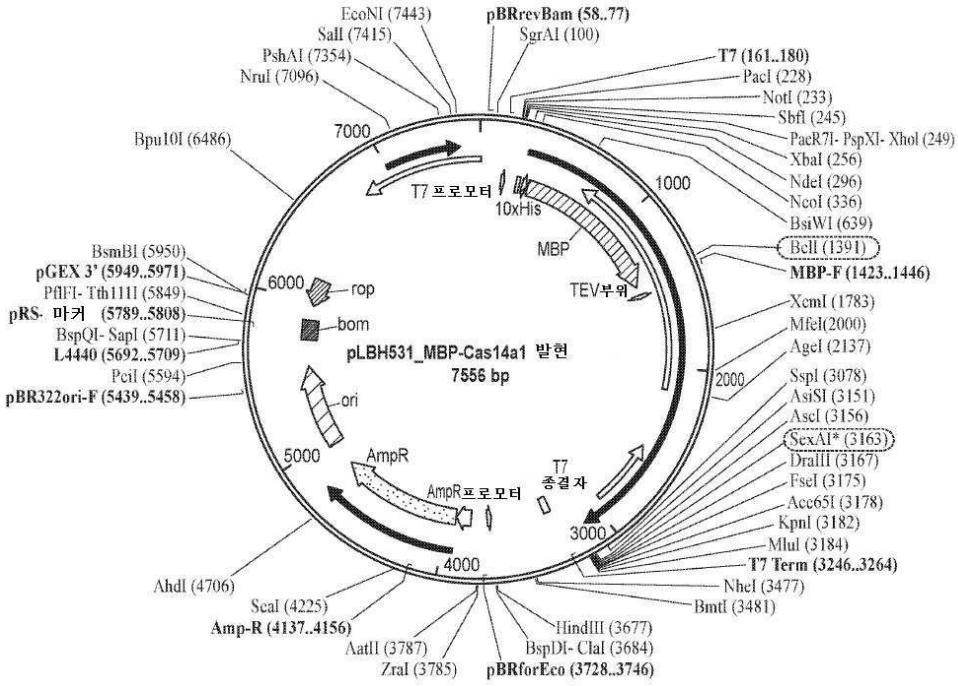
도면25c


Tet-R (4575..4594)
EcoNI (4457)
HincII (18)
BglII (83)
SnaBI (4429)
NruI (4386)
SphI (375)
AfeI (4262)
AleI (4128)
NdeI (4085)
AflII (4043)
AatII (4042)
ZraI (4040)
CAT-R (3826..3845)
BsrDI (3680)
PflMI* (3468)
PasI (3462)
NcoI (3392)
ScaI (3280)
SacI (3241)
Eco53kI (3239)
SpeI (3125)
EarI (3083)
BstBI (2950)
BssSaI (2590)
pBR322ori-F (2558..2577)
SacII (2465)
BsrBI (2387)
L4440 (2344..2361)
AvrII (2323)
PshAI (2278)
PluTI (2099)
SfoI (2097)
NarI (2096)
KasI (2095)
NaeI (2036)
NgoMIV (2034)
Bsu36I (1880)
BsrGI (1822)
BglI (1488)
BmgBI (1250)
BciVI (1268)
FspI (1174)
BamHI (1120)
PsPFI (940)
BseYI (936)
BlpI (929)
BstZ17I (774)
AccI (773)
MfeI (639)
tetR/tetA 프로모터스
RBS
REV부위
TetR
tet 오퍼레이터
cat 프로모터
CmR
pLBH559_Tet-HisCas14a1 좌위
4670 bp
람다 t0 종결자
T7Te 종결자
4500
500
4000
1000
3500
1500
3000
2000
2500

도면25d



도면25e



서 열 목 록

SEQUENCE LISTING

<110> The Regents of the University of California

<120> CASZ COMPOSITIONS AND METHODS OF USE

<130> WO/2019/089820

<140> PCT/US2018/058545

<141> 2018-10-31

<150> US 62/580,395

<151> 2017-11-01

<160> 373

<170> PatentIn version 3.5

<210> 1

<211> 500

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 1

Met Glu Val Gln Lys Thr Val Met Lys Thr Leu Ser Leu Arg Ile Leu
1 5 10 15

Arg Pro Leu Tyr Ser Gln Glu Ile Glu Lys Glu Ile Lys Glu Glu Lys
20 25 30

Glu Arg Arg Lys Gln Ala Gly Gly Thr Gly Glu Leu Asp Gly Gly Phe
35 40 45

Tyr Lys Lys Leu Glu Lys Lys His Ser Glu Met Phe Ser Phe Asp Arg
50 55 60

Leu Asn Leu Leu Leu Asn Gln Leu Gln Arg Glu Ile Ala Lys Val Tyr
65 70 75 80

Asn His Ala Ile Ser Glu Leu Tyr Ile Ala Thr Ile Ala Gln Gly Asn
85 90 95

Lys Ser Asn Lys His Tyr Ile Ser Ser Ile Val Tyr Asn Arg Ala Tyr
100 105 110

Gly Tyr Phe Tyr Asn Ala Tyr Ile Ala Leu Gly Ile Cys Ser Lys Val
115 120 125

Glu Ala Asn Phe Arg Ser Asn Glu Leu Leu Thr Gln Gln Ser Ala Leu
130 135 140

Pro Thr Ala Lys Ser Asp Asn Phe Pro Ile Val Leu His Lys Gln Lys

145 150 155 160

Gly Ala Glu Gly Glu Asp Gly Gly Phe Arg Ile Ser Thr Glu Gly Ser

165 170 175

Asp Leu Ile Phe Glu Ile Pro Ile Pro Phe Tyr Glu Tyr Asn Gly Glu

180 185 190

Asn Arg Lys Glu Pro Tyr Lys Trp Val Lys Lys Gly Gly Gln Lys Pro

195 200 205

Val Leu Lys Leu Ile Leu Ser Thr Phe Arg Arg Gln Arg Asn Lys Gly

210 215 220

Trp Ala Lys Asp Glu Gly Thr Asp Ala Glu Ile Arg Lys Val Thr Glu

225 230 235 240

Gly Lys Tyr Gln Val Ser Gln Ile Glu Ile Asn Arg Gly Lys Lys Leu

245 250 255

Gly Glu His Gln Lys Trp Phe Ala Asn Phe Ser Ile Glu Gln Pro Ile

260 265 270

Tyr Glu Arg Lys Pro Asn Arg Ser Ile Val Gly Gly Leu Asp Val Gly

275 280 285

Ile Arg Ser Pro Leu Val Cys Ala Ile Asn Asn Ser Phe Ser Arg Tyr

290 295 300

Ser Val Asp Ser Asn Asp Val Phe Lys Phe Ser Lys Gln Val Phe Ala

305 310 315 320

Phe Arg Arg Arg Leu Leu Ser Lys Asn Ser Leu Lys Arg Lys Gly His

325 330 335

Gly Ala Ala His Lys Leu Glu Pro Ile Thr Glu Met Thr Glu Lys Asn

340 345 350

Asp Lys Phe Arg Lys Lys Ile Ile Glu Arg Trp Ala Lys Glu Val Thr

355 360 365

Asn Phe Phe Val Lys Asn Gln Val Gly Ile Val Gln Ile Glu Asp Leu

370 375 380

Ser Thr Met Lys Asp Arg Glu Asp His Phe Phe Asn Gln Tyr Leu Arg

385 390 395 400

Gly Phe Trp Pro Tyr Tyr Gln Met Gln Thr Leu Ile Glu Asn Lys Leu
405 410 415
Lys Glu Tyr Gly Ile Glu Val Lys Arg Val Gln Ala Lys Tyr Thr Ser
420 425 430
Gln Leu Cys Ser Asn Pro Asn Cys Arg Tyr Trp Asn Asn Tyr Phe Asn
435 440 445
Phe Glu Tyr Arg Lys Val Asn Lys Phe Pro Lys Phe Lys Cys Glu Lys
450 455 460

Cys Asn Leu Glu Ile Ser Ala Asp Tyr Asn Ala Ala Arg Asn Leu Ser
465 470 475 480
Thr Pro Asp Ile Glu Lys Phe Val Ala Lys Ala Thr Lys Gly Ile Asn
485 490 495
Leu Pro Glu Lys
500

<210> 2
<211> 507
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 2
Met Glu Glu Ala Lys Thr Val Ser Lys Thr Leu Ser Leu Arg Ile Leu
1 5 10 15

Arg Pro Leu Tyr Ser Ala Glu Ile Glu Lys Glu Ile Lys Glu Glu Lys
20 25 30
Glu Arg Arg Lys Gln Gly Gly Lys Ser Gly Glu Leu Asp Ser Gly Phe
35 40 45
Tyr Lys Lys Leu Glu Lys Lys His Thr Gln Met Phe Gly Trp Asp Lys
50 55 60
Leu Asn Leu Met Leu Ser Gln Leu Gln Arg Gln Ile Ala Arg Val Phe
65 70 75 80

Asn Gln Ser Ile Ser Glu Leu Tyr Ile Glu Thr Val Ile Gln Gly Lys
85 90 95

Lys Ser Asn Lys His Tyr Thr Ser Lys Ile Val Tyr Asn Arg Ala Tyr
100 105 110
Ser Val Phe Tyr Asn Ala Tyr Leu Ala Leu Gly Ile Thr Ser Lys Val
115 120 125
Glu Ala Asn Phe Arg Ser Thr Glu Leu Leu Met Gln Lys Ser Ser Leu
130 135 140

Pro Thr Ala Lys Ser Asp Asn Phe Pro Ile Leu Leu His Lys Gln Lys
145 150 155 160
Gly Val Glu Gly Glu Glu Gly Gly Phe Lys Ile Ser Ala Asp Gly Asn
165 170 175
Asp Leu Ile Phe Glu Ile Pro Ile Pro Phe Tyr Glu Tyr Asp Ser Ala
180 185 190
Asn Lys Lys Glu Pro Phe Lys Trp Ile Lys Lys Gly Gly Gln Lys Pro
195 200 205

Thr Ile Lys Leu Ile Leu Ser Thr Phe Arg Arg Gln Arg Asn Lys Gly
210 215 220
Trp Ala Lys Asp Glu Gly Thr Asp Ala Glu Ile Arg Lys Val Ile Glu
225 230 235 240
Gly Lys Tyr Gln Val Ser His Ile Glu Ile Asn Arg Gly Lys Lys Leu
245 250 255
Gly Asp His Gln Lys Trp Phe Val Asn Phe Thr Ile Glu Gln Pro Ile
260 265 270

Tyr Glu Arg Lys Leu Asp Lys Asn Ile Ile Gly Gly Ile Asp Val Gly
275 280 285
Ile Lys Ser Pro Leu Val Cys Ala Val Asn Asn Ser Phe Ala Arg Tyr
290 295 300
Ser Val Asp Ser Asn Asp Val Leu Lys Phe Ser Lys Gln Ala Phe Ala
305 310 315 320
Phe Arg Arg Arg Leu Leu Ser Lys Asn Ser Leu Lys Arg Ser Gly His
325 330 335

Gly Ser Lys Asn Lys Leu Asp Pro Ile Thr Arg Met Thr Glu Lys Asn

340 345 350

Asp Arg Phe Arg Lys Lys Ile Ile Glu Arg Trp Ala Lys Glu Val Thr

355 360 365

Asn Phe Phe Ile Lys Asn Gln Val Gly Thr Val Gln Ile Glu Asp Leu

370 375 380

Ser Thr Met Lys Asp Arg Gln Asp Asn Phe Phe Asn Gln Tyr Leu Arg

385 390 395 400

Gly Phe Trp Pro Tyr Tyr Gln Met Gln Asn Leu Ile Glu Asn Lys Leu

405 410 415

Lys Glu Tyr Gly Ile Glu Thr Lys Arg Ile Lys Ala Arg Tyr Thr Ser

420 425 430

Gln Leu Cys Ser Asn Pro Ser Cys Arg His Trp Asn Ser Tyr Phe Ser

435 440 445

Phe Asp His Arg Lys Thr Asn Asn Phe Pro Lys Phe Lys Cys Glu Lys

450 455 460

Cys Ala Leu Glu Ile Ser Ala Asp Tyr Asn Ala Ala Arg Asn Ile Ser

465 470 475 480

Thr Pro Asp Ile Glu Lys Phe Val Ala Lys Ala Thr Lys Gly Ile Asn

485 490 495

Leu Pro Asp Lys Asn Glu Asn Val Ile Leu Glu

500 505

<210> 3

<211> 529

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 3

Met Ala Lys Asn Thr Ile Thr Lys Thr Leu Lys Leu Arg Ile Val Arg

1 5 10 15

Pro Tyr Asn Ser Ala Glu Val Glu Lys Ile Val Ala Asp Glu Lys Asn

20 25 30

Asn Arg Glu Lys Ile Ala Leu Glu Lys Asn Lys Asp Lys Val Lys Glu

35 40 45

Ala Cys Ser Lys His Leu Lys Val Ala Ala Tyr Cys Thr Thr Gln Val

50 55 60

Glu Arg Asn Ala Cys Leu Phe Cys Lys Ala Arg Lys Leu Asp Asp Lys

65 70 75 80

Phe Tyr Gln Lys Leu Arg Gly Gln Phe Pro Asp Ala Val Phe Trp Gln

85 90 95

Glu Ile Ser Glu Ile Phe Arg Gln Leu Gln Lys Gln Ala Ala Glu Ile

100 105 110

Tyr Asn Gln Ser Leu Ile Glu Leu Tyr Tyr Glu Ile Phe Ile Lys Gly

115 120 125

Lys Gly Ile Ala Asn Ala Ser Ser Val Glu His Tyr Leu Ser Asp Val

130 135 140

Cys Tyr Thr Arg Ala Ala Glu Leu Phe Lys Asn Ala Ala Ile Ala Ser

145 150 155 160

Gly Leu Arg Ser Lys Ile Lys Ser Asn Phe Arg Leu Lys Glu Leu Lys

165 170 175

Asn Met Lys Ser Gly Leu Pro Thr Thr Lys Ser Asp Asn Phe Pro Ile

180 185 190

Pro Leu Val Lys Gln Lys Gly Gly Gln Tyr Thr Gly Phe Glu Ile Ser

195 200 205

Asn His Asn Ser Asp Phe Ile Ile Lys Ile Pro Phe Gly Arg Trp Gln

210 215 220

Val Lys Lys Glu Ile Asp Lys Tyr Arg Pro Trp Glu Lys Phe Asp Phe

225 230 235 240

Glu Gln Val Gln Lys Ser Pro Lys Pro Ile Ser Leu Leu Leu Ser Thr

245 250 255

Gln Arg Arg Lys Arg Asn Lys Gly Trp Ser Lys Asp Glu Gly Thr Glu

260 265 270

Ala Glu Ile Lys Lys Val Met Asn Gly Asp Tyr Gln Thr Ser Tyr Ile

275 280 285

Glu Val Lys Arg Gly Ser Lys Ile Gly Glu Lys Ser Ala Trp Met Leu
290 295 300
Asn Leu Ser Ile Asp Val Pro Lys Ile Asp Lys Gly Val Asp Pro Ser
305 310 315 320
Ile Ile Gly Gly Ile Asp Val Gly Val Lys Ser Pro Leu Val Cys Ala

325 330 335
Ile Asn Asn Ala Phe Ser Arg Tyr Ser Ile Ser Asp Asn Asp Leu Phe
340 345 350
His Phe Asn Lys Lys Met Phe Ala Arg Arg Arg Ile Leu Leu Lys Lys
355 360 365
Asn Arg His Lys Arg Ala Gly His Gly Ala Lys Asn Lys Leu Lys Pro
370 375 380
Ile Thr Ile Leu Thr Glu Lys Ser Glu Arg Phe Arg Lys Lys Leu Ile

385 390 395 400
Glu Arg Trp Ala Cys Glu Ile Ala Asp Phe Phe Ile Lys Asn Lys Val
405 410 415
Gly Thr Val Gln Met Glu Asn Leu Glu Ser Met Lys Arg Lys Glu Asp
420 425 430
Ser Tyr Phe Asn Ile Arg Leu Arg Gly Phe Trp Pro Tyr Ala Glu Met
435 440 445
Gln Asn Lys Ile Glu Phe Lys Leu Lys Gln Tyr Gly Ile Glu Ile Arg

450 455 460
Lys Val Ala Pro Asn Asn Thr Ser Lys Thr Cys Ser Lys Cys Gly His
465 470 475 480
Leu Asn Asn Tyr Phe Asn Phe Glu Tyr Arg Lys Lys Asn Lys Phe Pro
485 490 495
His Phe Lys Cys Glu Lys Cys Asn Phe Lys Glu Asn Ala Asp Tyr Asn
500 505 510
Ala Ala Leu Asn Ile Ser Asn Pro Lys Leu Lys Ser Thr Lys Glu Glu

515 520 525
Pro

<210> 4

<211> 726

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 4

Met Glu Arg Gln Lys Val Pro Gln Ile Arg Lys Ile Val Arg Val Val
1 5 10 15

Pro Leu Arg Ile Leu Arg Pro Lys Tyr Ser Asp Val Ile Glu Asn Ala
20 25 30

Leu Lys Lys Phe Lys Glu Lys Gly Asp Asp Thr Asn Thr Asn Asp Phe
35 40 45

Trp Arg Ala Ile Arg Asp Arg Asp Thr Glu Phe Phe Arg Lys Glu Leu
50 55 60

Asn Phe Ser Glu Asp Glu Ile Asn Gln Leu Glu Arg Asp Thr Leu Phe
65 70 75 80

Arg Val Gly Leu Asp Asn Arg Val Leu Phe Ser Tyr Phe Asp Phe Leu
85 90 95

Gln Glu Lys Leu Met Lys Asp Tyr Asn Lys Ile Ile Ser Lys Leu Phe
100 105 110

Ile Asn Arg Gln Ser Lys Ser Ser Phe Glu Asn Asp Leu Thr Asp Glu
115 120 125

Glu Val Glu Glu Leu Ile Glu Lys Asp Val Thr Pro Phe Tyr Gly Ala
130 135 140

Tyr Ile Gly Lys Gly Ile Lys Ser Val Ile Lys Ser Asn Leu Gly Gly
145 150 155 160

Lys Phe Ile Lys Ser Val Lys Ile Asp Arg Glu Thr Lys Lys Val Thr
165 170 175

Lys Leu Thr Ala Ile Asn Ile Gly Leu Met Gly Leu Pro Val Ala Lys
180 185 190

Ser Asp Thr Phe Pro Ile Lys Ile Ile Lys Thr Asn Pro Asp Tyr Ile

195 200 205

Thr Phe Gln Lys Ser Thr Lys Glu Asn Leu Gln Lys Ile Glu Asp Tyr

210 215 220

Glu Thr Gly Ile Glu Tyr Gly Asp Leu Leu Val Gln Ile Thr Ile Pro

225 230 235 240

Trp Phe Lys Asn Glu Asn Lys Asp Phe Ser Leu Ile Lys Thr Lys Glu

245 250 255

Ala Ile Glu Tyr Tyr Lys Leu Asn Gly Val Gly Lys Lys Asp Leu Leu

260 265 270

Asn Ile Asn Leu Val Leu Thr Thr Tyr His Ile Arg Lys Lys Lys Ser

275 280 285

Trp Gln Ile Asp Gly Ser Ser Gln Ser Leu Val Arg Glu Met Ala Asn

290 295 300

Gly Glu Leu Glu Glu Lys Trp Lys Ser Phe Phe Asp Thr Phe Ile Lys

305 310 315 320

Lys Tyr Gly Asp Glu Gly Lys Ser Ala Leu Val Lys Arg Arg Val Asn

325 330 335

Lys Lys Ser Arg Ala Lys Gly Glu Lys Gly Arg Glu Leu Asn Leu Asp

340 345 350

Glu Arg Ile Lys Arg Leu Tyr Asp Ser Ile Lys Ala Lys Ser Phe Pro

355 360 365

Ser Glu Ile Asn Leu Ile Pro Glu Asn Tyr Lys Trp Lys Leu His Phe

370 375 380

Ser Ile Glu Ile Pro Pro Met Val Asn Asp Ile Asp Ser Asn Leu Tyr

385 390 395 400

Gly Gly Ile Asp Phe Gly Glu Gln Asn Ile Ala Thr Leu Cys Val Lys

405 410 415

Asn Ile Glu Lys Asp Asp Tyr Asp Phe Leu Thr Ile Tyr Gly Asn Asp

420 425 430

Leu Leu Lys His Ala Gln Ala Ser Tyr Ala Arg Arg Arg Ile Met Arg

435 440 445

Val Gln Asp Glu Tyr Lys Ala Arg Gly His Gly Lys Ser Arg Lys Thr
450 455 460
Lys Ala Gln Glu Asp Tyr Ser Glu Arg Met Gln Lys Leu Arg Gln Lys
465 470 475 480
Ile Thr Glu Arg Leu Val Lys Gln Ile Ser Asp Phe Phe Leu Trp Arg

485 490 495
Asn Lys Phe His Met Ala Val Cys Ser Leu Arg Tyr Glu Asp Leu Asn
500 505 510
Thr Leu Tyr Lys Gly Glu Ser Val Lys Ala Lys Arg Met Arg Gln Phe
515 520 525
Ile Asn Lys Gln Gln Leu Phe Asn Gly Ile Glu Arg Lys Leu Lys Asp
530 535 540
Tyr Asn Ser Glu Ile Tyr Val Asn Ser Arg Tyr Pro His Tyr Thr Ser

545 550 555 560
Arg Leu Cys Ser Lys Cys Gly Lys Leu Asn Leu Tyr Phe Asp Phe Leu
565 570 575
Lys Phe Arg Thr Lys Asn Ile Ile Ile Arg Lys Asn Pro Asp Gly Ser
580 585 590
Glu Ile Lys Tyr Met Pro Phe Phe Ile Cys Glu Phe Cys Gly Trp Lys
595 600 605
Gln Ala Gly Asp Lys Asn Ala Ser Ala Asn Ile Ala Asp Lys Asp Tyr

610 615 620
Gln Asp Lys Leu Asn Lys Glu Lys Glu Phe Cys Asn Ile Arg Lys Pro
625 630 635 640
Lys Ser Lys Lys Glu Asp Ile Gly Glu Glu Asn Glu Glu Glu Arg Asp
645 650 655
Tyr Ser Arg Arg Phe Asn Arg Asn Ser Phe Ile Tyr Asn Ser Leu Lys
660 665 670
Lys Asp Asn Lys Leu Asn Gln Glu Lys Leu Phe Asp Glu Trp Lys Asn

675 680 685
Gln Leu Lys Arg Lys Ile Asp Gly Arg Asn Lys Phe Glu Pro Lys Glu

690 695 700

Tyr Lys Asp Arg Phe Ser Tyr Leu Phe Ala Tyr Tyr Gln Glu Ile Ile

705 710 715 720

Lys Asn Glu Ser Glu Ser

725

<210> 5

<211> 517

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 5

Met Val Pro Thr Glu Leu Ile Thr Lys Thr Leu Gln Leu Arg Val Ile

1 5 10 15

Arg Pro Leu Tyr Phe Glu Glu Ile Glu Lys Glu Leu Ala Glu Leu Lys

20 25 30

Glu Gln Lys Glu Lys Glu Phe Glu Glu Thr Asn Ser Leu Leu Leu Glu

35 40 45

Ser Lys Lys Ile Asp Ala Lys Ser Leu Lys Lys Leu Lys Arg Lys Ala

50 55 60

Arg Ser Ser Ala Ala Val Glu Phe Trp Lys Ile Ala Lys Glu Lys Tyr

65 70 75 80

Pro Asp Ile Leu Thr Lys Pro Glu Met Glu Phe Ile Phe Ser Glu Met

85 90 95

Gln Lys Met Met Ala Arg Phe Tyr Asn Lys Ser Met Thr Asn Ile Phe

100 105 110

Ile Glu Met Asn Asn Asp Glu Lys Val Asn Pro Leu Ser Leu Ile Ser

115 120 125

Lys Ala Ser Thr Glu Ala Asn Gln Val Ile Lys Cys Ser Ser Ile Ser

130 135 140

Ser Gly Leu Asn Arg Lys Ile Ala Gly Ser Ile Asn Lys Thr Lys Phe

145 150 155 160

Lys Gln Val Arg Asp Gly Leu Ile Ser Leu Pro Thr Ala Arg Thr Glu

165 170 175

Thr Phe Pro Ile Ser Phe Tyr Lys Ser Thr Ala Asn Lys Asp Glu Ile

180 185 190

Pro Ile Ser Lys Ile Asn Leu Pro Ser Glu Glu Glu Ala Asp Leu Thr

195 200 205

Ile Thr Leu Pro Phe Pro Phe Phe Glu Ile Lys Lys Glu Lys Lys Gly

210 215 220

Gln Lys Ala Tyr Ser Tyr Phe Asn Ile Ile Glu Lys Ser Gly Arg Ser

225 230 235 240

Asn Asn Lys Ile Asp Leu Leu Leu Ser Thr His Arg Arg Gln Arg Arg

245 250 255

Lys Gly Trp Lys Glu Glu Gly Gly Thr Ser Ala Glu Ile Arg Arg Leu

260 265 270

Met Glu Gly Glu Phe Asp Lys Glu Trp Glu Ile Tyr Leu Gly Glu Ala

275 280 285

Glu Lys Ser Glu Lys Ala Lys Asn Asp Leu Ile Lys Asn Met Thr Arg

290 295 300

Gly Lys Leu Ser Lys Asp Ile Lys Glu Gln Leu Glu Asp Ile Gln Val

305 310 315 320

Lys Tyr Phe Ser Asp Asn Asn Val Glu Ser Trp Asn Asp Leu Ser Lys

325 330 335

Glu Gln Lys Gln Glu Leu Ser Lys Leu Arg Lys Lys Lys Val Glu Glu

340 345 350

Leu Lys Asp Trp Lys His Val Lys Glu Ile Leu Lys Thr Arg Ala Lys

355 360 365

Ile Gly Trp Val Glu Leu Lys Arg Gly Lys Arg Gln Arg Asp Arg Asn

370 375 380

Lys Trp Phe Val Asn Ile Thr Ile Thr Arg Pro Pro Phe Ile Asn Lys

385 390 395 400

Glu Leu Asp Asp Thr Lys Phe Gly Gly Ile Asp Leu Gly Val Lys Val

405 410 415

Pro Phe Val Cys Ala Val His Gly Ser Pro Ala Arg Leu Ile Ile Lys
420 425 430
Glu Asn Glu Ile Leu Gln Phe Asn Lys Met Val Ser Ala Arg Asn Arg
435 440 445
Gln Ile Thr Lys Asp Ser Glu Gln Arg Lys Gly Arg Gly Lys Lys Asn

450 455 460
Lys Phe Ile Lys Lys Glu Ile Phe Asn Glu Arg Asn Glu Leu Phe Arg
465 470 475 480
Lys Lys Ile Ile Glu Arg Trp Ala Asn Gln Ile Val Lys Phe Phe Glu
485 490 495
Asp Gln Lys Cys Ala Thr Val Gln Ile Glu Asn Leu Glu Ser Phe Asp
500 505 510
Arg Thr Ser Tyr Lys
515

<210
> 6
<211> 481
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 6
Met Lys Ser Asp Thr Lys Asp Lys Lys Ile Ile Ile His Gln Thr Lys
1 5 10 15
Thr Leu Ser Leu Arg Ile Val Lys Pro Gln Ser Ile Pro Met Glu Glu
20 25 30
Phe Thr Asp Leu Val Arg Tyr His Gln Met Ile Ile Phe Pro Val Tyr
35 40 45
Asn Asn Gly Ala Ile Asp Leu Tyr Lys Lys Leu Phe Lys Ala Lys Ile

50 55 60
Gln Lys Gly Asn Glu Ala Arg Ala Ile Lys Tyr Phe Met Asn Lys Ile
65 70 75 80
Val Tyr Ala Pro Ile Ala Asn Thr Val Lys Asn Ser Tyr Ile Ala Leu
85 90 95

Gly Tyr Ser Thr Lys Met Gln Ser Ser Phe Ser Gly Lys Arg Leu Trp
100 105 110
Asp Leu Arg Phe Gly Glu Ala Thr Pro Pro Thr Ile Lys Ala Asp Phe

115 120 125
Pro Leu Pro Phe Tyr Asn Gln Ser Gly Phe Lys Val Ser Ser Glu Asn
130 135 140
Gly Glu Phe Ile Ile Gly Ile Pro Phe Gly Gln Tyr Thr Lys Lys Thr
145 150 155 160
Val Ser Asp Ile Glu Lys Lys Thr Ser Phe Ala Trp Asp Lys Phe Thr
165 170 175
Leu Glu Asp Thr Thr Lys Lys Thr Leu Ile Glu Leu Leu Leu Ser Thr

180 185 190
Lys Thr Arg Lys Met Asn Glu Gly Trp Lys Asn Asn Glu Gly Thr Glu
195 200 205
Ala Glu Ile Lys Arg Val Met Asp Gly Thr Tyr Gln Val Thr Ser Leu
210 215 220
Glu Ile Leu Gln Arg Asp Asp Ser Trp Phe Val Asn Phe Asn Ile Ala
225 230 235 240
Tyr Asp Ser Leu Lys Lys Gln Pro Asp Arg Asp Lys Ile Ala Gly Ile

245 250 255
His Met Gly Ile Thr Arg Pro Leu Thr Ala Val Ile Tyr Asn Asn Lys
260 265 270
Tyr Arg Ala Leu Ser Ile Tyr Pro Asn Thr Val Met His Leu Thr Gln
275 280 285
Lys Gln Leu Ala Arg Ile Lys Glu Gln Arg Thr Asn Ser Lys Tyr Ala
290 295 300
Thr Gly Gly His Gly Arg Asn Ala Lys Val Thr Gly Thr Asp Thr Leu

305 310 315 320
Ser Glu Ala Tyr Arg Gln Arg Arg Lys Lys Ile Ile Glu Asp Trp Ile
325 330 335
Ala Ser Ile Val Lys Phe Ala Ile Asn Asn Glu Ile Gly Thr Ile Tyr

340 345 350

Leu Glu Asp Ile Ser Asn Thr Asn Ser Phe Phe Ala Ala Arg Glu Gln

355 360 365

Lys Leu Ile Tyr Leu Glu Asp Ile Ser Asn Thr Asn Ser Phe Leu Ser

370 375 380

Thr Tyr Lys Tyr Pro Ile Ser Ala Ile Ser Asp Thr Leu Gln His Lys

385 390 395 400

Leu Glu Glu Lys Ala Ile Gln Val Ile Arg Lys Lys Ala Tyr Tyr Val

405 410 415

Asn Gln Ile Cys Ser Leu Cys Gly His Tyr Asn Lys Gly Phe Thr Tyr

420 425 430

Gln Phe Arg Arg Lys Asn Lys Phe Pro Lys Met Lys Cys Gln Gly Cys

435 440 445

Leu Glu Ala Thr Ser Thr Glu Phe Asn Ala Ala Ala Asn Val Ala Asn

450 455 460

Pro Asp Tyr Glu Lys Leu Leu Ile Lys His Gly Leu Leu Gln Leu Lys

465 470 475 480

Lys

<210> 7

<211> 358

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 7

Met Ser Thr Ile Thr Arg Gln Val Arg Leu Ser Pro Thr Pro Glu Gln

1 5 10 15

Ser Arg Leu Leu Met Ala His Cys Gln Gln Tyr Ile Ser Thr Val Asn

20 25 30

Val Leu Val Ala Ala Phe Asp Ser Glu Val Leu Thr Gly Lys Val Ser

35 40 45

Thr Lys Asp Phe Arg Ala Ala Leu Pro Ser Ala Val Lys Asn Gln Ala

50 55 60

Leu Arg Asp Ala Gln Ser Val Phe Lys Arg Ser Val Glu Leu Gly Cys

65 70 75 80

Leu Pro Val Leu Lys Lys Pro His Cys Gln Trp Asn Asn Gln Asn Trp

85 90 95

Arg Val Glu Gly Asp Gln Leu Ile Leu Pro Ile Cys Lys Asp Gly Lys

100 105 110

Thr Gln Gln Glu Arg Phe Arg Cys Ala Ala Val Ala Leu Glu Gly Lys

115 120 125

Ala Gly Ile Leu Arg Ile Lys Lys Lys Arg Gly Lys Trp Ile Ala Asp

130 135 140

Leu Thr Val Thr Gln Glu Asp Ala Pro Glu Ser Ser Gly Ser Ala Ile

145 150 155 160

Met Gly Val Asp Leu Gly Ile Lys Val Pro Ala Val Ala His Ile Gly

165 170 175

Gly Lys Gly Thr Arg Phe Phe Gly Asn Gly Arg Ser Gln Arg Ser Met

180 185 190

Arg Arg Arg Phe Tyr Ala Arg Arg Lys Thr Leu Gln Lys Ala Lys Lys

195 200 205

Leu Arg Ala Val Arg Lys Ser Lys Gly Lys Glu Ala Arg Trp Met Lys

210 215 220

Thr Ile Asn His Gln Leu Ser Arg Gln Ile Val Asn His Ala His Ala

225 230 235 240

Leu Gly Val Gly Thr Ile Lys Ile Glu Ala Leu Gln Gly Ile Arg Lys

245 250 255

Gly Thr Thr Arg Lys Ser Arg Gly Ala Ala Ala Arg Lys Asn Asn Arg

260 265 270

Met Thr Asn Thr Trp Ser Phe Ser Gln Leu Thr Leu Phe Ile Thr Tyr

275 280 285

Lys Ala Gln Arg Gln Gly Ile Thr Val Glu Gln Val Asp Pro Ala Tyr

290 295 300

Thr Ser Gln Asp Cys Pro Ala Cys Arg Ala Arg Asn Gly Ala Gln Asp
305 310 315 320
Arg Thr Tyr Val Cys Ser Glu Cys Gly Trp Arg Gly His Arg Asp Thr

325 330 335
Val Gly Ala Ile Asn Ile Ser Arg Arg Ala Gly Leu Ser Gly His Arg
340 345 350
Arg Gly Ala Thr Gly Ala
355

<210> 8
<211> 507
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 8
Met Ile Ala Gln Lys Thr Ile Lys Ile Lys Leu Asn Pro Thr Lys Glu
1 5 10 15
Gln Ile Ile Lys Leu Asn Ser Ile Ile Glu Glu Tyr Ile Lys Val Ser

20 25 30
Asn Phe Thr Ala Lys Lys Ile Ala Glu Ile Gln Glu Ser Phe Thr Asp
35 40 45
Ser Gly Leu Thr Gln Gly Thr Cys Ser Glu Cys Gly Lys Glu Lys Thr
50 55 60
Tyr Arg Lys Tyr His Leu Leu Lys Lys Asp Asn Lys Leu Phe Cys Ile
65 70 75 80
Thr Cys Tyr Lys Arg Lys Tyr Ser Gln Phe Thr Leu Gln Lys Val Glu

85 90 95
Phe Gln Asn Lys Thr Gly Leu Arg Asn Val Ala Lys Leu Pro Lys Thr
100 105 110
Tyr Tyr Thr Asn Ala Ile Arg Phe Ala Ser Asp Thr Phe Ser Gly Phe
115 120 125
Asp Glu Ile Ile Lys Lys Lys Gln Asn Arg Leu Asn Ser Ile Gln Asn
130 135 140

Arg Leu Asn Phe Trp Lys Glu Leu Leu Tyr Asn Pro Ser Asn Arg Asn

145 150 155 160

Glu Ile Lys Ile Lys Val Val Lys Tyr Ala Pro Lys Thr Asp Thr Arg

165 170 175

Glu His Pro His Tyr Tyr Ser Glu Ala Glu Ile Lys Gly Arg Ile Lys

180 185 190

Arg Leu Glu Lys Gln Leu Lys Lys Phe Lys Met Pro Lys Tyr Pro Glu

195 200 205

Phe Thr Ser Glu Thr Ile Ser Leu Gln Arg Glu Leu Tyr Ser Trp Lys

210 215 220

Asn Pro Asp Glu Leu Lys Ile Ser Ser Ile Thr Asp Lys Asn Glu Ser

225 230 235 240

Met Asn Tyr Tyr Gly Lys Glu Tyr Leu Lys Arg Tyr Ile Asp Leu Ile

245 250 255

Asn Ser Gln Thr Pro Gln Ile Leu Leu Glu Lys Glu Asn Asn Ser Phe

260 265 270

Tyr Leu Cys Phe Pro Ile Thr Lys Asn Ile Glu Met Pro Lys Ile Asp

275 280 285

Asp Thr Phe Glu Pro Val Gly Ile Asp Trp Gly Ile Thr Arg Asn Ile

290 295 300

Ala Val Val Ser Ile Leu Asp Ser Lys Thr Lys Lys Pro Lys Phe Val

305 310 315 320

Lys Phe Tyr Ser Ala Gly Tyr Ile Leu Gly Lys Arg Lys His Tyr Lys

325 330 335

Ser Leu Arg Lys His Phe Gly Gln Lys Lys Arg Gln Asp Lys Ile Asn

340 345 350

Lys Leu Gly Thr Lys Glu Asp Arg Phe Ile Asp Ser Asn Ile His Lys

355 360 365

Leu Ala Phe Leu Ile Val Lys Glu Ile Arg Asn His Ser Asn Lys Pro

370 375 380

Ile Ile Leu Met Glu Asn Ile Thr Asp Asn Arg Glu Glu Ala Glu Lys

385 390 395 400

Ser Met Arg Gln Asn Ile Leu Leu His Ser Val Lys Ser Arg Leu Gln

405 410 415

Asn Tyr Ile Ala Tyr Lys Ala Leu Trp Asn Asn Ile Pro Thr Asn Leu

420 425 430

Val Lys Pro Glu His Thr Ser Gln Ile Cys Asn Arg Cys Gly His Gln

435 440 445

Asp Arg Glu Asn Arg Pro Lys Gly Ser Lys Leu Phe Lys Cys Val Lys

450 455 460

Cys Asn Tyr Met Ser Asn Ala Asp Phe Asn Ala Ser Ile Asn Ile Ala

465 470 475 480

Arg Lys Phe Tyr Ile Gly Glu Tyr Glu Pro Phe Tyr Lys Asp Asn Glu

485 490 495

Lys Met Lys Ser Gly Val Asn Ser Ile Ser Met

500 505

<210> 9

<211> 534

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 9

Leu Lys Leu Ser Glu Gln Glu Asn Ile Thr Thr Gly Val Lys Phe Lys

1 5 10 15

Leu Lys Leu Asp Lys Glu Thr Ser Glu Gly Leu Asn Asp Tyr Phe Asp

20 25 30

Glu Tyr Gly Lys Ala Ile Asn Phe Ala Ile Lys Val Ile Gln Lys Glu

35 40 45

Leu Ala Glu Asp Arg Phe Ala Gly Lys Val Arg Leu Asp Glu Asn Lys

50 55 60

Lys Pro Leu Leu Asn Glu Asp Gly Lys Lys Ile Trp Asp Phe Pro Asn

65 70 75 80

Glu Phe Cys Ser Cys Gly Lys Gln Val Asn Arg Tyr Val Asn Gly Lys
85 90 95
Ser Leu Cys Gln Glu Cys Tyr Lys Asn Lys Phe Thr Glu Tyr Gly Ile
100 105 110
Arg Lys Arg Met Tyr Ser Ala Lys Gly Arg Lys Ala Glu Gln Asp Ile
115 120 125
Asn Ile Lys Asn Ser Thr Asn Lys Ile Ser Lys Thr His Phe Asn Tyr
130 135 140

Ala Ile Arg Glu Ala Phe Ile Leu Asp Lys Ser Ile Lys Lys Gln Arg
145 150 155 160
Lys Glu Arg Phe Arg Arg Leu Arg Glu Met Lys Lys Lys Leu Gln Glu
165 170 175
Phe Ile Glu Ile Arg Asp Gly Asn Lys Ile Leu Cys Pro Lys Ile Glu
180 185 190
Lys Gln Arg Val Glu Arg Tyr Ile His Pro Ser Trp Ile Asn Lys Glu
195 200 205

Lys Lys Leu Glu Asp Phe Arg Gly Tyr Ser Met Ser Asn Val Leu Gly
210 215 220
Lys Ile Lys Ile Leu Asp Arg Asn Ile Lys Arg Glu Glu Lys Ser Leu
225 230 235 240
Lys Glu Lys Gly Gln Ile Asn Phe Lys Ala Arg Arg Leu Met Leu Asp
245 250 255
Lys Ser Val Lys Phe Leu Asn Asp Asn Lys Ile Ser Phe Thr Ile Ser
260 265 270

Lys Asn Leu Pro Lys Glu Tyr Glu Leu Asp Leu Pro Glu Lys Glu Lys
275 280 285
Arg Leu Asn Trp Leu Lys Glu Lys Ile Lys Ile Ile Lys Asn Gln Lys
290 295 300
Pro Lys Tyr Ala Tyr Leu Leu Arg Lys Asp Asp Asn Phe Tyr Leu Gln
305 310 315 320
Tyr Thr Leu Glu Thr Glu Phe Asn Leu Lys Glu Asp Tyr Ser Gly Ile

325 330 335

Val Gly Ile Asp Arg Gly Val Ser His Ile Ala Val Tyr Thr Phe Val
340 345 350
His Asn Asn Gly Lys Asn Glu Arg Pro Leu Phe Leu Asn Ser Ser Glu
355 360 365
Ile Leu Arg Leu Lys Asn Leu Gln Lys Glu Arg Asp Arg Phe Leu Arg
370 375 380
Arg Lys His Asn Lys Lys Arg Lys Lys Ser Asn Met Arg Asn Ile Glu
385 390 395 400

Lys Lys Ile Gln Leu Ile Leu His Asn Tyr Ser Lys Gln Ile Val Asp
405 410 415
Phe Ala Lys Asn Lys Asn Ala Phe Ile Val Phe Glu Lys Leu Glu Lys
420 425 430
Pro Lys Lys Asn Arg Ser Lys Met Ser Lys Lys Ser Gln Tyr Lys Leu
435 440 445
Ser Gln Phe Thr Phe Lys Lys Leu Ser Asp Leu Val Asp Tyr Lys Ala
450 455 460

Lys Arg Glu Gly Ile Lys Val Leu Tyr Ile Ser Pro Glu Tyr Thr Ser
465 470 475 480
Lys Glu Cys Ser His Cys Gly Glu Lys Val Asn Thr Gln Arg Pro Phe
485 490 495
Asn Gly Asn Ser Ser Leu Phe Lys Cys Asn Lys Cys Gly Val Glu Leu
500 505 510
Asn Ala Asp Tyr Asn Ala Ser Ile Asn Ile Ala Lys Lys Gly Leu Asn
515 520 525

Ile Leu Asn Ser Thr Asn
530

<210> 10
<211> 537
<212> PRT
<213> Artificial sequence

<220><223> Synthetic sequence

<400> 10

Met Glu Glu Ser Ile Ile Thr Gly Val Lys Phe Lys Leu Arg Ile Asp
1 5 10 15

Lys Glu Thr Thr Lys Lys Leu Asn Glu Tyr Phe Asp Glu Tyr Gly Lys
20 25 30

Ala Ile Asn Phe Ala Val Lys Ile Ile Gln Lys Glu Leu Ala Asp Asp
35 40 45

Arg Phe Ala Gly Lys Ala Lys Leu Asp Gln Asn Lys Asn Pro Ile Leu
50 55 60

Asp Glu Asn Gly Lys Lys Ile Tyr Glu Phe Pro Asp Glu Phe Cys Ser
65 70 75 80

Cys Gly Lys Gln Val Asn Lys Tyr Val Asn Asn Lys Pro Phe Cys Gln
85 90 95

Glu Cys Tyr Lys Ile Arg Phe Thr Glu Asn Gly Ile Arg Lys Arg Met
100 105 110

Tyr Ser Ala Lys Gly Arg Lys Ala Glu His Lys Ile Asn Ile Leu Asn
115 120 125

Ser Thr Asn Lys Ile Ser Lys Thr His Phe Asn Tyr Ala Ile Arg Glu
130 135 140

Ala Phe Ile Leu Asp Lys Ser Ile Lys Lys Gln Arg Lys Lys Arg Asn
145 150 155 160

Glu Arg Leu Arg Glu Ser Lys Lys Arg Leu Gln Gln Phe Ile Asp Met
165 170 175

Arg Asp Gly Lys Arg Glu Ile Cys Pro Thr Ile Lys Gly Gln Lys Val
180 185 190

Asp Arg Phe Ile His Pro Ser Trp Ile Thr Lys Asp Lys Lys Leu Glu
195 200 205

Asp Phe Arg Gly Tyr Thr Leu Ser Ile Ile Asn Ser Lys Ile Lys Ile
210 215 220

Leu Asp Arg Asn Ile Lys Arg Glu Glu Lys Ser Leu Lys Glu Lys Gly

225 230 235 240

Gln Ile Ile Phe Lys Ala Lys Arg Leu Met Leu Asp Lys Ser Ile Arg
245 250 255
Phe Val Gly Asp Arg Lys Val Leu Phe Thr Ile Ser Lys Thr Leu Pro
260 265 270
Lys Glu Tyr Glu Leu Asp Leu Pro Ser Lys Glu Lys Arg Leu Asn Trp
275 280 285
Leu Lys Glu Lys Ile Glu Ile Ile Lys Asn Gln Lys Pro Lys Tyr Ala
290 295 300

Tyr Leu Leu Arg Lys Asn Ile Glu Ser Glu Lys Lys Pro Asn Tyr Glu
305 310 315 320
Tyr Tyr Leu Gln Tyr Thr Leu Glu Ile Lys Pro Glu Leu Lys Asp Phe
325 330 335
Tyr Asp Gly Ala Ile Gly Ile Asp Arg Gly Ile Asn His Ile Ala Val
340 345 350
Cys Thr Phe Ile Ser Asn Asp Gly Lys Val Thr Pro Pro Lys Phe Phe
355 360 365

Ser Ser Gly Glu Ile Leu Arg Leu Lys Asn Leu Gln Lys Glu Arg Asp
370 375 380
Arg Phe Leu Leu Arg Lys His Asn Lys Asn Arg Lys Lys Gly Asn Met
385 390 395 400
Arg Val Ile Glu Asn Lys Ile Asn Leu Ile Leu His Arg Tyr Ser Lys
405 410 415
Gln Ile Val Asp Met Ala Lys Lys Leu Asn Ala Ser Ile Val Phe Glu
420 425 430

Glu Leu Gly Arg Ile Gly Lys Ser Arg Thr Lys Met Lys Lys Ser Gln
435 440 445
Arg Tyr Lys Leu Ser Leu Phe Ile Phe Lys Lys Leu Ser Asp Leu Val
450 455 460
Asp Tyr Lys Ser Arg Arg Glu Gly Ile Arg Val Thr Tyr Val Pro Pro
465 470 475 480

Glu Tyr Thr Ser Lys Glu Cys Ser His Cys Gly Glu Lys Val Asn Thr
485 490 495

Gln Arg Pro Phe Asn Gly Asn Tyr Ser Leu Phe Lys Cys Asn Lys Cys
500 505 510
Gly Ile Gln Leu Asn Ser Asp Tyr Asn Ala Ser Ile Asn Ile Ala Lys
515 520 525
Lys Gly Leu Lys Ile Pro Asn Ser Thr
530 535

<210> 11
<211> 540
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 11
Leu Trp Thr Ile Val Ile Gly Asp Phe Ile Glu Met Pro Lys Gln Asp

1 5 10 15
Leu Val Thr Thr Gly Ile Lys Phe Lys Leu Asp Val Asp Lys Glu Thr
20 25 30
Arg Lys Lys Leu Asp Asp Tyr Phe Asp Glu Tyr Gly Lys Ala Ile Asn
35 40 45
Phe Ala Val Lys Ile Ile Gln Lys Asn Leu Lys Glu Asp Arg Phe Ala
50 55 60
Gly Lys Ile Ala Leu Gly Glu Asp Lys Lys Pro Leu Leu Asp Lys Asp

65 70 75 80
Gly Lys Lys Ile Tyr Asn Tyr Pro Asn Glu Ser Cys Ser Cys Gly Asn
85 90 95
Gln Val Arg Arg Tyr Val Asn Ala Lys Pro Phe Cys Val Asp Cys Tyr
100 105 110
Lys Leu Lys Phe Thr Glu Asn Gly Ile Arg Lys Arg Met Tyr Ser Ala
115 120 125
Arg Gly Arg Lys Ala Asp Ser Asp Ile Asn Ile Lys Asn Ser Thr Asn

130 135 140

Lys Ile Ser Lys Thr His Phe Asn Tyr Ala Ile Arg Glu Gly Phe Ile

145 150 155 160

Leu Asp Lys Ser Leu Lys Lys Gln Arg Ser Lys Arg Ile Lys Lys Leu

165 170 175

Leu Glu Leu Lys Arg Lys Leu Gln Glu Phe Ile Asp Ile Arg Gln Gly

180 185 190

Gln Met Val Leu Cys Pro Lys Ile Lys Asn Gln Arg Val Asp Lys Phe

195 200 205

Ile His Pro Ser Trp Leu Lys Arg Asp Lys Lys Leu Glu Glu Phe Arg

210 215 220

Gly Tyr Ser Leu Ser Val Val Glu Gly Lys Ile Lys Ile Phe Asn Arg

225 230 235 240

Asn Ile Leu Arg Glu Glu Asp Ser Leu Arg Gln Arg Gly His Val Asn

245 250 255

Phe Lys Ala Asn Arg Ile Met Leu Asp Lys Ser Val Arg Phe Leu Asp

260 265 270

Gly Gly Lys Val Asn Phe Asn Leu Asn Lys Gly Leu Pro Lys Glu Tyr

275 280 285

Leu Leu Asp Leu Pro Lys Lys Glu Asn Lys Leu Ser Trp Leu Asn Glu

290 295 300

Lys Ile Ser Leu Ile Lys Leu Gln Lys Pro Lys Tyr Ala Tyr Leu Leu

305 310 315 320

Arg Arg Glu Gly Ser Phe Phe Ile Gln Tyr Thr Ile Glu Asn Val Pro

325 330 335

Lys Thr Phe Ser Asp Tyr Leu Gly Ala Ile Gly Ile Asp Arg Gly Ile

340 345 350

Ser His Ile Ala Val Cys Thr Phe Val Ser Lys Asn Gly Val Asn Lys

355 360 365

Ala Pro Val Phe Phe Ser Ser Gly Glu Ile Leu Lys Leu Lys Ser Leu

370 375 380

Gln Lys Gln Arg Asp Leu Phe Leu Arg Gly Lys His Asn Lys Ile Arg

385 390 395 400

Lys Lys Ser Asn Met Arg Asn Ile Asp Asn Lys Ile Asn Leu Ile Leu

405 410 415

His Lys Tyr Ser Arg Asn Ile Val Asn Leu Ala Lys Ser Glu Lys Ala

420 425 430

Phe Ile Val Phe Glu Lys Leu Glu Lys Ile Lys Lys Ser Arg Phe Lys

435 440 445

Met Ser Lys Ser Leu Gln Tyr Lys Leu Ser Gln Phe Thr Phe Lys Lys

450 455 460

Leu Ser Asp Leu Val Glu Tyr Lys Ala Lys Ile Glu Gly Ile Lys Val

465 470 475 480

Asp Tyr Val Pro Pro Glu Tyr Thr Ser Lys Glu Cys Ser His Cys Gly

485 490 495

Glu Lys Val Asp Thr Gln Arg Pro Phe Asn Gly Asn Ser Ser Leu Phe

500 505 510

Lys Cys Asn Lys Cys Arg Val Gln Leu Asn Ala Asp Tyr Asn Ala Ser

515 520 525

Ile Asn Ile Ala Lys Lys Ser Leu Asn Ile Ser Asn

530 535 540

<210> 12

<211> 542

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 12

Met Ser Lys Thr Thr Ile Ser Val Lys Leu Lys Ile Ile Asp Leu Ser

1 5 10 15

Ser Glu Lys Lys Glu Phe Leu Asp Asn Tyr Phe Asn Glu Tyr Ala Lys

20 25 30

Ala Thr Thr Phe Cys Gln Leu Arg Ile Arg Arg Leu Leu Arg Asn Thr

35 40 45

His Trp Leu Gly Lys Lys Glu Lys Ser Ser Lys Lys Trp Ile Phe Glu

50 55 60

Ser Gly Ile Cys Asp Leu Cys Gly Glu Asn Lys Glu Leu Val Asn Glu

65 70 75 80

Asp Arg Asn Ser Gly Glu Pro Ala Lys Ile Cys Lys Arg Cys Tyr Asn

85 90 95

Gly Arg Tyr Gly Asn Gln Met Ile Arg Lys Leu Phe Val Ser Thr Lys

100 105 110

Lys Arg Glu Val Gln Glu Asn Met Asp Ile Arg Arg Val Ala Lys Leu

115 120 125

Asn Asn Thr His Tyr His Arg Ile Pro Glu Glu Ala Phe Asp Met Ile

130 135 140

Lys Ala Ala Asp Thr Ala Glu Lys Arg Arg Lys Lys Asn Val Glu Tyr

145 150 155 160

Asp Lys Lys Arg Gln Met Glu Phe Ile Glu Met Phe Asn Asp Glu Lys

165 170 175

Lys Arg Ala Ala Arg Pro Lys Lys Pro Asn Glu Arg Glu Thr Arg Tyr

180 185 190

Val His Ile Ser Lys Leu Glu Ser Pro Ser Lys Gly Tyr Thr Leu Asn

195 200 205

Gly Ile Lys Arg Lys Ile Asp Gly Met Gly Lys Lys Ile Glu Arg Ala

210 215 220

Glu Lys Gly Leu Ser Arg Lys Lys Ile Phe Gly Tyr Gln Gly Asn Arg

225 230 235 240

Ile Lys Leu Asp Ser Asn Trp Val Arg Phe Asp Leu Ala Glu Ser Glu

245 250 255

Ile Thr Ile Pro Ser Leu Phe Lys Glu Met Lys Leu Arg Ile Thr Gly

260 265 270

Pro Thr Asn Val His Ser Lys Ser Gly Gln Ile Tyr Phe Ala Glu Trp

275 280 285

Phe Glu Arg Ile Asn Lys Gln Pro Asn Asn Tyr Cys Tyr Leu Ile Arg
290 295 300
Lys Thr Ser Ser Asn Gly Lys Tyr Glu Tyr Tyr Leu Gln Tyr Thr Tyr
305 310 315 320
Glu Ala Glu Val Glu Ala Asn Lys Glu Tyr Ala Gly Cys Leu Gly Val
325 330 335
Asp Ile Gly Cys Ser Lys Leu Ala Ala Ala Val Tyr Tyr Asp Ser Lys
340 345 350

Asn Lys Lys Ala Gln Lys Pro Ile Glu Ile Phe Thr Asn Pro Ile Lys
355 360 365
Lys Ile Lys Met Arg Arg Glu Lys Leu Ile Lys Leu Leu Ser Arg Val
370 375 380
Lys Val Arg His Arg Arg Arg Lys Leu Met Gln Leu Ser Lys Thr Glu
385 390 395 400
Pro Ile Ile Asp Tyr Thr Cys His Lys Thr Ala Arg Lys Ile Val Glu
405 410 415

Met Ala Asn Thr Ala Lys Ala Phe Ile Ser Met Glu Asn Leu Glu Thr
420 425 430
Gly Ile Lys Gln Lys Gln Gln Ala Arg Glu Thr Lys Lys Gln Lys Phe
435 440 445
Tyr Arg Asn Met Phe Leu Phe Arg Lys Leu Ser Lys Leu Ile Glu Tyr
450 455 460
Lys Ala Leu Leu Lys Gly Ile Lys Ile Val Tyr Val Lys Pro Asp Tyr
465 470 475 480

Thr Ser Gln Thr Cys Ser Ser Cys Gly Ala Asp Lys Glu Lys Thr Glu
485 490 495
Arg Pro Ser Gln Ala Ile Phe Arg Cys Leu Asn Pro Thr Cys Arg Tyr
500 505 510
Tyr Gln Arg Asp Ile Asn Ala Asp Phe Asn Ala Ala Val Asn Ile Ala
515 520 525
Lys Lys Ala Leu Asn Asn Thr Glu Val Val Thr Thr Leu Leu

530 535 540

<210> 13

<211> 564

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 13

Met Ala Arg Ala Lys Asn Gln Pro Tyr Gln Lys Leu Thr Thr Thr Thr
1 5 10 15

Gly Ile Lys Phe Lys Leu Asp Leu Ser Glu Glu Glu Gly Lys Arg Phe
20 25 30

Asp Glu Tyr Phe Ser Glu Tyr Ala Lys Ala Val Asn Phe Cys Ala Lys
35 40 45

Val Ile Tyr Gln Leu Arg Lys Asn Leu Lys Phe Ala Gly Lys Lys Glu

50 55 60

Leu Ala Ala Lys Glu Trp Lys Phe Glu Ile Ser Asn Cys Asp Phe Cys
65 70 75 80

Asn Lys Gln Lys Glu Ile Tyr Tyr Lys Asn Ile Ala Asn Gly Gln Lys
85 90 95

Val Cys Lys Gly Cys His Arg Thr Asn Phe Ser Asp Asn Ala Ile Arg
100 105 110

Lys Lys Met Ile Pro Val Lys Gly Arg Lys Val Glu Ser Lys Phe Asn

115 120 125

Ile His Asn Thr Thr Lys Lys Ile Ser Gly Thr His Arg His Trp Ala
130 135 140

Phe Glu Asp Ala Ala Asp Ile Ile Glu Ser Met Asp Lys Gln Arg Lys
145 150 155 160

Glu Lys Gln Lys Arg Leu Arg Arg Glu Lys Arg Lys Leu Ser Tyr Phe
165 170 175

Phe Glu Leu Phe Gly Asp Pro Ala Lys Arg Tyr Glu Leu Pro Lys Val

180 185 190

Gly Lys Gln Arg Val Pro Arg Tyr Leu His Lys Ile Ile Asp Lys Asp
195 200 205
Ser Leu Thr Lys Lys Arg Gly Tyr Ser Leu Ser Tyr Ile Lys Asn Lys
210 215 220
Ile Lys Ile Ser Glu Arg Asn Ile Glu Arg Asp Glu Lys Ser Leu Arg
225 230 235 240
Lys Ala Ser Pro Ile Ala Phe Gly Ala Arg Lys Ile Lys Met Ser Lys

245 250 255
Leu Asp Pro Lys Arg Ala Phe Asp Leu Glu Asn Asn Val Phe Lys Ile
260 265 270
Pro Gly Lys Val Ile Lys Gly Gln Tyr Lys Phe Phe Gly Thr Asn Val
275 280 285
Ala Asn Glu His Gly Lys Lys Phe Tyr Lys Asp Arg Ile Ser Lys Ile
290 295 300
Leu Ala Gly Lys Pro Lys Tyr Phe Tyr Leu Leu Arg Lys Lys Val Ala

305 310 315 320
Glu Ser Asp Gly Asn Pro Ile Phe Glu Tyr Tyr Val Gln Trp Ser Ile
325 330 335
Asp Thr Glu Thr Pro Ala Ile Thr Ser Tyr Asp Asn Ile Leu Gly Ile
340 345 350
Asp Ala Gly Ile Thr Asn Leu Ala Thr Thr Val Leu Ile Pro Lys Asn
355 360 365
Leu Ser Ala Glu His Cys Ser His Cys Gly Asn Asn His Val Lys Pro

370 375 380
Ile Phe Thr Lys Phe Phe Ser Gly Lys Glu Leu Lys Ala Ile Lys Ile
385 390 395 400
Lys Ser Arg Lys Gln Lys Tyr Phe Leu Arg Gly Lys His Asn Lys Leu
405 410 415
Val Lys Ile Lys Arg Ile Arg Pro Ile Glu Gln Lys Val Asp Gly Tyr
420 425 430
Cys His Val Val Ser Lys Gln Ile Val Glu Met Ala Lys Glu Arg Asn

435 440 445

Ser Cys Ile Ala Leu Glu Lys Leu Glu Lys Pro Lys Lys Ser Lys Phe

450 455 460

Arg Gln Arg Arg Arg Glu Lys Tyr Ala Val Ser Met Phe Val Phe Lys

465 470 475 480

Lys Leu Ala Thr Phe Ile Lys Tyr Lys Ala Ala Arg Glu Gly Ile Glu

485 490 495

Ile Ile Pro Val Glu Pro Glu Gly Thr Ser Tyr Thr Cys Ser His Cys

500 505 510

Lys Asn Ala Gln Asn Asn Gln Arg Pro Tyr Phe Lys Pro Asn Ser Lys

515 520 525

Lys Ser Trp Thr Ser Met Phe Lys Cys Gly Lys Cys Gly Ile Glu Leu

530 535 540

Asn Ser Asp Tyr Asn Ala Ala Phe Asn Ile Ala Gln Lys Ala Leu Asn

545 550 555 560

Met Thr Ser Ala

<210> 14

<211> 610

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 14

Met Asp Glu Lys His Phe Phe Cys Ser Tyr Cys Asn Lys Glu Leu Lys

1 5 10 15

Ile Ser Lys Asn Leu Ile Asn Lys Ile Ser Lys Gly Ser Ile Arg Glu

20 25 30

Asp Glu Ala Val Ser Lys Ala Ile Ser Ile His Asn Lys Lys Glu His

35 40 45

Ser Leu Ile Leu Gly Ile Lys Phe Lys Leu Phe Ile Glu Asn Lys Leu

50 55 60

Asp Lys Lys Lys Leu Asn Glu Tyr Phe Asp Asn Tyr Ser Lys Ala Val

65 70 75 80

Thr Phe Ala Ala Arg Ile Phe Asp Lys Ile Arg Ser Pro Tyr Lys Phe

85 90 95

Ile Gly Leu Lys Asp Lys Asn Thr Lys Lys Trp Thr Phe Pro Lys Ala

100 105 110

Lys Cys Val Phe Cys Leu Glu Glu Lys Glu Val Ala Tyr Ala Asn Glu

115 120 125

Lys Asp Asn Ser Lys Ile Cys Thr Glu Cys Tyr Leu Lys Glu Phe Gly

130 135 140

Glu Asn Gly Ile Arg Lys Lys Ile Tyr Ser Thr Arg Gly Arg Lys Val

145 150 155 160

Glu Pro Lys Tyr Asn Ile Phe Asn Ser Thr Lys Glu Leu Ser Ser Thr

165 170 175

His Tyr Asn Tyr Ala Ile Arg Asp Ala Phe Gln Leu Leu Asp Ala Leu

180 185 190

Lys Lys Gln Arg Gln Lys Lys Leu Lys Ser Ile Phe Asn Gln Lys Leu

195 200 205

Arg Leu Lys Glu Phe Glu Asp Ile Phe Ser Asp Pro Gln Lys Arg Ile

210 215 220

Glu Leu Ser Leu Lys Pro His Gln Arg Glu Lys Arg Tyr Ile His Leu

225 230 235 240

Ser Lys Ser Gly Gln Glu Ser Ile Asn Arg Gly Tyr Thr Leu Arg Phe

245 250 255

Val Arg Gly Lys Ile Lys Ser Leu Thr Arg Asn Ile Glu Arg Glu Glu

260 265 270

Lys Ser Leu Arg Lys Lys Thr Pro Ile His Phe Lys Gly Asn Arg Leu

275 280 285

Met Ile Phe Pro Ala Gly Ile Lys Phe Asp Phe Ala Ser Asn Lys Val

290 295 300

Lys Ile Ser Ile Ser Lys Asn Leu Pro Asn Glu Phe Asn Phe Ser Gly

305 310 315 320

Thr Asn Val Lys Asn Glu His Gly Lys Ser Phe Phe Lys Ser Arg Ile
325 330 335
Glu Leu Ile Lys Thr Gln Lys Pro Lys Tyr Ala Tyr Val Leu Arg Lys
340 345 350
Ile Lys Arg Glu Tyr Ser Lys Leu Arg Asn Tyr Glu Ile Glu Lys Ile
355 360 365
Arg Leu Glu Asn Pro Asn Ala Asp Leu Cys Asp Phe Tyr Leu Gln Tyr

370 375 380
Thr Ile Glu Thr Glu Ser Arg Asn Asn Glu Glu Ile Asn Gly Ile Ile
385 390 395 400
Gly Ile Asp Arg Gly Ile Thr Asn Leu Ala Cys Leu Val Leu Leu Lys
405 410 415
Lys Gly Asp Lys Lys Pro Ser Gly Val Lys Phe Tyr Lys Gly Asn Lys
420 425 430
Ile Leu Gly Met Lys Ile Ala Tyr Arg Lys His Leu Tyr Leu Leu Lys

435 440 445
Gly Lys Arg Asn Lys Leu Arg Lys Gln Arg Gln Ile Arg Ala Ile Glu
450 455 460
Pro Lys Ile Asn Leu Ile Leu His Gln Ile Ser Lys Asp Ile Val Lys
465 470 475 480
Ile Ala Lys Glu Lys Asn Phe Ala Ile Ala Leu Glu Gln Leu Glu Lys
485 490 495
Pro Lys Lys Ala Arg Phe Ala Gln Arg Lys Lys Glu Lys Tyr Lys Leu

500 505 510
Ala Leu Phe Thr Phe Lys Asn Leu Ser Thr Leu Ile Glu Tyr Lys Ser
515 520 525
Lys Arg Glu Gly Ile Pro Val Ile Tyr Val Pro Pro Glu Lys Thr Ser
530 535 540
Gln Met Cys Ser His Cys Ala Ile Asn Gly Asp Glu His Val Asp Thr
545 550 555 560
Gln Arg Pro Tyr Lys Lys Pro Asn Ala Gln Lys Pro Ser Tyr Ser Leu

565 570 575

Phe Lys Cys Asn Lys Cys Gly Ile Glu Leu Asn Ala Asp Tyr Asn Ala

580 585 590

Ala Phe Asn Ile Ala Gln Lys Gly Leu Lys Thr Leu Met Leu Asn His

595 600 605

Ser His

610

<210> 15

<211> 369

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 15

Met Leu Gln Thr Leu Leu Val Lys Leu Asp Pro Ser Lys Glu Gln Tyr

1 5 10 15

Lys Met Leu Tyr Glu Thr Met Glu Arg Phe Asn Glu Ala Cys Asn Gln

20 25 30

Ile Ala Glu Thr Val Phe Ala Ile His Ser Ala Asn Lys Ile Glu Val

35 40 45

Gln Lys Thr Val Tyr Tyr Pro Ile Arg Glu Lys Phe Gly Leu Ser Ala

50 55 60

Gln Leu Thr Ile Leu Ala Ile Arg Lys Val Cys Glu Ala Tyr Lys Arg

65 70 75 80

Asp Lys Ser Ile Lys Pro Glu Phe Arg Leu Asp Gly Ala Leu Val Tyr

85 90 95

Asp Gln Arg Val Leu Ser Trp Lys Gly Leu Asp Lys Val Ser Leu Val

100 105 110

Thr Leu Gln Gly Arg Gln Ile Ile Pro Ile Lys Phe Gly Asp Tyr Gln

115 120 125

Lys Ala Arg Met Asp Arg Ile Arg Gly Gln Ala Asp Leu Ile Leu Val

130 135 140

Lys Gly Val Phe Tyr Leu Cys Val Val Val Glu Val Ser Glu Glu Ser

145 150 155 160

Pro Tyr Asp Pro Lys Gly Val Leu Gly Val Asp Leu Gly Ile Lys Asn

165 170 175

Leu Ala Val Asp Ser Asp Gly Glu Val His Ser Gly Glu Gln Thr Thr

180 185 190

Asn Thr Arg Glu Arg Leu Asp Ser Leu Lys Ala Arg Leu Gln Ser Lys

195 200 205

Gly Thr Lys Ser Ala Lys Arg His Leu Lys Lys Leu Ser Gly Arg Met

210 215 220

Ala Lys Phe Ser Lys Asp Val Asn His Cys Ile Ser Lys Lys Leu Val

225 230 235 240

Ala Lys Ala Lys Gly Thr Leu Met Ser Ile Ala Leu Glu Asp Leu Gln

245 250 255

Gly Ile Arg Asp Arg Val Thr Val Arg Lys Ala Gln Arg Arg Asn Leu

260 265 270

His Thr Trp Asn Phe Gly Leu Leu Arg Met Phe Val Asp Tyr Lys Ala

275 280 285

Lys Ile Ala Gly Val Pro Leu Val Phe Val Asp Pro Arg Asn Thr Ser

290 295 300

Arg Thr Cys Pro Ser Cys Gly His Val Ala Lys Ala Asn Arg Pro Thr

305 310 315 320

Arg Asp Glu Phe Arg Cys Val Ser Cys Gly Phe Ala Gly Ala Ala Asp

325 330 335

His Ile Ala Ala Met Asn Ile Ala Phe Arg Ala Glu Val Ser Gln Pro

340 345 350

Ile Val Thr Arg Phe Phe Val Gln Ser Gln Ala Pro Ser Phe Arg Val

355 360 365

Gly

<210> 16

<211> 552

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 16

Met Asp Glu Glu Pro Asp Ser Ala Glu Pro Asn Leu Ala Pro Ile Ser

1 5 10 15

Val Lys Leu Lys Leu Val Lys Leu Asp Gly Glu Lys Leu Ala Ala Leu

20 25 30

Asn Asp Tyr Phe Asn Glu Tyr Ala Lys Ala Val Asn Phe Cys Glu Leu

35 40 45

Lys Met Gln Lys Ile Arg Lys Asn Leu Val Asn Ile Arg Gly Thr Tyr

50 55 60

Leu Lys Glu Lys Lys Ala Trp Ile Asn Gln Thr Gly Glu Cys Cys Ile

65 70 75 80

Cys Lys Lys Ile Asp Glu Leu Arg Cys Glu Asp Lys Asn Pro Asp Ile

85 90 95

Asn Gly Lys Ile Cys Lys Lys Cys Tyr Asn Gly Arg Tyr Gly Asn Gln

100 105 110

Met Ile Arg Lys Leu Phe Val Ser Thr Asn Lys Arg Ala Val Pro Lys

115 120 125

Ser Leu Asp Ile Arg Lys Val Ala Arg Leu His Asn Thr His Tyr His

130 135 140

Arg Ile Pro Pro Glu Ala Ala Asp Ile Ile Lys Ala Ile Glu Thr Ala

145 150 155 160

Glu Arg Lys Arg Arg Asn Arg Ile Leu Phe Asp Glu Arg Arg Tyr Asn

165 170 175

Glu Leu Lys Asp Ala Leu Glu Asn Glu Glu Lys Arg Val Ala Arg Pro

180 185 190

Lys Lys Pro Lys Glu Arg Glu Val Arg Tyr Val Pro Ile Ser Lys Lys

195 200 205

Asp Thr Pro Ser Lys Gly Tyr Thr Met Asn Ala Leu Val Arg Lys Val

210 215 220

Ser Gly Met Ala Lys Lys Ile Glu Arg Ala Lys Arg Asn Leu Asn Lys
225 230 235 240
Arg Lys Lys Ile Glu Tyr Leu Gly Arg Arg Ile Leu Leu Asp Lys Asn
245 250 255
Trp Val Arg Phe Asp Phe Asp Lys Ser Glu Ile Ser Ile Pro Thr Met

260 265 270
Lys Glu Phe Phe Gly Glu Met Arg Phe Glu Ile Thr Gly Pro Ser Asn
275 280 285
Val Met Ser Pro Asn Gly Arg Glu Tyr Phe Thr Lys Trp Phe Asp Arg
290 295 300
Ile Lys Ala Gln Pro Asp Asn Tyr Cys Tyr Leu Leu Arg Lys Glu Ser
305 310 315 320
Glu Asp Glu Thr Asp Phe Tyr Leu Gln Tyr Thr Trp Arg Pro Asp Ala

325 330 335
His Pro Lys Lys Asp Tyr Thr Gly Cys Leu Gly Ile Asp Ile Gly Gly
340 345 350
Ser Lys Leu Ala Ser Ala Val Tyr Phe Asp Ala Asp Lys Asn Arg Ala
355 360 365
Lys Gln Pro Ile Gln Ile Phe Ser Asn Pro Ile Gly Lys Trp Lys Thr
370 375 380
Lys Arg Gln Lys Val Ile Lys Val Leu Ser Lys Ala Ala Val Arg His

385 390 395 400
Lys Thr Lys Lys Leu Glu Ser Leu Arg Asn Ile Glu Pro Arg Ile Asp
405 410 415
Val His Cys His Arg Ile Ala Arg Lys Ile Val Gly Met Ala Leu Ala
420 425 430
Ala Asn Ala Phe Ile Ser Met Glu Asn Leu Glu Gly Gly Ile Arg Glu
435 440 445
Lys Gln Lys Ala Lys Glu Thr Lys Lys Gln Lys Phe Ser Arg Asn Met

450 455 460
Phe Val Phe Arg Lys Leu Ser Lys Leu Ile Glu Tyr Lys Ala Leu Met

465 470 475 480

Glu Gly Val Lys Val Val Tyr Ile Val Pro Asp Tyr Thr Ser Gln Leu

485 490 495

Cys Ser Ser Cys Gly Thr Asn Asn Thr Lys Arg Pro Lys Gln Ala Ile

500 505 510

Phe Met Cys Gln Asn Thr Glu Cys Arg Tyr Phe Gly Lys Asn Ile Asn

515 520 525

Ala Asp Phe Asn Ala Ala Ile Asn Ile Ala Lys Lys Ala Leu Asn Arg

530 535 540

Lys Asp Ile Val Arg Glu Leu Ser

545 550

<210> 17

<211> 534

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 17

Met Glu Lys Asn Asn Ser Glu Gln Thr Ser Ile Thr Thr Gly Ile Lys

1 5 10 15

Phe Lys Leu Lys Leu Asp Lys Glu Thr Lys Glu Lys Leu Asn Asn Tyr

20 25 30

Phe Asp Glu Tyr Gly Lys Ala Ile Asn Phe Ala Val Arg Ile Ile Gln

35 40 45

Met Gln Leu Asn Asp Asp Arg Leu Ala Gly Lys Tyr Lys Arg Asp Glu

50 55 60

Lys Gly Lys Pro Ile Leu Gly Glu Asp Gly Lys Lys Ile Leu Glu Ile

65 70 75 80

Pro Asn Asp Phe Cys Ser Cys Gly Asn Gln Val Asn His Tyr Val Asn

85 90 95

Gly Val Ser Phe Cys Gln Glu Cys Tyr Lys Lys Arg Phe Ser Glu Asn

100 105 110

Gly Ile Arg Lys Arg Met Tyr Ser Ala Lys Gly Arg Lys Ala Glu Gln

115 120 125

Asp Ile Asn Ile Lys Asn Ser Thr Asn Lys Ile Ser Lys Thr His Phe

130 135 140

Asn Tyr Ala Ile Arg Glu Ala Phe Asn Leu Asp Lys Ser Ile Lys Lys

145 150 155 160

Gln Arg Glu Lys Arg Phe Lys Lys Leu Lys Asp Met Lys Arg Lys Leu

165 170 175

Gln Glu Phe Leu Glu Ile Arg Asp Gly Lys Arg Val Ile Cys Pro Lys

180 185 190

Ile Glu Lys Gln Lys Val Glu Arg Tyr Ile His Pro Ser Trp Ile Asn

195 200 205

Lys Glu Lys Lys Leu Glu Glu Phe Arg Gly Tyr Ser Leu Ser Ile Val

210 215 220

Asn Ser Lys Ile Lys Ser Phe Asp Arg Asn Ile Gln Arg Glu Glu Lys

225 230 235 240

Ser Leu Lys Glu Lys Gly Gln Ile Asn Phe Lys Ala Gln Arg Leu Met

245 250 255

Leu Asp Lys Ser Val Lys Phe Leu Lys Asp Asn Lys Val Ser Phe Thr

260 265 270

Ile Ser Lys Glu Leu Pro Lys Thr Phe Glu Leu Asp Leu Pro Lys Lys

275 280 285

Glu Lys Lys Leu Asn Trp Leu Asn Glu Lys Leu Glu Ile Ile Lys Asn

290 295 300

Gln Lys Pro Lys Tyr Ala Tyr Leu Leu Arg Lys Glu Asn Asn Ile Phe

305 310 315 320

Leu Gln Tyr Thr Leu Asp Ser Ile Pro Glu Ile His Ser Glu Tyr Ser

325 330 335

Gly Ala Val Gly Ile Asp Arg Gly Val Ser His Ile Ala Val Tyr Thr

340 345 350

Phe Leu Asp Lys Asp Gly Lys Asn Glu Arg Pro Phe Phe Leu Ser Ser

355 360 365

Ser Gly Ile Leu Arg Leu Lys Asn Leu Gln Lys Glu Arg Asp Lys Phe
370 375 380
Leu Arg Lys Lys His Asn Lys Ile Arg Lys Lys Gly Asn Met Arg Asn
385 390 395 400

Ile Glu Gln Lys Ile Asn Leu Ile Leu His Glu Tyr Ser Lys Gln Ile
405 410 415
Val Asn Phe Ala Lys Asp Lys Asn Ala Phe Ile Val Phe Glu Leu Leu
420 425 430
Glu Lys Pro Lys Lys Ser Arg Glu Arg Met Ser Lys Lys Ile Gln Tyr
435 440 445
Lys Leu Ser Gln Phe Thr Phe Lys Lys Leu Ser Asp Leu Val Asp Tyr
450 455 460

Lys Ala Lys Arg Glu Gly Ile Lys Val Ile Tyr Val Glu Pro Ala Tyr
465 470 475 480
Thr Ser Lys Asp Cys Ser His Cys Gly Glu Arg Val Asn Thr Gln Arg
485 490 495
Pro Phe Asn Gly Asn Phe Ser Leu Phe Lys Cys Asn Lys Cys Gly Ile
500 505 510
Val Leu Asn Ser Asp Tyr Asn Ala Ser Leu Asn Ile Ala Arg Lys Gly
515 520 525

Leu Asn Ile Ser Ala Asn
530

<210> 18
<211> 577
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 18
Met Ala Glu Glu Lys Phe Phe Phe Cys Glu Lys Cys Asn Lys Asp Ile
1 5 10 15
Lys Ile Pro Lys Asn Tyr Ile Asn Lys Gln Gly Ala Glu Glu Lys Ala
20 25 30

Arg Ala Lys His Glu His Arg Val His Ala Leu Ile Leu Gly Ile Lys
35 40 45

Phe Lys Ile Tyr Pro Lys Lys Glu Asp Ile Ser Lys Leu Asn Asp Tyr
50 55 60
Phe Asp Glu Tyr Ala Lys Ala Val Thr Phe Thr Ala Lys Ile Val Asp
65 70 75 80
Lys Leu Lys Ala Pro Phe Leu Phe Ala Gly Lys Arg Asp Lys Asp Thr
85 90 95
Ser Lys Lys Lys Trp Val Phe Pro Val Asp Lys Cys Ser Phe Cys Lys
100 105 110

Glu Lys Thr Glu Ile Asn Tyr Arg Thr Lys Gln Gly Lys Asn Ile Cys
115 120 125
Asn Ser Cys Tyr Leu Thr Glu Phe Gly Glu Gln Gly Leu Leu Glu Lys
130 135 140
Ile Tyr Ala Thr Lys Gly Arg Lys Val Ser Ser Ser Phe Asn Leu Phe
145 150 155 160
Asn Ser Thr Lys Lys Leu Thr Gly Thr His Asn Asn Tyr Val Val Lys
165 170 175

Glu Ser Leu Gln Leu Leu Asp Ala Leu Lys Lys Gln Arg Ser Lys Arg
180 185 190
Leu Lys Lys Leu Ser Asn Thr Arg Arg Lys Leu Lys Gln Phe Glu Glu
195 200 205
Met Phe Glu Lys Glu Asp Lys Arg Phe Gln Leu Pro Leu Lys Glu Lys
210 215 220
Gln Arg Glu Leu Arg Phe Ile His Val Ser Gln Lys Asp Arg Ala Thr
225 230 235 240

Glu Phe Lys Gly Tyr Thr Met Asn Lys Ile Lys Ser Lys Ile Lys Val
245 250 255
Leu Arg Arg Asn Ile Glu Arg Glu Gln Arg Ser Leu Asn Arg Lys Ser
260 265 270
Pro Val Phe Phe Arg Gly Thr Arg Ile Arg Leu Ser Pro Ser Val Gln

275 280 285

Phe Asp Asp Lys Asp Asn Lys Ile Lys Leu Thr Leu Ser Lys Glu Leu

290 295 300

Pro Lys Glu Tyr Ser Phe Ser Gly Leu Asn Val Ala Asn Glu His Gly

305 310 315 320

Arg Lys Phe Phe Ala Glu Lys Leu Lys Leu Ile Lys Glu Asn Lys Ser

325 330 335

Lys Tyr Ala Tyr Leu Leu Arg Arg Gln Val Asn Lys Asn Asn Lys Lys

340 345 350

Pro Ile Tyr Asp Tyr Tyr Leu Gln Tyr Thr Val Glu Phe Leu Pro Asn

355 360 365

Ile Ile Thr Asn Tyr Asn Gly Ile Leu Gly Ile Asp Arg Gly Ile Asn

370 375 380

Thr Leu Ala Cys Ile Val Leu Leu Glu Asn Lys Lys Glu Lys Pro Ser

385 390 395 400

Phe Val Lys Phe Phe Ser Gly Lys Gly Ile Leu Asn Leu Lys Asn Lys

405 410 415

Arg Arg Lys Gln Leu Tyr Phe Leu Lys Gly Val His Asn Lys Tyr Arg

420 425 430

Lys Gln Gln Lys Ile Arg Pro Ile Glu Pro Arg Ile Asp Gln Ile Leu

435 440 445

His Asp Ile Ser Lys Gln Ile Ile Asp Leu Ala Lys Glu Lys Arg Val

450 455 460

Ala Ile Ser Leu Glu Gln Leu Glu Lys Pro Gln Lys Pro Lys Phe Arg

465 470 475 480

Gln Ser Arg Lys Ala Lys Tyr Lys Leu Ser Gln Phe Asn Phe Lys Thr

485 490 495

Leu Ser Asn Tyr Ile Asp Tyr Lys Ala Lys Lys Glu Gly Ile Arg Val

500 505 510

Ile Tyr Ile Ala Pro Glu Met Thr Ser Gln Asn Cys Ser Arg Cys Ala

515 520 525

Met Lys Asn Asp Leu His Val Asn Thr Gln Arg Pro Tyr Lys Asn Thr
530 535 540
Ser Ser Leu Phe Lys Cys Asn Lys Cys Gly Val Glu Leu Asn Ala Asp
545 550 555 560

Tyr Asn Ala Ala Phe Asn Ile Ala Gln Lys Gly Leu Lys Ile Leu Asn
565 570 575
Ser

<210> 19
<211> 613
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 19
Met Ile Ser Leu Lys Leu Lys Leu Leu Pro Asp Glu Glu Gln Lys Lys
1 5 10 15
Leu Leu Asp Glu Met Phe Trp Lys Trp Ala Ser Ile Cys Thr Arg Val
20 25 30

Gly Phe Gly Arg Ala Asp Lys Glu Asp Leu Lys Pro Pro Lys Asp Ala
35 40 45
Glu Gly Val Trp Phe Ser Leu Thr Gln Leu Asn Gln Ala Asn Thr Asp
50 55 60
Ile Asn Asp Leu Arg Glu Ala Met Lys His Gln Lys His Arg Leu Glu
65 70 75 80
Tyr Glu Lys Asn Arg Leu Glu Ala Gln Arg Asp Asp Thr Gln Asp Ala
85 90 95

Leu Lys Asn Pro Asp Arg Arg Glu Ile Ser Thr Lys Arg Lys Asp Leu
100 105 110
Phe Arg Pro Lys Ala Ser Val Glu Lys Gly Phe Leu Lys Leu Lys Tyr
115 120 125
His Gln Glu Arg Tyr Trp Val Arg Arg Leu Lys Glu Ile Asn Lys Leu
130 135 140

Ile Glu Arg Lys Thr Lys Thr Leu Ile Lys Ile Glu Lys Gly Arg Ile
145 150 155 160

Lys Phe Lys Ala Thr Arg Ile Thr Leu His Gln Gly Ser Phe Lys Ile
165 170 175

Arg Phe Gly Asp Lys Pro Ala Phe Leu Ile Lys Ala Leu Ser Gly Lys
180 185 190

Asn Gln Ile Asp Ala Pro Phe Val Val Val Pro Glu Gln Pro Ile Cys
195 200 205

Gly Ser Val Val Asn Ser Lys Lys Tyr Leu Asp Glu Ile Thr Thr Asn
210 215 220

Phe Leu Ala Tyr Ser Val Asn Ala Met Leu Phe Gly Leu Ser Arg Ser
225 230 235 240

Glu Glu Met Leu Leu Lys Ala Lys Arg Pro Glu Lys Ile Lys Lys Lys
245 250 255

Glu Glu Lys Leu Ala Lys Lys Gln Ser Ala Phe Glu Asn Lys Lys Lys
260 265 270

Glu Leu Gln Lys Leu Leu Gly Arg Glu Leu Thr Gln Gln Glu Glu Ala
275 280 285

Ile Ile Glu Glu Thr Arg Asn Gln Phe Phe Gln Asp Phe Glu Val Lys
290 295 300

Ile Thr Lys Gln Tyr Ser Glu Leu Leu Ser Lys Ile Ala Asn Glu Leu
305 310 315 320

Lys Gln Lys Asn Asp Phe Leu Lys Val Asn Lys Tyr Pro Ile Leu Leu
325 330 335

Arg Lys Pro Leu Lys Lys Ala Lys Ser Lys Lys Ile Asn Asn Leu Ser
340 345 350

Pro Ser Glu Trp Lys Tyr Tyr Leu Gln Phe Gly Val Lys Pro Leu Leu
355 360 365

Lys Gln Lys Ser Arg Arg Lys Ser Arg Asn Val Leu Gly Ile Asp Arg
370 375 380

Gly Leu Lys His Leu Leu Ala Val Thr Val Leu Glu Pro Asp Lys Lys

385 390 395 400

Thr Phe Val Trp Asn Lys Leu Tyr Pro Asn Pro Ile Thr Gly Trp Lys

405 410 415

Trp Arg Arg Arg Lys Leu Leu Arg Ser Leu Lys Arg Leu Lys Arg Arg

420 425 430

Ile Lys Ser Gln Lys His Glu Thr Ile His Glu Asn Gln Thr Arg Lys

435 440 445

Lys Leu Lys Ser Leu Gln Gly Arg Ile Asp Asp Leu Leu His Asn Ile

450 455 460

Ser Arg Lys Ile Val Glu Thr Ala Lys Glu Tyr Asp Ala Val Ile Val

465 470 475 480

Val Glu Asp Leu Gln Ser Met Arg Gln His Gly Arg Ser Lys Gly Asn

485 490 495

Arg Leu Lys Thr Leu Asn Tyr Ala Leu Ser Leu Phe Asp Tyr Ala Asn

500 505 510

Val Met Gln Leu Ile Lys Tyr Lys Ala Gly Ile Glu Gly Ile Gln Ile

515 520 525

Tyr Asp Val Lys Pro Ala Gly Thr Ser Gln Asn Cys Ala Tyr Cys Leu

530 535 540

Leu Ala Gln Arg Asp Ser His Glu Tyr Lys Arg Ser Gln Glu Asn Ser

545 550 555 560

Lys Ile Gly Val Cys Leu Asn Pro Asn Cys Gln Asn His Lys Lys Gln

565 570 575

Ile Asp Ala Asp Leu Asn Ala Ala Arg Val Ile Ala Ser Cys Tyr Ala

580 585 590

Leu Lys Ile Asn Asp Ser Gln Pro Phe Gly Thr Arg Lys Arg Phe Lys

595 600 605

Lys Arg Thr Thr Asn

610

<210> 20

<211> 615

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 20

Met Glu Thr Leu Ser Leu Lys Leu Lys Leu Asn Pro Ser Lys Glu Gln
1 5 10 15

Leu Leu Val Leu Asp Lys Met Phe Trp Lys Trp Ala Ser Ile Cys Thr
20 25 30

Arg Leu Gly Leu Lys Lys Ala Glu Met Ser Asp Leu Glu Pro Pro Lys
35 40 45

Asp Ala Glu Gly Val Trp Phe Ser Lys Thr Gln Leu Asn Gln Ala Asn
50 55 60

Thr Asp Val Asn Asp Leu Arg Lys Ala Met Gln His Gln Gly Lys Arg
65 70 75 80

Ile Glu Tyr Glu Leu Asp Lys Val Glu Asn Arg Arg Asn Glu Ile Gln
85 90 95

Glu Met Leu Glu Lys Pro Asp Arg Arg Asp Ile Ser Pro Asn Arg Lys
100 105 110

Asp Leu Phe Arg Pro Lys Ala Ala Val Glu Lys Gly Tyr Leu Lys Leu
115 120 125

Lys Tyr His Lys Leu Gly Tyr Trp Ser Lys Glu Leu Lys Thr Ala Asn
130 135 140

Lys Leu Ile Glu Arg Lys Arg Lys Thr Leu Ala Lys Ile Asp Ala Gly
145 150 155 160

Lys Met Lys Phe Lys Pro Thr Arg Ile Ser Leu His Thr Asn Ser Phe
165 170 175

Arg Ile Lys Phe Gly Glu Glu Pro Lys Ile Ala Leu Ser Thr Thr Ser
180 185 190

Lys His Glu Lys Ile Glu Leu Pro Leu Ile Thr Ser Leu Gln Arg Pro
195 200 205

Leu Lys Thr Ser Cys Ala Lys Lys Ser Lys Thr Tyr Leu Asp Ala Ala
210 215 220

Ile Leu Asn Phe Leu Ala Tyr Ser Thr Asn Ala Ala Leu Phe Gly Leu
225 230 235 240

Ser Arg Ser Glu Glu Met Leu Leu Lys Ala Lys Lys Pro Glu Lys Ile
245 250 255

Glu Lys Arg Asp Arg Lys Leu Ala Thr Lys Arg Glu Ser Phe Asp Lys
260 265 270

Lys Leu Lys Thr Leu Glu Lys Leu Leu Glu Arg Lys Leu Ser Glu Lys
275 280 285

Glu Lys Ser Val Phe Lys Arg Lys Gln Thr Glu Phe Phe Asp Lys Phe
290 295 300

Cys Ile Thr Leu Asp Glu Thr Tyr Val Glu Ala Leu His Arg Ile Ala
305 310 315 320

Glu Glu Leu Val Ser Lys Asn Lys Tyr Leu Glu Ile Lys Lys Tyr Pro
325 330 335

Val Leu Leu Arg Lys Pro Glu Ser Arg Leu Arg Ser Lys Lys Leu Lys
340 345 350

Asn Leu Lys Pro Glu Asp Trp Thr Tyr Tyr Ile Gln Phe Gly Phe Gln
355 360 365

Pro Leu Leu Asp Thr Pro Lys Pro Ile Lys Thr Lys Thr Val Leu Gly
370 375 380

Ile Asp Arg Gly Val Arg His Leu Leu Ala Val Ser Ile Phe Asp Pro
385 390 395 400

Arg Thr Lys Thr Phe Thr Phe Asn Arg Leu Tyr Ser Asn Pro Ile Val
405 410 415

Asp Trp Lys Trp Arg Arg Arg Lys Leu Leu Arg Ser Ile Lys Arg Leu
420 425 430

Lys Arg Arg Leu Lys Ser Glu Lys His Val His Leu His Glu Asn Gln
435 440 445

Phe Lys Ala Lys Leu Arg Ser Leu Glu Gly Arg Ile Glu Asp His Phe
450 455 460

His Asn Leu Ser Lys Glu Ile Val Asp Leu Ala Lys Glu Asn Asn Ser

465 470 475 480

Val Ile Val Val Glu Asn Leu Gly Gly Met Arg Gln His Gly Arg Gly

485 490 495

Arg Gly Lys Trp Leu Lys Ala Leu Asn Tyr Ala Leu Ser His Phe Asp

500 505 510

Tyr Ala Lys Val Met Gln Leu Ile Lys Tyr Lys Ala Glu Leu Ala Gly

515 520 525

Val Phe Val Tyr Asp Val Ala Pro Ala Gly Thr Ser Ile Asn Cys Ala

530 535 540

Tyr Cys Leu Leu Asn Asp Lys Asp Ala Ser Asn Tyr Thr Arg Gly Lys

545 550 555 560

Val Ile Asn Gly Lys Lys Asn Thr Lys Ile Gly Glu Cys Lys Thr Cys

565 570 575

Lys Lys Glu Phe Asp Ala Asp Leu Asn Ala Ala Arg Val Ile Ala Leu

580 585 590

Cys Tyr Glu Lys Arg Leu Asn Asp Pro Gln Pro Phe Gly Thr Arg Lys

595 600 605

Gln Phe Lys Pro Lys Lys Pro

610 615

<210> 21

<211> 775

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 21

Met Lys Ala Leu Lys Leu Gln Leu Ile Pro Thr Arg Lys Gln Tyr Lys

1 5 10 15

Ile Leu Asp Glu Met Phe Trp Lys Trp Ala Ser Leu Ala Asn Arg Val

20 25 30

Ser Gln Lys Gly Glu Ser Lys Glu Thr Leu Ala Pro Lys Lys Asp Ile

35 40 45

Gln Lys Ile Gln Phe Asn Ala Thr Gln Leu Asn Gln Ile Glu Lys Asp

50 55 60

Ile Lys Asp Leu Arg Gly Ala Met Lys Glu Gln Gln Lys Gln Lys Glu
65 70 75 80
Arg Leu Leu Leu Gln Ile Gln Glu Arg Arg Ser Thr Ile Ser Glu Met
85 90 95
Leu Asn Asp Asp Asn Asn Lys Glu Arg Asp Pro His Arg Pro Leu Asn
100 105 110
Phe Arg Pro Lys Gly Trp Arg Lys Phe His Thr Ser Lys His Trp Val
115 120 125

Gly Glu Leu Ser Lys Ile Leu Arg Gln Glu Asp Arg Val Lys Lys Thr
130 135 140
Ile Glu Arg Ile Val Ala Gly Lys Ile Ser Phe Lys Pro Lys Arg Ile
145 150 155 160
Gly Ile Trp Ser Ser Asn Tyr Lys Ile Asn Phe Phe Lys Arg Lys Ile
165 170 175
Ser Ile Asn Pro Leu Asn Ser Lys Gly Phe Glu Leu Thr Leu Met Thr
180 185 190

Glu Pro Thr Gln Asp Leu Ile Gly Lys Asn Gly Gly Lys Ser Val Leu
195 200 205
Asn Asn Lys Arg Tyr Leu Asp Asp Ser Ile Lys Ser Leu Leu Met Phe
210 215 220
Ala Leu His Ser Arg Phe Phe Gly Leu Asn Asn Thr Asp Thr Tyr Leu
225 230 235 240
Leu Gly Gly Lys Ile Asn Pro Ser Leu Val Lys Tyr Tyr Lys Lys Asn
245 250 255

Gln Asp Met Gly Glu Phe Gly Arg Glu Ile Val Glu Lys Phe Glu Arg
260 265 270
Lys Leu Lys Gln Glu Ile Asn Glu Gln Gln Lys Lys Ile Ile Met Ser
275 280 285
Gln Ile Lys Glu Gln Tyr Ser Asn Arg Asp Ser Ala Phe Asn Lys Asp
290 295 300

Tyr Leu Gly Leu Ile Asn Glu Phe Ser Glu Val Phe Asn Gln Arg Lys
305 310 315 320

Ser Glu Arg Ala Glu Tyr Leu Leu Asp Ser Phe Glu Asp Lys Ile Lys
325 330 335

Gln Ile Lys Gln Glu Ile Gly Glu Ser Leu Asn Ile Ser Asp Trp Asp
340 345 350

Phe Leu Ile Asp Glu Ala Lys Lys Ala Tyr Gly Tyr Glu Glu Gly Phe
355 360 365

Thr Glu Tyr Val Tyr Ser Lys Arg Tyr Leu Glu Ile Leu Asn Lys Ile
370 375 380

Val Lys Ala Val Leu Ile Thr Asp Ile Tyr Phe Asp Leu Arg Lys Tyr
385 390 395 400

Pro Ile Leu Leu Arg Lys Pro Leu Asp Lys Ile Lys Lys Ile Ser Asn
405 410 415

Leu Lys Pro Asp Glu Trp Ser Tyr Tyr Ile Gln Phe Gly Tyr Asp Ser
420 425 430

Ile Asn Pro Val Gln Leu Met Ser Thr Asp Lys Phe Leu Gly Ile Asp
435 440 445

Arg Gly Leu Thr His Leu Leu Ala Tyr Ser Val Phe Asp Lys Glu Lys
450 455 460

Lys Glu Phe Ile Ile Asn Gln Leu Glu Pro Asn Pro Ile Met Gly Trp
465 470 475 480

Lys Trp Lys Leu Arg Lys Val Lys Arg Ser Leu Gln His Leu Glu Arg
485 490 495

Arg Ile Arg Ala Gln Lys Met Val Lys Leu Pro Glu Asn Gln Met Lys
500 505 510

Lys Lys Leu Lys Ser Ile Glu Pro Lys Ile Glu Val His Tyr His Asn
515 520 525

Ile Ser Arg Lys Ile Val Asn Leu Ala Lys Asp Tyr Asn Ala Ser Ile
530 535 540

Val Val Glu Ser Leu Glu Gly Gly Gly Leu Lys Gln His Gly Arg Lys

545 550 555 560

Lys Asn Ala Arg Asn Arg Ser Leu Asn Tyr Ala Leu Ser Leu Phe Asp
565 570 575

Tyr Gly Lys Ile Ala Ser Leu Ile Lys Tyr Lys Ala Asp Leu Glu Gly
580 585 590

Val Pro Met Tyr Glu Val Leu Pro Ala Tyr Thr Ser Gln Gln Cys Ala
595 600 605

Lys Cys Val Leu Glu Lys Gly Ser Phe Val Asp Pro Glu Ile Ile Gly
610 615 620

Tyr Val Glu Asp Ile Gly Ile Lys Gly Ser Leu Leu Asp Ser Leu Phe
625 630 635 640

Glu Gly Thr Glu Leu Ser Ser Ile Gln Val Leu Lys Lys Ile Lys Asn
645 650 655

Lys Ile Glu Leu Ser Ala Arg Asp Asn His Asn Lys Glu Ile Asn Leu
660 665 670

Ile Leu Lys Tyr Asn Phe Lys Gly Leu Val Ile Val Arg Gly Gln Asp
675 680 685

Lys Glu Glu Ile Ala Glu His Pro Ile Lys Glu Ile Asn Gly Lys Phe
690 695 700

Ala Ile Leu Asp Phe Val Tyr Lys Arg Gly Lys Glu Lys Val Gly Lys
705 710 715 720

Lys Gly Asn Gln Lys Val Arg Tyr Thr Gly Asn Lys Lys Val Gly Tyr
725 730 735

Cys Ser Lys His Gly Gln Val Asp Ala Asp Leu Asn Ala Ser Arg Val
740 745 750

Ile Ala Leu Cys Lys Tyr Leu Asp Ile Asn Asp Pro Ile Leu Phe Gly
755 760 765

Glu Gln Arg Lys Ser Phe Lys
770 775

<210> 22

<211> 777

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 22

Met Val Thr Arg Ala Ile Lys Leu Lys Leu Asp Pro Thr Lys Asn Gln
1 5 10 15

Tyr Lys Leu Leu Asn Glu Met Phe Trp Lys Trp Ala Ser Leu Ala Asn
20 25 30

Arg Phe Ser Gln Lys Gly Ala Ser Lys Glu Thr Leu Ala Pro Lys Asp
35 40 45

Gly Thr Gln Lys Ile Gln Phe Asn Ala Thr Gln Leu Asn Gln Ile Lys
50 55 60

Lys Asp Val Asp Asp Leu Arg Gly Ala Met Glu Lys Gln Gly Lys Gln
65 70 75 80

Lys Glu Arg Leu Leu Ile Gln Ile Gln Glu Arg Leu Leu Thr Ile Ser
85 90 95

Glu Ile Leu Arg Asp Asp Ser Lys Lys Glu Lys Asp Pro His Arg Pro
100 105 110

Gln Asn Phe Arg Pro Phe Gly Trp Arg Arg Phe His Thr Ser Ala Tyr
115 120 125

Trp Ser Ser Glu Ala Ser Lys Leu Thr Arg Gln Val Asp Arg Val Arg
130 135 140

Arg Thr Ile Glu Arg Ile Lys Ala Gly Lys Ile Asn Phe Lys Pro Lys
145 150 155 160

Arg Ile Gly Leu Trp Ser Ser Thr Tyr Lys Ile Asn Phe Leu Lys Lys
165 170 175

Lys Ile Asn Ile Ser Pro Leu Lys Ser Lys Ser Phe Glu Leu Asp Leu
180 185 190

Ile Thr Glu Pro Gln Gln Lys Ile Ile Gly Lys Glu Gly Gly Lys Ser
195 200 205

Val Ala Asn Ser Lys Lys Tyr Leu Asp Asp Ser Ile Lys Ser Leu Leu
210 215 220

Ile Phe Ala Ile Lys Ser Arg Leu Phe Gly Leu Asn Asn Lys Asp Lys
225 230 235 240

Pro Leu Phe Glu Asn Ile Ile Thr Pro Asn Leu Val Arg Tyr His Lys
245 250 255

Lys Gly Gln Glu Gln Glu Asn Phe Lys Lys Glu Val Ile Lys Lys Phe
260 265 270

Glu Asn Lys Leu Lys Lys Glu Ile Ser Gln Lys Gln Lys Glu Ile Ile
275 280 285

Phe Ser Gln Ile Glu Arg Gln Tyr Glu Asn Arg Asp Ala Thr Phe Ser
290 295 300

Glu Asp Tyr Leu Arg Ala Ile Ser Glu Phe Ser Glu Ile Phe Asn Gln
305 310 315 320

Arg Lys Lys Glu Arg Ala Lys Glu Leu Leu Asn Ser Phe Asn Glu Lys
325 330 335

Ile Arg Gln Leu Lys Lys Glu Val Asn Gly Asn Ile Ser Glu Glu Asp
340 345 350

Leu Lys Ile Leu Glu Val Glu Ala Glu Lys Ala Tyr Asn Tyr Glu Asn
355 360 365

Gly Phe Ile Glu Trp Glu Tyr Ser Glu Gln Phe Leu Gly Val Leu Glu
370 375 380

Lys Ile Ala Arg Ala Val Leu Ile Ser Asp Asn Tyr Phe Asp Leu Lys
385 390 395 400

Lys Tyr Pro Ile Leu Ile Arg Lys Pro Thr Asn Lys Ser Lys Lys Ile
405 410 415

Thr Asn Leu Lys Pro Glu Glu Trp Asp Tyr Tyr Ile Gln Phe Gly Tyr
420 425 430

Gly Leu Ile Asn Ser Pro Met Lys Ile Glu Thr Lys Asn Phe Met Gly
435 440 445

Ile Asp Arg Gly Leu Thr His Leu Leu Ala Tyr Ser Ile Phe Asp Arg
450 455 460

Asp Ser Glu Lys Phe Thr Ile Asn Gln Leu Glu Leu Asn Pro Ile Lys

465 470 475 480

Gly Trp Lys Trp Lys Leu Arg Lys Val Lys Arg Ser Leu Gln His Leu

485 490 495

Glu Arg Arg Met Arg Ala Gln Lys Gly Val Lys Leu Pro Glu Asn Gln

500 505 510

Met Lys Lys Arg Leu Lys Ser Ile Glu Pro Lys Ile Glu Ser Tyr Tyr

515 520 525

His Asn Leu Ser Arg Lys Ile Val Asn Leu Ala Lys Ala Asn Asn Ala

530 535 540

Ser Ile Val Val Glu Ser Leu Glu Gly Gly Gly Leu Lys Gln His Gly

545 550 555 560

Arg Lys Lys Asn Ser Arg His Arg Ala Leu Asn Tyr Ala Leu Ser Leu

565 570 575

Phe Asp Tyr Gly Lys Ile Ala Ser Leu Ile Lys Tyr Lys Ser Asp Leu

580 585 590

Glu Gly Val Pro Met Tyr Glu Val Leu Pro Ala Tyr Thr Ser Gln Gln

595 600 605

Cys Ala Lys Cys Val Leu Lys Lys Gly Ser Phe Val Glu Pro Glu Ile

610 615 620

Ile Gly Tyr Ile Glu Glu Ile Gly Phe Lys Glu Asn Leu Leu Thr Leu

625 630 635 640

Leu Phe Glu Asp Thr Gly Leu Ser Ser Val Gln Val Leu Lys Lys Ser

645 650 655

Lys Asn Lys Met Thr Leu Ser Ala Arg Asp Lys Glu Gly Lys Met Val

660 665 670

Asp Leu Val Leu Lys Tyr Asn Phe Lys Gly Leu Val Ile Ser Gln Glu

675 680 685

Lys Lys Lys Glu Glu Ile Val Glu Phe Pro Ile Lys Glu Ile Asp Gly

690 695 700

Lys Phe Ala Val Leu Asp Ser Ala Tyr Lys Arg Gly Lys Glu Arg Ile

705 710 715 720

Ser Lys Lys Gly Asn Gln Lys Leu Val Tyr Thr Gly Asn Lys Lys Val
725 730 735
Gly Tyr Cys Ser Val His Gly Gln Val Asp Ala Asp Leu Asn Ala Ser
740 745 750

Arg Val Ile Ala Leu Cys Lys Tyr Leu Gly Ile Asn Glu Pro Ile Val
755 760 765
Phe Gly Glu Gln Arg Lys Ser Phe Lys
770 775

<210> 23
<211> 610
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 23

Leu Asp Leu Ile Thr Glu Pro Ile Gln Pro His Lys Ser Ser Ser Leu
1 5 10 15
Arg Ser Lys Glu Phe Leu Glu Tyr Gln Ile Ser Asp Phe Leu Asn Phe
20 25 30
Ser Leu His Ser Leu Phe Phe Gly Leu Ala Ser Asn Glu Gly Pro Leu
35 40 45
Val Asp Phe Lys Ile Tyr Asp Lys Ile Val Ile Pro Lys Pro Glu Glu
50 55 60
Arg Phe Pro Lys Lys Glu Ser Glu Glu Gly Lys Lys Leu Asp Ser Phe
65 70 75 80
Asp Lys Arg Val Glu Glu Tyr Tyr Ser Asp Lys Leu Glu Lys Lys Ile
85 90 95
Glu Arg Lys Leu Asn Thr Glu Glu Lys Asn Val Ile Asp Arg Glu Lys
100 105 110
Thr Arg Ile Trp Gly Glu Val Asn Lys Leu Glu Glu Ile Arg Ser Ile
115 120 125
Ile Asp Glu Ile Asn Glu Ile Lys Lys Gln Lys His Ile Ser Glu Lys
130 135 140

Ser Lys Leu Leu Gly Glu Lys Trp Lys Lys Val Asn Asn Ile Gln Glu

145 150 155 160

Thr Leu Leu Ser Gln Glu Tyr Val Ser Leu Ile Ser Asn Leu Ser Asp

165 170 175

Glu Leu Thr Asn Lys Lys Lys Glu Leu Leu Ala Lys Lys Tyr Ser Lys

180 185 190

Phe Asp Asp Lys Ile Lys Lys Ile Lys Glu Asp Tyr Gly Leu Glu Phe

195 200 205

Asp Glu Asn Thr Ile Lys Lys Glu Gly Glu Lys Ala Phe Leu Asn Pro

210 215 220

Asp Lys Phe Ser Lys Tyr Gln Phe Ser Ser Ser Tyr Leu Lys Leu Ile

225 230 235 240

Gly Glu Ile Ala Arg Ser Leu Ile Thr Tyr Lys Gly Phe Leu Asp Leu

245 250 255

Asn Lys Tyr Pro Ile Ile Phe Arg Lys Pro Ile Asn Lys Val Lys Lys

260 265 270

Ile His Asn Leu Glu Pro Asp Glu Trp Lys Tyr Tyr Ile Gln Phe Gly

275 280 285

Tyr Glu Gln Ile Asn Asn Pro Lys Leu Glu Thr Glu Asn Ile Leu Gly

290 295 300

Ile Asp Arg Gly Leu Thr His Ile Leu Ala Tyr Ser Val Phe Glu Pro

305 310 315 320

Arg Ser Ser Lys Phe Ile Leu Asn Lys Leu Glu Pro Asn Pro Ile Glu

325 330 335

Gly Trp Lys Trp Lys Leu Arg Lys Leu Arg Arg Ser Ile Gln Asn Leu

340 345 350

Glu Arg Arg Trp Arg Ala Gln Asp Asn Val Lys Leu Pro Glu Asn Gln

355 360 365

Met Lys Lys Asn Leu Arg Ser Ile Glu Asp Lys Val Glu Asn Leu Tyr

370 375 380

His Asn Leu Ser Arg Lys Ile Val Asp Leu Ala Lys Glu Lys Asn Ala

385 390 395 400

Cys Ile Val Phe Glu Lys Leu Glu Gly Gln Gly Met Lys Gln His Gly

405 410 415

Arg Lys Lys Ser Asp Arg Leu Arg Gly Leu Asn Tyr Lys Leu Ser Leu

420 425 430

Phe Asp Tyr Gly Lys Ile Ala Lys Leu Ile Lys Tyr Lys Ala Glu Ile

435 440 445

Glu Gly Ile Pro Ile Tyr Arg Ile Asp Ser Ala Tyr Thr Ser Gln Asn

450 455 460

Cys Ala Lys Cys Val Leu Glu Ser Arg Arg Phe Ala Gln Pro Glu Glu

465 470 475 480

Ile Ser Cys Leu Asp Asp Phe Lys Glu Gly Asp Asn Leu Asp Lys Arg

485 490 495

Ile Leu Glu Gly Thr Gly Leu Val Glu Ala Lys Ile Tyr Lys Lys Leu

500 505 510

Leu Lys Glu Lys Lys Glu Asp Phe Glu Ile Glu Glu Asp Ile Ala Met

515 520 525

Phe Asp Thr Lys Lys Val Ile Lys Glu Asn Lys Glu Lys Thr Val Ile

530 535 540

Leu Asp Tyr Val Tyr Thr Arg Arg Lys Glu Ile Ile Gly Thr Asn His

545 550 555 560

Lys Lys Asn Ile Lys Gly Ile Ala Lys Tyr Thr Gly Asn Thr Lys Ile

565 570 575

Gly Tyr Cys Met Lys His Gly Gln Val Asp Ala Asp Leu Asn Ala Ser

580 585 590

Arg Thr Ile Ala Leu Cys Lys Asn Phe Asp Ile Asn Asn Pro Glu Ile

595 600 605

Trp Lys

610

<210> 24

<211> 632

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 24

Met Ser Asp Glu Ser Leu Val Ser Ser Glu Asp Lys Leu Ala Ile Lys
1 5 10 15
Ile Lys Ile Val Pro Asn Ala Glu Gln Ala Lys Met Leu Asp Glu Met
20 25 30
Phe Lys Lys Trp Ser Ser Ile Cys Asn Arg Ile Ser Arg Gly Lys Glu
35 40 45
Asp Ile Glu Thr Leu Arg Pro Asp Glu Gly Lys Glu Leu Gln Phe Asn
50 55 60
Ser Thr Gln Leu Asn Ser Ala Thr Met Asp Val Ser Asp Leu Lys Lys
65 70 75 80
Ala Met Ala Arg Gln Gly Glu Arg Leu Glu Ala Glu Val Ser Lys Leu
85 90 95
Arg Gly Arg Tyr Glu Thr Ile Asp Ala Ser Leu Arg Asp Pro Ser Arg
100 105 110
Arg His Thr Asn Pro Gln Lys Pro Ser Ser Phe Tyr Pro Ser Asp Trp
115 120 125
Asp Ile Ser Gly Arg Leu Thr Pro Arg Phe His Thr Ala Arg His Tyr
130 135 140
Ser Thr Glu Leu Arg Lys Leu Lys Ala Lys Glu Asp Lys Met Leu Lys
145 150 155 160
Thr Ile Asn Lys Ile Lys Asn Gly Lys Ile Val Phe Lys Pro Lys Arg
165 170 175
Ile Thr Leu Trp Pro Ser Ser Val Asn Met Ala Phe Lys Gly Ser Arg
180 185 190
Leu Leu Leu Lys Pro Phe Ala Asn Gly Phe Glu Met Glu Leu Pro Ile
195 200 205
Val Ile Ser Pro Gln Lys Thr Ala Asp Gly Lys Ser Gln Lys Ala Ser
210 215 220

Ala Glu Tyr Met Arg Asn Ala Leu Leu Gly Leu Ala Gly Tyr Ser Ile

225 230 235 240

Asn Gln Leu Leu Phe Gly Met Asn Arg Ser Gln Lys Met Leu Ala Asn

245 250 255

Ala Lys Lys Pro Glu Lys Val Glu Lys Phe Leu Glu Gln Met Lys Asn

260 265 270

Lys Asp Ala Asn Phe Asp Lys Lys Ile Lys Ala Leu Glu Gly Lys Trp

275 280 285

Leu Leu Asp Arg Lys Leu Lys Glu Ser Glu Lys Ser Ser Ile Ala Val

290 295 300

Val Arg Thr Lys Phe Phe Lys Ser Gly Lys Val Glu Leu Asn Glu Asp

305 310 315 320

Tyr Leu Lys Leu Leu Lys His Met Ala Asn Glu Ile Leu Glu Arg Asp

325 330 335

Gly Phe Val Asn Leu Asn Lys Tyr Pro Ile Leu Ser Arg Lys Pro Met

340 345 350

Lys Arg Tyr Lys Gln Lys Asn Ile Asp Asn Leu Lys Pro Asn Met Trp

355 360 365

Lys Tyr Tyr Ile Gln Phe Gly Tyr Glu Pro Ile Phe Glu Arg Lys Ala

370 375 380

Ser Gly Lys Pro Lys Asn Ile Met Gly Ile Asp Arg Gly Leu Thr His

385 390 395 400

Leu Leu Ala Val Ala Val Phe Ser Pro Asp Gln Gln Lys Phe Leu Phe

405 410 415

Asn His Leu Glu Ser Asn Pro Ile Met His Trp Lys Trp Lys Leu Arg

420 425 430

Lys Ile Arg Arg Ser Ile Gln His Met Glu Arg Arg Ile Arg Ala Glu

435 440 445

Lys Asn Lys His Ile His Glu Ala Gln Leu Lys Lys Arg Leu Gly Ser

450 455 460

Ile Glu Glu Lys Thr Glu Gln His Tyr His Ile Val Ser Ser Lys Ile

465 470 475 480

Ile Asn Trp Ala Ile Glu Tyr Glu Ala Ala Ile Val Leu Glu Ser Leu

485 490 495

Ser His Met Lys Gln Arg Gly Gly Lys Lys Ser Val Arg Thr Arg Ala

500 505 510

Leu Asn Tyr Ala Leu Ser Leu Phe Asp Tyr Glu Lys Val Ala Arg Leu

515 520 525

Ile Thr Tyr Lys Ala Arg Ile Arg Gly Ile Pro Val Tyr Asp Val Leu

530 535 540

Pro Gly Met Thr Ser Lys Thr Cys Ala Thr Cys Leu Leu Asn Gly Ser

545 550 555 560

Gln Gly Ala Tyr Val Arg Gly Leu Glu Thr Thr Lys Ala Ala Gly Lys

565 570 575

Ala Thr Lys Arg Lys Asn Met Lys Ile Gly Lys Cys Met Val Cys Asn

580 585 590

Ser Ser Glu Asn Ser Met Ile Asp Ala Asp Leu Asn Ala Ala Arg Val

595 600 605

Ile Ala Ile Cys Lys Tyr Lys Asn Leu Asn Asp Pro Gln Pro Ala Gly

610 615 620

Ser Arg Lys Val Phe Lys Arg Phe

625 630

<210> 25

<211> 625

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 25

Met Leu Ala Leu Lys Leu Lys Ile Met Pro Thr Glu Lys Gln Ala Glu

1 5 10 15

Ile Leu Asp Ala Met Phe Trp Lys Trp Ala Ser Ile Cys Ser Arg Ile

20 25 30

Ala Lys Met Lys Lys Lys Val Ser Val Lys Glu Asn Lys Lys Glu Leu
35 40 45
Ser Lys Lys Ile Pro Ser Asn Ser Asp Ile Trp Phe Ser Lys Thr Gln
50 55 60
Leu Cys Gln Ala Glu Val Asp Val Gly Asp His Lys Lys Ala Leu Lys
65 70 75 80
Asn Phe Glu Lys Arg Gln Glu Ser Leu Leu Asp Glu Leu Lys Tyr Lys
85 90 95

Val Lys Ala Ile Asn Glu Val Ile Asn Asp Glu Ser Lys Arg Glu Ile
100 105 110
Asp Pro Asn Asn Pro Ser Lys Phe Arg Ile Lys Asp Ser Thr Lys Lys
115 120 125
Gly Asn Leu Asn Ser Pro Lys Phe Phe Thr Leu Lys Lys Trp Gln Lys
130 135 140
Ile Leu Gln Glu Asn Glu Lys Arg Ile Lys Lys Lys Glu Ser Thr Ile
145 150 155 160

Glu Lys Leu Lys Arg Gly Asn Ile Phe Phe Asn Pro Thr Lys Ile Ser
165 170 175
Leu His Glu Glu Glu Tyr Ser Ile Asn Phe Gly Ser Ser Lys Leu Leu
180 185 190
Leu Asn Cys Phe Tyr Lys Tyr Asn Lys Lys Ser Gly Ile Asn Ser Asp
195 200 205
Gln Leu Glu Asn Lys Phe Asn Glu Phe Gln Asn Gly Leu Asn Ile Ile
210 215 220

Cys Ser Pro Leu Gln Pro Ile Arg Gly Ser Ser Lys Arg Ser Phe Glu
225 230 235 240
Phe Ile Arg Asn Ser Ile Ile Asn Phe Leu Met Tyr Ser Leu Tyr Ala
245 250 255
Lys Leu Phe Gly Ile Pro Arg Ser Val Lys Ala Leu Met Lys Ser Asn
260 265 270
Lys Asp Glu Asn Lys Leu Lys Leu Glu Glu Lys Leu Lys Lys Lys Lys

275 280 285

Ser Ser Phe Asn Lys Thr Val Lys Glu Phe Glu Lys Met Ile Gly Arg
290 295 300
Lys Leu Ser Asp Asn Glu Ser Lys Ile Leu Asn Asp Glu Ser Lys Lys
305 310 315 320
Phe Phe Glu Ile Ile Lys Ser Asn Asn Lys Tyr Ile Pro Ser Glu Glu
325 330 335
Tyr Leu Lys Leu Leu Lys Asp Ile Ser Glu Glu Ile Tyr Asn Ser Asn
340 345 350

Ile Asp Phe Lys Pro Tyr Lys Tyr Ser Ile Leu Ile Arg Lys Pro Leu
355 360 365
Ser Lys Phe Lys Ser Lys Lys Leu Tyr Asn Leu Lys Pro Thr Asp Tyr
370 375 380
Lys Tyr Tyr Leu Gln Leu Ser Tyr Glu Pro Phe Ser Lys Gln Leu Ile
385 390 395 400
Ala Thr Lys Thr Ile Leu Gly Ile Asp Arg Gly Leu Lys His Leu Leu
405 410 415

Ala Val Ser Val Phe Asp Pro Ser Gln Asn Lys Phe Val Tyr Asn Lys
420 425 430
Leu Ile Lys Asn Pro Val Phe Lys Trp Lys Lys Arg Tyr His Asp Leu
435 440 445
Lys Arg Ser Ile Arg Asn Arg Glu Arg Arg Ile Arg Ala Leu Thr Gly
450 455 460
Val His Ile His Glu Asn Gln Leu Ile Lys Lys Leu Lys Ser Met Lys
465 470 475 480

Asn Lys Ile Asn Val Leu Tyr His Asn Val Ser Lys Asn Ile Val Asp
485 490 495
Leu Ala Lys Lys Tyr Glu Ser Thr Ile Val Leu Glu Arg Leu Glu Asn
500 505 510
Leu Lys Gln His Gly Arg Ser Lys Gly Lys Arg Tyr Lys Lys Leu Asn
515 520 525

Tyr Val Leu Ser Asn Phe Asp Tyr Lys Lys Ile Glu Ser Leu Ile Ser
530 535 540

Tyr Lys Ala Lys Lys Glu Gly Val Pro Val Ser Asn Ile Asn Pro Lys
545 550 555 560
Tyr Thr Ser Lys Thr Cys Ala Lys Cys Leu Leu Glu Val Asn Gln Leu
565 570 575
Ser Glu Leu Lys Asn Glu Tyr Asn Arg Asp Ser Lys Asn Ser Lys Ile
580 585 590
Gly Ile Cys Asn Ile His Gly Gln Ile Asp Ala Asp Leu Asn Ala Ala
595 600 605

Arg Val Ile Ala Leu Cys Tyr Ser Lys Asn Leu Asn Glu Pro His Phe
610 615 620
Lys
625

<210> 26
<211> 517
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 26
Val Ile Asn Leu Phe Gly Tyr Lys Phe Ala Leu Tyr Pro Asn Lys Thr
1 5 10 15
Gln Glu Glu Leu Leu Asn Lys His Leu Gly Glu Cys Gly Trp Leu Tyr
20 25 30

Asn Lys Ala Ile Glu Gln Asn Glu Tyr Tyr Lys Ala Asp Ser Asn Ile
35 40 45
Glu Glu Ala Gln Lys Lys Phe Glu Leu Leu Pro Asp Lys Asn Ser Asp
50 55 60
Glu Ala Lys Val Leu Arg Gly Asn Ile Ser Lys Asp Asn Tyr Val Tyr
65 70 75 80
Arg Thr Leu Val Lys Lys Lys Lys Ser Glu Ile Asn Val Gln Ile Arg

85 90 95

Lys Ala Val Val Leu Arg Pro Ala Glu Thr Ile Arg Asn Leu Ala Lys
100 105 110
Val Lys Lys Lys Gly Leu Ser Val Gly Arg Leu Lys Phe Ile Pro Ile
115 120 125
Arg Glu Trp Asp Val Leu Pro Phe Lys Gln Ser Asp Gln Ile Arg Leu
130 135 140
Glu Glu Asn Tyr Leu Ile Leu Glu Pro Tyr Gly Arg Leu Lys Phe Lys
145 150 155 160

Met His Arg Pro Leu Leu Gly Lys Pro Lys Thr Phe Cys Ile Lys Arg
165 170 175
Thr Ala Thr Asp Arg Trp Thr Ile Ser Phe Ser Thr Glu Tyr Asp Asp
180 185 190
Ser Asn Met Arg Lys Asn Asp Gly Gly Gln Val Gly Ile Asp Val Gly
195 200 205
Leu Lys Thr His Leu Arg Leu Ser Asn Glu Asn Pro Asp Glu Asp Pro
210 215 220

Arg Tyr Pro Asn Pro Lys Ile Trp Lys Arg Tyr Asp Arg Arg Leu Thr
225 230 235 240
Ile Leu Gln Arg Arg Ile Ser Lys Ser Lys Lys Leu Gly Lys Asn Arg
245 250 255
Thr Arg Leu Arg Leu Arg Leu Ser Arg Leu Trp Glu Lys Ile Arg Asn
260 265 270
Ser Arg Ala Asp Leu Ile Gln Asn Glu Thr Tyr Glu Ile Leu Ser Glu
275 280 285

Asn Lys Leu Ile Ala Ile Glu Asp Leu Asn Val Lys Gly Met Gln Glu
290 295 300
Lys Lys Asp Lys Lys Gly Arg Lys Gly Arg Thr Arg Ala Gln Glu Lys
305 310 315 320
Gly Leu His Arg Ser Ile Ser Asp Ala Ala Phe Ser Glu Phe Arg Arg
325 330 335

Val Leu Glu Tyr Lys Ala Lys Arg Phe Gly Ser Glu Val Lys Pro Val
340 345 350

Ser Ala Ile Asp Ser Ser Lys Glu Cys His Asn Cys Gly Asn Lys Lys
355 360 365
Gly Met Pro Leu Glu Ser Arg Ile Tyr Glu Cys Pro Lys Cys Gly Leu
370 375 380
Lys Ile Asp Arg Asp Leu Asn Ser Ala Lys Val Ile Leu Ala Arg Ala
385 390 395 400
Thr Gly Val Arg Pro Gly Ser Asn Ala Arg Ala Asp Thr Lys Ile Ser
405 410 415

Ala Thr Ala Gly Ala Ser Val Gln Thr Glu Gly Thr Val Ser Glu Asp
420 425 430
Phe Arg Gln Gln Met Glu Thr Ser Asp Gln Lys Pro Met Gln Gly Glu
435 440 445
Gly Ser Lys Glu Pro Pro Met Asn Pro Glu His Lys Ser Ser Gly Arg
450 455 460
Gly Ser Lys His Val Asn Ile Gly Cys Lys Asn Lys Val Gly Leu Tyr
465 470 475 480

Asn Glu Asp Glu Asn Ser Arg Ser Thr Glu Lys Gln Ile Met Asp Glu
485 490 495
Asn Arg Ser Thr Thr Glu Asp Met Val Glu Ile Gly Ala Leu His Ser
500 505 510
Pro Val Leu Thr Thr
515

<210> 27
<211> 410
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 27
Met Ile Ala Ser Ile Asp Tyr Glu Ala Val Ser Gln Ala Leu Ile Val

1 5 10 15

Phe Glu Phe Lys Ala Lys Gly Lys Asp Ser Gln Tyr Gln Ala Ile Asp
20 25 30
Glu Ala Ile Arg Ser Tyr Arg Phe Ile Arg Asn Ser Cys Leu Arg Tyr
35 40 45
Trp Met Asp Asn Lys Lys Val Gly Lys Tyr Asp Leu Asn Lys Tyr Cys
50 55 60
Lys Val Leu Ala Lys Gln Tyr Pro Phe Ala Asn Lys Leu Asn Ser Gln
65 70 75 80

Ala Arg Gln Ser Ala Ala Glu Cys Ser Trp Ser Ala Ile Ser Arg Phe
85 90 95
Tyr Asp Asn Cys Lys Arg Lys Val Ser Gly Lys Lys Gly Phe Pro Lys
100 105 110
Phe Lys Lys His Ala Arg Ser Val Glu Tyr Lys Thr Ser Gly Trp Lys
115 120 125
Leu Ser Glu Asn Arg Lys Ala Ile Thr Phe Thr Asp Lys Asn Gly Ile
130 135 140

Gly Lys Leu Lys Leu Lys Gly Thr Tyr Asp Leu His Phe Ser Gln Leu
145 150 155 160
Glu Asp Met Lys Arg Val Arg Leu Val Arg Arg Ala Asp Gly Tyr Tyr
165 170 175
Val Gln Phe Cys Ile Ser Val Asp Val Lys Val Glu Thr Glu Pro Thr
180 185 190
Gly Lys Ala Ile Gly Leu Asp Val Gly Ile Lys Tyr Phe Leu Ala Asp
195 200 205

Ser Ser Gly Asn Thr Ile Glu Asn Pro Gln Phe Tyr Arg Lys Ala Glu
210 215 220
Lys Lys Leu Asn Arg Ala Asn Arg Arg Lys Ser Lys Lys Tyr Ile Arg
225 230 235 240
Gly Val Lys Pro Gln Ser Lys Asn Tyr His Lys Ala Arg Cys Arg Tyr
245 250 255

Ala Arg Lys His Leu Arg Val Ser Arg Gln Arg Lys Glu Tyr Cys Lys
260 265 270

Arg Val Ala Tyr Cys Val Ile His Ser Asn Asp Val Val Ala Tyr Glu
275 280 285
Asp Leu Asn Val Lys Gly Met Val Lys Asn Arg His Leu Ala Lys Ser
290 295 300
Ile Ser Asp Val Ala Trp Ser Thr Phe Arg His Trp Leu Glu Tyr Phe
305 310 315 320
Ala Ile Lys Tyr Gly Lys Leu Thr Ile Pro Val Ala Pro His Asn Thr
325 330 335

Ser Gln Asn Cys Ser Asn Cys Asp Lys Lys Val Pro Lys Ser Leu Ser
340 345 350
Thr Arg Thr His Ile Cys His His Cys Gly Tyr Ser Glu Asp Arg Asp
355 360 365
Val Asn Ala Ala Lys Asn Ile Leu Lys Lys Ala Leu Ser Thr Val Gly
370 375 380
Gln Thr Gly Ser Leu Lys Leu Gly Glu Ile Glu Pro Leu Leu Val Leu
385 390 395 400

Glu Gln Ser Cys Thr Arg Lys Phe Asp Leu
405 410

<210> 28
<211> 486
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 28
Leu Ala Glu Glu Asn Thr Leu His Leu Thr Leu Ala Met Ser Leu Pro
1 5 10 15
Leu Asn Asp Leu Pro Glu Asn Arg Thr Arg Ser Glu Leu Trp Arg Arg
20 25 30
Gln Trp Leu Pro Gln Lys Lys Leu Ser Leu Leu Leu Gly Val Asn Gln

35 40 45

Ser Val Arg Lys Ala Ala Ala Asp Cys Leu Arg Trp Phe Glu Pro Tyr

50 55 60

Gln Glu Leu Leu Trp Trp Glu Pro Thr Asp Pro Asp Gly Lys Lys Leu

65 70 75 80

Leu Asp Lys Glu Gly Arg Pro Ile Lys Arg Thr Ala Gly His Met Arg

85 90 95

Val Leu Arg Lys Leu Glu Glu Ile Ala Pro Phe Arg Gly Tyr Gln Leu

100 105 110

Gly Ser Ala Val Lys Asn Gly Leu Arg His Lys Val Ala Asp Leu Leu

115 120 125

Leu Ser Tyr Ala Lys Arg Lys Leu Asp Pro Gln Phe Thr Asp Lys Thr

130 135 140

Ser Tyr Pro Ser Ile Gly Asp Gln Phe Pro Ile Val Trp Thr Gly Ala

145 150 155 160

Phe Val Cys Tyr Glu Gln Ser Ile Thr Gly Gln Leu Tyr Leu Tyr Leu

165 170 175

Pro Leu Phe Pro Arg Gly Ser His Gln Glu Asp Ile Thr Asn Asn Tyr

180 185 190

Asp Pro Asp Arg Gly Pro Ala Leu Gln Val Phe Gly Glu Lys Glu Ile

195 200 205

Ala Arg Leu Ser Arg Ser Thr Ser Gly Leu Leu Leu Pro Leu Gln Phe

210 215 220

Asp Lys Trp Gly Glu Ala Thr Phe Ile Arg Gly Glu Asn Asn Pro Pro

225 230 235 240

Thr Trp Lys Ala Thr His Arg Arg Ser Asp Lys Lys Trp Leu Ser Glu

245 250 255

Val Leu Leu Arg Glu Lys Asp Phe Gln Pro Lys Arg Val Glu Leu Leu

260 265 270

Val Arg Asn Gly Arg Ile Phe Val Asn Val Ala Cys Glu Ile Pro Thr

275 280 285

Lys Pro Leu Leu Glu Val Glu Asn Phe Met Gly Val Ser Phe Gly Leu

290 295 300

Glu His Leu Val Thr Val Val Val Ile Asn Arg Asp Gly Asn Val Val
305 310 315 320
His Gln Arg Gln Glu Pro Ala Arg Arg Tyr Glu Lys Thr Tyr Phe Ala
325 330 335
Arg Leu Glu Arg Leu Arg Arg Arg Gly Gly Pro Phe Ser Gln Glu Leu
340 345 350
Glu Thr Phe His Tyr Arg Gln Val Ala Gln Ile Val Glu Glu Ala Leu

355 360 365

Arg Phe Lys Ser Val Pro Ala Val Glu Gln Val Gly Asn Ile Pro Lys
370 375 380
Gly Arg Tyr Asn Pro Arg Leu Asn Leu Arg Leu Ser Tyr Trp Pro Phe
385 390 395 400
Gly Lys Leu Ala Asp Leu Thr Ser Tyr Lys Ala Val Lys Glu Gly Leu
405 410 415
Pro Lys Pro Tyr Ser Val Tyr Ser Ala Thr Ala Lys Met Leu Cys Ser

420 425 430

Thr Cys Gly Ala Ala Asn Lys Glu Gly Asp Gln Pro Ile Ser Leu Lys
435 440 445
Gly Pro Thr Val Tyr Cys Gly Asn Cys Gly Thr Arg His Asn Thr Gly
450 455 460
Phe Asn Thr Ala Leu Asn Leu Ala Arg Arg Ala Gln Glu Leu Phe Val
465 470 475 480
Lys Gly Val Val Ala Arg
485

<210> 29
<211> 602
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence

<400> 29

Met Ser Gln Ser Leu Leu Lys Trp His Asp Met Ala Gly Arg Asp Lys
1 5 10 15
Asp Ala Ser Arg Ser Leu Gln Lys Ser Ala Val Glu Gly Val Leu Leu
20 25 30
His Leu Thr Ala Ser His Arg Val Ala Leu Glu Met Leu Glu Lys Ser
35 40 45
Val Ser Gln Thr Val Ala Val Thr Met Glu Ala Ala Gln Gln Arg Leu
50 55 60
Val Ile Val Leu Glu Asp Asp Pro Thr Lys Ala Thr Ser Arg Lys Arg
65 70 75 80
Val Ile Ser Ala Asp Leu Gln Phe Thr Arg Glu Glu Phe Gly Ser Leu
85 90 95
Pro Asn Trp Ala Gln Lys Leu Ala Ser Thr Cys Pro Glu Ile Ala Thr
100 105 110
Lys Tyr Ala Asp Lys His Ile Asn Ser Ile Arg Ile Ala Trp Gly Val
115 120 125
Ala Lys Glu Ser Thr Asn Gly Asp Ala Val Glu Gln Lys Leu Gln Trp
130 135 140
Gln Ile Arg Leu Leu Asp Val Thr Met Phe Leu Gln Gln Leu Val Leu
145 150 155 160
Gln Leu Ala Asp Lys Ala Leu Leu Glu Gln Ile Pro Ser Ser Ile Arg
165 170 175
Gly Gly Ile Gly Gln Glu Val Ala Gln Gln Val Thr Ser His Ile Gln
180 185 190
Leu Leu Asp Ser Gly Thr Val Leu Lys Ala Glu Leu Pro Thr Ile Ser
195 200 205
Asp Arg Asn Ser Glu Leu Ala Arg Lys Gln Trp Glu Asp Ala Ile Gln
210 215 220
Thr Val Cys Thr Tyr Ala Leu Pro Phe Ser Arg Glu Arg Ala Arg Ile
225 230 235 240

Leu Asp Pro Gly Lys Tyr Ala Ala Glu Asp Pro Arg Gly Asp Arg Leu

245 250 255

Ile Asn Ile Asp Pro Met Trp Ala Arg Val Leu Lys Gly Pro Thr Val

260 265 270

Lys Ser Leu Pro Leu Leu Phe Val Ser Gly Ser Ser Ile Arg Ile Val

275 280 285

Lys Leu Thr Leu Pro Arg Lys His Ala Ala Gly His Lys His Thr Phe

290 295 300

Thr Ala Thr Tyr Leu Val Leu Pro Val Ser Arg Glu Trp Ile Asn Ser

305 310 315 320

Leu Pro Gly Thr Val Gln Glu Lys Val Gln Trp Trp Lys Lys Pro Asp

325 330 335

Val Leu Ala Thr Gln Glu Leu Leu Val Gly Lys Gly Ala Leu Lys Lys

340 345 350

Ser Ala Asn Thr Leu Val Ile Pro Ile Ser Ala Gly Lys Lys Arg Phe

355 360 365

Phe Asn His Ile Leu Pro Ala Leu Gln Arg Gly Phe Pro Leu Gln Trp

370 375 380

Gln Arg Ile Val Gly Arg Ser Tyr Arg Arg Pro Ala Thr His Arg Lys

385 390 395 400

Trp Phe Ala Gln Leu Thr Ile Gly Tyr Thr Asn Pro Ser Ser Leu Pro

405 410 415

Glu Met Ala Leu Gly Ile His Phe Gly Met Lys Asp Ile Leu Trp Trp

420 425 430

Ala Leu Ala Asp Lys Gln Gly Asn Ile Leu Lys Asp Gly Ser Ile Pro

435 440 445

Gly Asn Ser Ile Leu Asp Phe Ser Leu Gln Glu Lys Gly Lys Ile Glu

450 455 460

Arg Gln Gln Lys Ala Gly Lys Asn Val Ala Gly Lys Lys Tyr Gly Lys

465 470 475 480

Ser Leu Leu Asn Ala Thr Tyr Arg Val Val Asn Gly Val Leu Glu Phe

485 490 495

Ser Lys Gly Ile Ser Ala Glu His Ala Ser Gln Pro Ile Gly Leu Gly

500 505 510

Leu Glu Thr Ile Arg Phe Val Asp Lys Ala Ser Gly Ser Ser Pro Val

515 520 525

Asn Ala Arg His Ser Asn Trp Asn Tyr Gly Gln Leu Ser Gly Ile Phe

530 535 540

Ala Asn Lys Ala Gly Pro Ala Gly Phe Ser Val Thr Glu Ile Thr Leu

545 550 555 560

Lys Lys Ala Gln Arg Asp Leu Ser Asp Ala Glu Gln Ala Arg Val Leu

565 570 575

Ala Ile Glu Ala Thr Lys Arg Phe Ala Ser Arg Ile Lys Arg Leu Ala

580 585 590

Thr Lys Arg Lys Asp Asp Thr Leu Phe Val

595 600

<210> 30

<211> 494

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 30

Val Glu Pro Val Glu Lys Glu Arg Phe Tyr Tyr Arg Thr Tyr Thr Phe

1 5 10 15

Arg Leu Asp Gly Gln Pro Arg Thr Gln Asn Leu Thr Thr Gln Ser Gly

20 25 30

Trp Gly Leu Leu Thr Lys Ala Val Leu Asp Asn Thr Lys His Tyr Trp

35 40 45

Glu Ile Val His His Ala Arg Ile Ala Asn Gln Pro Ile Val Phe Glu

50 55 60

Asn Pro Val Ile Asp Glu Gln Gly Asn Pro Lys Leu Asn Lys Leu Gly

65 70 75 80

Gln Pro Arg Phe Trp Lys Arg Pro Ile Ser Asp Ile Val Asn Gln Leu
85 90 95
Arg Ala Leu Phe Glu Asn Gln Asn Pro Tyr Gln Leu Gly Ser Ser Leu
100 105 110
Ile Gln Gly Thr Tyr Trp Asp Val Ala Glu Asn Leu Ala Ser Trp Tyr
115 120 125
Ala Leu Asn Lys Glu Tyr Leu Ala Gly Thr Ala Thr Trp Gly Glu Pro
130 135 140

Ser Phe Pro Glu Pro His Pro Leu Thr Glu Ile Asn Gln Trp Met Pro
145 150 155 160
Leu Thr Phe Ser Ser Gly Lys Val Val Arg Leu Leu Lys Asn Ala Ser
165 170 175
Gly Arg Tyr Phe Ile Gly Leu Pro Ile Leu Gly Glu Asn Asn Pro Cys
180 185 190
Tyr Arg Met Arg Thr Ile Glu Lys Leu Ile Pro Cys Asp Gly Lys Gly
195 200 205

Arg Val Thr Ser Gly Ser Leu Ile Leu Phe Pro Leu Val Gly Ile Tyr
210 215 220
Ala Gln Gln His Arg Arg Met Thr Asp Ile Cys Glu Ser Ile Arg Thr
225 230 235 240
Glu Lys Gly Lys Leu Ala Trp Ala Gln Val Ser Ile Asp Tyr Val Arg
245 250 255
Glu Val Asp Lys Arg Arg Arg Met Arg Arg Thr Arg Lys Ser Gln Gly
260 265 270

Trp Ile Gln Gly Pro Trp Gln Glu Val Phe Ile Leu Arg Leu Val Leu
275 280 285
Ala His Lys Ala Pro Lys Leu Tyr Lys Pro Arg Cys Phe Ala Gly Ile
290 295 300
Ser Leu Gly Pro Lys Thr Leu Ala Ser Cys Val Ile Leu Asp Gln Asp
305 310 315 320
Glu Arg Val Val Glu Lys Gln Gln Trp Ser Gly Ser Glu Leu Leu Ser

325 330 335

Leu Ile His Gln Gly Glu Glu Arg Leu Arg Ser Leu Arg Glu Gln Ser
340 345 350
Lys Pro Thr Trp Asn Ala Ala Tyr Arg Lys Gln Leu Lys Ser Leu Ile
355 360 365
Asn Thr Gln Val Phe Thr Ile Val Thr Phe Leu Arg Glu Arg Gly Ala
370 375 380
Ala Val Arg Leu Glu Ser Ile Ala Arg Val Arg Lys Ser Thr Pro Ala
385 390 395 400

Pro Pro Val Asn Phe Leu Leu Ser His Trp Ala Tyr Arg Gln Ile Thr
405 410 415
Glu Arg Leu Lys Asp Leu Ala Ile Arg Asn Gly Met Pro Leu Thr His
420 425 430
Ser Asn Gly Ser Tyr Gly Val Arg Phe Thr Cys Ser Gln Cys Gly Ala
435 440 445
Thr Asn Gln Gly Ile Lys Asp Pro Thr Lys Tyr Lys Val Asp Ile Glu
450 455 460

Ser Glu Thr Phe Leu Cys Ser Ile Cys Ser His Arg Glu Ile Ala Ala
465 470 475 480
Val Asn Thr Ala Thr Asn Leu Ala Lys Gln Leu Leu Asp Glu
485 490

<210> 31
<211> 526
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 31
Met Asn Asp Thr Glu Thr Ser Glu Thr Leu Thr Ser His Arg Thr Val
1 5 10 15
Cys Ala His Leu His Val Val Gly Glu Thr Gly Ser Leu Pro Arg Leu

20 25 30

Val Glu Ala Ala Leu Ala Glu Leu Ile Thr Leu Asn Gly Arg Ala Thr
35 40 45
Gln Ala Leu Leu Ser Leu Ala Lys Asn Gly Leu Val Leu Arg Arg Asp
50 55 60
Lys Glu Glu Asn Leu Ile Ala Ala Glu Leu Thr Leu Pro Cys Arg Lys
65 70 75 80
Asn Lys Tyr Ala Asp Val Ala Ala Lys Ala Gly Glu Pro Ile Leu Ala

85 90 95
Thr Arg Ile Asn Asn Lys Gly Lys Leu Val Thr Lys Lys Trp Tyr Gly
100 105 110
Glu Gly Asn Ser Tyr His Ile Val Arg Phe Thr Pro Glu Thr Gly Met
115 120 125
Phe Thr Val Arg Val Phe Asp Arg Tyr Ala Phe Asp Glu Glu Leu Leu
130 135 140
His Leu His Ser Glu Val Val Phe Gly Ser Asp Leu Pro Lys Gly Ile

145 150 155 160
Lys Ala Lys Thr Asp Ser Leu Pro Ala Asn Phe Leu Gln Ala Val Phe
165 170 175
Thr Ser Phe Leu Glu Leu Pro Phe Gln Gly Phe Pro Asp Ile Val Val
180 185 190
Lys Pro Ala Met Lys Gln Ala Ala Glu Gln Leu Leu Ser Tyr Val Gln
195 200 205
Leu Glu Ala Gly Glu Asn Gln Gln Ala Glu Tyr Pro Asp Thr Asn Glu

210 215 220
Arg Asp Pro Glu Leu Arg Leu Val Glu Trp Gln Lys Ser Leu His Glu
225 230 235 240
Leu Ser Val Arg Thr Glu Pro Phe Glu Phe Val Arg Ala Arg Asp Ile
245 250 255
Asp Tyr Tyr Ala Glu Thr Asp Arg Arg Gly Asn Arg Phe Val Asn Ile
260 265 270
Thr Pro Glu Trp Thr Lys Phe Ala Glu Ser Pro Phe Ala Arg Arg Leu

275 280 285

Pro Leu Lys Ile Pro Pro Glu Phe Cys Ile Leu Leu Arg Arg Lys Thr

290 295 300

Glu Gly His Ala Lys Ile Pro Asn Arg Ile Tyr Leu Gly Leu Gln Ile

305 310 315 320

Phe Asp Gly Val Thr Pro Asp Ser Thr Leu Gly Val Leu Ala Thr Ala

325 330 335

Glu Asp Gly Lys Leu Phe Trp Trp His Asp His Leu Asp Glu Phe Ser

340 345 350

Asn Leu Glu Gly Lys Pro Glu Pro Lys Leu Lys Asn Lys Pro Gln Leu

355 360 365

Leu Met Val Ser Leu Glu Tyr Asp Arg Glu Gln Arg Phe Glu Glu Ser

370 375 380

Val Gly Gly Asp Arg Lys Ile Cys Leu Val Thr Leu Lys Glu Thr Arg

385 390 395 400

Asn Phe Arg Arg Gly Trp Asn Gly Arg Ile Leu Gly Ile His Phe Gln

405 410 415

His Asn Pro Val Ile Thr Trp Ala Leu Met Asp His Asp Ala Glu Val

420 425 430

Leu Glu Lys Gly Phe Ile Glu Gly Asn Ala Phe Leu Gly Lys Ala Leu

435 440 445

Asp Lys Gln Ala Leu Asn Glu Tyr Leu Gln Lys Gly Gly Lys Trp Val

450 455 460

Gly Asp Arg Ser Phe Gly Asn Lys Leu Lys Gly Ile Thr His Thr Leu

465 470 475 480

Ala Ser Leu Ile Val Arg Leu Ala Arg Glu Lys Asp Ala Trp Ile Ala

485 490 495

Leu Glu Glu Ile Ser Trp Val Gln Lys Gln Ser Ala Asp Ser Val Ala

500 505 510

Asn His Glu Ile Val Glu Gln Pro His His Ser Leu Thr Arg

515 520 525

<210> 32

<211> 649

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 32

Met Asn Asp Thr Glu Thr Ser Glu Thr Leu Thr Ser His Arg Thr Val
1 5 10 15

Cys Ala His Leu His Val Val Gly Glu Thr Gly Ser Leu Pro Arg Leu
20 25 30

Val Glu Ala Ala Leu Ala Glu Leu Ile Thr Leu Asn Gly Arg Ala Thr
35 40 45

Gln Ala Leu Leu Ser Leu Ala Lys Asn Gly Leu Val Leu Arg Arg Asp
50 55 60

Lys Glu Glu Asn Leu Ile Ala Ala Glu Leu Thr Leu Pro Cys Arg Lys
65 70 75 80

Asn Lys Tyr Ala Asp Val Ala Ala Lys Ala Gly Glu Pro Ile Leu Ala
85 90 95

Thr Arg Ile Asn Asn Lys Gly Lys Leu Val Thr Lys Lys Trp Tyr Gly
100 105 110

Glu Gly Asn Ser Tyr His Ile Val Arg Phe Thr Pro Glu Thr Gly Met
115 120 125

Phe Thr Val Arg Val Phe Asp Arg Tyr Ala Phe Asp Glu Glu Leu Leu
130 135 140

His Leu His Ser Glu Val Val Phe Gly Ser Asp Leu Pro Lys Gly Ile
145 150 155 160

Lys Ala Lys Thr Asp Ser Leu Pro Ala Asn Phe Leu Gln Ala Val Phe
165 170 175

Thr Ser Phe Leu Glu Leu Pro Phe Gln Gly Phe Pro Asp Ile Val Val
180 185 190

Lys Pro Ala Met Lys Gln Ala Ala Glu Gln Leu Leu Ser Tyr Val Gln
195 200 205

Leu Glu Ala Gly Glu Asn Gln Gln Ala Glu Tyr Pro Asp Thr Asn Glu
210 215 220
Arg Asp Pro Glu Leu Arg Leu Val Glu Trp Gln Lys Ser Leu His Glu
225 230 235 240
Leu Ser Val Arg Thr Glu Pro Phe Glu Phe Val Arg Ala Arg Asp Ile
245 250 255

Asp Tyr Tyr Ala Glu Thr Asp Arg Arg Gly Asn Arg Phe Val Asn Ile
260 265 270
Thr Pro Glu Trp Thr Lys Phe Ala Glu Ser Pro Phe Ala Arg Arg Leu
275 280 285
Pro Leu Lys Ile Pro Pro Glu Phe Cys Ile Leu Leu Arg Arg Lys Thr
290 295 300
Glu Gly His Ala Lys Ile Pro Asn Arg Ile Tyr Leu Gly Leu Gln Ile
305 310 315 320

Phe Asp Gly Val Thr Pro Asp Ser Thr Leu Gly Val Leu Ala Thr Ala
325 330 335
Glu Asp Gly Lys Leu Phe Trp Trp His Asp His Leu Asp Glu Phe Ser
340 345 350
Asn Leu Glu Gly Lys Pro Glu Pro Lys Leu Lys Asn Lys Pro Gln Leu
355 360 365
Leu Met Val Ser Leu Glu Tyr Asp Arg Glu Gln Arg Phe Glu Glu Ser
370 375 380

Val Gly Gly Asp Arg Lys Ile Cys Leu Val Thr Leu Lys Glu Thr Arg
385 390 395 400
Asn Phe Arg Arg Gly Arg His Gly His Thr Arg Thr Asp Arg Leu Pro
405 410 415
Ala Gly Asn Thr Leu Trp Arg Ala Asp Phe Ala Thr Ser Ala Glu Val
420 425 430
Ala Ala Pro Lys Trp Asn Gly Arg Ile Leu Gly Ile His Phe Gln His
435 440 445

Asn Pro Val Ile Thr Trp Ala Leu Met Asp His Asp Ala Glu Val Leu

450 455 460

Glu Lys Gly Phe Ile Glu Gly Asn Ala Phe Leu Gly Lys Ala Leu Asp
465 470 475 480

Lys Gln Ala Leu Asn Glu Tyr Leu Gln Lys Gly Gly Lys Trp Val Gly
485 490 495

Asp Arg Ser Phe Gly Asn Lys Leu Lys Gly Ile Thr His Thr Leu Ala
500 505 510

Ser Leu Ile Val Arg Leu Ala Arg Glu Lys Asp Ala Trp Ile Ala Leu
515 520 525

Glu Glu Ile Ser Trp Val Gln Lys Gln Ser Ala Asp Ser Val Ala Asn
530 535 540

Arg Arg Phe Ser Met Trp Asn Tyr Ser Arg Leu Ala Thr Leu Ile Glu
545 550 555 560

Trp Leu Gly Thr Asp Ile Ala Thr Arg Asp Cys Gly Thr Ala Ala Pro
565 570 575

Leu Ala His Lys Val Ser Asp Tyr Leu Thr His Phe Thr Cys Pro Glu
580 585 590

Cys Gly Ala Cys Arg Lys Ala Gly Gln Lys Lys Glu Ile Ala Asp Thr
595 600 605

Val Arg Ala Gly Asp Ile Leu Thr Cys Arg Lys Cys Gly Phe Ser Gly
610 615 620

Pro Ile Pro Asp Asn Phe Ile Ala Glu Phe Val Ala Lys Lys Ala Leu
625 630 635 640

Glu Arg Met Leu Lys Lys Lys Pro Val
645

<210> 33

<211> 414

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 33

Met Ala Lys Arg Asn Phe Gly Glu Lys Ser Glu Ala Leu Tyr Arg Ala

1 5 10 15

Val Arg Phe Glu Val Arg Pro Ser Lys Glu Glu Leu Ser Ile Leu Leu

20 25 30

Ala Val Ser Glu Val Leu Arg Met Leu Phe Asn Ser Ala Leu Ala Glu

35 40 45

Arg Gln Gln Val Phe Thr Glu Phe Ile Ala Ser Leu Tyr Ala Glu Leu

50 55 60

Lys Ser Ala Ser Val Pro Glu Glu Ile Ser Glu Ile Arg Lys Lys Leu

65 70 75 80

Arg Glu Ala Tyr Lys Glu His Ser Ile Ser Leu Phe Asp Gln Ile Asn

85 90 95

Ala Leu Thr Ala Arg Arg Val Glu Asp Glu Ala Phe Ala Ser Val Thr

100 105 110

Arg Asn Trp Gln Glu Glu Thr Leu Asp Ala Leu Asp Gly Ala Tyr Lys

115 120 125

Ser Phe Leu Ser Leu Arg Arg Lys Gly Asp Tyr Asp Ala His Ser Pro

130 135 140

Arg Ser Arg Asp Ser Gly Phe Phe Gln Lys Ile Pro Gly Arg Ser Gly

145 150 155 160

Phe Lys Ile Gly Glu Gly Arg Ile Ala Leu Ser Cys Gly Ala Gly Arg

165 170 175

Lys Leu Ser Phe Pro Ile Pro Asp Tyr Gln Gln Gly Arg Leu Ala Glu

180 185 190

Thr Thr Lys Leu Lys Lys Phe Glu Leu Tyr Arg Asp Gln Pro Asn Leu

195 200 205

Ala Lys Ser Gly Arg Phe Trp Ile Ser Val Val Tyr Glu Leu Pro Lys

210 215 220

Pro Glu Ala Thr Thr Cys Gln Ser Glu Gln Val Ala Phe Val Ala Leu

225 230 235 240

Gly Ala Ser Ser Ile Gly Val Val Ser Gln Arg Gly Glu Glu Val Ile

245 250 255

Ala Leu Trp Arg Ser Asp Lys His Trp Val Pro Lys Ile Glu Ala Val
260 265 270
Glu Glu Arg Met Lys Arg Arg Val Lys Gly Ser Arg Gly Trp Leu Arg
275 280 285
Leu Leu Asn Ser Gly Lys Arg Arg Met His Met Ile Ser Ser Arg Gln

290 295 300
His Val Gln Asp Glu Arg Glu Ile Val Asp Tyr Leu Val Arg Asn His
305 310 315 320
Gly Ser His Phe Val Val Thr Glu Leu Val Val Arg Ser Lys Glu Gly
325 330 335
Lys Leu Ala Asp Ser Ser Lys Pro Glu Arg Gly Gly Ser Leu Gly Leu
340 345 350
Asn Trp Ala Ala Gln Asn Thr Gly Ser Leu Ser Arg Leu Val Arg Gln

355 360 365
Leu Glu Glu Lys Val Lys Glu His Gly Gly Ser Val Arg Lys His Lys
370 375 380
Leu Thr Leu Thr Glu Ala Pro Pro Ala Arg Gly Ala Glu Asn Lys Leu
385 390 395 400
Trp Met Ala Arg Lys Leu Arg Glu Ser Phe Leu Lys Glu Val
405 410

<210> 34

<211> 413

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 34

Leu Ala Lys Asn Asp Glu Lys Glu Leu Leu Tyr Gln Ser Val Lys Phe
1 5 10 15
Glu Ile Tyr Pro Asp Glu Ser Lys Ile Arg Val Leu Thr Arg Val Ser
20 25 30
Asn Ile Leu Val Leu Val Trp Asn Ser Ala Leu Gly Glu Arg Arg Ala
35 40 45

Arg Phe Glu Leu Tyr Ile Ala Pro Leu Tyr Glu Glu Leu Lys Lys Phe
50 55 60

Pro Arg Lys Ser Ala Glu Ser Asn Ala Leu Arg Gln Lys Ile Arg Glu
65 70 75 80
Gly Tyr Lys Glu His Ile Pro Thr Phe Phe Asp Gln Leu Lys Lys Leu
85 90 95
Leu Thr Pro Met Arg Lys Glu Asp Pro Ala Leu Leu Gly Ser Val Pro
100 105 110
Arg Ala Tyr Gln Glu Glu Thr Leu Asn Thr Leu Asn Gly Ser Phe Val
115 120 125

Ser Phe Met Thr Leu Arg Arg Asn Asn Asp Met Asp Ala Lys Pro Pro
130 135 140
Lys Gly Arg Ala Glu Asp Arg Phe His Glu Ile Ser Gly Arg Ser Gly
145 150 155 160
Phe Lys Ile Asp Gly Ser Glu Phe Val Leu Ser Thr Lys Glu Gln Lys
165 170 175
Leu Arg Phe Pro Ile Pro Asn Tyr Gln Leu Glu Lys Leu Lys Glu Ala
180 185 190

Lys Gln Ile Lys Lys Phe Thr Leu Tyr Gln Ser Arg Asp Arg Arg Phe
195 200 205
Trp Ile Ser Ile Ala Tyr Glu Ile Glu Leu Pro Asp Gln Arg Pro Phe
210 215 220
Asn Pro Glu Glu Val Ile Tyr Ile Ala Phe Gly Ala Ser Ser Ile Gly
225 230 235 240
Val Ile Ser Pro Glu Gly Glu Lys Val Ile Asp Phe Trp Arg Pro Asp
245 250 255

Lys His Trp Lys Pro Lys Ile Lys Glu Val Glu Asn Arg Met Arg Ser
260 265 270
Cys Lys Lys Gly Ser Arg Ala Trp Lys Lys Arg Ala Ala Ala Arg Arg
275 280 285
Lys Met Tyr Ala Met Thr Gln Arg Gln Gln Lys Leu Asn His Arg Glu

290 295 300

Ile Val Ala Ser Leu Leu Arg Leu Gly Phe His Phe Val Val Thr Glu
305 310 315 320

Tyr Thr Val Arg Ser Lys Pro Gly Lys Leu Ala Asp Gly Ser Asn Pro
325 330 335

Lys Arg Gly Gly Ala Pro Gln Gly Phe Asn Trp Ser Ala Gln Asn Thr
340 345 350

Gly Ser Phe Gly Glu Phe Ile Leu Trp Leu Lys Gln Lys Val Lys Glu
355 360 365

Gln Gly Gly Thr Val Gln Thr Phe Arg Leu Val Leu Gly Gln Ser Glu
370 375 380

Arg Pro Glu Lys Arg Gly Arg Asp Asn Lys Ile Glu Met Val Arg Leu
385 390 395 400

Leu Arg Glu Lys Tyr Leu Glu Ser Gln Thr Ile Val Val
405 410

<210> 35

<211> 449

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 35

Met Ala Lys Gly Lys Lys Lys Glu Gly Lys Pro Leu Tyr Arg Ala Val
1 5 10 15

Arg Phe Glu Ile Phe Pro Thr Ser Asp Gln Ile Thr Leu Phe Leu Arg

20 25 30

Val Ser Lys Asn Leu Gln Gln Val Trp Asn Glu Ala Trp Gln Glu Arg
35 40 45

Gln Ser Cys Tyr Glu Gln Phe Phe Gly Ser Ile Tyr Glu Arg Ile Gly
50 55 60

Gln Ala Lys Lys Arg Ala Gln Glu Ala Gly Phe Ser Glu Val Trp Glu
65 70 75 80

Asn Glu Ala Lys Lys Gly Leu Asn Lys Lys Leu Arg Gln Gln Glu Ile

85 90 95

Ser Met Gln Leu Val Ser Glu Lys Glu Ser Leu Leu Gln Glu Leu Ser

100 105 110

Ile Ala Phe Gln Glu His Gly Val Thr Leu Tyr Asp Gln Ile Asn Gly

115 120 125

Leu Thr Ala Arg Arg Ile Ile Gly Glu Phe Ala Leu Ile Pro Arg Asn

130 135 140

Trp Gln Glu Glu Thr Leu Asp Ser Leu Asp Gly Ser Phe Lys Ser Phe

145 150 155 160

Leu Ala Leu Arg Lys Asn Gly Asp Pro Asp Ala Lys Pro Pro Arg Gln

165 170 175

Arg Val Ser Glu Asn Ser Phe Tyr Lys Ile Pro Gly Arg Ser Gly Phe

180 185 190

Lys Val Ser Asn Gly Gln Ile Tyr Leu Ser Phe Gly Lys Ile Gly Gln

195 200 205

Thr Leu Thr Ser Val Ile Pro Glu Phe Gln Leu Lys Arg Leu Glu Thr

210 215 220

Ala Ile Lys Leu Lys Lys Phe Glu Leu Cys Arg Asp Glu Arg Asp Met

225 230 235 240

Ala Lys Pro Gly Arg Phe Trp Ile Ser Val Ala Tyr Glu Ile Pro Lys

245 250 255

Pro Glu Lys Val Pro Val Val Ser Lys Gln Ile Thr Tyr Leu Ala Ile

260 265 270

Gly Ala Ser Arg Leu Gly Val Val Ser Pro Lys Gly Glu Phe Cys Leu

275 280 285

Asn Leu Pro Arg Ser Asp Tyr His Trp Lys Pro Gln Ile Asn Ala Leu

290 295 300

Gln Glu Arg Leu Glu Gly Val Val Lys Gly Ser Arg Lys Trp Lys Lys

305 310 315 320

Arg Met Ala Ala Cys Thr Arg Met Phe Ala Lys Leu Gly His Gln Gln

325 330 335

Lys Gln His Gly Gln Tyr Glu Val Val Lys Lys Leu Leu Arg His Gly

340 345 350

Val His Phe Val Val Thr Glu Leu Lys Val Arg Ser Lys Pro Gly Ala

355 360 365

Leu Ala Asp Ala Ser Lys Ser Asp Arg Lys Gly Ser Pro Thr Gly Pro

370 375 380

Asn Trp Ser Ala Gln Asn Thr Gly Asn Ile Ala Arg Leu Ile Gln Lys

385 390 395 400

Leu Thr Asp Lys Ala Ser Glu His Gly Gly Thr Val Ile Lys Arg Asn

405 410 415

Pro Pro Leu Leu Ser Leu Glu Glu Arg Gln Leu Pro Asp Ala Gln Arg

420 425 430

Lys Ile Phe Ile Ala Lys Lys Leu Arg Glu Glu Phe Leu Ala Asp Gln

435 440 445

Lys

<210> 36

<211> 711

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 36

Met Ala Lys Arg Glu Lys Lys Asp Asp Val Val Leu Arg Gly Thr Lys

1 5 10 15

Met Arg Ile Tyr Pro Thr Asp Arg Gln Val Thr Leu Met Asp Met Trp

20 25 30

Arg Arg Arg Cys Ile Ser Leu Trp Asn Leu Leu Leu Asn Leu Glu Thr

35 40 45

Ala Ala Tyr Gly Ala Lys Asn Thr Arg Ser Lys Leu Gly Trp Arg Ser

50 55 60

Ile Trp Ala Arg Val Val Glu Glu Asn His Ala Lys Ala Leu Ile Val

65 70 75 80

Tyr Gln His Gly Lys Cys Lys Lys Asp Gly Ser Phe Val Leu Lys Arg

85 90 95

Asp Gly Thr Val Lys His Pro Pro Arg Glu Arg Phe Pro Gly Asp Arg

100 105 110

Lys Ile Leu Leu Gly Leu Phe Asp Ala Leu Arg His Thr Leu Asp Lys

115 120 125

Gly Ala Lys Cys Lys Cys Asn Val Asn Gln Pro Tyr Ala Leu Thr Arg

130 135 140

Ala Trp Leu Asp Glu Thr Gly His Gly Ala Arg Thr Ala Asp Ile Ile

145 150 155 160

Ala Trp Leu Lys Asp Phe Lys Gly Glu Cys Asp Cys Thr Ala Ile Ser

165 170 175

Thr Ala Ala Lys Tyr Cys Pro Ala Pro Pro Thr Ala Glu Leu Leu Thr

180 185 190

Lys Ile Lys Arg Ala Ala Pro Ala Asp Asp Leu Pro Val Asp Gln Ala

195 200 205

Ile Leu Leu Asp Leu Phe Gly Ala Leu Arg Gly Gly Leu Lys Gln Lys

210 215 220

Glu Cys Asp His Thr His Ala Arg Thr Val Ala Tyr Phe Glu Lys His

225 230 235 240

Glu Leu Ala Gly Arg Ala Glu Asp Ile Leu Ala Trp Leu Ile Ala His

245 250 255

Gly Gly Thr Cys Asp Cys Lys Ile Val Glu Glu Ala Ala Asn His Cys

260 265 270

Pro Gly Pro Arg Leu Phe Ile Trp Glu His Glu Leu Ala Met Ile Met

275 280 285

Ala Arg Leu Lys Ala Glu Pro Arg Thr Glu Trp Ile Gly Asp Leu Pro

290 295 300

Ser His Ala Ala Gln Thr Val Val Lys Asp Leu Val Lys Ala Leu Gln

305 310 315 320

Thr Met Leu Lys Glu Arg Ala Lys Ala Ala Ala Gly Asp Glu Ser Ala

325 330 335

Arg Lys Thr Gly Phe Pro Lys Phe Lys Lys Gln Ala Tyr Ala Ala Gly

340 345 350

Ser Val Tyr Phe Pro Asn Thr Thr Met Phe Phe Asp Val Ala Ala Gly

355 360 365

Arg Val Gln Leu Pro Asn Gly Cys Gly Ser Met Arg Cys Glu Ile Pro

370 375 380

Arg Gln Leu Val Ala Glu Leu Leu Glu Arg Asn Leu Lys Pro Gly Leu

385 390 395 400

Val Ile Gly Ala Gln Leu Gly Leu Leu Gly Gly Arg Ile Trp Arg Gln

405 410 415

Gly Asp Arg Trp Tyr Leu Ser Cys Gln Trp Glu Arg Pro Gln Pro Thr

420 425 430

Leu Leu Pro Lys Thr Gly Arg Thr Ala Gly Val Lys Ile Ala Ala Ser

435 440 445

Ile Val Phe Thr Thr Tyr Asp Asn Arg Gly Gln Thr Lys Glu Tyr Pro

450 455 460

Met Pro Pro Ala Asp Lys Lys Leu Thr Ala Val His Leu Val Ala Gly

465 470 475 480

Lys Gln Asn Ser Arg Ala Leu Glu Ala Gln Lys Glu Lys Glu Lys Lys

485 490 495

Leu Lys Ala Arg Lys Glu Arg Leu Arg Leu Gly Lys Leu Glu Lys Gly

500 505 510

His Asp Pro Asn Ala Leu Lys Pro Leu Lys Arg Pro Arg Val Arg Arg

515 520 525

Ser Lys Leu Phe Tyr Lys Ser Ala Ala Arg Leu Ala Ala Cys Glu Ala

530 535 540

Ile Glu Arg Asp Arg Arg Asp Gly Phe Leu His Arg Val Thr Asn Glu

545 550 555 560

Ile Val His Lys Phe Asp Ala Val Ser Val Gln Lys Met Ser Val Ala
565 570 575
Pro Met Met Arg Arg Gln Lys Gln Lys Glu Lys Gln Ile Glu Ser Lys

580 585 590
Lys Asn Glu Ala Lys Lys Glu Asp Asn Gly Ala Ala Lys Lys Pro Arg
595 600 605
Asn Leu Lys Pro Val Arg Lys Leu Leu Arg His Val Ala Met Ala Arg
610 615 620
Gly Arg Gln Phe Leu Glu Tyr Lys Tyr Asn Asp Leu Arg Gly Pro Gly
625 630 635 640
Ser Val Leu Ile Ala Asp Arg Leu Glu Pro Glu Val Gln Glu Cys Ser

645 650 655
Arg Cys Gly Thr Lys Asn Pro Gln Met Lys Asp Gly Arg Arg Leu Leu
660 665 670
Arg Cys Ile Gly Val Leu Pro Asp Gly Thr Asp Cys Asp Ala Val Leu
675 680 685
Pro Arg Asn Arg Asn Ala Ala Arg Asn Ala Glu Lys Arg Leu Arg Lys
690 695 700
His Arg Glu Ala His Asn Ala
705 710

<210> 37
<211> 574
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 37
Met Asn Glu Val Leu Pro Ile Pro Ala Val Gly Glu Asp Ala Ala Asp
1 5 10 15
Thr Ile Met Arg Gly Ser Lys Met Arg Ile Tyr Pro Ser Val Arg Gln
20 25 30
Ala Ala Thr Met Asp Leu Trp Arg Arg Arg Cys Ile Gln Leu Trp Asn
35 40 45

Leu Leu Leu Glu Leu Glu Gln Ala Ala Tyr Ser Gly Glu Asn Arg Arg

50 55 60

Thr Gln Ile Gly Trp Arg Ser Ile Trp Ala Thr Val Val Glu Asp Ser

65 70 75 80

His Ala Glu Ala Val Arg Val Ala Arg Glu Gly Lys Lys Arg Lys Asp

85 90 95

Gly Thr Phe Arg Lys Ala Pro Ser Gly Lys Glu Ile Pro Pro Leu Asp

100 105 110

Pro Ala Met Leu Ala Lys Ile Gln Arg Gln Met Asn Gly Ala Val Asp

115 120 125

Val Asp Pro Lys Thr Gly Glu Val Thr Pro Ala Gln Pro Arg Leu Phe

130 135 140

Met Trp Glu His Glu Leu Gln Lys Ile Met Ala Arg Leu Lys Gln Ala

145 150 155 160

Pro Arg Thr His Trp Ile Asp Asp Leu Pro Ser His Ala Ala Gln Ser

165 170 175

Val Val Lys Asp Leu Ile Lys Ala Leu Gln Ala Met Leu Arg Glu Arg

180 185 190

Lys Lys Arg Ala Ser Gly Ile Gly Gly Arg Asp Thr Gly Phe Pro Lys

195 200 205

Phe Lys Lys Asn Arg Tyr Ala Ala Gly Ser Val Tyr Phe Ala Asn Thr

210 215 220

Gln Leu Arg Phe Glu Ala Lys Arg Gly Lys Ala Gly Asp Pro Asp Ala

225 230 235 240

Val Arg Gly Glu Phe Ala Arg Val Lys Leu Pro Asn Gly Val Gly Trp

245 250 255

Met Glu Cys Arg Met Pro Arg His Ile Asn Ala Ala His Ala Tyr Ala

260 265 270

Gln Ala Thr Leu Met Gly Gly Arg Ile Trp Arg Gln Gly Glu Asn Trp

275 280 285

Tyr Leu Ser Cys Gln Trp Lys Met Pro Lys Pro Ala Pro Leu Pro Arg

290 295 300

Ala Gly Arg Thr Ala Ala Ile Lys Ile Ala Ala Ala Ile Pro Ile Thr

305 310 315 320

Thr Val Asp Asn Arg Gly Gln Thr Arg Glu Tyr Ala Met Pro Pro Ile

325 330 335

Asp Arg Glu Arg Ile Ala Ala His Ala Ala Ala Gly Arg Ala Gln Ser

340 345 350

Arg Ala Leu Glu Ala Arg Lys Arg Arg Ala Lys Lys Arg Glu Ala Tyr

355 360 365

Ala Lys Lys Arg His Ala Lys Lys Leu Glu Arg Gly Ile Ala Ala Lys

370 375 380

Pro Pro Gly Arg Ala Arg Ile Lys Leu Ser Pro Gly Phe Tyr Ala Ala

385 390 395 400

Ala Ala Lys Leu Ala Lys Leu Glu Ala Glu Asp Ala Asn Ala Arg Glu

405 410 415

Ala Trp Leu His Glu Ile Thr Thr Gln Ile Val Arg Asn Phe Asp Val

420 425 430

Ile Ala Val Pro Arg Met Glu Val Ala Lys Leu Met Lys Lys Pro Glu

435 440 445

Pro Pro Glu Glu Lys Glu Glu Gln Val Lys Ala Pro Trp Gln Gly Lys

450 455 460

Arg Arg Ser Leu Lys Ala Ala Arg Val Met Met Arg Arg Thr Ala Met

465 470 475 480

Ala Leu Ile Gln Thr Thr Leu Lys Tyr Lys Ala Val Asp Leu Arg Gly

485 490 495

Pro Gln Ala Tyr Glu Glu Ile Ala Pro Leu Asp Val Thr Ala Ala Ala

500 505 510

Cys Ser Gly Cys Gly Val Leu Lys Pro Glu Trp Lys Met Ala Arg Ala

515 520 525

Lys Gly Arg Glu Ile Met Arg Cys Gln Glu Pro Leu Pro Gly Gly Lys

530 535 540

Thr Cys Asn Thr Val Leu Thr Tyr Thr Arg Asn Ser Ala Arg Val Ile
545 550 555 560
Gly Arg Glu Leu Ala Val Arg Leu Ala Glu Arg Gln Lys Ala

565 570

<210> 38
<211> 400
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 38
Met Thr Thr Gln Lys Thr Tyr Asn Phe Cys Phe Tyr Asp Gln Arg Phe
1 5 10 15
Phe Glu Leu Ser Lys Glu Ala Gly Glu Val Tyr Ser Arg Ser Leu Glu
20 25 30
Glu Phe Trp Lys Ile Tyr Asp Glu Thr Gly Val Trp Leu Ser Lys Phe
35 40 45

Asp Leu Gln Lys His Met Arg Asn Lys Leu Glu Arg Lys Leu Leu His
50 55 60
Ser Asp Ser Phe Leu Gly Ala Met Gln Gln Val His Ala Asn Leu Ala
65 70 75 80
Ser Trp Lys Gln Ala Lys Lys Val Val Pro Asp Ala Cys Pro Pro Arg
85 90 95
Lys Pro Lys Phe Leu Gln Ala Ile Leu Phe Lys Lys Ser Gln Ile Lys
100 105 110

Tyr Lys Asn Gly Phe Leu Arg Leu Thr Leu Gly Thr Glu Lys Glu Phe
115 120 125
Leu Tyr Leu Lys Trp Asp Ile Asn Ile Pro Leu Pro Ile Tyr Gly Ser
130 135 140
Val Thr Tyr Ser Lys Thr Arg Gly Trp Lys Ile Asn Leu Cys Leu Glu
145 150 155 160
Thr Glu Val Glu Gln Lys Asn Leu Ser Glu Asn Lys Tyr Leu Ser Ile

165 170 175

Asp Leu Gly Val Lys Arg Val Ala Thr Ile Phe Asp Gly Glu Asn Thr
180 185 190
Ile Thr Leu Ser Gly Lys Lys Phe Met Gly Leu Met His Tyr Arg Asn
195 200 205
Lys Leu Asn Gly Lys Thr Gln Ser Arg Leu Ser His Lys Lys Lys Gly
210 215 220
Ser Asn Asn Tyr Lys Lys Ile Gln Arg Ala Lys Arg Lys Thr Thr Asp
225 230 235 240

Arg Leu Leu Asn Ile Gln Lys Glu Met Leu His Lys Tyr Ser Ser Phe
245 250 255
Ile Val Asn Tyr Ala Ile Arg Asn Asp Ile Gly Asn Ile Ile Ile Gly
260 265 270
Asp Asn Ser Ser Thr His Asp Ser Pro Asn Met Arg Gly Lys Thr Asn
275 280 285
Gln Lys Ile Ser Gln Asn Pro Glu Gln Lys Leu Lys Asn Tyr Ile Lys
290 295 300

Tyr Lys Phe Glu Ser Ile Ser Gly Arg Val Asp Ile Val Pro Glu Pro
305 310 315 320
Tyr Thr Ser Arg Lys Cys Pro His Cys Lys Asn Ile Lys Lys Ser Ser
325 330 335
Pro Lys Gly Arg Thr Tyr Lys Cys Lys Lys Cys Gly Phe Ile Phe Asp
340 345 350
Arg Asp Gly Val Gly Ala Ile Asn Ile Tyr Asn Glu Asn Val Ser Phe
355 360 365

Gly Gln Ile Ile Ser Pro Gly Arg Ile Arg Ser Leu Thr Glu Pro Ile
370 375 380
Gly Met Lys Phe His Asn Glu Ile Tyr Phe Lys Ser Tyr Val Ala Ala
385 390 395 400
<210> 39
<211> 743

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 39

Met Ser Val Arg Ser Phe Gln Ala Arg Val Glu Cys Asp Lys Gln Thr
1 5 10 15

Met Glu His Leu Trp Arg Thr His Lys Val Phe Asn Glu Arg Leu Pro
20 25 30

Glu Ile Ile Lys Ile Leu Phe Lys Met Lys Arg Gly Glu Cys Gly Gln
35 40 45

Asn Asp Lys Gln Lys Ser Leu Tyr Lys Ser Ile Ser Gln Ser Ile Leu
50 55 60

Glu Ala Asn Ala Gln Asn Ala Asp Tyr Leu Leu Asn Ser Val Ser Ile
65 70 75 80

Lys Gly Trp Lys Pro Gly Thr Ala Lys Lys Tyr Arg Asn Ala Ser Phe
85 90 95

Thr Trp Ala Asp Asp Ala Ala Lys Leu Ser Ser Gln Gly Ile His Val
100 105 110

Tyr Asp Lys Lys Gln Val Leu Gly Asp Leu Pro Gly Met Met Ser Gln
115 120 125

Met Val Cys Arg Gln Ser Val Glu Ala Ile Ser Gly His Ile Glu Leu
130 135 140

Thr Lys Lys Trp Glu Lys Glu His Asn Glu Trp Leu Lys Glu Lys Glu
145 150 155 160

Lys Trp Glu Ser Glu Asp Glu His Lys Lys Tyr Leu Asp Leu Arg Glu
165 170 175

Lys Phe Glu Gln Phe Glu Gln Ser Ile Gly Gly Lys Ile Thr Lys Arg
180 185 190

Arg Gly Arg Trp His Leu Tyr Leu Lys Trp Leu Ser Asp Asn Pro Asp
195 200 205

Phe Ala Ala Trp Arg Gly Asn Lys Ala Val Ile Asn Pro Leu Ser Glu

210 215 220

Lys Ala Gln Ile Arg Ile Asn Lys Ala Lys Pro Asn Lys Lys Asn Ser

225 230 235 240

Val Glu Arg Asp Glu Phe Phe Lys Ala Asn Pro Glu Met Lys Ala Leu

245 250 255

Asp Asn Leu His Gly Tyr Tyr Glu Arg Asn Phe Val Arg Arg Arg Lys

260 265 270

Thr Lys Lys Asn Pro Asp Gly Phe Asp His Lys Pro Thr Phe Thr Leu

275 280 285

Pro His Pro Thr Ile His Pro Arg Trp Phe Val Phe Asn Lys Pro Lys

290 295 300

Thr Asn Pro Glu Gly Tyr Arg Lys Leu Ile Leu Pro Lys Lys Ala Gly

305 310 315 320

Asp Leu Gly Ser Leu Glu Met Arg Leu Leu Thr Gly Glu Lys Asn Lys

325 330 335

Gly Asn Tyr Pro Asp Asp Trp Ile Ser Val Lys Phe Lys Ala Asp Pro

340 345 350

Arg Leu Ser Leu Ile Arg Pro Val Lys Gly Arg Arg Val Val Arg Lys

355 360 365

Gly Lys Glu Gln Gly Gln Thr Lys Glu Thr Asp Ser Tyr Glu Phe Phe

370 375 380

Asp Lys His Leu Lys Lys Trp Arg Pro Ala Lys Leu Ser Gly Val Lys

385 390 395 400

Leu Ile Phe Pro Asp Lys Thr Pro Lys Ala Ala Tyr Leu Tyr Phe Thr

405 410 415

Cys Asp Ile Pro Asp Glu Pro Leu Thr Glu Thr Ala Lys Lys Ile Gln

420 425 430

Trp Leu Glu Thr Gly Asp Val Thr Lys Lys Gly Lys Lys Arg Lys Lys

435 440 445

Lys Val Leu Pro His Gly Leu Val Ser Cys Ala Val Asp Leu Ser Met

450 455 460

Arg Arg Gly Thr Thr Gly Phe Ala Thr Leu Cys Arg Tyr Glu Asn Gly

465 470 475 480

Lys Ile His Ile Leu Arg Ser Arg Asn Leu Trp Val Gly Tyr Lys Glu

485 490 495

Gly Lys Gly Cys His Pro Tyr Arg Trp Thr Glu Gly Pro Asp Leu Gly

500 505 510

His Ile Ala Lys His Lys Arg Glu Ile Arg Ile Leu Arg Ser Lys Arg

515 520 525

Gly Lys Pro Val Lys Gly Glu Glu Ser His Ile Asp Leu Gln Lys His

530 535 540

Ile Asp Tyr Met Gly Glu Asp Arg Phe Lys Lys Ala Ala Arg Thr Ile

545 550 555 560

Val Asn Phe Ala Leu Asn Thr Glu Asn Ala Ala Ser Lys Asn Gly Phe

565 570 575

Tyr Pro Arg Ala Asp Val Leu Leu Leu Glu Asn Leu Glu Gly Leu Ile

580 585 590

Pro Asp Ala Glu Lys Glu Arg Gly Ile Asn Arg Ala Leu Ala Gly Trp

595 600 605

Asn Arg Arg His Leu Val Glu Arg Val Ile Glu Met Ala Lys Asp Ala

610 615 620

Gly Phe Lys Arg Arg Val Phe Glu Ile Pro Pro Tyr Gly Thr Ser Gln

625 630 635 640

Val Cys Ser Lys Cys Gly Ala Leu Gly Arg Arg Tyr Ser Ile Ile Arg

645 650 655

Glu Asn Asn Arg Arg Glu Ile Arg Phe Gly Tyr Val Glu Lys Leu Phe

660 665 670

Ala Cys Pro Asn Cys Gly Tyr Cys Ala Asn Ala Asp His Asn Ala Ser

675 680 685

Val Asn Leu Asn Arg Arg Phe Leu Ile Glu Asp Ser Phe Lys Ser Tyr

690 695 700

Tyr Asp Trp Lys Arg Leu Ser Glu Lys Lys Gln Lys Glu Glu Ile Glu

705 710 715 720

Thr Ile Glu Ser Lys Leu Met Asp Lys Leu Cys Ala Met His Lys Ile

725 730 735

Ser Arg Gly Ser Ile Ser Lys

740

<210> 40

<211> 769

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 40

Met His Leu Trp Arg Thr His Cys Val Phe Asn Gln Arg Leu Pro Ala

1 5 10 15

Leu Leu Lys Arg Leu Phe Ala Met Arg Arg Gly Glu Val Gly Gly Asn

20 25 30

Glu Ala Gln Arg Gln Val Tyr Gln Arg Val Ala Gln Phe Val Leu Ala

35 40 45

Arg Asp Ala Lys Asp Ser Val Asp Leu Leu Asn Ala Val Ser Leu Arg

50 55 60

Lys Arg Ser Ala Asn Ser Ala Phe Lys Lys Lys Ala Thr Ile Ser Cys

65 70 75 80

Asn Gly Gln Ala Arg Glu Val Thr Gly Glu Glu Val Phe Ala Glu Ala

85 90 95

Val Ala Leu Ala Ser Lys Gly Val Phe Ala Tyr Asp Lys Asp Asp Met

100 105 110

Arg Ala Gly Leu Pro Asp Ser Leu Phe Gln Pro Leu Thr Arg Asp Ala

115 120 125

Val Ala Cys Met Arg Ser His Glu Glu Leu Val Ala Thr Trp Lys Lys

130 135 140

Glu Tyr Arg Glu Trp Arg Asp Arg Lys Ser Glu Trp Glu Ala Glu Pro

145 150 155 160

Glu His Ala Leu Tyr Leu Asn Leu Arg Pro Lys Phe Glu Glu Gly Glu

165 170 175

Ala Ala Arg Gly Gly Arg Phe Arg Lys Arg Ala Glu Arg Asp His Ala

180 185 190

Tyr Leu Asp Trp Leu Glu Ala Asn Pro Gln Leu Ala Ala Trp Arg Arg

195 200 205

Lys Ala Pro Pro Ala Val Val Pro Ile Asp Glu Ala Gly Lys Arg Arg

210 215 220

Ile Ala Arg Ala Lys Ala Trp Lys Gln Ala Ser Val Arg Ala Glu Glu

225 230 235 240

Phe Trp Lys Arg Asn Pro Glu Leu His Ala Leu His Lys Ile His Val

245 250 255

Gln Tyr Leu Arg Glu Phe Val Arg Pro Arg Arg Thr Arg Arg Asn Lys

260 265 270

Arg Arg Glu Gly Phe Lys Gln Arg Pro Thr Phe Thr Met Pro Asp Pro

275 280 285

Val Arg His Pro Arg Trp Cys Leu Phe Asn Ala Pro Gln Thr Ser Pro

290 295 300

Gln Gly Tyr Arg Leu Leu Arg Leu Pro Gln Ser Arg Arg Thr Val Gly

305 310 315 320

Ser Val Glu Leu Arg Leu Leu Thr Gly Pro Ser Asp Gly Ala Gly Phe

325 330 335

Pro Asp Ala Trp Val Asn Val Arg Phe Lys Ala Asp Pro Arg Leu Ala

340 345 350

Gln Leu Arg Pro Val Lys Val Pro Arg Thr Val Thr Arg Gly Lys Asn

355 360 365

Lys Gly Ala Lys Val Glu Ala Asp Gly Phe Arg Tyr Tyr Asp Asp Gln

370 375 380

Leu Leu Ile Glu Arg Asp Ala Gln Val Ser Gly Val Lys Leu Leu Phe

385 390 395 400

Arg Asp Ile Arg Met Ala Pro Phe Ala Asp Lys Pro Ile Glu Asp Arg

405 410 415

Leu Leu Ser Ala Thr Pro Tyr Leu Val Phe Ala Val Glu Ile Lys Asp

420 425 430

Glu Ala Arg Thr Glu Arg Ala Lys Ala Ile Arg Phe Asp Glu Thr Ser

435 440 445

Glu Leu Thr Lys Ser Gly Lys Lys Arg Lys Thr Leu Pro Ala Gly Leu

450 455 460

Val Ser Val Ala Val Asp Leu Asp Thr Arg Gly Val Gly Phe Leu Thr

465 470 475 480

Arg Ala Val Ile Gly Val Pro Glu Ile Gln Gln Thr His His Gly Val

485 490 495

Arg Leu Leu Gln Ser Arg Tyr Val Ala Val Gly Gln Val Glu Ala Arg

500 505 510

Ala Ser Gly Glu Ala Glu Trp Ser Pro Gly Pro Asp Leu Ala His Ile

515 520 525

Ala Arg His Lys Arg Glu Ile Arg Arg Leu Arg Gln Leu Arg Gly Lys

530 535 540

Pro Val Lys Gly Glu Arg Ser His Val Arg Leu Gln Ala His Ile Asp

545 550 555 560

Arg Met Gly Glu Asp Arg Phe Lys Lys Ala Ala Arg Lys Ile Val Asn

565 570 575

Glu Ala Leu Arg Gly Ser Asn Pro Ala Ala Gly Asp Pro Tyr Thr Arg

580 585 590

Ala Asp Val Leu Leu Tyr Glu Ser Leu Glu Thr Leu Leu Pro Asp Ala

595 600 605

Glu Arg Glu Arg Gly Ile Asn Arg Ala Leu Leu Arg Trp Asn Arg Ala

610 615 620

Lys Leu Ile Glu His Leu Lys Arg Met Cys Asp Asp Ala Gly Ile Arg

625 630 635 640

His Phe Pro Val Ser Pro Phe Gly Thr Ser Gln Val Cys Ser Lys Cys

645 650 655

Gly Ala Leu Gly Arg Arg Tyr Ser Leu Ala Arg Glu Asn Gly Arg Ala
660 665 670
Val Ile Arg Phe Gly Trp Val Glu Arg Leu Phe Ala Cys Pro Asn Pro

675 680 685
Glu Cys Pro Gly Arg Arg Pro Asp Arg Pro Asp Arg Pro Phe Thr Cys
690 695 700
Asn Ser Asp His Asn Ala Ser Val Asn Leu His Arg Val Phe Ala Leu
705 710 715 720
Gly Asp Gln Ala Val Ala Ala Phe Arg Ala Leu Ala Pro Arg Asp Ser
725 730 735
Pro Ala Arg Thr Leu Ala Val Lys Arg Val Glu Asp Thr Leu Arg Pro

740 745 750
Gln Leu Met Arg Val His Lys Leu Ala Asp Ala Gly Val Asp Ser Pro
755 760 765
Phe

<210> 41
<211> 666
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 41
Met Ala Thr Leu Val Tyr Arg Tyr Gly Val Arg Ala His Gly Ser Ala
1 5 10 15
Arg Gln Gln Asp Ala Val Val Ser Asp Pro Ala Met Leu Glu Gln Leu

20 25 30
Arg Leu Gly His Glu Leu Arg Asn Ala Leu Val Gly Val Gln His Arg
35 40 45
Tyr Glu Asp Gly Lys Arg Ala Val Trp Ser Gly Phe Ala Ser Val Ala
50 55 60
Ala Ala Asp His Arg Val Thr Thr Gly Glu Thr Ala Val Ala Glu Leu
65 70 75 80

Glu Lys Gln Ala Arg Ala Glu His Ser Ala Asp Arg Thr Ala Ala Thr

85 90 95

Arg Gln Gly Thr Ala Glu Ser Leu Lys Ala Ala Arg Ala Ala Val Lys

100 105 110

Gln Ala Arg Ala Asp Arg Lys Ala Ala Met Ala Ala Val Ala Glu Gln

115 120 125

Ala Lys Pro Lys Ile Gln Ala Leu Gly Asp Asp Arg Asp Ala Glu Ile

130 135 140

Lys Asp Leu Tyr Arg Arg Phe Cys Gln Asp Gly Val Leu Leu Pro Arg

145 150 155 160

Cys Gly Arg Cys Ala Gly Asp Leu Arg Ser Asp Gly Asp Cys Thr Asp

165 170 175

Cys Gly Ala Ala His Glu Pro Arg Lys Leu Tyr Trp Ala Thr Tyr Asn

180 185 190

Ala Ile Arg Glu Asp His Gln Thr Ala Val Lys Leu Val Glu Ala Lys

195 200 205

Arg Lys Ala Gly Gln Pro Ala Arg Leu Arg Phe Arg Arg Trp Thr Gly

210 215 220

Asp Gly Thr Leu Thr Val Gln Leu Gln Arg Met His Gly Pro Ala Cys

225 230 235 240

Arg Cys Val Thr Cys Ala Glu Lys Leu Thr Arg Arg Ala Arg Lys Thr

245 250 255

Asp Pro Gln Ala Pro Ala Val Ala Ala Asp Pro Ala Tyr Pro Pro Thr

260 265 270

Asp Pro Pro Arg Asp Pro Ala Leu Leu Ala Ser Gly Gln Gly Lys Trp

275 280 285

Arg Asn Val Leu Gln Leu Gly Thr Trp Ile Pro Pro Gly Glu Trp Ser

290 295 300

Ala Met Ser Arg Ala Glu Arg Arg Arg Val Gly Arg Ser His Ile Gly

305 310 315 320

Trp Gln Leu Gly Gly Gly Arg Gln Leu Thr Leu Pro Val Gln Leu His

325 330 335

Arg Gln Met Pro Ala Asp Ala Asp Val Ala Met Ala Gln Leu Thr Arg

340 345 350

Val Arg Val Gly Gly Arg His Arg Met Ser Val Ala Leu Thr Ala Lys

355 360 365

Leu Pro Asp Pro Pro Gln Val Gln Gly Leu Pro Pro Val Ala Leu His

370 375 380

Leu Gly Trp Arg Gln Arg Pro Asp Gly Ser Leu Arg Val Ala Thr Trp

385 390 395 400

Ala Cys Pro Gln Pro Leu Asp Leu Pro Pro Ala Val Ala Asp Val Val

405 410 415

Val Ser His Gly Gly Arg Trp Gly Glu Val Ile Met Pro Ala Arg Trp

420 425 430

Leu Ala Asp Ala Glu Val Pro Pro Arg Leu Leu Gly Arg Arg Asp Lys

435 440 445

Ala Met Glu Pro Val Leu Glu Ala Leu Ala Asp Trp Leu Glu Ala His

450 455 460

Thr Glu Ala Cys Thr Ala Arg Met Thr Pro Ala Leu Val Arg Arg Trp

465 470 475 480

Arg Ser Gln Gly Arg Leu Ala Gly Leu Thr Asn Arg Trp Arg Gly Gln

485 490 495

Pro Pro Thr Gly Ser Ala Glu Ile Leu Thr Tyr Leu Glu Ala Trp Arg

500 505 510

Ile Gln Asp Lys Leu Leu Trp Glu Arg Glu Ser His Leu Arg Arg Arg

515 520 525

Leu Ala Ala Arg Arg Asp Asp Ala Trp Arg Arg Val Ala Ser Trp Leu

530 535 540

Ala Arg His Ala Gly Val Leu Val Val Asp Asp Ala Asp Ile Ala Glu

545 550 555 560

Leu Arg Arg Arg Asp Asp Pro Ala Asp Thr Asp Pro Thr Met Pro Ala

565 570 575

Ser Ala Ala Gln Ala Ala Arg Ala Arg Ala Ala Leu Ala Ala Pro Gly
580 585 590
Arg Leu Arg His Leu Ala Thr Ile Thr Ala Thr Arg Asp Gly Leu Gly

595 600 605
Val His Thr Val Ala Ser Ala Gly Leu Thr Arg Leu His Arg Lys Cys
610 615 620
Gly His Gln Ala Gln Pro Asp Pro Arg Tyr Ala Ala Ser Ala Val Val
625 630 635 640
Thr Cys Pro Gly Cys Gly Asn Gly Tyr Asp Gln Asp Tyr Asn Ala Ala
645 650 655
Met Leu Met Leu Asp Arg Gln Gln Gln Pro
660 665

<210> 42
<211> 564
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 42
Met Ser Arg Val Glu Leu His Arg Ala Tyr Lys Phe Arg Leu Tyr Pro
1 5 10 15
Thr Pro Ala Gln Val Ala Glu Leu Ala Glu Trp Glu Arg Gln Leu Arg
20 25 30
Arg Leu Tyr Asn Leu Ala His Ser Gln Arg Leu Ala Ala Met Gln Arg
35 40 45
His Val Arg Pro Lys Ser Pro Gly Val Leu Lys Ser Glu Cys Leu Ser

50 55 60
Cys Gly Ala Val Ala Val Ala Glu Ile Gly Thr Asp Gly Lys Ala Lys
65 70 75 80
Lys Thr Val Lys His Ala Val Gly Cys Ser Val Leu Glu Cys Arg Ser
85 90 95
Cys Gly Gly Ser Pro Asp Ala Glu Gly Arg Thr Ala His Thr Ala Ala
100 105 110

Cys Ser Phe Val Asp Tyr Tyr Arg Gln Gly Arg Glu Met Thr Gln Leu

115 120 125

Leu Glu Glu Asp Asp Gln Leu Ala Arg Val Val Cys Ser Ala Arg Gln

130 135 140

Glu Thr Leu Arg Asp Leu Glu Lys Ala Trp Gln Arg Trp His Lys Met

145 150 155 160

Pro Gly Phe Gly Lys Pro His Phe Lys Lys Arg Ile Asp Ser Cys Arg

165 170 175

Ile Tyr Phe Ser Thr Pro Lys Ser Trp Ala Val Asp Leu Gly Tyr Leu

180 185 190

Ser Phe Thr Gly Val Ala Ser Ser Val Gly Arg Ile Lys Ile Arg Gln

195 200 205

Asp Arg Val Trp Pro Gly Asp Ala Lys Phe Ser Ser Cys His Val Val

210 215 220

Arg Asp Val Asp Glu Trp Tyr Ala Val Phe Pro Leu Thr Phe Thr Lys

225 230 235 240

Glu Ile Glu Lys Pro Lys Gly Gly Ala Val Gly Ile Asn Arg Gly Ala

245 250 255

Val His Ala Ile Ala Asp Ser Thr Gly Arg Val Val Asp Ser Pro Lys

260 265 270

Phe Tyr Ala Arg Ser Leu Gly Val Ile Arg His Arg Ala Arg Leu Leu

275 280 285

Asp Arg Lys Val Pro Phe Gly Arg Ala Val Lys Pro Ser Pro Thr Lys

290 295 300

Tyr His Gly Leu Pro Lys Ala Asp Ile Asp Ala Ala Ala Ala Arg Val

305 310 315 320

Asn Ala Ser Pro Gly Arg Leu Val Tyr Glu Ala Arg Ala Arg Gly Ser

325 330 335

Ile Ala Ala Ala Glu Ala His Leu Ala Ala Leu Val Leu Pro Ala Pro

340 345 350

Arg Gln Thr Ser Gln Leu Pro Ser Glu Gly Arg Asn Arg Glu Arg Ala

355 360 365

Arg Arg Phe Leu Ala Leu Ala His Gln Arg Val Arg Arg Gln Arg Glu

370 375 380

Trp Phe Leu His Asn Glu Ser Ala His Tyr Ala Gln Ser Tyr Thr Lys

385 390 395 400

Ile Ala Ile Glu Asp Trp Ser Thr Lys Glu Met Thr Ser Ser Glu Pro

405 410 415

Arg Asp Ala Glu Glu Met Lys Arg Val Thr Arg Ala Arg Asn Arg Ser

420 425 430

Ile Leu Asp Val Gly Trp Tyr Glu Leu Gly Arg Gln Ile Ala Tyr Lys

435 440 445

Ser Glu Ala Thr Gly Ala Glu Phe Ala Lys Val Asp Pro Gly Leu Arg

450 455 460

Glu Thr Glu Thr His Val Pro Glu Ala Ile Val Arg Glu Arg Asp Val

465 470 475 480

Asp Val Ser Gly Met Leu Arg Gly Glu Ala Gly Ile Ser Gly Thr Cys

485 490 495

Ser Arg Cys Gly Gly Leu Leu Arg Ala Ser Ala Ser Gly His Ala Asp

500 505 510

Ala Glu Cys Glu Val Cys Leu His Val Glu Val Gly Asp Val Asn Ala

515 520 525

Ala Val Asn Val Leu Lys Arg Ala Met Phe Pro Gly Ala Ala Pro Pro

530 535 540

Ser Lys Glu Lys Ala Lys Val Thr Ile Gly Ile Lys Gly Arg Lys Lys

545 550 555 560

Lys Arg Ala Ala

<210> 43

<211> 565

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 43

Met Ser Arg Val Glu Leu His Arg Ala Tyr Lys Phe Arg Leu Tyr Pro
1 5 10 15

Thr Pro Val Gln Val Ala Glu Leu Ser Glu Trp Glu Arg Gln Leu Arg
20 25 30

Arg Leu Tyr Asn Leu Gly His Glu Gln Arg Leu Leu Thr Leu Thr Arg
35 40 45

His Leu Arg Pro Lys Ser Pro Gly Val Leu Lys Gly Glu Cys Leu Ser

50 55 60

Cys Asp Ser Thr Gln Val Gln Glu Val Gly Ala Asp Gly Arg Pro Lys
65 70 75 80

Thr Thr Val Arg His Ala Glu Gln Cys Pro Thr Leu Ala Cys Arg Ser
85 90 95

Cys Gly Ala Leu Arg Asp Ala Glu Gly Arg Thr Ala His Thr Val Ala
100 105 110

Cys Ala Phe Val Asp Tyr Tyr Arg Gln Gly Arg Glu Met Thr Glu Leu

115 120 125

Leu Ala Ala Asp Asp Gln Leu Ala Arg Val Val Cys Ser Ala Arg Gln
130 135 140

Glu Val Leu Arg Asp Leu Asp Lys Ala Trp Gln Arg Trp Arg Lys Met
145 150 155 160

Pro Gly Phe Gly Lys Pro Arg Phe Lys Arg Arg Thr Asp Ser Cys Arg
165 170 175

Ile Tyr Phe Ser Thr Pro Lys Ala Trp Lys Leu Glu Gly Gly His Leu

180 185 190

Ser Phe Thr Gly Ala Ala Thr Thr Val Gly Ala Ile Lys Met Arg Gln
195 200 205

Asp Arg Asn Trp Pro Ala Ser Val Gln Phe Ser Ser Cys His Val Val
210 215 220

Arg Asp Val Asp Glu Trp Tyr Ala Val Phe Pro Leu Thr Phe Val Ala
225 230 235 240

Glu Val Ala Arg Pro Lys Gly Gly Ala Val Gly Ile Asn Arg Gly Ala

245 250 255

Val His Ala Ile Ala Asp Ser Thr Gly Arg Val Val Asp Ser Pro Arg

260 265 270

Tyr Tyr Ala Arg Ala Leu Gly Val Ile Arg His Arg Ala Arg Leu Phe

275 280 285

Asp Arg Lys Val Pro Ser Gly His Ala Val Lys Pro Ser Pro Thr Lys

290 295 300

Tyr Arg Gly Leu Ser Ala Ile Glu Val Asp Arg Val Ala Arg Ala Thr

305 310 315 320

Gly Phe Thr Pro Gly Arg Val Val Thr Glu Ala Leu Asn Arg Gly Gly

325 330 335

Val Ala Tyr Ala Glu Cys Ala Leu Ala Ala Ile Ala Val Leu Gly His

340 345 350

Gly Pro Glu Arg Pro Leu Thr Ser Asp Gly Arg Asn Arg Glu Lys Ala

355 360 365

Arg Lys Phe Leu Ala Leu Ala His Gln Arg Val Arg Arg Gln Arg Glu

370 375 380

Trp Phe Leu His Asn Glu Ser Ala His Tyr Ala Arg Thr Tyr Ser Lys

385 390 395 400

Ile Ala Ile Glu Asp Trp Ser Thr Lys Glu Met Thr Ala Ser Glu Pro

405 410 415

Gln Gly Glu Glu Thr Arg Arg Val Thr Arg Ser Arg Asn Arg Ser Ile

420 425 430

Leu Asp Val Gly Trp Tyr Glu Leu Gly Arg Gln Leu Ala Tyr Lys Thr

435 440 445

Glu Ala Thr Gly Ala Glu Phe Ala Gln Val Asp Pro Gly Leu Lys Glu

450 455 460

Thr Glu Thr Asn Val Pro Lys Ala Ile Ala Asp Ala Arg Asp Val Asp

465 470 475 480

Val Ser Gly Met Leu Arg Gly Glu Ala Gly Ile Ser Gly Thr Cys Ser

485 490 495

Lys Cys Gly Gly Leu Leu Arg Ala Pro Ala Ser Gly His Ala Asp Ala

500 505 510

Glu Cys Glu Ile Cys Leu Asn Val Glu Val Gly Asp Val Asn Ala Ala

515 520 525

Val Asn Val Leu Lys Arg Ala Met Phe Pro Gly Asp Ala Pro Pro Ala

530 535 540

Ser Gly Glu Lys Pro Lys Val Ser Ile Gly Ile Lys Gly Arg Gln Lys

545 550 555 560

Lys Lys Lys Ala Ala

565

<210

> 44

<211> 499

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 44

Met Glu Ala Ile Ala Thr Gly Met Ser Pro Glu Arg Arg Val Glu Leu

1 5 10 15

Gly Ile Leu Pro Gly Ser Val Glu Leu Lys Arg Ala Tyr Lys Phe Arg

20 25 30

Leu Tyr Pro Met Lys Val Gln Gln Ala Glu Leu Ser Glu Trp Glu Arg

35 40 45

Gln Leu Arg Arg Leu Tyr Asn Leu Ala His Glu Gln Arg Leu Ala Ala

50 55 60

Leu Leu Arg Tyr Arg Asp Trp Asp Phe Gln Lys Gly Ala Cys Pro Ser

65 70 75 80

Cys Arg Val Ala Val Pro Gly Val His Thr Ala Ala Cys Asp His Val

85 90 95

Asp Tyr Phe Arg Gln Ala Arg Glu Met Thr Gln Leu Leu Glu Val Asp

100 105 110

Ala Gln Leu Ser Arg Val Ile Cys Cys Ala Arg Gln Glu Val Leu Arg

115 120 125

Asp Leu Asp Lys Ala Trp Gln Arg Trp Arg Lys Lys Leu Gly Gly Arg

130 135 140

Pro Arg Phe Lys Arg Arg Thr Asp Ser Cys Arg Ile Tyr Leu Ser Thr

145 150 155 160

Pro Lys His Trp Glu Ile Ala Gly Arg Tyr Leu Arg Leu Ser Gly Leu

165 170 175

Ala Ser Ser Val Gly Glu Ile Arg Ile Glu Gln Asp Arg Ala Phe Pro

180 185 190

Glu Gly Ala Leu Leu Ser Ser Cys Ser Ile Val Arg Asp Val Asp Glu

195 200 205

Trp Tyr Ala Cys Leu Pro Leu Thr Phe Thr Gln Pro Ile Glu Arg Ala

210 215 220

Pro His Arg Ser Val Gly Leu Asn Arg Gly Val Val His Ala Leu Ala

225 230 235 240

Asp Ser Asp Gly Arg Val Val Asp Ser Pro Lys Phe Phe Glu Arg Ala

245 250 255

Leu Ala Thr Val Gln Lys Arg Ser Arg Asp Leu Ala Arg Lys Val Ser

260 265 270

Gly Ser Arg Asn Ala His Lys Ala Arg Ile Lys Leu Ala Lys Ala His

275 280 285

Gln Arg Val Arg Arg Gln Arg Ala Ala Phe Leu His Gln Glu Ser Ala

290 295 300

Tyr Tyr Ser Lys Gly Phe Asp Leu Val Ala Leu Glu Asp Met Ser Val

305 310 315 320

Arg Lys Met Thr Ala Thr Ala Gly Glu Ala Pro Glu Met Gly Arg Gly

325 330 335

Ala Gln Arg Asp Leu Asn Arg Gly Ile Leu Asp Val Gly Trp Tyr Glu

340 345 350

Leu Ala Arg Gln Ile Asp Tyr Lys Arg Leu Ala His Gly Gly Glu Leu

355 360 365

Leu Arg Val Asp Pro Gly Gln Thr Thr Pro Leu Ala Cys Val Thr Glu

370 375 380

Glu Gln Pro Ala Arg Gly Ile Ser Ser Ala Cys Ala Val Cys Gly Ile

385 390 395 400

Pro Leu Ala Arg Pro Ala Ser Gly Asn Ala Arg Met Arg Cys Thr Ala

405 410 415

Cys Gly Ser Ser Gln Val Gly Asp Val Asn Ala Ala Glu Asn Val Leu

420 425 430

Thr Arg Ala Leu Ser Ser Ala Pro Ser Gly Pro Lys Ser Pro Lys Ala

435 440 445

Ser Ile Lys Ile Lys Gly Arg Gln Lys Arg Leu Gly Thr Pro Ala Asn

450 455 460

Arg Ala Gly Glu Ala Ser Gly Gly Asp Pro Pro Val Arg Gly Pro Val

465 470 475 480

Glu Gly Gly Thr Leu Ala Tyr Val Val Glu Pro Val Ser Glu Ser Gln

485 490 495

Ser Asp Thr

<210> 45

<211> 560

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 45

Met Thr Val Arg Thr Tyr Lys Tyr Arg Ala Tyr Pro Thr Pro Glu Gln

1 5 10 15

Ala Glu Ala Leu Thr Ser Trp Leu Arg Phe Ala Ser Gln Leu Tyr Asn

20 25 30

Ala Ala Leu Glu His Arg Lys Asn Ala Trp Gly Arg His Asp Ala His

35 40 45

Gly Arg Gly Phe Arg Phe Trp Asp Gly Asp Ala Ala Pro Arg Lys Lys

50 55 60

Ser Asp Pro Pro Gly Arg Trp Val Tyr Arg Gly Gly Gly Gly Ala His

65 70 75 80

Ile Ser Lys Asn Asp Gln Gly Lys Leu Leu Thr Glu Phe Arg Arg Glu

85 90 95

His Ala Glu Leu Leu Pro Pro Gly Met Pro Ala Leu Val Gln His Glu

100 105 110

Val Leu Ala Arg Leu Glu Arg Ser Met Ala Ala Phe Phe Gln Arg Ala

115 120 125

Thr Lys Gly Gln Lys Ala Gly Tyr Pro Arg Trp Arg Ser Glu His Arg

130 135 140

Tyr Asp Ser Leu Thr Phe Gly Leu Thr Ser Pro Ser Lys Glu Arg Phe

145 150 155 160

Asp Pro Glu Thr Gly Glu Ser Leu Gly Arg Gly Lys Thr Val Gly Ala

165 170 175

Gly Thr Tyr His Asn Gly Asp Leu Arg Leu Thr Gly Leu Gly Glu Leu

180 185 190

Arg Ile Leu Glu His Arg Arg Ile Pro Met Gly Ala Ile Pro Lys Ser

195 200 205

Val Ile Val Arg Arg Ser Gly Lys Arg Trp Phe Val Ser Ile Ala Met

210 215 220

Glu Met Pro Ser Val Glu Pro Ala Ala Ser Gly Arg Pro Ala Val Gly

225 230 235 240

Leu Asp Met Gly Val Val Thr Trp Gly Thr Ala Phe Thr Ala Asp Thr

245 250 255

Ser Ala Ala Ala Ala Leu Val Ala Asp Leu Arg Arg Met Ala Thr Asp

260 265 270

Pro Ser Asp Cys Arg Arg Leu Glu Glu Leu Glu Arg Glu Ala Ala Gln

275 280 285

Leu Ser Glu Val Leu Ala His Cys Arg Ala Arg Gly Leu Asp Pro Ala

290 295 300

Arg Pro Arg Arg Cys Pro Lys Glu Leu Thr Lys Leu Tyr Arg Arg Ser

305 310 315 320

Leu His Arg Leu Gly Glu Leu Asp Arg Ala Cys Ala Arg Ile Arg Arg

325 330 335

Arg Leu Gln Ala Ala His Asp Ile Ala Glu Pro Val Pro Asp Glu Ala

340 345 350

Gly Ser Ala Val Leu Ile Glu Gly Ser Asn Ala Gly Met Arg His Ala

355 360 365

Arg Arg Val Ala Arg Thr Gln Arg Arg Val Ala Arg Arg Thr Arg Ala

370 375 380

Gly His Ala His Ser Asn Arg Arg Lys Lys Ala Val Gln Ala Tyr Ala

385 390 395 400

Arg Ala Lys Glu Arg Glu Arg Ser Ala Arg Gly Asp His Arg His Lys

405 410 415

Val Ser Arg Ala Leu Val Arg Gln Phe Glu Glu Ile Ser Val Glu Ala

420 425 430

Leu Asp Ile Lys Gln Leu Thr Val Ala Pro Glu His Asn Pro Asp Pro

435 440 445

Gln Pro Asp Leu Pro Ala His Val Gln Arg Arg Arg Asn Arg Gly Glu

450 455 460

Leu Asp Ala Ala Trp Gly Ala Phe Phe Ala Ala Leu Asp Tyr Lys Ala

465 470 475 480

Ala Asp Ala Gly Gly Arg Val Ala Arg Lys Pro Ala Pro His Thr Thr

485 490 495

Gln Glu Cys Ala Arg Cys Gly Thr Leu Val Pro Lys Pro Ile Ser Leu

500 505 510

Arg Val His Arg Cys Pro Ala Cys Gly Tyr Thr Ala Pro Arg Thr Val

515 520 525

Asn Ser Ala Arg Asn Val Leu Gln Arg Pro Leu Glu Glu Pro Gly Arg

530 535 540

Ala Gly Pro Ser Gly Ala Asn Gly Arg Gly Val Pro His Ala Val Ala
545 550 555 560

<210> 46

<211> 404

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 46

Met Asn Cys Arg Tyr Arg Tyr Arg Ile Tyr Pro Thr Pro Gly Gln Arg
1 5 10 15

Gln Ser Leu Ala Arg Leu Phe Gly Cys Val Arg Val Val Trp Asn Asp
20 25 30

Ala Leu Phe Leu Cys Arg Gln Ser Glu Lys Leu Pro Lys Asn Ser Glu
35 40 45

Leu Gln Lys Leu Cys Ile Thr Gln Ala Lys Lys Thr Glu Ala Arg Gly
50 55 60

Trp Leu Gly Gln Val Ser Ala Ile Pro Leu Gln Gln Ser Val Ala Asp
65 70 75 80

Leu Gly Val Ala Phe Lys Asn Phe Phe Gln Ser Arg Ser Gly Lys Arg
85 90 95

Lys Gly Lys Lys Val Asn Pro Pro Arg Val Lys Arg Arg Asn Asn Arg
100 105 110

Gln Gly Ala Arg Phe Thr Arg Gly Gly Phe Lys Val Lys Thr Ser Lys
115 120 125

Val Tyr Leu Ala Arg Ile Gly Asp Ile Lys Ile Lys Trp Ser Arg Pro
130 135 140

Leu Pro Ser Glu Pro Ser Ser Val Thr Val Ile Lys Asp Cys Ala Gly
145 150 155 160

Gln Tyr Phe Leu Ser Phe Val Val Glu Val Lys Pro Glu Ile Lys Pro
165 170 175

Pro Lys Asn Pro Ser Ile Gly Ile Asp Leu Gly Leu Lys Thr Phe Ala

180 185 190

Ser Cys Ser Asn Gly Glu Lys Ile Asp Ser Pro Asp Tyr Ser Arg Leu
195 200 205

Tyr Arg Lys Leu Lys Arg Cys Gln Arg Arg Leu Ala Lys Arg Gln Arg
210 215 220

Gly Ser Lys Arg Arg Glu Arg Met Arg Val Lys Val Ala Lys Leu Asn
225 230 235 240

Ala Gln Ile Arg Asp Lys Arg Lys Asp Phe Leu His Lys Leu Ser Thr
245 250 255

Lys Val Val Asn Glu Asn Gln Val Ile Ala Leu Glu Asp Leu Asn Val
260 265 270

Gly Gly Met Leu Lys Asn Arg Lys Leu Ser Arg Ala Ile Ser Gln Ala
275 280 285

Gly Trp Tyr Glu Phe Arg Ser Leu Cys Glu Gly Lys Ala Glu Lys His
290 295 300

Asn Arg Asp Phe Arg Val Ile Ser Arg Trp Glu Pro Thr Ser Gln Val
305 310 315 320

Cys Ser Glu Cys Gly Tyr Arg Trp Gly Lys Ile Asp Leu Ser Val Arg
325 330 335

Ser Ile Val Cys Ile Asn Cys Gly Val Glu His Asp Arg Asp Asp Asn
340 345 350

Ala Ser Val Asn Ile Glu Gln Ala Gly Leu Lys Val Gly Val Gly His
355 360 365

Thr His Asp Ser Lys Arg Thr Gly Ser Ala Cys Lys Thr Ser Asn Gly
370 375 380

Ala Val Cys Val Glu Pro Ser Thr His Arg Glu Tyr Val Gln Leu Thr
385 390 395 400

Leu Phe Asp Trp

<210> 47

<211> 392

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 47

Met Lys Ser Arg Trp Thr Phe Arg Cys Tyr Pro Thr Pro Glu Gln Glu
1 5 10 15

Gln His Leu Ala Arg Thr Phe Gly Cys Val Arg Phe Val Trp Asn Trp
20 25 30

Ala Leu Arg Ala Arg Thr Asp Ala Phe Arg Ala Gly Glu Arg Ile Gly
35 40 45

Tyr Pro Ala Thr Asp Lys Ala Leu Thr Leu Leu Lys Gln Gln Pro Glu
50 55 60

Thr Val Trp Leu Asn Glu Val Ser Ser Val Cys Leu Gln Gln Ala Leu
65 70 75 80

Arg Asp Leu Gln Val Ala Phe Ser Asn Phe Phe Asp Lys Arg Ala Ala
85 90 95

His Pro Ser Phe Lys Arg Lys Glu Ala Arg Gln Ser Ala Asn Tyr Thr
100 105 110

Glu Arg Gly Phe Ser Phe Asp His Glu Arg Arg Ile Leu Lys Leu Ala
115 120 125

Lys Ile Gly Ala Ile Lys Val Lys Trp Ser Arg Lys Ala Ile Pro His
130 135 140

Pro Ser Ser Ile Arg Leu Ile Arg Thr Ala Ser Gly Lys Tyr Phe Val
145 150 155 160

Ser Leu Val Val Glu Thr Gln Pro Ala Pro Met Pro Glu Thr Gly Glu
165 170 175

Ser Val Gly Val Asp Phe Gly Val Ala Arg Leu Ala Thr Leu Ser Asn
180 185 190

Gly Glu Arg Ile Ser Asn Pro Lys His Gly Ala Lys Trp Gln Arg Arg
195 200 205

Leu Ala Phe Tyr Gln Lys Arg Leu Ala Arg Ala Thr Lys Gly Ser Lys

210 215 220

Arg Arg Met Arg Ile Lys Arg His Val Ala Arg Ile His Glu Lys Ile
225 230 235 240
Gly Asn Ser Arg Ser Asp Thr Leu His Lys Leu Ser Thr Asp Leu Val
245 250 255
Thr Arg Phe Asp Leu Ile Cys Val Glu Asp Leu Asn Leu Arg Gly Met
260 265 270
Val Lys Asn His Ser Leu Ala Arg Ser Leu His Asp Ala Ser Ile Gly
275 280 285

Ser Ala Ile Arg Met Ile Glu Glu Lys Ala Glu Arg Tyr Gly Lys Asn
290 295 300
Val Val Lys Ile Asp Arg Trp Phe Pro Ser Ser Lys Thr Cys Ser Asp
305 310 315 320
Cys Gly His Ile Val Glu Gln Leu Pro Leu Asn Val Arg Glu Trp Thr
325 330 335
Cys Pro Glu Cys Gly Thr Thr His Asp Arg Asp Ala Asn Ala Ala Ala
340 345 350

Asn Ile Leu Ala Val Gly Gln Thr Val Ser Ala His Gly Gly Thr Val
355 360 365
Arg Arg Ser Arg Ala Lys Ala Ser Glu Arg Lys Ser Gln Arg Ser Ala
370 375 380
Asn Arg Gln Gly Val Asn Arg Ala
385 390

<210> 48
<211> 10
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 48
Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
1 5 10

<210> 49

<211> 10

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 49

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala

1 5 10

<210> 50

<211> 10

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 50

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala

1 5 10

<210> 51

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 51

gttgcattcc ttcattcgtc tattcgggtt ctgcaac 37

<210> 52

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 52

gttgcattcc ttcattcgtc tatccgggtt ctgcaag 37

<210> 53

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 53

gttgcagaac ccgaatagac gaatgaagga atgcaac 37

<210> 54

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 54

ctatcatatt cagaacaaag ggattaagga atgcaac 37

<210> 55

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 55

ctttcatact cagaacaaag ggattaagga atgcaac 37

<210> 56

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 56

gtctacaact cattgataga aatcaatgag ttagaca 37

<210> 57

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 57

gttataaagg cggggatcgc gaccgagcga ttgaaag 37

<210> 58

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 58

gttgcattcc ttaattcatt ttctcaatat cggaaac 37

<210> 59

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 59

gttgcagaaa tagaataaag gaattaagga atgcaac 37

<210> 60

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 60

ctttcatact cagaacaaag ggattaagga atgcaac 37

<210> 61

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 61

atttcatact cagaacaaag ggattaagga atgcaac 37

<210> 62

<211> 38

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 62

gtttcagcgc acgaattaac gagatgagag atgcaact 38

<210> 63

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 63

cttgcagaag ctgaatagac gaatcaagga atgcaac 37

<210> 64

<211> 39

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 64

cacttgcagg ccttgaatag aggagttaag gaatgcaac 39

<210> 65

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 65

gtctccatga ctgaaaagtc gtggccgaat tgaaac 36

<210> 66

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 66

gttgcagcgc ccgaactgac gagacgagag atgcaac 37

<210> 67

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 67

gttgcgcgaa tagaataaag gaattaagga atgcaac 37

<210> 68

<211> 38

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 68

agttgcattc cttaatccct ctgttcagtt tgtgcaat 38

<210> 69

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 69

gttgcattcc tagtttctct aattagcact gtgcaac 37

<210> 70

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 70

gttgcggcgc gcgaataaac gagactagga atgcaac 37

<210> 71

<211> 41

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 71

actagttgca ttccttaatc cctttgttct gaatatgcta g 41

<210> 72

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 72

ctttcatatt cagaacaaag ggattaagga atgcaac 37

<210> 73

<211> 39

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 73

gttgcagtcc ttaaccccta gtttctgaat atgaaagat 39

<210> 74

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 74

gttgcagccc ccgaactaac gagatgagag atgcaac 37

<210> 75

<211> 38

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 75

cttgcagaac aatcatatat gactaatcag actgcaac 38

<210> 76

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 76

gttgcactca ccggtgctca cgacgtaggg atgcaac 37

<210> 77

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 77

gtccctactc gctagggaaa ctaattgaat ggaaac 36

<210> 78

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 78

gttgcattcg ggtgcaaaac agggagtaga gtgtaac 37

<210> 79

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 79

cttccaaact cgagccagtg gggagagaag tggca 35

<210> 80

<211> 38

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 80

cctgtagacc ggtctcattc tgagaggggt atgcaact 38

<210> 81

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 81

gtctcgagac cctacagatt ttggagaggg gtgggac 37

<210> 82

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 82

gtcccacccc tctccaaaat ctgtagggtc tcgagac 37

<210> 83

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 83

gtagcaggac tctcctcgag agaaacaggg gtatgct 37

<210> 84

<211> 40

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 84

gtacaatacc tctcctttaa gagagggagg ggtacgctac 40

<210> 85

<211> 31

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 85

ccccctcgtt tccttcaggg gattcctttc c 31

<210> 86

<211> 28

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 86

ggttcccccg ggcgcgggtg gggtggcg 28

<210> 87

<211> 28

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 87

ggctgctccg ggtgcgcgtg gagcgagg 28

<210> 88

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 88

gttttatacc ctttagaatt taaactgtct aaaag 35

<210> 89

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 89

attgcaccgg ccaacgcaaa tctgattgat ggacac 36

<210> 90

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 90

gccgcagcgg ccgacgcggc cctgatcgat ggacac 36

<210> 91

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 91

gtcgaaatgc ccgcgcgggg gcgtcgtacc cgcgac 36

<210> 92

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 92

ggctagcccg tgcgcgcagg gacgagtgg 29

<210> 93

<211> 24

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 93

gcccgtgcgc gcagggacga gtgg 24

<210> 94

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 94

gttgcagcgg ccgacggagc gcgagcgtgg atgccac 37

<210> 95

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 95

ccatcgcccc gcgcgcacgt ggatgagcc 29

<210> 96

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 96

ctttagactt ctccggaagt cgaattaatg gaaac 35

<210> 97

<211> 28

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 97

gggcgccccg cgcgagcggg ggttgaag 28

<210> 98

<211> 10

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 98

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala

1 5 10

<210> 99

<211> 10

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 99

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala

1 5 10

<210> 100

<211> 10

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 100

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala

1 5 10

<210> 101

<211> 84

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 101

Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala
1 5 10 15

Ser Arg Gly Gln Ser Ala Ala Met Ala Pro Phe Gly Gly Leu Lys Ser
20 25 30

Met Thr Gly Phe Pro Val Arg Lys Val Asn Thr Asp Ile Thr Ser Ile
35 40 45

Thr Ser Asn Gly Gly Arg Val Lys Cys Met Gln Val Trp Pro Pro Ile
50 55 60

Gly Lys Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Pro Leu Thr Arg
65 70 75 80

Asp Ser Arg Ala

<210> 102

<211> 57

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 102

Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala
1 5 10 15

Ser Arg Gly Gln Ser Ala Ala Met Ala Pro Phe Gly Gly Leu Lys Ser
20 25 30

Met Thr Gly Phe Pro Val Arg Lys Val Asn Thr Asp Ile Thr Ser Ile
35 40 45

Thr Ser Asn Gly Gly Arg Val Lys Ser

50 55

<210> 103

<211> 85

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 103

Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala
1 5 10 15

Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala
20 25 30

Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser
35 40 45

Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Glu Lys
50 55 60

Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Asp Leu Thr Asp Ser Gly
65 70 75 80

Gly Arg Val Asn Cys
85

<210> 104

<211> 76

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 104

Met Ala Gln Val Ser Arg Ile Cys Asn Gly Val Gln Asn Pro Ser Leu
1 5 10 15

Ile Ser Asn Leu Ser Lys Ser Ser Gln Arg Lys Ser Pro Leu Ser Val
20 25 30

Ser Leu Lys Thr Gln Gln His Pro Arg Ala Tyr Pro Ile Ser Ser Ser
35 40 45

Trp Gly Leu Lys Lys Ser Gly Met Thr Leu Ile Gly Ser Glu Leu Arg
50 55 60

Pro Leu Lys Val Met Ser Ser Val Ser Thr Ala Cys

65 70 75

<210> 105

<211> 76

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 105

Met Ala Gln Val Ser Arg Ile Cys Asn Gly Val Trp Asn Pro Ser Leu

1 5 10 15

Ile Ser Asn Leu Ser Lys Ser Ser Gln Arg Lys Ser Pro Leu Ser Val

20 25 30

Ser Leu Lys Thr Gln Gln His Pro Arg Ala Tyr Pro Ile Ser Ser Ser

35 40 45

Trp Gly Leu Lys Lys Ser Gly Met Thr Leu Ile Gly Ser Glu Leu Arg

50 55 60

Pro Leu Lys Val Met Ser Ser Val Ser Thr Ala Cys

65 70 75

<210> 106

<211> 72

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 106

Met Ala Gln Ile Asn Asn Met Ala Gln Gly Ile Gln Thr Leu Asn Pro

1 5 10 15

Asn Ser Asn Phe His Lys Pro Gln Val Pro Lys Ser Ser Ser Phe Leu

20 25 30

Val Phe Gly Ser Lys Lys Leu Lys Asn Ser Ala Asn Ser Met Leu Val

35 40 45

Leu Lys Lys Asp Ser Ile Phe Met Gln Leu Phe Cys Ser Phe Arg Ile

50 55 60

Ser Ala Ser Val Ala Thr Ala Cys

65 70

<210> 107

<211> 69

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 107

Met Ala Ala Leu Val Thr Ser Gln Leu Ala Thr Ser Gly Thr Val Leu

1 5 10 15

Ser Val Thr Asp Arg Phe Arg Arg Pro Gly Phe Gln Gly Leu Arg Pro

20 25 30

Arg Asn Pro Ala Asp Ala Ala Leu Gly Met Arg Thr Val Gly Ala Ser

35 40 45

Ala Ala Pro Lys Gln Ser Arg Lys Pro His Arg Phe Asp Arg Arg Cys

50 55 60

Leu Ser Met Val Val

65

<210

> 108

<211> 77

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 108

Met Ala Ala Leu Thr Thr Ser Gln Leu Ala Thr Ser Ala Thr Gly Phe

1 5 10 15

Gly Ile Ala Asp Arg Ser Ala Pro Ser Ser Leu Leu Arg His Gly Phe

20 25 30

Gln Gly Leu Lys Pro Arg Ser Pro Ala Gly Gly Asp Ala Thr Ser Leu

35 40 45

Ser Val Thr Thr Ser Ala Arg Ala Thr Pro Lys Gln Gln Arg Ser Val

50 55 60

Gln Arg Gly Ser Arg Arg Phe Pro Ser Val Val Val Cys

65 70 75

<210> 109

<211> 57

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 109

Met Ala Ser Ser Val Leu Ser Ser Ala Ala Val Ala Thr Arg Ser Asn

1 5 10 15

Val Ala Gln Ala Asn Met Val Ala Pro Phe Thr Gly Leu Lys Ser Ala

20 25 30

Ala Ser Phe Pro Val Ser Arg Lys Gln Asn Leu Asp Ile Thr Ser Ile

35 40 45

Ala Ser Asn Gly Gly Arg Val Gln Cys

50 55

<210> 110

<211> 65

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 110

Met Glu Ser Leu Ala Ala Thr Ser Val Phe Ala Pro Ser Arg Val Ala

1 5 10 15

Val Pro Ala Ala Arg Ala Leu Val Arg Ala Gly Thr Val Val Pro Thr

20 25 30

Arg Arg Thr Ser Ser Thr Ser Gly Thr Ser Gly Val Lys Cys Ser Ala

35 40 45

Ala Val Thr Pro Gln Ala Ser Pro Val Ile Ser Arg Ser Ala Ala Ala

50 55 60

Ala

65

<210> 111

<211> 72

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 111

Met Gly Ala Ala Ala Thr Ser Met Gln Ser Leu Lys Phe Ser Asn Arg

1 5 10 15

Leu Val Pro Pro Ser Arg Arg Leu Ser Pro Val Pro Asn Asn Val Thr

20 25 30

Cys Asn Asn Leu Pro Lys Ser Ala Ala Pro Val Arg Thr Val Lys Cys

35 40 45

Cys Ala Ser Ser Trp Asn Ser Thr Ile Asn Gly Ala Ala Ala Thr Thr

50 55 60

Asn Gly Ala Ser Ala Ala Ser Ser

65 70

<210> 112

<211> 20

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<220><221> misc_feature

<222> (4)..(4)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (8)..(8)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (11)..(11)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (15)..(15)

<223> Xaa can be any naturally occurring amino acid

<220><221

> misc_feature

<222> (19)..(19)

<223> Xaa can be any naturally occurring amino acid

<400> 112

Gly Leu Phe Xaa Ala Leu Leu Xaa Leu Leu Xaa Ser Leu Trp Xaa Leu

1 5 10 15

Leu Leu Xaa Ala

20

<210> 113

<211> 20

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 113

Gly Leu Phe His Ala Leu Leu His Leu Leu His Ser Leu Trp His Leu

1 5 10 15

Leu Leu His Ala

20

<210> 114

<211> 7

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 114

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 115

<211> 16

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 115

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 116

<211> 9

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 116

Pro Ala Ala Lys Arg Val Lys Leu Asp

1 5

<210> 117

<211> 11

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 117

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro

1 5 10

<210> 118

<211> 38

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 118

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly

1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro

20 25 30

Arg Asn Gln Gly Gly Tyr

35

<210> 119

<211> 42

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 119

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40

<210> 120

<211> 8

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 120

Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 121

<211> 8

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 121

Pro Pro Lys Lys Ala Arg Glu Asp
1 5

<210> 122

<211> 8

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 122

Pro Gln Pro Lys Lys Lys Pro Leu
1 5

<210> 123

<211> 12

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 123

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro

1 5 10

<210> 124

<211> 5

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 124

Asp Arg Leu Arg Arg

1 5

<210> 125

<211> 7

<212

> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 125

Pro Lys Gln Lys Lys Arg Lys

1 5

<210> 126

<211> 10

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 126

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu

1 5 10

<210> 127

<211> 10

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 127

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg

1 5 10

<210> 128

<211> 20

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 128

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys

1 5 10 15

Lys Ser Lys Lys

20

<210> 129

<211> 17

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 129

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys

1 5 10 15

Lys

<210> 130

<211> 11

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 130

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5 10

<210> 131

<211> 12

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 131

Arg Arg Gln Arg Arg Thr Ser Lys Leu Met Lys Arg

1 5 10

<210> 132

<211> 27

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 132

Gly Trp Thr Leu Asn Ser Ala Gly Tyr Leu Leu Gly Lys Ile Asn Leu

1 5 10 15

Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu

20 25

<210> 133

<211> 33

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 133

Lys Ala Leu Ala Trp Glu Ala Lys Leu Ala Lys Ala Leu Ala Lys Ala

1 5 10 15

Leu Ala Lys His Leu Ala Lys Ala Leu Ala Lys Ala Leu Lys Cys Glu

20 25 30

Ala

<210> 134

<211> 16

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 134

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys

1 5 10 15

<210> 135

<211> 9

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 135

Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5

<210> 136

<211> 8

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 136

Arg Lys Lys Arg Arg Gln Arg Arg

1 5

<210> 137

<211> 11

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 137

Tyr Ala Arg Ala Ala Ala Arg Gln Ala Arg Ala

1 5 10

<210> 138

<211> 11

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 138

Thr His Arg Leu Pro Arg Arg Arg Arg Arg Arg

1 5 10

<210> 139

<211> 11

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 139

Gly Gly Arg Arg Ala Arg Arg Arg Arg Arg Arg

1 5 10

<210> 140

<211> 5

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 140

Gly Ser Gly Gly Ser

1 5

<210> 141

<211> 6

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400

> 141

Gly Gly Ser Gly Gly Ser

1 5

<210> 142

<211> 4

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 142

Gly Gly Gly Ser

1

<210> 143

<211> 4

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 143

Gly Gly Ser Gly

1

<210> 144

<211> 5

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 144

Gly Gly Ser Gly Gly

1 5

<210> 145

<211> 5

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 145

Gly Ser Gly Ser Gly

1 5

<210> 146

<211> 5

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 146

Gly Ser Gly Gly Gly

1 5

<210> 147

<211> 5

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 147

Gly Gly Gly Ser Gly

1 5

<210> 148

<211> 5

<212> PRT

<213>

Artificial sequence

<220><223> Synthetic sequence

<400> 148

Gly Ser Ser Ser Gly

1 5

<210> 149

<211> 4

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 149

Ala Ala Ala Ala

1

<210> 150

<211> 4

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 150

Ala Ala Ala Ala

1

<210> 151

<211> 80

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 151

cgattcctcc ctacagtagt taggtatagc cgaaaggtag agactaaatc tgtagttgga 60

gtgggccgct tgcatcggcc 80

<210> 152

<211> 122

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 152

tcgtctcgag ggttaccaaa attggcactt ctcgacttta ggccgatgca agcggcccac 60

tccactacag atttagtctc taccttgcgg ctatacctaa cttactgtag ggaggaatcg 120

tg 122

<

210> 153

<211> 91

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 153

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 60

gagtgaaggt gggctgcttg catcagccta a 91

<210> 154

<211> 114

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 154

cagaataata ctgacttact aagatatctt gagggtatac ccgaaaagat tggcgttgtt 60

gcaacgcaat aagatgtaaa tctgaaaagg tttggaatca tataaataat ttta 114

<210> 155

<211> 104

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 155

aagccaagat atggaatgcc attgtaatat tatggtgttg acttagttta gatttaaaca 60

atcttcgatg gctatatgcg gaaggtttgg cgtcgttgta acgc 104

<210> 156

<211> 213

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 156

cagtgtgcat agctataaca ctacgcaaag actgctaaag agcgatgtgc tctatcgcag 60

tctcaccttt aatggactta cggatctttt ggagcactaa gctccgctgc ggtgcaacac 120

cgccctttc ttgcctctgc ttgccctttc cggttattat agccgggaga gtgcggaaga 180

ttaccgctct agctcgcagc atgttactga gtc 213

<210> 157

<211> 94

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 157

gcaagtcatt cggggacact ttttgttatt taaagtgttt tagataaatc agtgtcatgc 60

tgaataacga cccgacctat aaataacata atcc 94

<210> 158

<211> 252

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 158

gtccttaagg tactacacat tacatgtgaa cgtggagcta ataatagaaa tattattaga 60

ctacacctta ttaataacgg taggagatct atatggtctt gaatggaata gtaattgtga 120

aattataatt tctgttctta gctacttaag atggctcgtt gcaagccact cgggggctct 180

cttgaagtca aagagcttta gacaaatcag tgtcaaactg aataacgacc cgaccatgac 240

ttcataatcc cg 252

<210> 159

<211> 152

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 159

ctcgaggcta tttcatactc agaacaaagg gattaaggaa tgcaacccat ttcaaaatct 60

gtggaattat cacaaccatt aacatttcat actcagaaca aagggattaa ggaatgcaac 120

ggacaccttg ctatgtcctt ttgatgtatg tg 152

<210> 160

<211> 61

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 160

ctcgaggcta tttcatactc agaacaaagc tcagaacaaa gggaataagg aatgcaacgg 60

a 61

<210> 161

<211> 101

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 161

ctcagaaaca aagggattaa ggaatgcaac ccatttcaaa atctgtggaa actcagaaca 60

aagggattaa ggaatgcaac ggacaccttg ctatgtcctt c 101

<210> 162

<211> 86

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 162

ctcagaacaa agggattaag gaatgcaacc catttcctca gaacaaaggg attaaggaat 60

gcaacggaca ccttgctatg tccttc 86

<210> 163

<211> 76

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 163

ctcagaacaa agggattaag gaatgcaacc catttctcag aacaaaggga ttaaggaatg 60

caacggacac cttgct 76

<210> 164

<211> 82

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 164

ctcagaacaa agggattaag gaatgcaacc catttcaaaa ctcagaacaa agggattaag 60

gaatgcaacg gacaccttgc ta 82

<210> 165

<211> 72

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 165

ctcagaacaa agggattaag gaatgcaacc catttcctca gaacaaaggg attaaggaat 60

gcaacggaca cc 72

<210> 166

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 166

ctcagaacaa agggattaag gaatgcaacc cgggattaag gaatgcaacg gacaccttg 59

<210> 167

<211> 50

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 167

ctcagaacaa agggattaag gaatgcaacc aacggacacc ttgctatgtc 50

<210> 168

<211> 45

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 168

ctcagaacaa agggattaag gaatgcaacc catttcaaaa tctgt 45

<210> 169

<211> 41

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 169

ctcagaacaa agggattaag gaatgcaacc catttcaaaa t 41

<210> 170

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 170

ctcagaacaa agggattaag gaatgcaacc catttc 36

<210> 171

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 171

ctcagaacaa agggattaag gaatgcaacc catttc 36

<210> 172

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 172

ctcagaacaa agggattaag gaatgcaacc catttc 36

<210> 173

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 173

ctcagaacaa agggattaag gaatgcaacc catttca 37

<210> 174

<211> 38

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 174

ctcagaacaa agggattaag gaatgcaacc catttcaa 38

<210> 175

<211> 46

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 175

ctcagaacaa agggattaag gaatgcaacc catttcaaaa tctgtg 46

<210> 176

<211> 46

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 176

ctcagaacaa agggattaag gaatgcaacc catttcaaaa tctgtg 46

<210> 177

<211> 44

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 177

ctcagaacaa agggattaag gaatgcaacc catttcaaaa tctg 44

<210> 178

<211> 726

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 178

Ser Glu Ser Glu Asn Lys Ile Ile Glu Gln Tyr Tyr Ala Phe Leu Tyr
1 5 10 15

Ser Phe Arg Asp Lys Tyr Glu Lys Pro Glu Phe Lys Asn Arg Gly Asp
20 25 30

Ile Lys Arg Lys Leu Gln Asn Lys Trp Glu Asp Phe Leu Lys Glu Gln
35 40 45

Asn Leu Lys Asn Asp Lys Lys Leu Ser Asn Tyr Ile Phe Ser Asn Arg
50 55 60

Asn Phe Arg Arg Ser Tyr Asp Arg Glu Glu Glu Asn Glu Glu Gly Ile
65 70 75 80

Asp Glu Lys Lys Ser Lys Pro Lys Arg Ile Asn Cys Phe Glu Lys Glu
85 90 95

Lys Asn Leu Lys Asp Gln Tyr Asp Lys Asp Ala Ile Asn Ala Ser Ala
100 105 110

Asn Lys Asp Gly Ala Gln Lys Trp Gly Cys Phe Glu Cys Ile Phe Phe
115 120 125

Pro Met Tyr Lys Ile Glu Ser Gly Asp Pro Asn Lys Arg Ile Ile Ile
130 135 140
Asn Lys Thr Arg Phe Lys Leu Phe Asp Phe Tyr Leu Asn Leu Lys Gly
145 150 155 160

Cys Lys Ser Cys Leu Arg Ser Thr Tyr His Pro Tyr Arg Ser Asn Val
165 170 175
Tyr Ile Glu Ser Asn Tyr Asp Lys Leu Lys Arg Glu Ile Gly Asn Phe
180 185 190
Leu Gln Gln Lys Asn Ile Phe Gln Arg Met Arg Lys Ala Lys Val Ser
195 200 205
Glu Gly Lys Tyr Leu Thr Asn Leu Asp Glu Tyr Arg Leu Ser Cys Val
210 215 220

Ala Met His Phe Lys Asn Arg Trp Leu Phe Phe Asp Ser Ile Gln Lys
225 230 235 240
Val Leu Arg Glu Thr Ile Lys Gln Arg Leu Lys Gln Met Arg Glu Ser
245 250 255
Tyr Asp Glu Gln Ala Lys Thr Lys Arg Ser Lys Gly His Gly Arg Ala
260 265 270
Lys Tyr Glu Asp Gln Val Arg Met Ile Arg Arg Arg Ala Tyr Ser Ala
275 280 285

Gln Ala His Lys Leu Leu Asp Asn Gly Tyr Ile Thr Leu Phe Asp Tyr
290 295 300
Asp Asp Lys Glu Ile Asn Lys Val Cys Leu Thr Ala Ile Asn Gln Glu
305 310 315 320
Gly Phe Asp Ile Gly Gly Tyr Leu Asn Ser Asp Ile Asp Asn Val Met
325 330 335
Pro Pro Ile Glu Ile Ser Phe His Leu Lys Trp Lys Tyr Asn Glu Pro
340 345 350

Ile Leu Asn Ile Glu Ser Pro Phe Ser Lys Ala Lys Ile Ser Asp Tyr
355 360 365
Leu Arg Lys Ile Arg Glu Asp Leu Asn Leu Glu Arg Gly Lys Glu Gly

370 375 380

Lys Ala Arg Ser Lys Lys Asn Val Arg Arg Lys Val Leu Ala Ser Lys

385 390 395 400

Gly Glu Asp Gly Tyr Lys Lys Ile Phe Thr Asp Phe Phe Ser Lys Trp

405 410 415

Lys Glu Glu Leu Glu Gly Asn Ala Met Glu Arg Val Leu Ser Gln Ser

420 425 430

Ser Gly Asp Ile Gln Trp Ser Lys Lys Lys Arg Ile His Tyr Thr Thr

435 440 445

Leu Val Leu Asn Ile Asn Leu Leu Asp Lys Lys Gly Val Gly Asn Leu

450 455 460

Lys Tyr Tyr Glu Ile Ala Glu Lys Thr Lys Ile Leu Ser Phe Asp Lys

465 470 475 480

Asn Glu Asn Lys Phe Trp Pro Ile Thr Ile Gln Val Leu Leu Asp Gly

485 490 495

Tyr Glu Ile Gly Thr Glu Tyr Asp Glu Ile Lys Gln Leu Asn Glu Lys

500 505 510

Thr Ser Lys Gln Phe Thr Ile Tyr Asp Pro Asn Thr Lys Ile Ile Lys

515 520 525

Ile Pro Phe Thr Asp Ser Lys Ala Val Pro Leu Gly Met Leu Gly Ile

530 535 540

Asn Ile Ala Thr Leu Lys Thr Val Lys Lys Thr Glu Arg Asp Ile Lys

545 550 555 560

Val Ser Lys Ile Phe Lys Gly Gly Leu Asn Ser Lys Ile Val Ser Lys

565 570 575

Ile Gly Lys Gly Ile Tyr Ala Gly Tyr Phe Pro Thr Val Asp Lys Glu

580 585 590

Ile Leu Glu Glu Val Glu Glu Asp Thr Leu Asp Asn Glu Phe Ser Ser

595 600 605

Lys Ser Gln Arg Asn Ile Phe Leu Lys Ser Ile Ile Lys Asn Tyr Asp

610 615 620

Lys Met Leu Lys Glu Gln Leu Phe Asp Phe Tyr Ser Phe Leu Val Arg
625 630 635 640
Asn Asp Leu Gly Val Arg Phe Leu Thr Asp Arg Glu Leu Gln Asn Ile
645 650 655
Glu Asp Glu Ser Phe Asn Leu Glu Lys Arg Phe Phe Glu Thr Asp Arg
660 665 670

Asp Arg Ile Ala Arg Trp Phe Asp Asn Thr Asn Thr Asp Asp Gly Lys
675 680 685
Glu Lys Phe Lys Lys Leu Ala Asn Glu Ile Val Asp Ser Tyr Lys Pro
690 695 700
Arg Leu Ile Arg Leu Pro Val Val Arg Val Ile Lys Arg Ile Gln Pro
705 710 715 720
Val Lys Gln Arg Glu Met
725

<210> 179
<211> 517
<212> PRT
<213> Artificial sequence

<220><223> Synthetic sequence
<400> 179
Lys Tyr Ser Thr Arg Asp Phe Ser Glu Leu Asn Glu Ile Gln Val Thr
1 5 10 15
Ala Cys Lys Gln Asp Glu Phe Phe Lys Val Ile Gln Asn Ala Trp Arg
20 25 30
Glu Ile Ile Lys Lys Arg Phe Leu Glu Asn Arg Glu Asn Phe Ile Glu
35 40 45
Lys Lys Ile Phe Lys Asn Lys Lys Gly Arg Gly Lys Arg Gln Glu Ser
50 55 60

Asp Lys Thr Ile Gln Arg Asn Arg Ala Ser Val Met Lys Asn Phe Gln
65 70 75 80
Leu Ile Glu Asn Glu Lys Ile Ile Leu Arg Ala Pro Ser Gly His Val
85 90 95

Ala Cys Val Phe Pro Val Lys Val Gly Leu Asp Ile Gly Gly Phe Lys
100 105 110
Thr Asp Asp Leu Glu Lys Asn Ile Phe Pro Pro Arg Thr Ile Thr Ile
115 120 125

Asn Val Phe Trp Lys Asn Arg Asp Arg Gln Arg Lys Gly Arg Lys Leu
130 135 140
Glu Val Trp Gly Ile Lys Ala Arg Thr Lys Leu Ile Glu Lys Val His
145 150 155 160
Lys Trp Asp Lys Leu Glu Glu Val Lys Lys Lys Arg Leu Lys Ser Leu
165 170 175
Glu Gln Lys Gln Glu Lys Ser Leu Asp Asn Trp Ser Glu Val Asn Asn
180 185 190

Asp Ser Phe Tyr Lys Val Gln Ile Asp Glu Leu Gln Glu Lys Ile Asp
195 200 205
Lys Ser Leu Lys Gly Arg Thr Met Asn Lys Ile Leu Asp Asn Lys Ala
210 215 220
Lys Glu Ser Lys Glu Ala Glu Gly Leu Tyr Ile Glu Trp Glu Lys Asp
225 230 235 240
Phe Glu Gly Glu Met Leu Arg Arg Ile Glu Ala Ser Thr Gly Gly Glu
245 250 255

Glu Lys Trp Gly Lys Arg Arg Gln Arg Arg His Thr Ser Leu Leu Leu
260 265 270
Asp Ile Lys Asn Asn Ser Arg Gly Ser Lys Glu Ile Ile Asn Phe Tyr
275 280 285
Ser Tyr Ala Lys Gln Gly Lys Lys Glu Lys Lys Ile Glu Phe Phe Pro
290 295 300
Phe Pro Leu Thr Ile Thr Leu Asp Ala Glu Glu Glu Ser Pro Leu Asn
305 310 315 320

Ile Lys Ser Ile Pro Ile Glu Asp Lys Asn Ala Thr Ser Lys Tyr Phe
325 330 335
Ser Ile Pro Phe Thr Glu Thr Arg Ala Thr Pro Leu Ser Ile Leu Gly

340 345 350

Asp Arg Val Gln Lys Phe Lys Thr Lys Asn Ile Ser Gly Ala Ile Lys

355 360 365

Arg Asn Leu Gly Ser Ser Ile Ser Ser Cys Lys Ile Val Gln Asn Ala

370 375 380

Glu Thr Ser Ala Lys Ser Ile Leu Ser Leu Pro Asn Val Lys Glu Asp

385 390 395 400

Asn Asn Met Glu Ile Phe Ile Asn Thr Met Ser Lys Asn Tyr Phe Arg

405 410 415

Ala Met Met Lys Gln Met Glu Ser Phe Ile Phe Glu Met Glu Pro Lys

420 425 430

Thr Leu Ile Asp Pro Tyr Lys Glu Lys Ala Ile Lys Trp Phe Glu Val

435 440 445

Ala Ala Ser Ser Arg Ala Lys Arg Lys Leu Lys Lys Leu Ser Lys Ala

450 455 460

Asp Ile Lys Lys Ser Glu Leu Leu Leu Ser Asn Thr Glu Glu Phe Glu

465 470 475 480

Lys Glu Lys Gln Glu Lys Leu Glu Ala Leu Glu Lys Glu Ile Glu Glu

485 490 495

Phe Tyr Leu Pro Arg Ile Val Arg Leu Gln Leu Thr Lys Thr Ile Leu

500 505 510

Glu Thr Pro Val Met

515

<210> 180

<211> 481

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 180

Lys Lys Leu Gln Leu Leu Gly His Lys Ile Leu Leu Lys Glu Tyr Asp

1 5 10 15

Pro Asn Ala Val Asn Ala Ala Ala Asn Phe Glu Thr Ser Thr Ala Glu

20 25 30

Leu Cys Gly Gln Cys Lys Met Lys Pro Phe Lys Asn Lys Arg Arg Phe

35 40 45

Gln Tyr Thr Phe Gly Lys Asn Tyr His Gly Cys Leu Ser Cys Ile Gln

50 55 60

Asn Val Tyr Tyr Ala Lys Lys Arg Ile Val Gln Ile Ala Lys Glu Glu

65 70 75 80

Leu Lys His Gln Leu Thr Asp Ser Ile Ala Ser Ile Pro Tyr Lys Tyr

85 90 95

Thr Ser Leu Phe Ser Asn Thr Asn Ser Ile Asp Glu Leu Tyr Ile Leu

100 105 110

Lys Gln Glu Arg Ala Ala Phe Phe Ser Asn Thr Asn Ser Ile Asp Glu

115 120 125

Leu Tyr Ile Thr Gly Ile Glu Asn Asn Ile Ala Phe Lys Val Ile Ser

130 135 140

Ala Ile Trp Asp Glu Ile Ile Lys Lys Arg Arg Gln Arg Tyr Ala Glu

145 150 155 160

Ser Leu Thr Asp Thr Gly Thr Val Lys Ala Asn Arg Gly His Gly Gly

165 170 175

Thr Ala Tyr Lys Ser Asn Thr Arg Gln Glu Lys Ile Arg Ala Leu Gln

180 185 190

Lys Gln Thr Leu His Met Val Thr Asn Pro Tyr Ile Ser Leu Ala Arg

195 200 205

Tyr Lys Asn Asn Tyr Ile Val Ala Thr Leu Pro Arg Thr Ile Gly Met

210 215 220

His Ile Gly Ala Ile Lys Asp Arg Asp Pro Gln Lys Lys Leu Ser Asp

225 230 235 240

Tyr Ala Ile Asn Phe Asn Val Phe Trp Ser Asp Asp Arg Gln Leu Ile

245 250 255

Glu Leu Ser Thr Val Gln Tyr Thr Gly Asp Met Val Arg Lys Ile Glu

260 265 270

Ala Glu Thr Gly Glu Asn Asn Lys Trp Gly Glu Asn Met Lys Arg Thr
275 280 285
Lys Thr Ser Leu Leu Leu Glu Ile Leu Thr Lys Lys Thr Thr Asp Glu
290 295 300

Leu Thr Phe Lys Asp Trp Ala Phe Ser Thr Lys Lys Glu Ile Asp Ser
305 310 315 320
Val Thr Lys Lys Thr Tyr Gln Gly Phe Pro Ile Gly Ile Ile Phe Glu
325 330 335
Gly Asn Glu Ser Ser Val Lys Phe Gly Ser Gln Asn Tyr Phe Pro Leu
340 345 350
Pro Phe Asp Ala Lys Ile Thr Pro Pro Thr Ala Glu Gly Phe Arg Leu
355 360 365

Asp Trp Leu Arg Lys Gly Ser Phe Ser Ser Gln Met Lys Thr Ser Tyr
370 375 380
Gly Leu Ala Ile Tyr Ser Asn Lys Val Thr Asn Ala Ile Pro Ala Tyr
385 390 395 400
Val Ile Lys Asn Met Phe Tyr Lys Ile Ala Arg Ala Glu Asn Gly Lys
405 410 415
Gln Ile Lys Ala Lys Phe Leu Lys Lys Tyr Leu Asp Ile Ala Gly Asn
420 425 430

Asn Tyr Val Pro Phe Ile Ile Met Gln His Tyr Arg Val Leu Asp Thr
435 440 445
Phe Glu Glu Met Pro Ile Ser Gln Pro Lys Val Ile Arg Leu Ser Leu
450 455 460
Thr Lys Thr Gln His Ile Ile Ile Lys Lys Asp Lys Thr Asp Ser Lys
465 470 475 480
Met

<210> 181
<211> 534
<212> PRT
<213> Artificial sequence

<220><223> Synthetic sequence

<

400> 181

Asn Thr Ser Asn Leu Ile Asn Leu Gly Lys Lys Ala Ile Asn Ile Ser
1 5 10 15

Ala Asn Tyr Asp Ala Asn Leu Glu Val Gly Cys Lys Asn Cys Lys Phe
20 25 30

Leu Ser Ser Asn Gly Asn Phe Pro Arg Gln Thr Asn Val Lys Glu Gly
35 40 45

Cys His Ser Cys Glu Lys Ser Thr Tyr Glu Pro Ser Ile Tyr Leu Val
50 55 60

Lys Ile Gly Glu Arg Lys Ala Lys Tyr Asp Val Leu Asp Ser Leu Lys
65 70 75 80

Lys Phe Thr Phe Gln Ser Leu Lys Tyr Gln Ser Lys Lys Ser Met Lys
85 90 95

Ser Arg Asn Lys Lys Pro Lys Glu Leu Lys Glu Phe Val Ile Phe Ala
100 105 110

Asn Lys Asn Lys Ala Phe Asp Val Ile Gln Lys Ser Tyr Asn His Leu
115 120 125

Ile Leu Gln Ile Lys Lys Glu Ile Asn Arg Met Asn Ser Lys Lys Arg
130 135 140

Lys Lys Asn His Lys Arg Arg Leu Phe Arg Asp Arg Glu Lys Gln Leu
145 150 155 160

Asn Lys Leu Arg Leu Ile Glu Ser Ser Asn Leu Phe Leu Pro Arg Glu
165 170 175

Asn Lys Gly Asn Asn His Val Phe Thr Tyr Val Ala Ile His Ser Val
180 185 190

Gly Arg Asp Ile Gly Val Ile Gly Ser Tyr Asp Glu Lys Leu Asn Phe
195 200 205

Glu Thr Glu Leu Thr Tyr Gln Leu Tyr Phe Asn Asp Asp Lys Arg Leu
210 215 220

Leu Tyr Ala Tyr Lys Pro Lys Gln Asn Lys Ile Ile Lys Ile Lys Glu

225 230 235 240

Lys Leu Trp Asn Leu Arg Lys Glu Lys Glu Pro Leu Asp Leu Glu Tyr

245 250 255

Glu Lys Pro Leu Asn Lys Ser Ile Thr Phe Ser Ile Lys Asn Asp Asn

260 265 270

Leu Phe Lys Val Ser Lys Asp Leu Met Leu Arg Arg Ala Lys Phe Asn

275 280 285

Ile Gln Gly Lys Glu Lys Leu Ser Lys Glu Glu Arg Lys Ile Asn Arg

290 295 300

Asp Leu Ile Lys Ile Lys Gly Leu Val Asn Ser Met Ser Tyr Gly Arg

305 310 315 320

Phe Asp Glu Leu Lys Lys Glu Lys Asn Ile Trp Ser Pro His Ile Tyr

325 330 335

Arg Glu Val Arg Gln Lys Glu Ile Lys Pro Cys Leu Ile Lys Asn Gly

340 345 350

Asp Arg Ile Glu Ile Phe Glu Gln Leu Lys Lys Lys Met Glu Arg Leu

355 360 365

Arg Arg Phe Arg Glu Lys Arg Gln Lys Lys Ile Ser Lys Asp Leu Ile

370 375 380

Phe Ala Glu Arg Ile Ala Tyr Asn Phe His Thr Lys Ser Ile Lys Asn

385 390 395 400

Thr Ser Asn Lys Ile Asn Ile Asp Gln Glu Ala Lys Arg Gly Lys Ala

405 410 415

Ser Tyr Met Arg Lys Arg Ile Gly Tyr Glu Thr Phe Lys Asn Lys Tyr

420 425 430

Cys Glu Gln Cys Leu Ser Lys Gly Asn Val Tyr Arg Asn Val Gln Lys

435 440 445

Gly Cys Ser Cys Phe Glu Asn Pro Phe Asp Trp Ile Lys Lys Gly Asp

450 455 460

Glu Asn Leu Leu Pro Lys Lys Asn Glu Asp Leu Arg Val Lys Gly Ala

465 470 475 480

Phe Arg Asp Glu Ala Leu Glu Lys Gln Ile Val Lys Ile Ala Phe Asn
485 490 495
Ile Ala Lys Gly Tyr Glu Asp Phe Tyr Asp Asn Leu Gly Glu Ser Thr
500 505 510

Glu Lys Asp Leu Lys Leu Lys Phe Lys Val Gly Thr Thr Ile Asn Glu
515 520 525
Gln Glu Ser Leu Lys Leu
530

<210> 182
<211> 537
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 182
Thr Ser Asn Pro Ile Lys Leu Gly Lys Lys Ala Ile Asn Ile Ser Ala
1 5 10 15
Asn Tyr Asp Ser Asn Leu Gln Ile Gly Cys Lys Asn Cys Lys Phe Leu
20 25 30

Ser Tyr Asn Gly Asn Phe Pro Arg Gln Thr Asn Val Lys Glu Gly Cys
35 40 45
His Ser Cys Glu Lys Ser Thr Tyr Glu Pro Pro Val Tyr Thr Val Arg
50 55 60
Ile Gly Glu Arg Arg Ser Lys Tyr Asp Val Leu Asp Ser Leu Lys Lys
65 70 75 80
Phe Ile Phe Leu Ser Leu Lys Tyr Arg Gln Ser Lys Lys Met Lys Thr
85 90 95

Arg Ser Lys Gly Ile Arg Gly Leu Glu Glu Phe Val Ile Ser Ala Asn
100 105 110
Leu Lys Lys Ala Met Asp Val Ile Gln Lys Ser Tyr Arg His Leu Ile
115 120 125
Leu Asn Ile Lys Asn Glu Ile Val Arg Met Asn Gly Lys Lys Arg Asn
130 135 140

Lys Asn His Lys Arg Leu Leu Phe Arg Asp Arg Glu Lys Gln Leu Asn
145 150 155 160

Lys Leu Arg Leu Ile Glu Gly Ser Ser Phe Phe Lys Pro Pro Thr Val
165 170 175
Lys Gly Asp Asn Ser Ile Phe Thr Cys Val Ala Ile His Asn Ile Gly
180 185 190
Arg Asp Ile Gly Ile Ala Gly Asp Tyr Phe Asp Lys Leu Glu Pro Lys
195 200 205
Ile Glu Leu Thr Tyr Gln Leu Tyr Tyr Glu Tyr Asn Pro Lys Lys Glu
210 215 220

Ser Glu Ile Asn Lys Arg Leu Leu Tyr Ala Tyr Lys Pro Lys Gln Asn
225 230 235 240
Lys Ile Ile Glu Ile Lys Glu Lys Leu Trp Asn Leu Arg Lys Glu Lys
245 250 255
Ser Pro Leu Asp Leu Glu Tyr Glu Lys Pro Leu Thr Lys Ser Ile Thr
260 265 270
Phe Leu Val Lys Arg Asp Gly Val Phe Arg Ile Ser Lys Asp Leu Met
275 280 285

Leu Arg Lys Ala Lys Phe Ile Ile Gln Gly Lys Glu Lys Leu Ser Lys
290 295 300
Glu Glu Arg Lys Ile Asn Arg Asp Leu Ile Lys Ile Lys Ser Asn Ile
305 310 315 320
Ile Ser Leu Thr Tyr Gly Arg Phe Asp Glu Leu Lys Lys Asp Lys Thr
325 330 335
Ile Trp Ser Pro His Ile Phe Arg Asp Val Lys Gln Gly Lys Ile Thr
340 345 350

Pro Cys Ile Glu Arg Lys Gly Asp Arg Met Asp Ile Phe Gln Gln Leu
355 360 365
Arg Lys Lys Ser Glu Arg Leu Arg Glu Asn Arg Lys Lys Arg Gln Lys
370 375 380
Lys Ile Ser Lys Asp Leu Ile Phe Ala Glu Arg Ile Ala Tyr Asn Phe

385 390 395 400

His Thr Lys Ser Ile Lys Asn Thr Ser Asn Leu Ile Asn Ile Lys His

405 410 415

Glu Ala Lys Arg Gly Lys Ala Ser Tyr Met Arg Lys Arg Ile Gly Asn

420 425 430

Glu Thr Phe Arg Ile Lys Tyr Cys Glu Gln Cys Phe Pro Lys Asn Asn

435 440 445

Val Tyr Lys Asn Val Gln Lys Gly Cys Ser Cys Phe Glu Asp Pro Phe

450 455 460

Glu Tyr Ile Lys Lys Gly Asn Glu Asp Leu Ile Pro Asn Lys Asn Gln

465 470 475 480

Asp Leu Lys Ala Lys Gly Ala Phe Arg Asp Asp Ala Leu Glu Lys Gln

485 490 495

Ile Ile Lys Val Ala Phe Asn Ile Ala Lys Gly Tyr Glu Asp Phe Tyr

500 505 510

Glu Asn Leu Lys Lys Thr Thr Glu Lys Asp Ile Arg Leu Lys Phe Lys

515 520 525

Val Gly Thr Ile Ile Ser Glu Glu Met

530 535

<210> 183

<211> 541

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 183

Asn Asn Ser Ile Asn Leu Ser Lys Lys Ala Ile Asn Ile Ser Ala Asn

1 5 10 15

Tyr Asp Ala Asn Leu Gln Val Arg Cys Lys Asn Cys Lys Phe Leu Ser

20 25 30

Ser Asn Gly Asn Phe Pro Arg Gln Thr Asp Val Lys Glu Gly Cys His

35 40 45

Ser Cys Glu Lys Ser Thr Tyr Glu Pro Pro Val Tyr Asp Val Lys Ile

50 55 60

Gly Glu Ile Lys Ala Lys Tyr Glu Val Leu Asp Ser Leu Lys Lys Phe
65 70 75 80

Thr Phe Gln Ser Leu Lys Tyr Gln Leu Ser Lys Ser Met Lys Phe Arg
85 90 95

Ser Lys Lys Ile Lys Glu Leu Lys Glu Phe Val Ile Phe Ala Lys Glu
100 105 110

Ser Lys Ala Leu Asn Val Ile Asn Arg Ser Tyr Lys His Leu Ile Leu

115 120 125

Asn Ile Lys Asn Asp Ile Asn Arg Met Asn Ser Lys Lys Arg Ile Lys
130 135 140

Asn His Lys Gly Arg Leu Phe Leu Asp Arg Gln Lys Gln Leu Ser Lys
145 150 155 160

Leu Lys Leu Ile Glu Gly Ser Ser Phe Phe Val Pro Ala Lys Asn Val
165 170 175

Gly Asn Lys Ser Val Phe Thr Cys Val Ala Ile His Ser Ile Gly Arg

180 185 190

Asp Ile Gly Ile Ala Gly Leu Tyr Asp Ser Phe Thr Lys Pro Val Asn
195 200 205

Glu Ile Thr Tyr Gln Ile Phe Phe Ser Gly Glu Arg Arg Leu Leu Tyr
210 215 220

Ala Tyr Lys Pro Lys Gln Leu Lys Ile Leu Ser Ile Lys Glu Asn Leu
225 230 235 240

Trp Ser Leu Lys Asn Glu Lys Lys Pro Leu Asp Leu Leu Tyr Glu Lys

245 250 255

Pro Leu Gly Lys Asn Leu Asn Phe Asn Val Lys Gly Gly Asp Leu Phe
260 265 270

Arg Val Ser Lys Asp Leu Met Ile Arg Asn Ala Lys Phe Asn Val His
275 280 285

Gly Arg Gln Arg Leu Ser Asp Glu Glu Arg Leu Ile Asn Arg Asn Phe

290 295 300

Ile Lys Ile Lys Gly Glu Val Val Ser Leu Ser Tyr Gly Arg Phe Glu

305 310 315 320

Glu Leu Lys Lys Asp Arg Lys Leu Trp Ser Pro His Ile Phe Lys Asp

325 330 335

Val Arg Gln Asn Lys Ile Lys Pro Cys Leu Val Met Gln Gly Gln Arg

340 345 350

Ile Asp Ile Phe Glu Gln Leu Lys Arg Lys Leu Glu Leu Leu Lys Lys

355 360 365

Ile Arg Lys Ser Arg Gln Lys Lys Leu Ser Lys Asp Leu Ile Phe Gly

370 375 380

Glu Arg Ile Ala Tyr Asn Phe His Thr Lys Ser Ile Lys Asn Thr Ser

385 390 395 400

Asn Lys Ile Asn Ile Asp Ser Asp Ala Lys Arg Gly Arg Ala Ser Tyr

405 410 415

Met Arg Lys Arg Ile Gly Asn Glu Thr Phe Lys Leu Lys Tyr Cys Asp

420 425 430

Val Cys Phe Pro Lys Ala Asn Val Tyr Arg Arg Val Gln Asn Gly Cys

435 440 445

Ser Cys Ser Glu Asn Pro Tyr Asn Tyr Ile Lys Lys Gly Asp Lys Asp

450 455 460

Leu Leu Pro Lys Lys Asp Glu Gly Leu Ala Ile Lys Gly Ala Phe Arg

465 470 475 480

Asp Glu Lys Leu Asn Lys Gln Ile Ile Lys Val Ala Phe Asn Ile Ala

485 490 495

Lys Gly Tyr Glu Asp Phe Tyr Asp Asp Leu Lys Lys Arg Thr Glu Lys

500 505 510

Asp Val Asp Leu Lys Phe Lys Ile Gly Thr Thr Val Leu Asp Gln Lys

515 520 525

Pro Met Glu Ile Phe Asp Gly Ile Val Ile Thr Trp Leu

530 535 540

<210> 184

<211> 542

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 184

Leu Leu Thr Thr Val Val Glu Thr Asn Asn Leu Ala Lys Lys Ala Ile
1 5 10 15

Asn Val Ala Ala Asn Phe Asp Ala Asn Ile Asp Arg Gln Tyr Tyr Arg
20 25 30

Cys Thr Pro Asn Leu Cys Arg Phe Ile Ala Gln Ser Pro Arg Glu Thr
35 40 45

Lys Glu Lys Asp Ala Gly Cys Ser Ser Cys Thr Gln Ser Thr Tyr Asp
50 55 60

Pro Lys Val Tyr Val Ile Lys Ile Gly Lys Leu Leu Ala Lys Tyr Glu
65 70 75 80

Ile Leu Lys Ser Leu Lys Arg Phe Leu Phe Met Asn Arg Tyr Phe Lys
85 90 95

Gln Lys Lys Thr Glu Arg Ala Gln Gln Lys Gln Lys Ile Gly Thr Glu
100 105 110

Leu Asn Glu Met Ser Ile Phe Ala Lys Ala Thr Asn Ala Met Glu Val
115 120 125

Ile Lys Arg Ala Thr Lys His Cys Thr Tyr Asp Ile Ile Pro Glu Thr
130 135 140

Lys Ser Leu Gln Met Leu Lys Arg Arg Arg His Arg Val Lys Val Arg
145 150 155 160

Ser Leu Leu Lys Ile Leu Lys Glu Arg Arg Met Lys Ile Lys Lys Ile
165 170 175

Pro Asn Thr Phe Ile Glu Ile Pro Lys Gln Ala Lys Lys Asn Lys Ser
180 185 190

Asp Tyr Tyr Val Ala Ala Ala Leu Lys Ser Cys Gly Ile Asp Val Gly

195 200 205

Leu Cys Gly Ala Tyr Glu Lys Asn Ala Glu Val Glu Ala Glu Tyr Thr
210 215 220
Tyr Gln Leu Tyr Tyr Glu Tyr Lys Gly Asn Ser Ser Thr Lys Arg Ile
225 230 235 240
Leu Tyr Cys Tyr Asn Asn Pro Gln Lys Asn Ile Arg Glu Phe Trp Glu
245 250 255
Ala Phe Tyr Ile Gln Gly Ser Lys Ser His Val Asn Thr Pro Gly Thr
260 265 270

Ile Arg Leu Lys Met Glu Lys Phe Leu Ser Pro Ile Thr Ile Glu Ser
275 280 285
Glu Ala Leu Asp Phe Arg Val Trp Asn Ser Asp Leu Lys Ile Arg Asn
290 295 300
Gly Gln Tyr Gly Phe Ile Lys Lys Arg Ser Leu Gly Lys Glu Ala Arg
305 310 315 320
Glu Ile Lys Lys Gly Met Gly Asp Ile Lys Arg Lys Ile Gly Asn Leu
325 330 335

Thr Tyr Gly Lys Ser Pro Ser Glu Leu Lys Ser Ile His Val Tyr Arg
340 345 350
Thr Glu Arg Glu Asn Pro Lys Lys Pro Arg Ala Ala Arg Lys Lys Glu
355 360 365
Asp Asn Phe Met Glu Ile Phe Glu Met Gln Arg Lys Lys Asp Tyr Glu
370 375 380
Val Asn Lys Lys Arg Arg Lys Glu Ala Thr Asp Ala Ala Lys Ile Met
385 390 395 400

Asp Phe Ala Glu Glu Pro Ile Arg His Tyr His Thr Asn Asn Leu Lys
405 410 415
Ala Val Arg Arg Ile Asp Met Asn Glu Gln Val Glu Arg Lys Lys Thr
420 425 430
Ser Val Phe Leu Lys Arg Ile Met Gln Asn Gly Tyr Arg Gly Asn Tyr
435 440 445

Cys Arg Lys Cys Ile Lys Ala Pro Glu Gly Ser Asn Arg Asp Glu Asn
450 455 460

Val Leu Glu Lys Asn Glu Gly Cys Leu Asp Cys Ile Gly Ser Glu Phe
465 470 475 480
Ile Trp Lys Lys Ser Ser Lys Glu Lys Lys Gly Leu Trp His Thr Asn
485 490 495
Arg Leu Leu Arg Arg Ile Arg Leu Gln Cys Phe Thr Thr Ala Lys Ala
500 505 510
Tyr Glu Asn Phe Tyr Asn Asp Leu Phe Glu Lys Lys Glu Ser Ser Leu
515 520 525

Asp Ile Ile Lys Leu Lys Val Ser Ile Thr Thr Lys Ser Met
530 535 540

<210> 185
<211> 564
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 185
Ala Ser Thr Met Asn Leu Ala Lys Gln Ala Ile Asn Phe Ala Ala Asn
1 5 10 15
Tyr Asp Ser Asn Leu Glu Ile Gly Cys Lys Gly Cys Lys Phe Met Ser
20 25 30
Thr Trp Ser Lys Lys Ser Asn Pro Lys Phe Tyr Pro Arg Gln Asn Asn

35 40 45
Gln Ala Asn Lys Cys His Ser Cys Thr Tyr Ser Thr Gly Glu Pro Glu
50 55 60
Val Pro Ile Ile Glu Ile Gly Glu Arg Ala Ala Lys Tyr Lys Ile Phe
65 70 75 80
Thr Ala Leu Lys Lys Phe Val Phe Met Ser Val Ala Tyr Lys Glu Arg
85 90 95
Arg Arg Gln Arg Phe Lys Ser Lys Lys Pro Lys Glu Leu Lys Glu Leu

100 105 110

Ala Ile Cys Ser Asn Arg Glu Lys Ala Met Glu Val Ile Gln Lys Ser

115 120 125

Val Val His Cys Tyr Gly Asp Val Lys Gln Glu Ile Pro Arg Ile Arg

130 135 140

Lys Ile Lys Val Leu Lys Asn His Lys Gly Arg Leu Phe Tyr Lys Gln

145 150 155 160

Lys Arg Ser Lys Ile Lys Ile Ala Lys Leu Glu Lys Gly Ser Phe Phe

165 170 175

Lys Thr Phe Ile Pro Lys Val His Asn Asn Gly Cys His Ser Cys His

180 185 190

Glu Ala Ser Leu Asn Lys Pro Ile Leu Val Thr Thr Ala Leu Asn Thr

195 200 205

Ile Gly Ala Asp Ile Gly Leu Ile Asn Asp Tyr Ser Thr Ile Ala Pro

210 215 220

Thr Glu Thr Asp Ile Ser Trp Gln Val Tyr Tyr Glu Phe Ile Pro Asn

225 230 235 240

Gly Asp Ser Glu Ala Val Lys Lys Arg Leu Leu Tyr Phe Tyr Lys Pro

245 250 255

Lys Gly Ala Leu Ile Lys Ser Ile Arg Asp Lys Tyr Phe Lys Lys Gly

260 265 270

His Glu Asn Ala Val Asn Thr Gly Phe Phe Lys Tyr Gln Gly Lys Ile

275 280 285

Val Lys Gly Pro Ile Lys Phe Val Asn Asn Glu Leu Asp Phe Ala Arg

290 295 300

Lys Pro Asp Leu Lys Ser Met Lys Ile Lys Arg Ala Gly Phe Ala Ile

305 310 315 320

Pro Ser Ala Lys Arg Leu Ser Lys Glu Asp Arg Glu Ile Asn Arg Glu

325 330 335

Ser Ile Lys Ile Lys Asn Lys Ile Tyr Ser Leu Ser Tyr Gly Arg Lys

340 345 350

Lys Thr Leu Ser Asp Lys Asp Ile Ile Lys His Leu Tyr Arg Pro Val

355 360 365

Arg Gln Lys Gly Val Lys Pro Leu Glu Tyr Arg Lys Ala Pro Asp Gly

370 375 380

Phe Leu Glu Phe Phe Tyr Ser Leu Lys Arg Lys Glu Arg Arg Leu Arg

385 390 395 400

Lys Gln Lys Glu Lys Arg Gln Lys Asp Met Ser Glu Ile Ile Asp Ala

405 410 415

Ala Asp Glu Phe Ala Trp His Arg His Thr Gly Ser Ile Lys Lys Thr

420 425 430

Thr Asn His Ile Asn Phe Lys Ser Glu Val Lys Arg Gly Lys Val Pro

435 440 445

Ile Met Lys Lys Arg Ile Ala Asn Asp Ser Phe Asn Thr Arg His Cys

450 455 460

Gly Lys Cys Val Lys Gln Gly Asn Ala Ile Asn Lys Tyr Tyr Ile Glu

465 470 475 480

Lys Gln Lys Asn Cys Phe Asp Cys Asn Ser Ile Glu Phe Lys Trp Glu

485 490 495

Lys Ala Ala Leu Glu Lys Lys Gly Ala Phe Lys Leu Asn Lys Arg Leu

500 505 510

Gln Tyr Ile Val Lys Ala Cys Phe Asn Val Ala Lys Ala Tyr Glu Ser

515 520 525

Phe Tyr Glu Asp Phe Arg Lys Gly Glu Glu Glu Ser Leu Asp Leu Lys

530 535 540

Phe Lys Ile Gly Thr Thr Thr Thr Leu Lys Gln Tyr Pro Gln Asn Lys

545 550 555 560

Ala Arg Ala Met

<210> 186

<211> 610

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 186

His Ser His Asn Leu Met Leu Thr Lys Leu Gly Lys Gln Ala Ile Asn
1 5 10 15

Phe Ala Ala Asn Tyr Asp Ala Asn Leu Glu Ile Gly Cys Lys Asn Cys
20 25 30

Lys Phe Leu Ser Tyr Ser Pro Lys Gln Ala Asn Pro Lys Lys Tyr Pro
35 40 45

Arg Gln Thr Asp Val His Glu Asp Gly Asn Ile Ala Cys His Ser Cys
50 55 60

Met Gln Ser Thr Lys Glu Pro Pro Val Tyr Ile Val Pro Ile Gly Glu
65 70 75 80

Arg Lys Ser Lys Tyr Glu Ile Leu Thr Ser Leu Asn Lys Phe Thr Phe
85 90 95

Leu Ala Leu Lys Tyr Lys Glu Lys Lys Arg Gln Ala Phe Arg Ala Lys
100 105 110

Lys Pro Lys Glu Leu Gln Glu Leu Ala Ile Ala Phe Asn Lys Glu Lys
115 120 125

Ala Ile Lys Val Ile Asp Lys Ser Ile Gln His Leu Ile Leu Asn Ile
130 135 140

Lys Pro Glu Ile Ala Arg Ile Gln Arg Gln Lys Arg Leu Lys Asn Arg
145 150 155 160

Lys Gly Lys Leu Leu Tyr Leu His Lys Arg Tyr Ala Ile Lys Met Gly
165 170 175

Leu Ile Lys Asn Gly Lys Tyr Phe Lys Val Gly Ser Pro Lys Lys Asp
180 185 190

Gly Lys Lys Leu Leu Val Leu Cys Ala Leu Asn Thr Ile Gly Arg Asp
195 200 205

Ile Gly Ile Ile Gly Asn Ile Glu Glu Asn Asn Arg Ser Glu Thr Glu
210 215 220

Ile Thr Tyr Gln Leu Tyr Phe Asp Cys Leu Asp Ala Asn Pro Asn Glu

225 230 235 240

Leu Arg Ile Lys Glu Ile Glu Tyr Asn Arg Leu Lys Ser Tyr Glu Arg

245 250 255

Lys Ile Lys Arg Leu Val Tyr Ala Tyr Lys Pro Lys Gln Thr Lys Ile

260 265 270

Leu Glu Ile Arg Ser Lys Phe Phe Ser Lys Gly His Glu Asn Lys Val

275 280 285

Asn Thr Gly Ser Phe Asn Phe Glu Asn Pro Leu Asn Lys Ser Ile Ser

290 295 300

Ile Lys Val Lys Asn Ser Ala Phe Asp Phe Lys Ile Gly Ala Pro Phe

305 310 315 320

Ile Met Leu Arg Asn Gly Lys Phe His Ile Pro Thr Lys Lys Arg Leu

325 330 335

Ser Lys Glu Glu Arg Glu Ile Asn Arg Thr Leu Ser Lys Ile Lys Gly

340 345 350

Arg Val Phe Arg Leu Thr Tyr Gly Arg Asn Ile Ser Glu Gln Gly Ser

355 360 365

Lys Ser Leu His Ile Tyr Arg Lys Glu Arg Gln His Pro Lys Leu Ser

370 375 380

Leu Glu Ile Arg Lys Gln Pro Asp Ser Phe Ile Asp Glu Phe Glu Lys

385 390 395 400

Leu Arg Leu Lys Gln Asn Phe Ile Ser Lys Leu Lys Lys Gln Arg Gln

405 410 415

Lys Lys Leu Ala Asp Leu Leu Gln Phe Ala Asp Arg Ile Ala Tyr Asn

420 425 430

Tyr His Thr Ser Ser Leu Glu Lys Thr Ser Asn Phe Ile Asn Tyr Lys

435 440 445

Pro Glu Val Lys Arg Gly Arg Thr Ser Tyr Ile Lys Lys Arg Ile Gly

450 455 460

Asn Glu Gly Phe Glu Lys Leu Tyr Cys Glu Thr Cys Ile Lys Ser Asn

465 470 475 480

Asp Lys Glu Asn Ala Tyr Ala Val Glu Lys Glu Glu Leu Cys Phe Val

485 490 495

Cys Lys Ala Lys Pro Phe Thr Trp Lys Lys Thr Asn Lys Asp Lys Leu

500 505 510

Gly Ile Phe Lys Tyr Pro Ser Arg Ile Lys Asp Phe Ile Arg Ala Ala

515 520 525

Phe Thr Val Ala Lys Ser Tyr Asn Asp Phe Tyr Glu Asn Leu Lys Lys

530 535 540

Lys Asp Leu Lys Asn Glu Ile Phe Leu Lys Phe Lys Ile Gly Leu Ile

545 550 555 560

Leu Ser His Glu Lys Lys Asn His Ile Ser Ile Ala Lys Ser Val Ala

565 570 575

Glu Asp Glu Arg Ile Ser Gly Lys Ser Ile Lys Asn Ile Leu Asn Lys

580 585 590

Ser Ile Lys Leu Glu Lys Asn Cys Tyr Ser Cys Phe Phe His Lys Glu

595 600 605

Asp Met

610

<210> 187

<211> 552

<212

> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 187

Ser Leu Glu Arg Val Ile Asp Lys Arg Asn Leu Ala Lys Lys Ala Ile

1 5 10 15

Asn Ile Ala Ala Asn Phe Asp Ala Asn Ile Asn Lys Gly Phe Tyr Arg

20 25 30

Cys Glu Thr Asn Gln Cys Met Phe Ile Ala Gln Lys Pro Arg Lys Thr

35 40 45

Asn Asn Thr Gly Cys Ser Ser Cys Leu Gln Ser Thr Tyr Asp Pro Val

50 55 60

Ile Tyr Val Val Lys Val Gly Glu Met Leu Ala Lys Tyr Glu Ile Leu
65 70 75 80

Lys Ser Leu Lys Arg Phe Val Phe Met Asn Arg Ser Phe Lys Gln Lys
85 90 95

Lys Thr Glu Lys Ala Lys Gln Lys Glu Arg Ile Gly Gly Glu Leu Asn
100 105 110

Glu Met Ser Ile Phe Ala Asn Ala Ala Leu Ala Met Gly Val Ile Lys

115 120 125

Arg Ala Ile Arg His Cys His Val Asp Ile Arg Pro Glu Ile Asn Arg
130 135 140

Leu Ser Glu Leu Lys Lys Thr Lys His Arg Val Ala Ala Lys Ser Leu
145 150 155 160

Val Lys Ile Val Lys Gln Arg Lys Thr Lys Trp Lys Gly Ile Pro Asn
165 170 175

Ser Phe Ile Gln Ile Pro Gln Lys Ala Arg Asn Lys Asp Ala Asp Phe

180 185 190

Tyr Val Ala Ser Ala Leu Lys Ser Gly Gly Ile Asp Ile Gly Leu Cys
195 200 205

Gly Thr Tyr Asp Lys Lys Pro His Ala Asp Pro Arg Trp Thr Tyr Gln
210 215 220

Leu Tyr Phe Asp Thr Glu Asp Glu Ser Glu Lys Arg Leu Leu Tyr Cys
225 230 235 240

Tyr Asn Asp Pro Gln Ala Lys Ile Arg Asp Phe Trp Lys Thr Phe Tyr

245 250 255

Glu Arg Gly Asn Pro Ser Met Val Asn Ser Pro Gly Thr Ile Glu Phe

260 265 270

Arg Met Glu Gly Phe Phe Glu Lys Met Thr Pro Ile Ser Ile Glu Ser

275 280 285

Lys Asp Phe Asp Phe Arg Val Trp Asn Lys Asp Leu Leu Ile Arg Arg

290 295 300

Gly Leu Tyr Glu Ile Lys Lys Arg Lys Asn Leu Asn Arg Lys Ala Arg

305 310 315 320

Glu Ile Lys Lys Ala Met Gly Ser Val Lys Arg Val Leu Ala Asn Met

325 330 335

Thr Tyr Gly Lys Ser Pro Thr Asp Lys Lys Ser Ile Pro Val Tyr Arg

340 345 350

Val Glu Arg Glu Lys Pro Lys Lys Pro Arg Ala Val Arg Lys Glu Glu

355 360 365

Asn Glu Leu Ala Asp Lys Leu Glu Asn Tyr Arg Arg Glu Asp Phe Leu

370 375 380

Ile Arg Asn Arg Arg Lys Arg Glu Ala Thr Glu Ile Ala Lys Ile Ile

385 390 395 400

Asp Ala Ala Glu Pro Pro Ile Arg His Tyr His Thr Asn His Leu Arg

405 410 415

Ala Val Lys Arg Ile Asp Leu Ser Lys Pro Val Ala Arg Lys Asn Thr

420 425 430

Ser Val Phe Leu Lys Arg Ile Met Gln Asn Gly Tyr Arg Gly Asn Tyr

435 440 445

Cys Lys Lys Cys Ile Lys Gly Asn Ile Asp Pro Asn Lys Asp Glu Cys

450 455 460

Arg Leu Glu Asp Ile Lys Lys Cys Ile Cys Cys Glu Gly Thr Gln Asn

465 470 475 480

Ile Trp Ala Lys Lys Glu Lys Leu Tyr Thr Gly Arg Ile Asn Val Leu

485 490 495

Asn Lys Arg Ile Lys Gln Met Lys Leu Glu Cys Phe Asn Val Ala Lys

500 505 510

Ala Tyr Glu Asn Phe Tyr Asp Asn Leu Ala Ala Leu Lys Glu Gly Asp

515 520 525

Leu Lys Val Leu Lys Leu Lys Val Ser Ile Pro Ala Leu Asn Pro Glu

530 535 540

Ala Ser Asp Pro Glu Glu Asp Met
545 550

<210> 188

<211> 534

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 188

Asn Ala Ser Ile Asn Leu Gly Lys Arg Ala Ile Asn Leu Ser Ala Asn
1 5 10 15

Tyr Asp Ser Asn Leu Val Ile Gly Cys Lys Asn Cys Lys Phe Leu Ser
20 25 30

Phe Asn Gly Asn Phe Pro Arg Gln Thr Asn Val Arg Glu Gly Cys His
35 40 45

Ser Cys Asp Lys Ser Thr Tyr Ala Pro Glu Val Tyr Ile Val Lys Ile
50 55 60

Gly Glu Arg Lys Ala Lys Tyr Asp Val Leu Asp Ser Leu Lys Lys Phe
65 70 75 80

Thr Phe Gln Ser Leu Lys Tyr Gln Ile Lys Lys Ser Met Arg Glu Arg
85 90 95

Ser Lys Lys Pro Lys Glu Leu Leu Glu Phe Val Ile Phe Ala Asn Lys
100 105 110

Asp Lys Ala Phe Asn Val Ile Gln Lys Ser Tyr Glu His Leu Ile Leu
115 120 125

Asn Ile Lys Gln Glu Ile Asn Arg Met Asn Gly Lys Lys Arg Ile Lys
130 135 140

Asn His Lys Lys Arg Leu Phe Lys Asp Arg Glu Lys Gln Leu Asn Lys
145 150 155 160

Leu Arg Leu Ile Gly Ser Ser Ser Leu Phe Phe Pro Arg Glu Asn Lys
165 170 175

Gly Asp Lys Asp Leu Phe Thr Tyr Val Ala Ile His Ser Val Gly Arg

180 185 190

Asp Ile Gly Val Ala Gly Ser Tyr Glu Ser His Ile Glu Pro Ile Ser
195 200 205
Asp Leu Thr Tyr Gln Leu Phe Ile Asn Asn Glu Lys Arg Leu Leu Tyr
210 215 220
Ala Tyr Lys Pro Lys Gln Asn Lys Ile Ile Glu Leu Lys Glu Asn Leu
225 230 235 240
Trp Asn Leu Lys Lys Glu Lys Lys Pro Leu Asp Leu Glu Phe Thr Lys
245 250 255

Pro Leu Glu Lys Ser Ile Thr Phe Ser Val Lys Asn Asp Lys Leu Phe
260 265 270
Lys Val Ser Lys Asp Leu Met Leu Arg Gln Ala Lys Phe Asn Ile Gln
275 280 285
Gly Lys Glu Lys Leu Ser Lys Glu Glu Arg Gln Ile Asn Arg Asp Phe
290 295 300
Ser Lys Ile Lys Ser Asn Val Ile Ser Leu Ser Tyr Gly Arg Phe Glu
305 310 315 320

Glu Leu Lys Lys Glu Lys Asn Ile Trp Ser Pro His Ile Tyr Arg Glu
325 330 335
Val Lys Gln Lys Glu Ile Lys Pro Cys Ile Val Arg Lys Gly Asp Arg
340 345 350
Ile Glu Leu Phe Glu Gln Leu Lys Arg Lys Met Asp Lys Leu Lys Lys
355 360 365
Phe Arg Lys Glu Arg Gln Lys Lys Ile Ser Lys Asp Leu Asn Phe Ala
370 375 380

Glu Arg Ile Ala Tyr Asn Phe His Thr Lys Ser Ile Lys Asn Thr Ser
385 390 395 400
Asn Lys Ile Asn Ile Asp Gln Glu Ala Lys Arg Gly Lys Ala Ser Tyr
405 410 415
Met Arg Lys Arg Ile Gly Asn Glu Ser Phe Arg Lys Lys Tyr Cys Glu
420 425 430

Gln Cys Phe Ser Val Gly Asn Val Tyr His Asn Val Gln Asn Gly Cys
435 440 445

Ser Cys Phe Asp Asn Pro Ile Glu Leu Ile Lys Lys Gly Asp Glu Gly
450 455 460
Leu Ile Pro Lys Gly Lys Glu Asp Arg Lys Tyr Lys Gly Ala Leu Arg
465 470 475 480
Asp Asp Asn Leu Gln Met Gln Ile Ile Arg Val Ala Phe Asn Ile Ala
485 490 495
Lys Gly Tyr Glu Asp Phe Tyr Asn Asn Leu Lys Glu Lys Thr Glu Lys
500 505 510

Asp Leu Lys Leu Lys Phe Lys Ile Gly Thr Thr Ile Ser Thr Gln Glu
515 520 525
Ser Asn Asn Lys Glu Met
530

<210> 189
<211> 577
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 189
Ser Asn Leu Ile Lys Leu Gly Lys Gln Ala Ile Asn Phe Ala Ala Asn
1 5 10 15
Tyr Asp Ala Asn Leu Glu Val Gly Cys Lys Asn Cys Lys Phe Leu Ser
20 25 30

Ser Thr Asn Lys Tyr Pro Arg Gln Thr Asn Val His Leu Asp Asn Lys
35 40 45
Met Ala Cys Arg Ser Cys Asn Gln Ser Thr Met Glu Pro Ala Ile Tyr
50 55 60
Ile Val Arg Ile Gly Glu Lys Lys Ala Lys Tyr Asp Ile Tyr Asn Ser
65 70 75 80
Leu Thr Lys Phe Asn Phe Gln Ser Leu Lys Tyr Lys Ala Lys Arg Ser

85 90 95

Gln Arg Phe Lys Pro Lys Gln Pro Lys Glu Leu Gln Glu Leu Ser Ile
100 105 110
Ala Val Arg Lys Glu Lys Ala Leu Asp Ile Ile Gln Lys Ser Ile Asp
115 120 125
His Leu Ile Gln Asp Ile Arg Pro Glu Ile Pro Arg Ile Lys Gln Gln
130 135 140
Lys Arg Tyr Lys Asn His Val Gly Lys Leu Phe Tyr Leu Gln Lys Arg
145 150 155 160

Arg Lys Asn Lys Leu Asn Leu Ile Gly Lys Gly Ser Phe Phe Lys Val
165 170 175
Phe Ser Pro Lys Glu Lys Lys Asn Glu Leu Leu Val Ile Cys Ala Leu
180 185 190
Thr Asn Ile Gly Arg Asp Ile Gly Leu Ile Gly Asn Tyr Asn Thr Ile
195 200 205
Ile Asn Pro Leu Phe Glu Val Thr Tyr Gln Leu Tyr Tyr Asp Tyr Ile
210 215 220

Pro Lys Lys Asn Asn Lys Asn Val Gln Arg Arg Leu Leu Tyr Ala Tyr
225 230 235 240
Lys Ser Lys Asn Glu Lys Ile Leu Lys Leu Lys Glu Ala Phe Phe Lys
245 250 255
Arg Gly His Glu Asn Ala Val Asn Leu Gly Ser Phe Ser Tyr Glu Lys
260 265 270
Pro Leu Glu Lys Ser Leu Thr Leu Lys Ile Lys Asn Asp Lys Asp Asp
275 280 285

Phe Gln Val Ser Pro Ser Leu Arg Ile Arg Thr Gly Arg Phe Phe Val
290 295 300
Pro Ser Lys Arg Asn Leu Ser Arg Gln Glu Arg Glu Ile Asn Arg Arg
305 310 315 320
Leu Val Lys Ile Lys Ser Lys Ile Lys Asn Met Thr Tyr Gly Lys Phe
325 330 335

Glu Thr Ala Arg Asp Lys Gln Ser Val His Ile Phe Arg Leu Glu Arg
340 345 350

Gln Lys Glu Lys Leu Pro Leu Gln Phe Arg Lys Asp Glu Lys Glu Phe
355 360 365
Met Glu Glu Phe Gln Lys Leu Lys Arg Arg Thr Asn Ser Leu Lys Lys
370 375 380
Leu Arg Lys Ser Arg Gln Lys Lys Leu Ala Asp Leu Leu Gln Leu Ser
385 390 395 400
Glu Lys Val Val Tyr Asn Asn His Thr Gly Thr Leu Lys Lys Thr Ser
405 410 415

Asn Phe Leu Asn Phe Ser Ser Ser Val Lys Arg Gly Lys Thr Ala Tyr
420 425 430
Ile Lys Glu Leu Leu Gly Gln Glu Gly Phe Glu Thr Leu Tyr Cys Ser
435 440 445
Asn Cys Ile Asn Lys Gly Gln Lys Thr Arg Tyr Asn Ile Glu Thr Lys
450 455 460
Glu Lys Cys Phe Ser Cys Lys Asp Val Pro Phe Val Trp Lys Lys Lys
465 470 475 480

Ser Thr Asp Lys Asp Arg Lys Gly Ala Phe Leu Phe Pro Ala Lys Leu
485 490 495
Lys Asp Val Ile Lys Ala Thr Phe Thr Val Ala Lys Ala Tyr Glu Asp
500 505 510
Phe Tyr Asp Asn Leu Lys Ser Ile Asp Glu Lys Lys Pro Tyr Ile Lys
515 520 525
Phe Lys Ile Gly Leu Ile Leu Ala His Val Arg His Glu His Lys Ala
530 535 540

Arg Ala Lys Glu Glu Ala Gly Gln Lys Asn Ile Tyr Asn Lys Pro Ile
545 550 555 560
Lys Ile Asp Lys Asn Cys Lys Glu Cys Phe Phe Phe Lys Glu Glu Ala
565 570 575
Met

<210> 190

<211> 613

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 190

Asn Thr Thr Arg Lys Lys Phe Arg Lys Arg Thr Gly Phe Pro Gln Ser
1 5 10 15

Asp Asn Ile Lys Leu Ala Tyr Cys Ser Ala Ile Val Arg Ala Ala Asn
20 25 30

Leu Asp Ala Asp Ile Gln Lys Lys His Asn Gln Cys Asn Pro Asn Leu
35 40 45

Cys Val Gly Ile Lys Ser Asn Glu Gln Ser Arg Lys Tyr Glu His Ser
50 55 60

Asp Arg Gln Ala Leu Leu Cys Tyr Ala Cys Asn Gln Ser Thr Gly Ala
65 70 75 80

Pro Lys Val Asp Tyr Ile Gln Ile Gly Glu Ile Gly Ala Lys Tyr Lys
85 90 95

Ile Leu Gln Met Val Asn Ala Tyr Asp Phe Leu Ser Leu Ala Tyr Asn
100 105 110

Leu Thr Lys Leu Arg Asn Gly Lys Ser Arg Gly His Gln Arg Met Ser
115 120 125

Gln Leu Asp Glu Val Val Ile Val Ala Asp Tyr Glu Lys Ala Thr Glu
130 135 140

Val Ile Lys Arg Ser Ile Asn His Leu Leu Asp Asp Ile Arg Gly Gln
145 150 155 160

Leu Ser Lys Leu Lys Lys Arg Thr Gln Asn Glu His Ile Thr Glu His
165 170 175

Lys Gln Ser Lys Ile Arg Arg Lys Leu Arg Lys Leu Ser Arg Leu Leu
180 185 190

Lys Arg Arg Arg Trp Lys Trp Gly Thr Ile Pro Asn Pro Tyr Leu Lys

195 200 205

Asn Trp Val Phe Thr Lys Lys Asp Pro Glu Leu Val Thr Val Ala Leu
210 215 220
Leu His Lys Leu Gly Arg Asp Ile Gly Leu Val Asn Arg Ser Lys Arg
225 230 235 240
Arg Ser Lys Gln Lys Leu Leu Pro Lys Val Gly Phe Gln Leu Tyr Tyr
245 250 255
Lys Trp Glu Ser Pro Ser Leu Asn Asn Ile Lys Lys Ser Lys Ala Lys
260 265 270

Lys Leu Pro Lys Arg Leu Leu Ile Pro Tyr Lys Asn Val Lys Leu Phe
275 280 285
Asp Asn Lys Gln Lys Leu Glu Asn Ala Ile Lys Ser Leu Leu Glu Ser
290 295 300
Tyr Gln Lys Thr Ile Lys Val Glu Phe Asp Gln Phe Phe Gln Asn Arg
305 310 315 320
Thr Glu Glu Ile Ile Ala Glu Glu Gln Gln Thr Leu Glu Arg Gly Leu
325 330 335

Leu Lys Gln Leu Glu Lys Lys Lys Asn Glu Phe Ala Ser Gln Lys Lys
340 345 350
Ala Leu Lys Glu Glu Lys Lys Lys Ile Lys Glu Pro Arg Lys Ala Lys
355 360 365
Leu Leu Met Glu Glu Ser Arg Ser Leu Gly Phe Leu Met Ala Asn Val
370 375 380
Ser Tyr Ala Leu Phe Asn Thr Thr Ile Glu Asp Leu Tyr Lys Lys Ser
385 390 395 400

Asn Val Val Ser Gly Cys Ile Pro Gln Glu Pro Val Val Val Phe Pro
405 410 415
Ala Asp Ile Gln Asn Lys Gly Ser Leu Ala Lys Ile Leu Phe Ala Pro
420 425 430
Lys Asp Gly Phe Arg Ile Lys Phe Ser Gly Gln His Leu Thr Ile Arg
435 440 445

Thr Ala Lys Phe Lys Ile Arg Gly Lys Glu Ile Lys Ile Leu Thr Lys
450 455 460

Thr Lys Arg Glu Ile Leu Lys Asn Ile Glu Lys Leu Arg Arg Val Trp
465 470 475 480
Tyr Arg Glu Gln His Tyr Lys Leu Lys Leu Phe Gly Lys Glu Val Ser
485 490 495
Ala Lys Pro Arg Phe Leu Asp Lys Arg Lys Thr Ser Ile Glu Arg Arg
500 505 510
Asp Pro Asn Lys Leu Ala Asp Gln Thr Asp Asp Arg Gln Ala Glu Leu
515 520 525

Arg Asn Lys Glu Tyr Glu Leu Arg His Lys Gln His Lys Met Ala Glu
530 535 540
Arg Leu Asp Asn Ile Asp Thr Asn Ala Gln Asn Leu Gln Thr Leu Ser
545 550 555 560
Phe Trp Val Gly Glu Ala Asp Lys Pro Pro Lys Leu Asp Glu Lys Asp
565 570 575
Ala Arg Gly Phe Gly Val Arg Thr Cys Ile Ser Ala Trp Lys Trp Phe
580 585 590

Met Glu Asp Leu Leu Lys Lys Gln Glu Glu Asp Pro Leu Leu Lys Leu
595 600 605
Lys Leu Ser Ile Met
610

<210> 191
<211> 615
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 191
Pro Lys Lys Pro Lys Phe Gln Lys Arg Thr Gly Phe Pro Gln Pro Asp
1 5 10 15
Asn Leu Arg Lys Glu Tyr Cys Leu Ala Ile Val Arg Ala Ala Asn Leu

20 25 30

Asp Ala Asp Phe Glu Lys Lys Cys Thr Lys Cys Glu Gly Ile Lys Thr
35 40 45
Asn Lys Lys Gly Asn Ile Val Lys Gly Arg Thr Tyr Asn Ser Ala Asp
50 55 60
Lys Asp Asn Leu Leu Cys Tyr Ala Cys Asn Ile Ser Thr Gly Ala Pro
65 70 75 80
Ala Val Asp Tyr Val Phe Val Gly Ala Leu Glu Ala Lys Tyr Lys Ile
85 90 95

Leu Gln Met Val Lys Ala Tyr Asp Phe His Ser Leu Ala Tyr Asn Leu
100 105 110
Ala Lys Leu Trp Lys Gly Arg Gly Arg Gly His Gln Arg Met Gly Gly
115 120 125
Leu Asn Glu Val Val Ile Val Ser Asn Asn Glu Lys Ala Leu Asp Val
130 135 140
Ile Glu Lys Ser Leu Asn His Phe His Asp Glu Ile Arg Gly Glu Leu
145 150 155 160

Ser Arg Leu Lys Ala Lys Phe Gln Asn Glu His Leu His Val His Lys
165 170 175
Glu Ser Lys Leu Arg Arg Lys Leu Arg Lys Ile Ser Arg Leu Leu Lys
180 185 190
Arg Arg Arg Trp Lys Trp Asp Val Ile Pro Asn Ser Tyr Leu Arg Asn
195 200 205
Phe Thr Phe Thr Lys Thr Arg Pro Asp Phe Ile Ser Val Ala Leu Leu
210 215 220

His Arg Val Gly Arg Asp Ile Gly Leu Val Thr Lys Thr Lys Ile Pro
225 230 235 240
Lys Pro Thr Asp Leu Leu Pro Gln Phe Gly Phe Gln Ile Tyr Tyr Thr
245 250 255
Trp Asp Glu Pro Lys Leu Asn Lys Leu Lys Lys Ser Arg Leu Arg Ser
260 265 270

Glu Pro Lys Arg Leu Leu Val Pro Tyr Lys Lys Ile Glu Leu Tyr Lys
275 280 285

Asn Lys Ser Val Leu Glu Glu Ala Ile Arg His Leu Ala Glu Val Tyr
290 295 300
Thr Glu Asp Leu Thr Ile Cys Phe Lys Asp Phe Phe Glu Thr Gln Lys
305 310 315 320
Arg Lys Phe Val Ser Lys Glu Lys Glu Ser Leu Lys Arg Glu Leu Leu
325 330 335
Lys Glu Leu Thr Lys Leu Lys Lys Asp Phe Ser Glu Arg Lys Thr Ala
340 345 350

Leu Lys Arg Asp Arg Lys Glu Ile Lys Glu Pro Lys Lys Ala Lys Leu
355 360 365
Leu Met Glu Glu Ser Arg Ser Leu Gly Phe Leu Ala Ala Asn Thr Ser
370 375 380
Tyr Ala Leu Phe Asn Leu Ile Ala Ala Asp Leu Tyr Thr Lys Ser Lys
385 390 395 400
Lys Ala Cys Ser Thr Lys Leu Pro Arg Gln Leu Ser Thr Ile Leu Pro
405 410 415

Leu Glu Ile Lys Glu His Lys Ser Thr Thr Ser Leu Ala Ile Lys Pro
420 425 430
Glu Glu Gly Phe Lys Ile Arg Phe Ser Asn Thr His Leu Ser Ile Arg
435 440 445
Thr Pro Lys Phe Lys Met Lys Gly Ala Asp Ile Lys Ala Leu Thr Lys
450 455 460
Arg Lys Arg Glu Ile Leu Lys Asn Ala Thr Lys Leu Glu Lys Ser Trp
465 470 475 480

Tyr Gly Leu Lys His Tyr Lys Leu Lys Leu Tyr Gly Lys Glu Val Ala
485 490 495
Ala Lys Pro Arg Phe Leu Asp Lys Arg Asn Pro Ser Ile Asp Arg Arg
500 505 510
Asp Pro Lys Glu Leu Met Glu Gln Ile Glu Asn Arg Arg Asn Glu Val

515 520 525

Lys Asp Leu Glu Tyr Glu Ile Arg Lys Gly Gln His Gln Met Ala Lys

530 535 540

Arg Leu Asp Asn Val Asp Thr Asn Ala Gln Asn Leu Gln Thr Lys Ser

545 550 555 560

Phe Trp Val Gly Glu Ala Asp Lys Pro Pro Glu Leu Asp Ser Met Glu

565 570 575

Ala Lys Lys Leu Gly Leu Arg Thr Cys Ile Ser Ala Trp Lys Trp Phe

580 585 590

Met Lys Asp Leu Val Leu Leu Gln Glu Lys Ser Pro Asn Leu Lys Leu

595 600 605

Lys Leu Ser Leu Thr Glu Met

610 615

<210> 192

<211> 775

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 192

Lys Phe Ser Lys Arg Gln Glu Gly Phe Leu Ile Pro Asp Asn Ile Asp

1 5 10 15

Leu Tyr Lys Cys Leu Ala Ile Val Arg Ser Ala Asn Leu Asp Ala Asp

20 25 30

Val Gln Gly His Lys Ser Cys Tyr Gly Val Lys Lys Asn Gly Thr Tyr

35 40 45

Arg Val Lys Gln Asn Gly Lys Lys Gly Val Lys Glu Lys Gly Arg Lys

50 55 60

Tyr Val Phe Asp Leu Ile Ala Phe Lys Gly Asn Ile Glu Lys Ile Pro

65 70 75 80

His Glu Ala Ile Glu Glu Lys Asp Gln Gly Arg Val Ile Val Leu Gly

85 90 95

Lys Phe Asn Tyr Lys Leu Ile Leu Asn Ile Glu Lys Asn His Asn Asp

100 105 110

Arg Ala Ser Leu Glu Ile Lys Asn Lys Ile Lys Lys Leu Val Gln Ile
115 120 125
Ser Ser Leu Glu Thr Gly Glu Phe Leu Ser Asp Leu Leu Ser Gly Lys
130 135 140
Ile Gly Ile Asp Glu Val Tyr Gly Ile Ile Glu Pro Asp Val Phe Ser
145 150 155 160
Gly Lys Glu Leu Val Cys Lys Ala Cys Gln Gln Ser Thr Tyr Ala Pro
165 170 175

Leu Val Glu Tyr Met Pro Val Gly Glu Leu Asp Ala Lys Tyr Lys Ile
180 185 190
Leu Ser Ala Ile Lys Gly Tyr Asp Phe Leu Ser Leu Ala Tyr Asn Leu
195 200 205
Ser Arg Asn Arg Ala Asn Lys Lys Arg Gly His Gln Lys Leu Gly Gly
210 215 220
Gly Glu Leu Ser Glu Val Val Ile Ser Ala Asn Tyr Asp Lys Ala Leu
225 230 235 240

Asn Val Ile Lys Arg Ser Ile Asn His Tyr His Val Glu Ile Lys Pro
245 250 255
Glu Ile Ser Lys Leu Lys Lys Lys Met Gln Asn Glu Pro Leu Lys Val
260 265 270
Met Lys Gln Ala Arg Ile Arg Arg Glu Leu His Gln Leu Ser Arg Lys
275 280 285
Val Lys Arg Leu Lys Trp Lys Trp Gly Met Ile Pro Asn Pro Glu Leu
290 295 300

Gln Asn Ile Ile Phe Glu Lys Lys Glu Lys Asp Phe Val Ser Tyr Ala
305 310 315 320
Leu Leu His Thr Leu Gly Arg Asp Ile Gly Leu Phe Lys Asp Thr Ser
325 330 335
Met Leu Gln Val Pro Asn Ile Ser Asp Tyr Gly Phe Gln Ile Tyr Tyr
340 345 350

Ser Trp Glu Asp Pro Lys Leu Asn Ser Ile Lys Lys Ile Lys Asp Leu
355 360 365

Pro Lys Arg Leu Leu Ile Pro Tyr Lys Arg Leu Asp Phe Tyr Ile Asp
370 375 380
Thr Ile Leu Val Ala Lys Val Ile Lys Asn Leu Ile Glu Leu Tyr Arg
385 390 395 400
Lys Ser Tyr Val Tyr Glu Thr Phe Gly Glu Glu Tyr Gly Tyr Ala Lys
405 410 415
Lys Ala Glu Asp Ile Leu Phe Asp Trp Asp Ser Ile Asn Leu Ser Glu
420 425 430

Gly Ile Glu Gln Lys Ile Gln Lys Ile Lys Asp Glu Phe Ser Asp Leu
435 440 445
Leu Tyr Glu Ala Arg Glu Ser Lys Arg Gln Asn Phe Val Glu Ser Phe
450 455 460
Glu Asn Ile Leu Gly Leu Tyr Asp Lys Asn Phe Ala Ser Asp Arg Asn
465 470 475 480
Ser Tyr Gln Glu Lys Ile Gln Ser Met Ile Ile Lys Lys Gln Gln Glu
485 490 495

Asn Ile Glu Gln Lys Leu Lys Arg Glu Phe Lys Glu Val Ile Glu Arg
500 505 510
Gly Phe Glu Gly Met Asp Gln Asn Lys Lys Tyr Tyr Lys Val Leu Ser
515 520 525
Pro Asn Ile Lys Gly Gly Leu Leu Tyr Thr Asp Thr Asn Asn Leu Gly
530 535 540
Phe Phe Arg Ser His Leu Ala Phe Met Leu Leu Ser Lys Ile Ser Asp
545 550 555 560

Asp Leu Tyr Arg Lys Asn Asn Leu Val Ser Lys Gly Gly Asn Lys Gly
565 570 575
Ile Leu Asp Gln Thr Pro Glu Thr Met Leu Thr Leu Glu Phe Gly Lys
580 585 590
Ser Asn Leu Pro Asn Ile Ser Ile Lys Arg Lys Phe Phe Asn Ile Lys

595 600 605

Tyr Asn Ser Ser Trp Ile Gly Ile Arg Lys Pro Lys Phe Ser Ile Lys

610 615 620

Gly Ala Val Ile Arg Glu Ile Thr Lys Lys Val Arg Asp Glu Gln Arg

625 630 635 640

Leu Ile Lys Ser Leu Glu Gly Val Trp His Lys Ser Thr His Phe Lys

645 650 655

Arg Trp Gly Lys Pro Arg Phe Asn Leu Pro Arg His Pro Asp Arg Glu

660 665 670

Lys Asn Asn Asp Asp Asn Leu Met Glu Ser Ile Thr Ser Arg Arg Glu

675 680 685

Gln Ile Gln Leu Leu Leu Arg Glu Lys Gln Lys Gln Gln Glu Lys Met

690 695 700

Ala Gly Arg Leu Asp Lys Ile Asp Lys Glu Ile Gln Asn Leu Gln Thr

705 710 715 720

Ala Asn Phe Gln Ile Lys Gln Ile Asp Lys Lys Pro Ala Leu Thr Glu

725 730 735

Lys Ser Glu Gly Lys Gln Ser Val Arg Asn Ala Leu Ser Ala Trp Lys

740 745 750

Trp Phe Met Glu Asp Leu Ile Lys Tyr Gln Lys Arg Thr Pro Ile Leu

755 760 765

Gln Leu Lys Leu Ala Lys Met

770 775

<210> 193

<211> 777

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 193

Lys Phe Ser Lys Arg Gln Glu Gly Phe Val Ile Pro Glu Asn Ile Gly

1 5 10 15

Leu Tyr Lys Cys Leu Ala Ile Val Arg Ser Ala Asn Leu Asp Ala Asp

20 25 30

Val Gln Gly His Val Ser Cys Tyr Gly Val Lys Lys Asn Gly Thr Tyr
35 40 45
Val Leu Lys Gln Asn Gly Lys Lys Ser Ile Arg Glu Lys Gly Arg Lys
50 55 60
Tyr Ala Ser Asp Leu Val Ala Phe Lys Gly Asp Ile Glu Lys Ile Pro
65 70 75 80
Phe Glu Val Ile Glu Glu Lys Lys Lys Glu Gln Ser Ile Val Leu Gly
85 90 95

Lys Phe Asn Tyr Lys Leu Val Leu Asp Val Met Lys Gly Glu Lys Asp
100 105 110
Arg Ala Ser Leu Thr Met Lys Asn Lys Ser Lys Lys Leu Val Gln Val
115 120 125
Ser Ser Leu Gly Thr Asp Glu Phe Leu Leu Thr Leu Leu Asn Glu Lys
130 135 140
Phe Gly Ile Glu Glu Ile Tyr Gly Ile Ile Glu Pro Glu Val Phe Ser
145 150 155 160

Gly Lys Lys Leu Val Cys Lys Ala Cys Gln Gln Ser Thr Tyr Ala Pro
165 170 175
Leu Val Glu Tyr Met Pro Val Gly Glu Leu Asp Ser Lys Tyr Lys Ile
180 185 190
Leu Ser Ala Ile Lys Gly Tyr Asp Phe Leu Ser Leu Ala Tyr Asn Leu
195 200 205
Ala Arg His Arg Ser Asn Lys Lys Arg Gly His Gln Lys Leu Gly Gly
210 215 220

Gly Glu Leu Ser Glu Val Val Ile Ser Ala Asn Asn Ala Lys Ala Leu
225 230 235 240
Asn Val Ile Lys Arg Ser Leu Asn His Tyr Tyr Ser Glu Ile Lys Pro
245 250 255
Glu Ile Ser Lys Leu Arg Lys Lys Met Gln Asn Glu Pro Leu Lys Val
260 265 270

Gly Lys Gln Ala Arg Met Arg Arg Glu Leu His Gln Leu Ser Arg Lys
275 280 285

Val Lys Arg Leu Lys Trp Lys Trp Gly Lys Ile Pro Asn Leu Glu Leu
290 295 300
Gln Asn Ile Thr Phe Lys Glu Ser Asp Arg Asp Phe Ile Ser Tyr Ala
305 310 315 320
Leu Leu His Thr Leu Gly Arg Asp Ile Gly Met Phe Asn Lys Thr Glu
325 330 335
Ile Lys Met Pro Ser Asn Ile Leu Gly Tyr Gly Phe Gln Ile Tyr Tyr
340 345 350

Asp Trp Glu Glu Pro Lys Leu Asn Thr Ile Lys Lys Ser Lys Asn Thr
355 360 365
Pro Lys Arg Ile Leu Ile Pro Tyr Lys Lys Leu Asp Phe Tyr Asn Asp
370 375 380
Ser Ile Leu Val Ala Arg Ala Ile Lys Glu Leu Val Gly Leu Phe Gln
385 390 395 400
Glu Ser Tyr Glu Trp Glu Ile Phe Gly Asn Glu Tyr Asn Tyr Ala Lys
405 410 415

Glu Ala Glu Val Glu Leu Ile Lys Leu Asp Glu Glu Ser Ile Asn Gly
420 425 430
Asn Val Glu Lys Lys Leu Gln Arg Ile Lys Glu Asn Phe Ser Asn Leu
435 440 445
Leu Glu Lys Ala Arg Glu Lys Lys Arg Gln Asn Phe Ile Glu Ser Phe
450 455 460
Glu Ser Ile Ala Arg Leu Tyr Asp Glu Ser Phe Thr Ala Asp Arg Asn
465 470 475 480

Glu Tyr Gln Arg Glu Ile Gln Ser Phe Ile Ile Glu Lys Gln Lys Gln
485 490 495
Ser Ile Glu Lys Lys Leu Lys Asn Glu Phe Lys Lys Ile Val Glu Lys
500 505 510
Lys Phe Asn Glu Gln Glu Gln Gly Lys Lys His Tyr Arg Val Leu Asn

515 520 525

Pro Thr Ile Ile Asn Glu Phe Leu Pro Lys Asp Lys Asn Asn Leu Gly

530 535 540

Phe Leu Arg Ser Lys Ile Ala Phe Ile Leu Leu Ser Lys Ile Ser Asp

545 550 555 560

Asp Leu Tyr Lys Lys Ser Asn Ala Val Ser Lys Gly Gly Glu Lys Gly

565 570 575

Ile Ile Lys Gln Gln Pro Glu Thr Ile Leu Asp Leu Glu Phe Ser Lys

580 585 590

Ser Lys Leu Pro Ser Ile Asn Ile Lys Lys Lys Leu Phe Asn Ile Lys

595 600 605

Tyr Thr Ser Ser Trp Leu Gly Ile Arg Lys Pro Lys Phe Asn Ile Lys

610 615 620

Gly Ala Lys Ile Arg Glu Ile Thr Arg Arg Val Arg Asp Val Gln Arg

625 630 635 640

Thr Leu Lys Ser Ala Glu Ser Ser Trp Tyr Ala Ser Thr His Phe Arg

645 650 655

Arg Trp Gly Phe Pro Arg Phe Asn Gln Pro Arg His Pro Asp Lys Glu

660 665 670

Lys Lys Ser Asp Asp Arg Leu Ile Glu Ser Ile Thr Leu Leu Arg Glu

675 680 685

Gln Ile Gln Ile Leu Leu Arg Glu Lys Gln Lys Gly Gln Lys Glu Met

690 695 700

Ala Gly Arg Leu Asp Asp Val Asp Lys Lys Ile Gln Asn Leu Gln Thr

705 710 715 720

Ala Asn Phe Gln Ile Lys Gln Thr Gly Asp Lys Pro Ala Leu Thr Glu

725 730 735

Lys Ser Ala Gly Lys Gln Ser Phe Arg Asn Ala Leu Ser Ala Trp Lys

740 745 750

Trp Phe Met Glu Asn Leu Leu Lys Tyr Gln Asn Lys Thr Pro Asp Leu

755 760 765

Lys Leu Lys Ile Ala Arg Thr Val Met

770 775

<210> 194

<211> 610

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 194

Lys Trp Ile Glu Pro Asn Asn Ile Asp Phe Asn Lys Cys Leu Ala Ile

1 5 10 15

Thr Arg Ser Ala Asn Leu Asp Ala Asp Val Gln Gly His Lys Met Cys

20 25 30

Tyr Gly Ile Lys Thr Asn Gly Thr Tyr Lys Ala Ile Gly Lys Ile Asn

35 40 45

Lys Lys His Asn Thr Gly Ile Ile Glu Lys Arg Arg Thr Tyr Val Tyr

50 55 60

Asp Leu Ile Val Thr Lys Glu Lys Asn Glu Lys Ile Val Lys Lys Thr

65 70 75 80

Asp Phe Met Ala Ile Asp Glu Glu Ile Glu Phe Asp Glu Lys Lys Glu

85 90 95

Lys Leu Leu Lys Lys Tyr Ile Lys Ala Glu Val Leu Gly Thr Gly Glu

100 105 110

Leu Ile Arg Lys Asp Leu Asn Asp Gly Glu Lys Phe Asp Asp Leu Cys

115 120 125

Ser Ile Glu Glu Pro Gln Ala Phe Arg Arg Ser Glu Leu Val Cys Lys

130 135 140

Ala Cys Asn Gln Ser Thr Tyr Ala Ser Asp Ile Arg Tyr Ile Pro Ile

145 150 155 160

Gly Glu Ile Glu Ala Lys Tyr Lys Ile Leu Lys Ala Ile Lys Gly Tyr

165 170 175

Asp Phe Leu Ser Leu Lys Tyr Asn Leu Gly Arg Leu Arg Asp Ser Lys

180 185 190

Lys Arg Gly His Gln Lys Met Gly Gln Gly Glu Leu Lys Glu Phe Val

195 200 205

Ile Cys Ala Asn Lys Glu Lys Ala Leu Asp Val Ile Lys Arg Ser Leu

210 215 220

Asn His Tyr Leu Asn Glu Val Lys Asp Glu Ile Ser Arg Leu Asn Lys

225 230 235 240

Lys Met Gln Asn Glu Pro Leu Lys Val Asn Asp Gln Ala Arg Trp Arg

245 250 255

Arg Glu Leu Asn Gln Ile Ser Arg Arg Leu Lys Arg Leu Lys Trp Lys

260 265 270

Trp Gly Glu Ile Pro Asn Pro Glu Leu Lys Asn Leu Ile Phe Lys Ser

275 280 285

Ser Arg Pro Glu Phe Val Ser Tyr Ala Leu Ile His Thr Leu Gly Arg

290 295 300

Asp Ile Gly Leu Ile Asn Glu Thr Glu Leu Lys Pro Asn Asn Ile Gln

305 310 315 320

Glu Tyr Gly Phe Gln Ile Tyr Tyr Lys Trp Glu Asp Pro Glu Leu Asn

325 330 335

His Ile Lys Lys Val Lys Asn Ile Pro Lys Arg Phe Ile Ile Pro Tyr

340 345 350

Lys Asn Leu Asp Leu Phe Gly Lys Tyr Thr Ile Leu Ser Arg Ala Ile

355 360 365

Glu Gly Ile Leu Lys Leu Tyr Ser Ser Ser Phe Gln Tyr Lys Ser Phe

370 375 380

Lys Asp Pro Asn Leu Phe Ala Lys Glu Gly Glu Lys Lys Ile Thr Asn

385 390 395 400

Glu Asp Phe Glu Leu Gly Tyr Asp Glu Lys Ile Lys Lys Ile Lys Asp

405 410 415

Asp Phe Lys Ser Tyr Lys Lys Ala Leu Leu Glu Lys Lys Lys Asn Thr

420 425 430

Leu Glu Asp Ser Leu Asn Ser Ile Leu Ser Val Tyr Glu Gln Ser Leu

435 440 445

Leu Thr Glu Gln Ile Asn Asn Val Lys Lys Trp Lys Glu Gly Leu Leu

450 455 460

Lys Ser Lys Glu Ser Ile His Lys Gln Lys Lys Ile Glu Asn Ile Glu

465 470 475 480

Asp Ile Ile Ser Arg Ile Glu Glu Leu Lys Asn Val Glu Gly Trp Ile

485 490 495

Arg Thr Lys Glu Arg Asp Ile Val Asn Lys Glu Glu Thr Asn Leu Lys

500 505 510

Arg Glu Ile Lys Lys Glu Leu Lys Asp Ser Tyr Tyr Glu Glu Val Arg

515 520 525

Lys Asp Phe Ser Asp Leu Lys Lys Gly Glu Glu Ser Glu Lys Lys Pro

530 535 540

Phe Arg Glu Glu Pro Lys Pro Ile Val Ile Lys Asp Tyr Ile Lys Phe

545 550 555 560

Asp Val Leu Pro Gly Glu Asn Ser Ala Leu Gly Phe Phe Leu Ser His

565 570 575

Leu Ser Phe Asn Leu Phe Asp Ser Ile Gln Tyr Glu Leu Phe Glu Lys

580 585 590

Ser Arg Leu Ser Ser Ser Lys His Pro Gln Ile Pro Glu Thr Ile Leu

595 600 605

Asp Leu

610

<210> 195

<211> 632

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 195

Phe Arg Lys Phe Val Lys Arg Ser Gly Ala Pro Gln Pro Asp Asn Leu

1 5 10 15

Asn Lys Tyr Lys Cys Ile Ala Ile Val Arg Ala Ala Asn Leu Asp Ala

20 25 30

Asp Ile Met Ser Asn Glu Ser Ser Asn Cys Val Met Cys Lys Gly Ile

35 40 45

Lys Met Asn Lys Arg Lys Thr Ala Lys Gly Ala Ala Lys Thr Thr Glu

50 55 60

Leu Gly Arg Val Tyr Ala Gly Gln Ser Gly Asn Leu Leu Cys Thr Ala

65 70 75 80

Cys Thr Lys Ser Thr Met Gly Pro Leu Val Asp Tyr Val Pro Ile Gly

85 90 95

Arg Ile Arg Ala Lys Tyr Thr Ile Leu Arg Ala Val Lys Glu Tyr Asp

100 105 110

Phe Leu Ser Leu Ala Tyr Asn Leu Ala Arg Thr Arg Val Ser Lys Lys

115 120 125

Gly Gly Arg Gln Lys Met His Ser Leu Ser Glu Leu Val Ile Ala Ala

130 135 140

Glu Tyr Glu Ile Ala Trp Asn Ile Ile Lys Ser Ser Val Ile His Tyr

145 150 155 160

His Gln Glu Thr Lys Glu Glu Ile Ser Gly Leu Arg Lys Lys Leu Gln

165 170 175

Ala Glu His Ile His Lys Asn Lys Glu Ala Arg Ile Arg Arg Glu Met

180 185 190

His Gln Ile Ser Arg Arg Ile Lys Arg Leu Lys Trp Lys Trp His Met

195 200 205

Ile Pro Asn Ser Glu Leu His Asn Phe Leu Phe Lys Gln Gln Asp Pro

210 215 220

Ser Phe Val Ala Val Ala Leu Leu His Thr Leu Gly Arg Asp Ile Gly

225 230 235 240

Met Ile Asn Lys Pro Lys Gly Ser Ala Lys Arg Glu Phe Ile Pro Glu

245 250 255

Tyr Gly Phe Gln Ile Tyr Tyr Lys Trp Met Asn Pro Lys Leu Asn Asp

260 265 270

Ile Asn Lys Gln Lys Tyr Arg Lys Met Pro Lys Arg Ser Leu Ile Pro

275 280 285

Tyr Lys Asn Leu Asn Val Phe Gly Asp Arg Glu Leu Ile Glu Asn Ala

290 295 300

Met His Lys Leu Leu Lys Leu Tyr Asp Glu Asn Leu Glu Val Lys Gly

305 310 315 320

Ser Lys Phe Phe Lys Thr Arg Val Val Ala Ile Ser Ser Lys Glu Ser

325 330 335

Glu Lys Leu Lys Arg Asp Leu Leu Trp Lys Gly Glu Leu Ala Lys Ile

340 345 350

Lys Lys Asp Phe Asn Ala Asp Lys Asn Lys Met Gln Glu Leu Phe Lys

355 360 365

Glu Val Lys Glu Pro Lys Lys Ala Asn Ala Leu Met Lys Gln Ser Arg

370 375 380

Asn Met Gly Phe Leu Leu Gln Asn Ile Ser Tyr Gly Ala Leu Gly Leu

385 390 395 400

Leu Ala Asn Arg Met Tyr Glu Ala Ser Ala Lys Gln Ser Lys Gly Asp

405 410 415

Ala Thr Lys Gln Pro Ser Ile Val Ile Pro Leu Glu Met Glu Phe Gly

420 425 430

Asn Ala Phe Pro Lys Leu Leu Leu Arg Ser Gly Lys Phe Ala Met Asn

435 440 445

Val Ser Ser Pro Trp Leu Thr Ile Arg Lys Pro Lys Phe Val Ile Lys

450 455 460

Gly Asn Lys Ile Lys Asn Ile Thr Lys Leu Met Lys Asp Glu Lys Ala

465 470 475 480

Lys Leu Lys Arg Leu Glu Thr Ser Tyr His Arg Ala Thr His Phe Arg

485 490 495

Pro Thr Leu Arg Gly Ser Ile Asp Trp Asp Ser Pro Tyr Phe Ser Ser

500 505 510

Pro Lys Gln Pro Asn Thr His Arg Arg Ser Pro Asp Arg Leu Ser Ala
515 520 525
Asp Ile Thr Glu Tyr Arg Gly Arg Leu Lys Ser Val Glu Ala Glu Leu

530 535 540
Arg Glu Gly Gln Arg Ala Met Ala Lys Lys Leu Asp Ser Val Asp Met
545 550 555 560
Thr Ala Ser Asn Leu Gln Thr Ser Asn Phe Gln Leu Glu Lys Gly Glu
565 570 575
Asp Pro Arg Leu Thr Glu Ile Asp Glu Lys Gly Arg Ser Ile Arg Asn
580 585 590
Cys Ile Ser Ser Trp Lys Lys Phe Met Glu Asp Leu Met Lys Ala Gln

595 600 605
Glu Ala Asn Pro Val Ile Lys Ile Lys Ile Ala Leu Lys Asp Glu Ser
610 615 620
Ser Val Leu Ser Glu Asp Ser Met
625 630

<210> 196
<211> 625
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 196
Lys Phe His Pro Glu Asn Leu Asn Lys Ser Tyr Cys Leu Ala Ile Val
1 5 10 15

Arg Ala Ala Asn Leu Asp Ala Asp Ile Gln Gly His Ile Asn Cys Ile
20 25 30
Gly Ile Lys Ser Asn Lys Ser Asp Arg Asn Tyr Glu Asn Lys Leu Glu
35 40 45
Ser Leu Gln Asn Val Glu Leu Leu Cys Lys Ala Cys Thr Lys Ser Thr
50 55 60
Tyr Lys Pro Asn Ile Asn Ser Val Pro Val Gly Glu Lys Lys Ala Lys

65 70 75 80

Tyr Ser Ile Leu Ser Glu Ile Lys Lys Tyr Asp Phe Asn Ser Leu Val
85 90 95
Tyr Asn Leu Lys Lys Tyr Arg Lys Gly Lys Ser Arg Gly His Gln Lys
100 105 110
Leu Asn Glu Leu Arg Glu Leu Val Ile Thr Ser Glu Tyr Lys Lys Ala
115 120 125
Leu Asp Val Ile Asn Lys Ser Val Asn His Tyr Leu Val Asn Ile Lys
130 135 140

Asn Lys Met Ser Lys Leu Lys Lys Ile Leu Gln Asn Glu His Ile His
145 150 155 160
Val Gly Thr Leu Ala Arg Ile Arg Arg Glu Arg Asn Arg Ile Ser Arg
165 170 175
Lys Leu Asp His Tyr Arg Lys Lys Trp Lys Phe Val Pro Asn Lys Ile
180 185 190
Leu Lys Asn Tyr Val Phe Lys Asn Gln Ser Pro Asp Phe Val Ser Val
195 200 205

Ala Leu Leu His Lys Leu Gly Arg Asp Ile Gly Leu Ile Thr Lys Thr
210 215 220
Ala Ile Leu Gln Lys Ser Phe Pro Glu Tyr Ser Leu Gln Leu Tyr Tyr
225 230 235 240
Lys Tyr Asp Thr Pro Lys Leu Asn Tyr Leu Lys Lys Ser Lys Phe Lys
245 250 255
Ser Leu Pro Lys Arg Ile Leu Ile Ser Tyr Lys Tyr Pro Lys Phe Asp
260 265 270

Ile Asn Ser Asn Tyr Ile Glu Glu Ser Ile Asp Lys Leu Leu Lys Leu
275 280 285
Tyr Glu Glu Ser Pro Ile Tyr Lys Asn Asn Ser Lys Ile Ile Glu Phe
290 295 300
Phe Lys Lys Ser Glu Asp Asn Leu Ile Lys Ser Glu Asn Asp Ser Leu
305 310 315 320

Lys Arg Gly Ile Met Lys Glu Phe Glu Lys Val Thr Lys Asn Phe Ser
325 330 335

Ser Lys Lys Lys Lys Leu Lys Glu Glu Leu Lys Leu Lys Asn Glu Asp
340 345 350

Lys Asn Ser Lys Met Leu Ala Lys Val Ser Arg Pro Ile Gly Phe Leu
355 360 365

Lys Ala Tyr Leu Ser Tyr Met Leu Phe Asn Ile Ile Ser Asn Arg Ile
370 375 380

Phe Glu Phe Ser Arg Lys Ser Ser Gly Arg Ile Pro Gln Leu Pro Ser
385 390 395 400

Cys Ile Ile Asn Leu Gly Asn Gln Phe Glu Asn Phe Lys Asn Glu Leu
405 410 415

Gln Asp Ser Asn Ile Gly Ser Lys Lys Asn Tyr Lys Tyr Phe Cys Asn
420 425 430

Leu Leu Leu Lys Ser Ser Gly Phe Asn Ile Ser Tyr Glu Glu Glu His
435 440 445

Leu Ser Ile Lys Thr Pro Asn Phe Phe Ile Asn Gly Arg Lys Leu Lys
450 455 460

Glu Ile Thr Ser Glu Lys Lys Lys Ile Arg Lys Glu Asn Glu Gln Leu
465 470 475 480

Ile Lys Gln Trp Lys Lys Leu Thr Phe Phe Lys Pro Ser Asn Leu Asn
485 490 495

Gly Lys Lys Thr Ser Asp Lys Ile Arg Phe Lys Ser Pro Asn Asn Pro
500 505 510

Asp Ile Glu Arg Lys Ser Glu Asp Asn Ile Val Glu Asn Ile Ala Lys
515 520 525

Val Lys Tyr Lys Leu Glu Asp Leu Leu Ser Glu Gln Arg Lys Glu Phe
530 535 540

Asn Lys Leu Ala Lys Lys His Asp Gly Val Asp Val Glu Ala Gln Cys
545 550 555 560

Leu Gln Thr Lys Ser Phe Trp Ile Asp Ser Asn Ser Pro Ile Lys Lys

565 570 575

Ser Leu Glu Lys Lys Asn Glu Lys Val Ser Val Lys Lys Lys Met Lys

580 585 590

Ala Ile Arg Ser Cys Ile Ser Ala Trp Lys Trp Phe Met Ala Asp Leu

595 600 605

Ile Glu Ala Gln Lys Glu Thr Pro Met Ile Lys Leu Lys Leu Ala Leu

610 615 620

Met

625

<210> 197

<211> 517

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 197

Thr Thr Leu Val Pro Ser His Leu Ala Gly Ile Glu Val Met Asp Glu

1 5 10 15

Thr Thr Ser Arg Asn Glu Asp Met Ile Gln Lys Glu Thr Ser Arg Ser

20 25 30

Asn Glu Asp Glu Asn Tyr Leu Gly Val Lys Asn Lys Cys Gly Ile Asn

35 40 45

Val His Lys Ser Gly Arg Gly Ser Ser Lys His Glu Pro Asn Met Pro

50 55 60

Pro Glu Lys Ser Gly Glu Gly Gln Met Pro Lys Gln Asp Ser Thr Glu

65 70 75 80

Met Gln Gln Arg Phe Asp Glu Ser Val Thr Gly Glu Thr Gln Val Ser

85 90 95

Ala Gly Ala Thr Ala Ser Ile Lys Thr Asp Ala Arg Ala Asn Ser Gly

100 105 110

Pro Arg Val Gly Thr Ala Arg Ala Leu Ile Val Lys Ala Ser Asn Leu

115 120 125

Asp Arg Asp Ile Lys Leu Gly Cys Lys Pro Cys Glu Tyr Ile Arg Ser

130 135 140

Glu Leu Pro Met Gly Lys Lys Asn Gly Cys Asn His Cys Glu Lys Ser
145 150 155 160
Ser Asp Ile Ala Ser Val Pro Lys Val Glu Ser Gly Phe Arg Lys Ala
165 170 175
Lys Tyr Glu Leu Val Arg Arg Phe Glu Ser Phe Ala Ala Asp Ser Ile
180 185 190
Ser Arg His Leu Gly Lys Glu Gln Ala Arg Thr Arg Gly Lys Arg Gly
195 200 205

Lys Lys Asp Lys Lys Glu Gln Met Gly Lys Val Asn Leu Asp Glu Ile
210 215 220
Ala Ile Leu Lys Asn Glu Ser Leu Ile Glu Tyr Thr Glu Asn Gln Ile
225 230 235 240
Leu Asp Ala Arg Ser Asn Arg Ile Lys Glu Trp Leu Arg Ser Leu Arg
245 250 255
Leu Arg Leu Arg Thr Arg Asn Lys Gly Leu Lys Lys Ser Lys Ser Ile
260 265 270

Arg Arg Gln Leu Ile Thr Leu Arg Arg Asp Tyr Arg Lys Trp Ile Lys
275 280 285
Pro Asn Pro Tyr Arg Pro Asp Glu Asp Pro Asn Glu Asn Ser Leu Arg
290 295 300
Leu His Thr Lys Leu Gly Val Asp Ile Gly Val Gln Gly Gly Asp Asn
305 310 315 320
Lys Arg Met Asn Ser Asp Asp Tyr Glu Thr Ser Phe Ser Ile Thr Trp
325 330 335

Arg Asp Thr Ala Thr Arg Lys Ile Cys Phe Thr Lys Pro Lys Gly Leu
340 345 350
Leu Pro Arg His Met Lys Phe Lys Leu Arg Gly Tyr Pro Glu Leu Ile
355 360 365
Leu Tyr Asn Glu Glu Leu Arg Ile Gln Asp Ser Gln Lys Phe Pro Leu
370 375 380

Val Asp Trp Glu Arg Ile Pro Ile Phe Lys Leu Arg Gly Val Ser Leu
385 390 395 400

Gly Lys Lys Lys Val Lys Ala Leu Asn Arg Ile Thr Glu Ala Pro Arg
405 410 415
Leu Val Val Ala Lys Arg Ile Gln Val Asn Ile Glu Ser Lys Lys Lys
420 425 430
Lys Val Leu Thr Arg Tyr Val Tyr Asn Asp Lys Ser Ile Asn Gly Arg
435 440 445
Leu Val Lys Ala Glu Asp Ser Asn Lys Asp Pro Leu Leu Glu Phe Lys
450 455 460

Lys Gln Ala Glu Glu Ile Asn Ser Asp Ala Lys Tyr Tyr Glu Asn Gln
465 470 475 480
Glu Ile Ala Lys Asn Tyr Leu Trp Gly Cys Glu Gly Leu His Lys Asn
485 490 495
Leu Leu Glu Glu Gln Thr Lys Asn Pro Tyr Leu Ala Phe Lys Tyr Gly
500 505 510
Phe Leu Asn Ile Val
515

<210> 198
<211> 410
<212> PRT
<213> Artificial sequence
<220
><223> Synthetic sequence
<400> 198
Leu Asp Phe Lys Arg Thr Cys Ser Gln Glu Leu Val Leu Leu Pro Glu
1 5 10 15
Ile Glu Gly Leu Lys Leu Ser Gly Thr Gln Gly Val Thr Ser Leu Ala
20 25 30
Lys Lys Leu Ile Asn Lys Ala Ala Asn Val Asp Arg Asp Glu Ser Tyr
35 40 45
Gly Cys His His Cys Ile His Thr Arg Thr Ser Leu Ser Lys Pro Val

50 55 60

Lys Lys Asp Cys Asn Ser Cys Asn Gln Ser Thr Asn His Pro Ala Val
65 70 75 80
Pro Ile Thr Leu Lys Gly Tyr Lys Ile Ala Phe Tyr Glu Leu Trp His
85 90 95
Arg Phe Thr Ser Trp Ala Val Asp Ser Ile Ser Lys Ala Leu His Arg
100 105 110
Asn Lys Val Met Gly Lys Val Asn Leu Asp Glu Tyr Ala Val Val Asp
115 120 125

Asn Ser His Ile Val Cys Tyr Ala Val Arg Lys Cys Tyr Glu Lys Arg
130 135 140
Gln Arg Ser Val Arg Leu His Lys Arg Ala Tyr Arg Cys Arg Ala Lys
145 150 155 160
His Tyr Asn Lys Ser Gln Pro Lys Val Gly Arg Ile Tyr Lys Lys Ser
165 170 175
Lys Arg Arg Asn Ala Arg Asn Leu Lys Lys Glu Ala Lys Arg Tyr Phe
180 185 190

Gln Pro Asn Glu Ile Thr Asn Gly Ser Ser Asp Ala Leu Phe Tyr Lys
195 200 205
Ile Gly Val Asp Leu Gly Ile Ala Lys Gly Thr Pro Glu Thr Glu Val
210 215 220
Lys Val Asp Val Ser Ile Cys Phe Gln Val Tyr Tyr Gly Asp Ala Arg
225 230 235 240
Arg Val Leu Arg Val Arg Lys Met Asp Glu Leu Gln Ser Phe His Leu
245 250 255

Asp Tyr Thr Gly Lys Leu Lys Leu Lys Gly Ile Gly Asn Lys Asp Thr
260 265 270
Phe Thr Ile Ala Lys Arg Asn Glu Ser Leu Lys Trp Gly Ser Thr Lys
275 280 285
Tyr Glu Val Ser Arg Ala His Lys Lys Phe Lys Pro Phe Gly Lys Lys
290 295 300

Gly Ser Val Lys Arg Lys Cys Asn Asp Tyr Phe Arg Ser Ile Ala Ser
305 310 315 320

Trp Ser Cys Glu Ala Ala Ser Gln Arg Ala Gln Ser Asn Leu Lys Asn
325 330 335
Ala Phe Pro Tyr Gln Lys Ala Leu Val Lys Cys Tyr Lys Asn Leu Asp
340 345 350
Tyr Lys Gly Val Lys Lys Asn Asp Met Trp Tyr Arg Leu Cys Ser Asn
355 360 365
Arg Ile Phe Arg Tyr Ser Arg Ile Ala Glu Asp Ile Ala Gln Tyr Gln
370 375 380

Ser Asp Lys Gly Lys Ala Lys Phe Glu Phe Val Ile Leu Ala Gln Ser
385 390 395 400
Val Ala Glu Tyr Asp Ile Ser Ala Ile Met
405 410

<210> 199
<211> 602
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 199
Val Phe Leu Thr Asp Asp Lys Arg Lys Thr Ala Leu Arg Lys Ile Arg
1 5 10 15
Ser Ala Phe Arg Lys Thr Ala Glu Ile Ala Leu Val Arg Ala Gln Glu

20 25 30
Ala Asp Ser Leu Asp Arg Gln Ala Lys Lys Leu Thr Ile Glu Thr Val
35 40 45
Ser Phe Gly Ala Pro Gly Ala Lys Asn Ala Phe Ile Gly Ser Leu Gln
50 55 60
Gly Tyr Asn Trp Asn Ser His Arg Ala Asn Val Pro Ser Ser Gly Ser
65 70 75 80
Ala Lys Asp Val Phe Arg Ile Thr Glu Leu Gly Leu Gly Ile Pro Gln

85 90 95

Ser Ala His Glu Ala Ser Ile Gly Lys Ser Phe Glu Leu Val Gly Asn

100 105 110

Val Val Arg Tyr Thr Ala Asn Leu Leu Ser Lys Gly Tyr Lys Lys Gly

115 120 125

Ala Val Asn Lys Gly Ala Lys Gln Gln Arg Glu Ile Lys Gly Lys Glu

130 135 140

Gln Leu Ser Phe Asp Leu Ile Ser Asn Gly Pro Ile Ser Gly Asp Lys

145 150 155 160

Leu Ile Asn Gly Gln Lys Asp Ala Leu Ala Trp Trp Leu Ile Asp Lys

165 170 175

Met Gly Phe His Ile Gly Leu Ala Met Glu Pro Leu Ser Ser Pro Asn

180 185 190

Thr Tyr Gly Ile Thr Leu Gln Ala Phe Trp Lys Arg His Thr Ala Pro

195 200 205

Arg Arg Tyr Ser Arg Gly Val Ile Arg Gln Trp Gln Leu Pro Phe Gly

210 215 220

Arg Gln Leu Ala Pro Leu Ile His Asn Phe Phe Arg Lys Lys Gly Ala

225 230 235 240

Ser Ile Pro Ile Val Leu Thr Asn Ala Ser Lys Lys Leu Ala Gly Lys

245 250 255

Gly Val Leu Leu Glu Gln Thr Ala Leu Val Asp Pro Lys Lys Trp Trp

260 265 270

Gln Val Lys Glu Gln Val Thr Gly Pro Leu Ser Asn Ile Trp Glu Arg

275 280 285

Ser Val Pro Leu Val Leu Tyr Thr Ala Thr Phe Thr His Lys His Gly

290 295 300

Ala Ala His Lys Arg Pro Leu Thr Leu Lys Val Ile Arg Ile Ser Ser

305 310 315 320

Gly Ser Val Phe Leu Leu Pro Leu Ser Lys Val Thr Pro Gly Lys Leu

325 330 335

Val Arg Ala Trp Met Pro Asp Ile Asn Ile Leu Arg Asp Gly Arg Pro

340 345 350

Asp Glu Ala Ala Tyr Lys Gly Pro Asp Leu Ile Arg Ala Arg Glu Arg

355 360 365

Ser Phe Pro Leu Ala Tyr Thr Cys Val Thr Gln Ile Ala Asp Glu Trp

370 375 380

Gln Lys Arg Ala Leu Glu Ser Asn Arg Asp Ser Ile Thr Pro Leu Glu

385 390 395 400

Ala Lys Leu Val Thr Gly Ser Asp Leu Leu Gln Ile His Ser Thr Val

405 410 415

Gln Gln Ala Val Glu Gln Gly Ile Gly Gly Arg Ile Ser Ser Pro Ile

420 425 430

Gln Glu Leu Leu Ala Lys Asp Ala Leu Gln Leu Val Leu Gln Gln Leu

435 440 445

Phe Met Thr Val Asp Leu Leu Arg Ile Gln Trp Gln Leu Lys Gln Glu

450 455 460

Val Ala Asp Gly Asn Thr Ser Glu Lys Ala Val Gly Trp Ala Ile Arg

465 470 475 480

Ile Ser Asn Ile His Lys Asp Ala Tyr Lys Thr Ala Ile Glu Pro Cys

485 490 495

Thr Ser Ala Leu Lys Gln Ala Trp Asn Pro Leu Ser Gly Phe Glu Glu

500 505 510

Arg Thr Phe Gln Leu Asp Ala Ser Ile Val Arg Lys Arg Ser Thr Ala

515 520 525

Lys Thr Pro Asp Asp Glu Leu Val Ile Val Leu Arg Gln Gln Ala Ala

530 535 540

Glu Met Thr Val Ala Val Thr Gln Ser Val Ser Lys Glu Leu Met Glu

545 550 555 560

Leu Ala Val Arg His Ser Ala Thr Leu His Leu Leu Val Gly Glu Val

565 570 575

Ala Ser Lys Gln Leu Ser Arg Ser Ala Asp Lys Asp Arg Gly Ala Met

580 585 590

Asp His Trp Lys Leu Leu Ser Gln Ser Met

595 600

<210> 200

<211> 494

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 200

Glu Asp Leu Leu Gln Lys Ala Leu Asn Thr Ala Thr Asn Val Ala Ala

1 5 10 15

Ile Glu Arg His Ser Cys Ile Ser Cys Leu Phe Thr Glu Ser Glu Ile

20 25 30

Asp Val Lys Tyr Lys Thr Pro Asp Lys Ile Gly Gln Asn Thr Ala Gly

35 40 45

Cys Gln Ser Cys Thr Phe Arg Val Gly Tyr Ser Gly Asn Ser His Thr

50 55 60

Leu Pro Met Gly Asn Arg Ile Ala Leu Asp Lys Leu Arg Glu Thr Ile

65 70 75 80

Gln Arg Tyr Ala Trp His Ser Leu Leu Phe Asn Val Pro Pro Ala Pro

85 90 95

Thr Ser Lys Arg Val Arg Ala Ile Ser Glu Leu Arg Val Ala Ala Gly

100 105 110

Arg Glu Arg Leu Phe Thr Val Ile Thr Phe Val Gln Thr Asn Ile Leu

115 120 125

Ser Lys Leu Gln Lys Arg Tyr Ala Ala Asn Trp Thr Pro Lys Ser Gln

130 135 140

Glu Arg Leu Ser Arg Leu Arg Glu Glu Gly Gln His Ile Leu Ser Leu

145 150 155 160

Leu Glu Ser Gly Ser Trp Gln Gln Lys Glu Val Val Arg Glu Asp Gln

165 170 175

Asp Leu Ile Val Cys Ser Ala Leu Thr Lys Pro Gly Leu Ser Ile Gly
180 185 190
Ala Phe Cys Arg Pro Lys Tyr Leu Lys Pro Ala Lys His Ala Leu Val
195 200 205
Leu Arg Leu Ile Phe Val Glu Gln Trp Pro Gly Gln Ile Trp Gly Gln
210 215 220
Ser Lys Arg Thr Arg Arg Met Arg Arg Arg Lys Asp Val Glu Arg Val
225 230 235 240

Tyr Asp Ile Ser Val Gln Ala Trp Ala Leu Lys Gly Lys Glu Thr Arg
245 250 255
Ile Ser Glu Cys Ile Asp Thr Met Arg Arg His Gln Gln Ala Tyr Ile
260 265 270
Gly Val Leu Pro Phe Leu Ile Leu Ser Gly Ser Thr Val Arg Gly Lys
275 280 285
Gly Asp Cys Pro Ile Leu Lys Glu Ile Thr Arg Met Arg Tyr Cys Pro
290 295 300

Asn Asn Glu Gly Leu Ile Pro Leu Gly Ile Phe Tyr Arg Gly Ser Ala
305 310 315 320
Asn Lys Leu Leu Arg Val Val Lys Gly Ser Ser Phe Thr Leu Pro Met
325 330 335
Trp Gln Asn Ile Glu Thr Leu Pro His Pro Glu Pro Phe Ser Pro Glu
340 345 350
Gly Trp Thr Ala Thr Gly Ala Leu Tyr Glu Lys Asn Leu Ala Tyr Trp
355 360 365

Ser Ala Leu Asn Glu Ala Val Asp Trp Tyr Thr Gly Gln Ile Leu Ser
370 375 380
Ser Gly Leu Gln Tyr Pro Asn Gln Asn Glu Phe Leu Ala Arg Leu Gln
385 390 395 400
Asn Val Ile Asp Ser Ile Pro Arg Lys Trp Phe Arg Pro Gln Gly Leu
405 410 415
Lys Asn Leu Lys Pro Asn Gly Gln Glu Asp Ile Val Pro Asn Glu Phe

420 425 430

Val Ile Pro Gln Asn Ala Ile Arg Ala His His Val Ile Glu Trp Tyr
435 440 445
His Lys Thr Asn Asp Leu Val Ala Lys Thr Leu Leu Gly Trp Gly Ser
450 455 460
Gln Thr Thr Leu Asn Gln Thr Arg Pro Gln Gly Asp Leu Arg Phe Thr
465 470 475 480
Tyr Thr Arg Tyr Tyr Phe Arg Glu Lys Glu Val Pro Glu Val
485 490

<210> 201
<211> 649
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 201
Val Pro Lys Lys Lys Leu Met Arg Glu Leu Ala Lys Lys Ala Val Phe
1 5 10 15
Glu Ala Ile Phe Asn Asp Pro Ile Pro Gly Ser Phe Gly Cys Lys Arg
20 25 30
Cys Thr Leu Ile Asp Gly Ala Arg Val Thr Asp Ala Ile Glu Lys Lys
35 40 45

Gln Gly Ala Lys Arg Cys Ala Gly Cys Glu Pro Cys Thr Phe His Thr
50 55 60
Leu Tyr Asp Ser Val Lys His Ala Leu Pro Ala Ala Thr Gly Cys Asp
65 70 75 80
Arg Thr Ala Ile Asp Thr Gly Leu Trp Glu Ile Leu Thr Ala Leu Arg
85 90 95
Ser Tyr Asn Trp Met Ser Phe Arg Arg Asn Ala Val Ser Asp Ala Ser
100 105 110

Gln Lys Gln Val Trp Ser Ile Glu Glu Leu Ala Ile Trp Ala Asp Lys
115 120 125

Glu Arg Ala Leu Arg Val Ile Leu Ser Ala Leu Thr His Thr Ile Gly
130 135 140
Lys Leu Lys Asn Gly Phe Ser Arg Asp Gly Val Trp Lys Gly Gly Lys
145 150 155 160
Gln Leu Tyr Glu Asn Leu Ala Gln Lys Asp Leu Ala Lys Gly Leu Phe
165 170 175

Ala Asn Gly Glu Ile Phe Gly Lys Glu Leu Val Glu Ala Asp His Asp
180 185 190
Met Leu Ala Trp Thr Ile Val Pro Asn His Gln Phe His Ile Gly Leu
195 200 205
Ile Arg Gly Asn Trp Lys Pro Ala Ala Val Glu Ala Ser Thr Ala Phe
210 215 220
Asp Ala Arg Trp Leu Thr Asn Gly Ala Pro Leu Arg Asp Thr Arg Thr
225 230 235 240

His Gly His Arg Gly Arg Arg Phe Asn Arg Thr Glu Lys Leu Thr Val
245 250 255
Leu Cys Ile Lys Arg Asp Gly Gly Val Ser Glu Glu Phe Arg Gln Glu
260 265 270
Arg Asp Tyr Glu Leu Ser Val Met Leu Leu Gln Pro Lys Asn Lys Leu
275 280 285
Lys Pro Glu Pro Lys Gly Glu Leu Asn Ser Phe Glu Asp Leu His Asp
290 295 300

His Trp Trp Phe Leu Lys Gly Asp Glu Ala Thr Ala Leu Val Gly Leu
305 310 315 320
Thr Ser Asp Pro Thr Val Gly Asp Phe Ile Gln Leu Gly Leu Tyr Ile
325 330 335
Arg Asn Pro Ile Lys Ala His Gly Glu Thr Lys Arg Arg Leu Leu Ile
340 345 350
Cys Phe Glu Pro Pro Ile Lys Leu Pro Leu Arg Arg Ala Phe Pro Ser
355 360 365

Glu Ala Phe Lys Thr Trp Glu Pro Thr Ile Asn Val Phe Arg Asn Gly

370 375 380

Arg Arg Asp Thr Glu Ala Tyr Tyr Asp Ile Asp Arg Ala Arg Val Phe

385 390 395 400

Glu Phe Pro Glu Thr Arg Val Ser Leu Glu His Leu Ser Lys Gln Trp

405 410 415

Glu Val Leu Arg Leu Glu Pro Asp Arg Glu Asn Thr Asp Pro Tyr Glu

420 425 430

Ala Gln Gln Asn Glu Gly Ala Glu Leu Gln Val Tyr Ser Leu Leu Gln

435 440 445

Glu Ala Ala Gln Lys Met Ala Pro Lys Val Val Ile Asp Pro Phe Gly

450 455 460

Gln Phe Pro Leu Glu Leu Phe Ser Thr Phe Val Ala Gln Leu Phe Asn

465 470 475 480

Ala Pro Leu Ser Asp Thr Lys Ala Lys Ile Gly Lys Pro Leu Asp Ser

485 490 495

Gly Phe Val Val Glu Ser His Leu His Leu Leu Glu Glu Asp Phe Ala

500 505 510

Tyr Arg Asp Phe Val Arg Val Thr Phe Met Gly Thr Glu Pro Thr Phe

515 520 525

Arg Val Ile His Tyr Ser Asn Gly Glu Gly Tyr Trp Lys Lys Thr Val

530 535 540

Leu Lys Gly Lys Asn Asn Ile Arg Thr Ala Leu Ile Pro Glu Gly Ala

545 550 555 560

Lys Ala Ala Val Asp Ala Tyr Lys Asn Lys Arg Cys Pro Leu Thr Leu

565 570 575

Glu Ala Ala Ile Leu Asn Glu Glu Lys Asp Arg Arg Leu Val Leu Gly

580 585 590

Asn Lys Ala Leu Ser Leu Leu Ala Gln Thr Ala Arg Gly Asn Leu Thr

595 600 605

Ile Leu Glu Ala Leu Ala Ala Glu Val Leu Arg Pro Leu Ser Gly Thr

610 615 620

Glu Gly Val Val His Leu His Ala Cys Val Thr Arg His Ser Thr Leu
625 630 635 640
Thr Glu Ser Thr Glu Thr Asp Asn Met
645

<210> 202
<211> 414
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 202

Val Glu Lys Leu Phe Ser Glu Arg Leu Lys Arg Ala Met Trp Leu Lys
1 5 10 15
Asn Glu Ala Gly Arg Ala Pro Pro Ala Glu Thr Leu Thr Leu Lys His
20 25 30
Lys Arg Val Ser Gly Gly His Glu Lys Val Lys Glu Glu Leu Gln Arg
35 40 45
Val Leu Arg Ser Leu Ser Gly Thr Asn Gln Ala Ala Trp Asn Leu Gly
50 55 60
Leu Ser Gly Gly Arg Glu Pro Lys Ser Ser Asp Ala Leu Lys Gly Glu
65 70 75 80
Lys Ser Arg Val Val Leu Glu Thr Val Val Phe His Ser Gly His Asn
85 90 95
Arg Val Leu Tyr Asp Val Ile Glu Arg Glu Asp Gln Val His Gln Arg
100 105 110
Ser Ser Ile Met His Met Arg Arg Lys Gly Ser Asn Leu Leu Arg Leu
115 120 125
Trp Gly Arg Ser Gly Lys Val Arg Arg Lys Met Arg Glu Glu Val Ala
130 135 140
Glu Ile Lys Pro Val Trp His Lys Asp Ser Arg Trp Leu Ala Ile Val
145 150 155 160
Glu Glu Gly Arg Gln Ser Val Val Gly Ile Ser Ser Ala Gly Leu Ala
165 170 175

Val Phe Ala Val Gln Glu Ser Gln Cys Thr Thr Ala Glu Pro Lys Pro
180 185 190
Leu Glu Tyr Val Val Ser Ile Trp Phe Arg Gly Ser Lys Ala Leu Asn
195 200 205
Pro Gln Asp Arg Tyr Leu Glu Phe Lys Lys Leu Lys Thr Thr Glu Ala

210 215 220
Leu Arg Gly Gln Gln Tyr Asp Pro Ile Pro Phe Ser Leu Lys Arg Gly
225 230 235 240
Ala Gly Cys Ser Leu Ala Ile Arg Gly Glu Gly Ile Lys Phe Gly Ser
245 250 255
Arg Gly Pro Ile Lys Gln Phe Phe Gly Ser Asp Arg Ser Arg Pro Ser
260 265 270
His Ala Asp Tyr Asp Gly Lys Arg Arg Leu Ser Leu Phe Ser Lys Tyr

275 280 285
Ala Gly Asp Leu Ala Asp Leu Thr Glu Glu Gln Trp Asn Arg Thr Val
290 295 300
Ser Ala Phe Ala Glu Asp Glu Val Arg Arg Ala Thr Leu Ala Asn Ile
305 310 315 320
Gln Asp Phe Leu Ser Ile Ser His Glu Lys Tyr Ala Glu Arg Leu Lys
325 330 335
Lys Arg Ile Glu Ser Ile Glu Glu Pro Val Ser Ala Ser Lys Leu Glu

340 345 350
Ala Tyr Leu Ser Ala Ile Phe Glu Thr Phe Val Gln Gln Arg Glu Ala
355 360 365
Leu Ala Ser Asn Phe Leu Met Arg Leu Val Glu Ser Val Ala Leu Leu
370 375 380
Ile Ser Leu Glu Glu Lys Ser Pro Arg Val Glu Phe Arg Val Ala Arg
385 390 395 400
Tyr Leu Ala Glu Ser Lys Glu Gly Phe Asn Arg Lys Ala Met

405 410
<210> 203

<211> 413

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 203

Val Val Ile Thr Gln Ser Glu Leu Tyr Lys Glu Arg Leu Leu Arg Val
1 5 10 15

Met Glu Ile Lys Asn Asp Arg Gly Arg Lys Glu Pro Arg Glu Ser Gln
20 25 30

Gly Leu Val Leu Arg Phe Thr Gln Val Thr Gly Gly Gln Glu Lys Val
35 40 45

Lys Gln Lys Leu Trp Leu Ile Phe Glu Gly Phe Ser Gly Thr Asn Gln
50 55 60

Ala Ser Trp Asn Phe Gly Gln Pro Ala Gly Gly Arg Lys Pro Asn Ser
65 70 75 80

Gly Asp Ala Leu Lys Gly Pro Lys Ser Arg Val Thr Tyr Glu Thr Val
85 90 95

Val Phe His Phe Gly Leu Arg Leu Leu Ser Ala Val Ile Glu Arg His
100 105 110

Asn Leu Lys Gln Gln Arg Gln Thr Met Ala Tyr Met Lys Arg Arg Ala
115 120 125

Ala Ala Arg Lys Lys Trp Ala Arg Ser Gly Lys Lys Cys Ser Arg Met
130 135 140

Arg Asn Glu Val Glu Lys Ile Lys Pro Lys Trp His Lys Asp Pro Arg
145 150 155 160

Trp Phe Asp Ile Val Lys Glu Gly Glu Pro Ser Ile Val Gly Ile Ser
165 170 175

Ser Ala Gly Phe Ala Ile Tyr Ile Val Glu Glu Pro Asn Phe Pro Arg
180 185 190

Gln Asp Pro Leu Glu Ile Glu Tyr Ala Ile Ser Ile Trp Phe Arg Arg
195 200 205

Asp Arg Ser Gln Tyr Leu Thr Phe Lys Lys Ile Gln Lys Ala Glu Lys

210 215 220

Leu Lys Glu Leu Gln Tyr Asn Pro Ile Pro Phe Arg Leu Lys Gln Glu
225 230 235 240

Lys Thr Ser Leu Val Phe Glu Ser Gly Asp Ile Lys Phe Gly Ser Arg
245 250 255

Gly Ser Ile Glu His Phe Arg Asp Glu Ala Arg Gly Lys Pro Pro Lys
260 265 270

Ala Asp Met Asp Asn Asn Arg Arg Leu Thr Met Phe Ser Val Phe Ser
275 280 285

Gly Asn Leu Thr Asn Leu Thr Glu Glu Gln Tyr Ala Arg Pro Val Ser
290 295 300

Gly Leu Leu Ala Pro Asp Glu Lys Arg Met Pro Thr Leu Leu Lys Lys
305 310 315 320

Leu Gln Asp Phe Phe Thr Pro Ile His Glu Lys Tyr Gly Glu Arg Ile
325 330 335

Lys Gln Arg Leu Ala Asn Ser Glu Ala Ser Lys Arg Pro Phe Lys Lys
340 345 350

Leu Glu Glu Tyr Leu Pro Ala Ile Tyr Leu Glu Phe Arg Ala Arg Arg
355 360 365

Glu Gly Leu Ala Ser Asn Trp Val Leu Val Leu Ile Asn Ser Val Arg
370 375 380

Thr Leu Val Arg Ile Lys Ser Glu Asp Pro Tyr Ile Glu Phe Lys Val
385 390 395 400

Ser Gln Tyr Leu Leu Glu Lys Glu Asp Asn Lys Ala Leu
405 410

<210> 204

<211> 449

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 204

Lys Gln Asp Ala Leu Phe Glu Glu Arg Leu Lys Lys Ala Ile Phe Ile

1 5 10 15

Lys Arg Gln Ala Asp Pro Leu Gln Arg Glu Glu Leu Ser Leu Leu Pro

20 25 30

Pro Asn Arg Lys Ile Val Thr Gly Gly His Glu Ser Ala Lys Asp Thr

35 40 45

Leu Lys Gln Ile Leu Arg Ala Ile Asn Gly Thr Asn Gln Ala Ser Trp

50 55 60

Asn Pro Gly Thr Pro Ser Gly Lys Arg Asp Ser Lys Ser Ala Asp Ala

65 70 75 80

Leu Ala Gly Pro Lys Ser Arg Val Lys Leu Glu Thr Val Val Phe His

85 90 95

Val Gly His Arg Leu Leu Lys Lys Val Val Glu Tyr Gln Gly His Gln

100 105 110

Lys Gln Gln His Gly Leu Lys Ala Phe Met Arg Thr Cys Ala Ala Met

115 120 125

Arg Lys Lys Trp Lys Arg Ser Gly Lys Val Val Gly Glu Leu Arg Glu

130 135 140

Gln Leu Ala Asn Ile Gln Pro Lys Trp His Tyr Asp Ser Arg Pro Leu

145 150 155 160

Asn Leu Cys Phe Glu Gly Lys Pro Ser Val Val Gly Leu Arg Ser Ala

165 170 175

Gly Ile Ala Leu Tyr Thr Ile Gln Lys Ser Val Val Pro Val Lys Glu

180 185 190

Pro Lys Pro Ile Glu Tyr Ala Val Ser Ile Trp Phe Arg Gly Pro Lys

195 200 205

Ala Met Asp Arg Glu Asp Arg Cys Leu Glu Phe Lys Lys Leu Lys Ile

210 215 220

Ala Thr Glu Leu Arg Lys Leu Gln Phe Glu Pro Ile Val Ser Thr Leu

225 230 235 240

Thr Gln Gly Ile Lys Gly Phe Ser Leu Tyr Ile Gln Gly Asn Ser Val

245 250 255

Lys Phe Gly Ser Arg Gly Pro Ile Lys Tyr Phe Ser Asn Glu Ser Val

260 265 270

Arg Gln Arg Pro Pro Lys Ala Asp Pro Asp Gly Asn Lys Arg Leu Ala

275 280 285

Leu Phe Ser Lys Phe Ser Gly Asp Leu Ser Asp Leu Thr Glu Glu Gln

290 295 300

Trp Asn Arg Pro Ile Leu Ala Phe Glu Gly Ile Ile Arg Arg Ala Thr

305 310 315 320

Leu Gly Asn Ile Gln Asp Tyr Leu Thr Val Gly His Glu Gln Phe Ala

325 330 335

Ile Ser Leu Glu Gln Leu Leu Ser Glu Lys Glu Ser Val Leu Gln Met

340 345 350

Ser Ile Glu Gln Gln Arg Leu Lys Lys Asn Leu Gly Lys Lys Ala Glu

355 360 365

Asn Glu Trp Val Glu Ser Phe Gly Ala Glu Gln Ala Arg Lys Lys Ala

370 375 380

Gln Gly Ile Arg Glu Tyr Ile Ser Gly Phe Phe Gln Glu Tyr Cys Ser

385 390 395 400

Gln Arg Glu Gln Trp Ala Glu Asn Trp Val Gln Gln Leu Asn Lys Ser

405 410 415

Val Arg Leu Phe Leu Thr Ile Gln Asp Ser Thr Pro Phe Ile Glu Phe

420 425 430

Arg Val Ala Arg Tyr Leu Pro Lys Gly Glu Lys Lys Lys Gly Lys Ala

435 440 445

Met

<210> 205

<211> 711

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 205

Ala Asn His Ala Glu Arg His Lys Arg Leu Arg Lys Glu Ala Asn Arg
1 5 10 15

Ala Ala Asn Arg Asn Arg Pro Leu Val Ala Asp Cys Asp Thr Gly Asp
20 25 30

Pro Leu Val Gly Ile Cys Arg Leu Leu Arg Arg Gly Asp Lys Met Gln
35 40 45

Pro Asn Lys Thr Gly Cys Arg Ser Cys Glu Gln Val Glu Pro Glu Leu
50 55 60

Arg Asp Ala Ile Leu Val Ser Gly Pro Gly Arg Leu Asp Asn Tyr Lys
65 70 75 80

Tyr Glu Leu Phe Gln Arg Gly Arg Ala Met Ala Val His Arg Leu Leu
85 90 95

Lys Arg Val Pro Lys Leu Asn Arg Pro Lys Lys Ala Ala Gly Asn Asp
100 105 110

Glu Lys Lys Ala Glu Asn Lys Lys Ser Glu Ile Gln Lys Glu Lys Gln
115 120 125

Lys Gln Arg Arg Met Met Pro Ala Val Ser Met Lys Gln Val Ser Val
130 135 140

Ala Asp Phe Lys His Val Ile Glu Asn Thr Val Arg His Leu Phe Gly
145 150 155 160

Asp Arg Arg Asp Arg Glu Ile Ala Glu Cys Ala Ala Leu Arg Ala Ala
165 170 175

Ser Lys Tyr Phe Leu Lys Ser Arg Arg Val Arg Pro Arg Lys Leu Pro
180 185 190

Lys Leu Ala Asn Pro Asp His Gly Lys Glu Leu Lys Gly Leu Arg Leu
195 200 205

Arg Glu Lys Arg Ala Lys Leu Lys Lys Glu Lys Glu Lys Gln Ala Glu
210 215 220

Leu Ala Arg Ser Asn Gln Lys Gly Ala Val Leu His Val Ala Thr Leu
225 230 235 240

Lys Lys Asp Ala Pro Pro Met Pro Tyr Glu Lys Thr Gln Gly Arg Asn

245 250 255

Asp Tyr Thr Thr Phe Val Ile Ser Ala Ala Ile Lys Val Gly Ala Thr

260 265 270

Arg Gly Thr Lys Pro Leu Leu Thr Pro Gln Pro Arg Glu Trp Gln Cys

275 280 285

Ser Leu Tyr Trp Arg Asp Gly Gln Arg Trp Ile Arg Gly Gly Leu Leu

290 295 300

Gly Leu Gln Ala Gly Ile Val Leu Gly Pro Lys Leu Asn Arg Glu Leu

305 310 315 320

Leu Glu Ala Val Leu Gln Arg Pro Ile Glu Cys Arg Met Ser Gly Cys

325 330 335

Gly Asn Pro Leu Gln Val Arg Gly Ala Ala Val Asp Phe Phe Met Thr

340 345 350

Thr Asn Pro Phe Tyr Val Ser Gly Ala Ala Tyr Ala Gln Lys Lys Phe

355 360 365

Lys Pro Phe Gly Thr Lys Arg Ala Ser Glu Asp Gly Ala Ala Ala Lys

370 375 380

Ala Arg Glu Lys Leu Met Thr Gln Leu Ala Lys Val Leu Asp Lys Val

385 390 395 400

Val Thr Gln Ala Ala His Ser Pro Leu Asp Gly Ile Trp Glu Thr Arg

405 410 415

Pro Glu Ala Lys Leu Arg Ala Met Ile Met Ala Leu Glu His Glu Trp

420 425 430

Ile Phe Leu Arg Pro Gly Pro Cys His Asn Ala Ala Glu Glu Val Ile

435 440 445

Lys Cys Asp Cys Thr Gly Gly His Ala Ile Leu Trp Ala Leu Ile Asp

450 455 460

Glu Ala Arg Gly Ala Leu Glu His Lys Glu Phe Tyr Ala Val Thr Arg

465 470 475 480

Ala His Thr His Asp Cys Glu Lys Gln Lys Leu Gly Gly Arg Leu Ala

485 490 495

Gly Phe Leu Asp Leu Leu Ile Ala Gln Asp Val Pro Leu Asp Asp Ala

500 505 510

Pro Ala Ala Arg Lys Ile Lys Thr Leu Leu Glu Ala Thr Pro Pro Ala

515 520 525

Pro Cys Tyr Lys Ala Ala Thr Ser Ile Ala Thr Cys Asp Cys Glu Gly

530 535 540

Lys Phe Asp Lys Leu Trp Ala Ile Ile Asp Ala Thr Arg Ala Gly His

545 550 555 560

Gly Thr Glu Asp Leu Trp Ala Arg Thr Leu Ala Tyr Pro Gln Asn Val

565 570 575

Asn Cys Lys Cys Lys Ala Gly Lys Asp Leu Thr His Arg Leu Ala Asp

580 585 590

Phe Leu Gly Leu Leu Ile Lys Arg Asp Gly Pro Phe Arg Glu Arg Pro

595 600 605

Pro His Lys Val Thr Gly Asp Arg Lys Leu Val Phe Ser Gly Asp Lys

610 615 620

Lys Cys Lys Gly His Gln Tyr Val Ile Leu Ala Lys Ala His Asn Glu

625 630 635 640

Glu Val Val Arg Ala Trp Ile Ser Arg Trp Gly Leu Lys Ser Arg Thr

645 650 655

Asn Lys Ala Gly Tyr Ala Ala Thr Glu Leu Asn Leu Leu Leu Asn Trp

660 665 670

Leu Ser Ile Cys Arg Arg Arg Trp Met Asp Met Leu Thr Val Gln Arg

675 680 685

Asp Thr Pro Tyr Ile Arg Met Lys Thr Gly Arg Leu Val Val Asp Asp

690 695 700

Lys Lys Glu Arg Lys Ala Met

705 710

<210> 206

<211> 574

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 206

Ala Lys Gln Arg Glu Ala Leu Arg Val Ala Leu Glu Arg Gly Ile Val
1 5 10 15

Arg Ala Ser Asn Arg Thr Tyr Thr Leu Val Thr Asn Cys Thr Lys Gly
20 25 30

Gly Pro Leu Pro Glu Gln Cys Arg Met Ile Glu Arg Gly Lys Ala Arg
35 40 45

Ala Met Lys Trp Glu Pro Lys Leu Val Gly Cys Gly Ser Cys Ala Ala
50 55 60

Ala Thr Val Asp Leu Pro Ala Ile Glu Glu Tyr Ala Gln Pro Gly Arg
65 70 75 80

Leu Asp Val Ala Lys Tyr Lys Leu Thr Thr Gln Ile Leu Ala Met Ala
85 90 95

Thr Arg Arg Met Met Val Arg Ala Ala Lys Leu Ser Arg Arg Lys Gly
100 105 110

Gln Trp Pro Ala Lys Val Gln Glu Glu Lys Glu Glu Pro Pro Glu Pro
115 120 125

Lys Lys Met Leu Lys Ala Val Glu Met Arg Pro Val Ala Ile Val Asp
130 135 140

Phe Asn Arg Val Ile Gln Thr Thr Ile Glu His Leu Trp Ala Glu Arg
145 150 155 160

Ala Asn Ala Asp Glu Ala Glu Leu Lys Ala Leu Lys Ala Ala Ala Ala
165 170 175

Tyr Phe Gly Pro Ser Leu Lys Ile Arg Ala Arg Gly Pro Pro Lys Ala
180 185 190

Ala Ile Gly Arg Glu Leu Lys Lys Ala His Arg Lys Lys Ala Tyr Ala
195 200 205

Glu Arg Lys Lys Ala Arg Arg Lys Arg Ala Glu Leu Ala Arg Ser Gln
210 215 220

Ala Arg Gly Ala Ala Ala His Ala Ala Ile Arg Glu Arg Asp Ile Pro

225 230 235 240

Pro Met Ala Tyr Glu Arg Thr Gln Gly Arg Asn Asp Val Thr Thr Ile

245 250 255

Pro Ile Ala Ala Ala Ile Lys Ile Ala Ala Thr Arg Gly Ala Arg Pro

260 265 270

Leu Pro Ala Pro Lys Pro Met Lys Trp Gln Cys Ser Leu Tyr Trp Asn

275 280 285

Glu Gly Gln Arg Trp Ile Arg Gly Gly Met Leu Thr Ala Gln Ala Tyr

290 295 300

Ala His Ala Ala Asn Ile His Arg Pro Met Arg Cys Glu Met Trp Gly

305 310 315 320

Val Gly Asn Pro Leu Lys Val Arg Ala Phe Glu Gly Arg Val Ala Asp

325 330 335

Pro Asp Gly Ala Lys Gly Arg Lys Ala Glu Phe Arg Leu Gln Thr Asn

340 345 350

Ala Phe Tyr Val Ser Gly Ala Ala Tyr Arg Asn Lys Lys Phe Lys Pro

355 360 365

Phe Gly Thr Asp Arg Gly Gly Ile Gly Ser Ala Arg Lys Lys Arg Glu

370 375 380

Arg Leu Met Ala Gln Leu Ala Lys Ile Leu Asp Lys Val Val Ser Gln

385 390 395 400

Ala Ala His Ser Pro Leu Asp Asp Ile Trp His Thr Arg Pro Ala Gln

405 410 415

Lys Leu Arg Ala Met Ile Lys Gln Leu Glu His Glu Trp Met Phe Leu

420 425 430

Arg Pro Gln Ala Pro Thr Val Glu Gly Thr Lys Pro Asp Val Asp Val

435 440 445

Ala Gly Asn Met Gln Arg Gln Ile Lys Ala Leu Met Ala Pro Asp Leu

450 455 460

Pro Pro Ile Glu Lys Gly Ser Pro Ala Lys Arg Phe Thr Gly Asp Lys

465 470 475 480

Arg Lys Lys Gly Glu Arg Ala Val Arg Val Ala Glu Ala His Ser Asp

485 490 495

Glu Val Val Thr Ala Trp Ile Ser Arg Trp Gly Ile Gln Thr Arg Arg

500 505 510

Asn Glu Gly Ser Tyr Ala Ala Gln Glu Leu Glu Leu Leu Leu Asn Trp

515 520 525

Leu Gln Ile Cys Arg Arg Arg Trp Leu Asp Met Thr Ala Ala Gln Arg

530 535 540

Val Ser Pro Tyr Ile Arg Met Lys Ser Gly Arg Met Ile Thr Asp Ala

545 550 555 560

Ala Asp Glu Gly Val Ala Pro Ile Pro Leu Val Glu Asn Met

565 570

<210> 207

<211> 743

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 207

Lys Ser Ile Ser Gly Arg Ser Ile Lys His Met Ala Cys Leu Lys Asp

1 5 10 15

Met Leu Lys Ser Glu Ile Thr Glu Ile Glu Glu Lys Gln Lys Lys Glu

20 25 30

Ser Leu Arg Lys Trp Asp Tyr Tyr Ser Lys Phe Ser Asp Glu Ile Leu

35 40 45

Phe Arg Arg Asn Leu Asn Val Ser Ala Asn His Asp Ala Asn Ala Cys

50 55 60

Tyr Gly Cys Asn Pro Cys Ala Phe Leu Lys Glu Val Tyr Gly Phe Arg

65 70 75 80

Ile Glu Arg Arg Asn Asn Glu Arg Ile Ile Ser Tyr Arg Arg Gly Leu

85 90 95

Ala Gly Cys Lys Ser Cys Val Gln Ser Thr Gly Tyr Pro Pro Ile Glu
100 105 110
Phe Val Arg Arg Lys Phe Gly Ala Asp Lys Ala Met Glu Ile Val Arg
115 120 125
Glu Val Leu His Arg Arg Asn Trp Gly Ala Leu Ala Arg Asn Ile Gly
130 135 140
Arg Glu Lys Glu Ala Asp Pro Ile Leu Gly Glu Leu Asn Glu Leu Leu
145 150 155 160

Leu Val Asp Ala Arg Pro Tyr Phe Gly Asn Lys Ser Ala Ala Asn Glu
165 170 175
Thr Asn Leu Ala Phe Asn Val Ile Thr Arg Ala Ala Lys Lys Phe Arg
180 185 190
Asp Glu Gly Met Tyr Asp Ile His Lys Gln Leu Asp Ile His Ser Glu
195 200 205
Glu Gly Lys Val Pro Lys Gly Arg Lys Ser Arg Leu Ile Arg Ile Glu
210 215 220

Arg Lys His Lys Ala Ile His Gly Leu Asp Pro Gly Glu Thr Trp Arg
225 230 235 240
Tyr Pro His Cys Gly Lys Gly Glu Lys Tyr Gly Val Trp Leu Asn Arg
245 250 255
Ser Arg Leu Ile His Ile Lys Gly Asn Glu Tyr Arg Cys Leu Thr Ala
260 265 270
Phe Gly Thr Thr Gly Arg Arg Met Ser Leu Asp Val Ala Cys Ser Val
275 280 285

Leu Gly His Pro Leu Val Lys Lys Lys Arg Lys Lys Gly Lys Lys Thr
290 295 300
Val Asp Gly Thr Glu Leu Trp Gln Ile Lys Lys Ala Thr Glu Thr Leu
305 310 315 320
Pro Glu Asp Pro Ile Asp Cys Thr Phe Tyr Leu Tyr Ala Ala Lys Pro
325 330 335
Thr Lys Asp Pro Phe Ile Leu Lys Val Gly Ser Leu Lys Ala Pro Arg

340 345 350

Trp Lys Lys Leu His Lys Asp Phe Phe Glu Tyr Ser Asp Thr Glu Lys
355 360 365
Thr Gln Gly Gln Glu Lys Gly Lys Arg Val Val Arg Arg Gly Lys Val
370 375 380
Pro Arg Ile Leu Ser Leu Arg Pro Asp Ala Lys Phe Lys Val Ser Ile
385 390 395 400
Trp Asp Asp Pro Tyr Asn Gly Lys Asn Lys Glu Gly Thr Leu Leu Arg
405 410 415

Met Glu Leu Ser Gly Leu Asp Gly Ala Lys Lys Pro Leu Ile Leu Lys
420 425 430
Arg Tyr Gly Glu Pro Asn Thr Lys Pro Lys Asn Phe Val Phe Trp Arg
435 440 445
Pro His Ile Thr Pro His Pro Leu Thr Phe Thr Pro Lys His Asp Phe
450 455 460
Gly Asp Pro Asn Lys Lys Thr Lys Arg Arg Arg Val Phe Asn Arg Glu
465 470 475 480

Tyr Tyr Gly His Leu Asn Asp Leu Ala Lys Met Glu Pro Asn Ala Lys
485 490 495
Phe Phe Glu Asp Arg Glu Val Ser Asn Lys Lys Asn Pro Lys Ala Lys
500 505 510
Asn Ile Arg Ile Gln Ala Lys Glu Ser Leu Pro Asn Ile Val Ala Lys
515 520 525
Asn Gly Arg Trp Ala Ala Phe Asp Pro Asn Asp Ser Leu Trp Lys Leu
530 535 540

Tyr Leu His Trp Arg Gly Arg Arg Lys Thr Ile Lys Gly Gly Ile Ser
545 550 555 560
Gln Glu Phe Gln Glu Phe Lys Glu Arg Leu Asp Leu Tyr Lys Lys His
565 570 575
Glu Asp Glu Ser Glu Trp Lys Glu Lys Glu Lys Leu Trp Glu Asn His
580 585 590

Glu Lys Glu Trp Lys Lys Thr Leu Glu Ile His Gly Ser Ile Ala Glu
595 600 605

Val Ser Gln Arg Cys Val Met Gln Ser Met Met Gly Pro Leu Asp Gly
610 615 620
Leu Val Gln Lys Lys Asp Tyr Val His Ile Gly Gln Ser Ser Leu Lys
625 630 635 640
Ala Ala Asp Asp Ala Trp Thr Phe Ser Ala Asn Arg Tyr Lys Lys Ala
645 650 655
Thr Gly Pro Lys Trp Gly Lys Ile Ser Val Ser Asn Leu Leu Tyr Asp
660 665 670

Ala Asn Gln Ala Asn Ala Glu Leu Ile Ser Gln Ser Ile Ser Lys Tyr
675 680 685
Leu Ser Lys Gln Lys Asp Asn Gln Gly Cys Glu Gly Arg Lys Met Lys
690 695 700
Phe Leu Ile Lys Ile Ile Glu Pro Leu Arg Glu Asn Phe Val Lys His
705 710 715 720
Thr Arg Trp Leu His Glu Met Thr Gln Lys Asp Cys Glu Val Arg Ala
725 730 735

Gln Phe Ser Arg Val Ser Met
740

<210> 208
<211> 769
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 208
Phe Pro Ser Asp Val Gly Ala Asp Ala Leu Lys His Val Arg Met Leu
1 5 10 15
Gln Pro Arg Leu Thr Asp Glu Val Arg Lys Val Ala Leu Thr Arg Ala
20 25 30
Pro Ser Asp Arg Pro Ala Leu Ala Arg Phe Ala Ala Val Ala Gln Asp

35 40 45

Gly Leu Ala Phe Val Arg His Leu Asn Val Ser Ala Asn His Asp Ser
50 55 60
Asn Cys Thr Phe Pro Arg Asp Pro Arg Asp Pro Arg Arg Gly Pro Cys
65 70 75 80
Glu Pro Asn Pro Cys Ala Phe Leu Arg Glu Val Trp Gly Phe Arg Ile
85 90 95
Val Ala Arg Gly Asn Glu Arg Ala Leu Ser Tyr Arg Arg Gly Leu Ala
100 105 110

Gly Cys Lys Ser Cys Val Gln Ser Thr Gly Phe Pro Ser Val Pro Phe
115 120 125
His Arg Ile Gly Ala Asp Asp Cys Met Arg Lys Leu His Glu Ile Leu
130 135 140
Lys Ala Arg Asn Trp Arg Leu Leu Ala Arg Asn Ile Gly Arg Glu Arg
145 150 155 160
Glu Ala Asp Pro Leu Leu Thr Glu Leu Ser Glu Tyr Leu Leu Val Asp
165 170 175

Ala Arg Thr Tyr Pro Asp Gly Ala Ala Pro Asn Ser Gly Arg Leu Ala
180 185 190
Glu Asn Val Ile Lys Arg Ala Ala Lys Lys Phe Arg Asp Glu Gly Met
195 200 205
Arg Asp Ile His Ala Gln Leu Arg Val His Ser Arg Glu Gly Lys Val
210 215 220
Pro Lys Gly Arg Leu Gln Arg Leu Arg Arg Ile Glu Arg Lys His Arg
225 230 235 240

Ala Ile His Ala Leu Asp Pro Gly Pro Ser Trp Glu Ala Glu Gly Ser
245 250 255
Ala Arg Ala Glu Val Gln Gly Val Ala Val Tyr Arg Ser Gln Leu Leu
260 265 270
Arg Val Gly His His Thr Gln Gln Ile Glu Pro Val Gly Ile Val Ala
275 280 285

Arg Thr Leu Phe Gly Val Gly Arg Thr Asp Leu Asp Val Ala Val Ser
290 295 300

Val Leu Gly Ala Pro Leu Thr Lys Arg Lys Lys Gly Ser Lys Thr Leu
305 310 315 320
Glu Ser Thr Glu Asp Phe Arg Ile Ala Lys Ala Arg Glu Thr Arg Ala
325 330 335
Glu Asp Lys Ile Glu Val Ala Phe Val Leu Tyr Pro Thr Ala Ser Leu
340 345 350
Leu Arg Asp Glu Ile Pro Lys Asp Ala Phe Pro Ala Met Arg Ile Asp
355 360 365

Arg Phe Leu Leu Lys Val Gly Ser Val Gln Ala Asp Arg Glu Ile Leu
370 375 380
Leu Gln Asp Asp Tyr Tyr Arg Phe Gly Asp Ala Glu Val Lys Ala Gly
385 390 395 400
Lys Asn Lys Gly Arg Thr Val Thr Arg Pro Val Lys Val Pro Arg Leu
405 410 415
Gln Ala Leu Arg Pro Asp Ala Lys Phe Arg Val Asn Val Trp Ala Asp
420 425 430

Pro Phe Gly Ala Gly Asp Ser Pro Gly Thr Leu Leu Arg Leu Glu Val
435 440 445
Ser Gly Val Thr Arg Arg Ser Gln Pro Leu Arg Leu Leu Arg Tyr Gly
450 455 460
Gln Pro Ser Thr Gln Pro Ala Asn Phe Leu Cys Trp Arg Pro His Arg
465 470 475 480
Val Pro Asp Pro Met Thr Phe Thr Pro Arg Gln Lys Phe Gly Glu Arg
485 490 495

Arg Lys Asn Arg Arg Thr Arg Arg Pro Arg Val Phe Glu Arg Leu Tyr
500 505 510
Gln Val His Ile Lys His Leu Ala His Leu Glu Pro Asn Arg Lys Trp
515 520 525
Phe Glu Glu Ala Arg Val Ser Ala Gln Lys Trp Ala Lys Ala Arg Ala

530 535 540

Ile Arg Arg Lys Gly Ala Glu Asp Ile Pro Val Val Ala Pro Pro Ala
545 550 555 560

Lys Arg Arg Trp Ala Ala Leu Gln Pro Asn Ala Glu Leu Trp Asp Leu
565 570 575

Tyr Ala His Asp Arg Glu Ala Arg Lys Arg Phe Arg Gly Gly Arg Ala
580 585 590

Ala Glu Gly Glu Glu Phe Lys Pro Arg Leu Asn Leu Tyr Leu Ala His
595 600 605

Glu Pro Glu Ala Glu Trp Glu Ser Lys Arg Asp Arg Trp Glu Arg Tyr
610 615 620

Glu Lys Lys Trp Thr Ala Val Leu Glu Glu His Ser Arg Met Cys Ala
625 630 635 640

Val Ala Asp Arg Thr Leu Pro Gln Phe Leu Ser Asp Pro Leu Gly Ala
645 650 655

Arg Met Asp Asp Lys Asp Tyr Ala Phe Val Gly Lys Ser Ala Leu Ala
660 665 670

Val Ala Glu Ala Phe Val Glu Glu Gly Thr Val Glu Arg Ala Gln Gly
675 680 685

Asn Cys Ser Ile Thr Ala Lys Lys Lys Phe Ala Ser Asn Ala Ser Arg
690 695 700

Lys Arg Leu Ser Val Ala Asn Leu Leu Asp Val Ser Asp Lys Ala Asp
705 710 715 720

Arg Ala Leu Val Phe Gln Ala Val Arg Gln Tyr Val Gln Arg Gln Ala
725 730 735

Glu Asn Gly Gly Val Glu Gly Arg Arg Met Ala Phe Leu Arg Lys Leu
740 745 750

Leu Ala Pro Leu Arg Gln Asn Phe Val Cys His Thr Arg Trp Leu His
755 760 765

Met

<210> 209

<211> 564

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 209

Ala Ala Arg Lys Lys Lys Arg Gly Lys Ile Gly Ile Thr Val Lys Ala
1 5 10 15

Lys Glu Lys Ser Pro Pro Ala Ala Gly Pro Phe Met Ala Arg Lys Leu
20 25 30

Val Asn Val Ala Ala Asn Val Asp Gly Val Glu Val His Leu Cys Val
35 40 45

Glu Cys Glu Ala Asp Ala His Gly Ser Ala Ser Ala Arg Leu Leu Gly
50 55 60

Gly Cys Arg Ser Cys Thr Gly Ser Ile Gly Ala Glu Gly Arg Leu Met
65 70 75 80

Gly Ser Val Asp Val Asp Arg Glu Arg Val Ile Ala Glu Pro Val His
85 90 95

Thr Glu Thr Glu Arg Leu Gly Pro Asp Val Lys Ala Phe Glu Ala Gly
100 105 110

Thr Ala Glu Ser Lys Tyr Ala Ile Gln Arg Gly Leu Glu Tyr Trp Gly
115 120 125

Val Asp Leu Ile Ser Arg Asn Arg Ala Arg Thr Val Arg Lys Met Glu
130 135 140

Glu Ala Asp Arg Pro Glu Ser Ser Thr Met Glu Lys Thr Ser Trp Asp
145 150 155 160

Glu Ile Ala Ile Lys Thr Tyr Ser Gln Ala Tyr His Ala Ser Glu Asn
165 170 175

His Leu Phe Trp Glu Arg Gln Arg Arg Val Arg Gln His Ala Leu Ala
180 185 190

Leu Phe Arg Arg Ala Arg Glu Arg Asn Arg Gly Glu Ser Pro Leu Gln
195 200 205

Ser Thr Gln Arg Pro Ala Pro Leu Val Leu Ala Ala Leu His Ala Glu
210 215 220

Ala Ala Ala Ile Ser Gly Arg Ala Arg Ala Glu Tyr Val Leu Arg Gly
225 230 235 240
Pro Ser Ala Asn Val Arg Ala Ala Ala Ala Asp Ile Asp Ala Lys Pro
245 250 255
Leu Gly His Tyr Lys Thr Pro Ser Pro Lys Val Ala Arg Gly Phe Pro
260 265 270
Val Lys Arg Asp Leu Leu Arg Ala Arg His Arg Ile Val Gly Leu Ser
275 280 285

Arg Ala Tyr Phe Lys Pro Ser Asp Val Val Arg Gly Thr Ser Asp Ala
290 295 300
Ile Ala His Val Ala Gly Arg Asn Ile Gly Val Ala Gly Gly Lys Pro
305 310 315 320
Lys Glu Ile Glu Lys Thr Phe Thr Leu Pro Phe Val Ala Tyr Trp Glu
325 330 335
Asp Val Asp Arg Val Val His Cys Ser Ser Phe Lys Ala Asp Gly Pro
340 345 350

Trp Val Arg Asp Gln Arg Ile Lys Ile Arg Gly Val Ser Ser Ala Val
355 360 365
Gly Thr Phe Ser Leu Tyr Gly Leu Asp Val Ala Trp Ser Lys Pro Thr
370 375 380
Ser Phe Tyr Ile Arg Cys Ser Asp Ile Arg Lys Lys Phe His Pro Lys
385 390 395 400
Gly Phe Gly Pro Met Lys His Trp Arg Gln Trp Ala Lys Glu Leu Asp
405 410 415

Arg Leu Thr Glu Gln Arg Ala Ser Cys Val Val Arg Ala Leu Gln Asp
420 425 430
Asp Glu Glu Leu Leu Gln Thr Met Glu Arg Gly Gln Arg Tyr Tyr Asp
435 440 445
Val Phe Ser Cys Ala Ala Thr His Ala Thr Arg Gly Glu Ala Asp Pro

450 455 460

Ser Gly Gly Cys Ser Arg Cys Glu Leu Val Ser Cys Gly Val Ala His
465 470 475 480

Lys Val Thr Lys Lys Ala Lys Gly Asp Thr Gly Ile Glu Ala Val Ala
485 490 495

Val Ala Gly Cys Ser Leu Cys Glu Ser Lys Leu Val Gly Pro Ser Lys
500 505 510

Pro Arg Val His Arg Gln Met Ala Ala Leu Arg Gln Ser His Ala Leu
515 520 525

Asn Tyr Leu Arg Arg Leu Gln Arg Glu Trp Glu Ala Leu Glu Ala Val
530 535 540

Gln Ala Pro Thr Pro Tyr Leu Arg Phe Lys Tyr Ala Arg His Leu Glu
545 550 555 560

Val Arg Ser Met

<210> 210

<211> 565

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 210

Ala Ala Lys Lys Lys Lys Gln Arg Gly Lys Ile Gly Ile Ser Val Lys
1 5 10 15

Pro Lys Glu Gly Ser Ala Pro Pro Ala Asp Gly Pro Phe Met Ala Arg
20 25 30

Lys Leu Val Asn Val Ala Ala Asn Val Asp Gly Val Glu Val Asn Leu
35 40 45

Cys Ile Glu Cys Glu Ala Asp Ala His Gly Ser Ala Pro Ala Arg Leu
50 55 60

Leu Gly Gly Cys Lys Ser Cys Thr Gly Ser Ile Gly Ala Glu Gly Arg
65 70 75 80

Leu Met Gly Ser Val Asp Val Asp Arg Ala Asp Ala Ile Ala Lys Pro

85 90 95

Val Asn Thr Glu Thr Glu Lys Leu Gly Pro Asp Val Gln Ala Phe Glu
100 105 110
Ala Gly Thr Ala Glu Thr Lys Tyr Ala Leu Gln Arg Gly Leu Glu Tyr
115 120 125
Trp Gly Val Asp Leu Ile Ser Arg Asn Arg Ser Arg Thr Val Arg Arg
130 135 140
Thr Glu Glu Gly Gln Pro Glu Ser Ala Thr Met Glu Lys Thr Ser Trp
145 150 155 160

Asp Glu Ile Ala Ile Lys Ser Tyr Thr Arg Ala Tyr His Ala Ser Glu
165 170 175
Asn His Leu Phe Trp Glu Arg Gln Arg Arg Val Arg Gln His Ala Leu
180 185 190
Ala Leu Phe Lys Arg Ala Lys Glu Arg Asn Arg Gly Asp Ser Thr Leu
195 200 205
Pro Arg Glu Pro Gly His Gly Leu Val Ala Ile Ala Ala Leu Ala Cys
210 215 220

Glu Ala Tyr Ala Val Gly Gly Arg Asn Leu Ala Glu Thr Val Val Arg
225 230 235 240
Gly Pro Thr Phe Gly Thr Ala Arg Ala Val Arg Asp Val Glu Ile Ala
245 250 255
Ser Leu Gly Arg Tyr Lys Thr Pro Ser Pro Lys Val Ala His Gly Ser
260 265 270
Pro Val Lys Arg Asp Phe Leu Arg Ala Arg His Arg Ile Val Gly Leu
275 280 285

Ala Arg Ala Tyr Tyr Arg Pro Ser Asp Val Val Arg Gly Thr Ser Asp
290 295 300
Ala Ile Ala His Val Ala Gly Arg Asn Ile Gly Val Ala Gly Gly Lys
305 310 315 320
Pro Arg Ala Val Glu Ala Val Phe Thr Leu Pro Phe Val Ala Tyr Trp
325 330 335

Glu Asp Val Asp Arg Val Val His Cys Ser Ser Phe Gln Val Ser Ala
340 345 350

Pro Trp Asn Arg Asp Gln Arg Met Lys Ile Ala Gly Val Thr Thr Ala
355 360 365
Ala Gly Thr Phe Ser Leu His Gly Gly Glu Leu Lys Trp Ala Lys Pro
370 375 380
Thr Ser Phe Tyr Ile Arg Cys Ser Asp Thr Arg Arg Lys Phe Arg Pro
385 390 395 400
Lys Gly Phe Gly Pro Met Lys Arg Trp Arg Gln Trp Ala Lys Asp Leu
405 410 415

Asp Arg Leu Val Glu Gln Arg Ala Ser Cys Val Val Arg Ala Leu Gln
420 425 430
Asp Asp Ala Ala Leu Leu Glu Thr Met Glu Arg Gly Gln Arg Tyr Tyr
435 440 445
Asp Val Phe Ala Cys Ala Val Thr His Ala Thr Arg Gly Glu Ala Asp
450 455 460
Arg Leu Ala Gly Cys Ser Arg Cys Ala Leu Thr Pro Cys Gln Glu Ala
465 470 475 480

His Arg Val Thr Thr Lys Pro Arg Gly Asp Ala Gly Val Glu Gln Val
485 490 495
Gln Thr Ser Asp Cys Ser Leu Cys Glu Gly Lys Leu Val Gly Pro Ser
500 505 510
Lys Pro Arg Leu His Arg Thr Leu Thr Leu Leu Arg Gln Glu His Gly
515 520 525
Leu Asn Tyr Leu Arg Arg Leu Gln Arg Glu Trp Glu Ser Leu Glu Ala
530 535 540

Val Gln Val Pro Thr Pro Tyr Leu Arg Phe Lys Tyr Ala Arg His Leu
545 550 555 560
Glu Val Arg Ser Met
565

<210> 211

<211> 499

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 211

Thr Asp Ser Gln Ser Glu Ser Val Pro Glu Val Val Tyr Ala Leu Thr
1 5 10 15

Gly Gly Glu Val Pro Gly Arg Val Pro Pro Asp Gly Gly Ser Ala Glu
20 25 30

Gly Ala Arg Asn Ala Pro Thr Gly Leu Arg Lys Gln Arg Gly Lys Ile
35 40 45

Lys Ile Ser Ala Lys Pro Ser Lys Pro Gly Ser Pro Ala Ser Ser Leu
50 55 60

Ala Arg Thr Leu Val Asn Glu Ala Ala Asn Val Asp Gly Val Gln Ser
65 70 75 80

Ser Gly Cys Ala Thr Cys Arg Met Arg Ala Asn Gly Ser Ala Pro Arg
85 90 95

Ala Leu Pro Ile Gly Cys Val Ala Cys Ala Ser Ser Ile Gly Arg Ala
100 105 110

Pro Gln Glu Glu Thr Val Cys Ala Leu Pro Thr Thr Gln Gly Pro Asp
115 120 125

Val Arg Leu Leu Glu Gly Gly His Ala Leu Arg Lys Tyr Asp Ile Gln
130 135 140

Arg Ala Leu Glu Tyr Trp Gly Val Asp Leu Ile Gly Arg Asn Leu Asp
145 150 155 160

Arg Gln Ala Gly Arg Gly Met Glu Pro Ala Glu Gly Ala Thr Ala Thr
165 170 175

Met Lys Arg Val Ser Met Asp Glu Leu Ala Val Leu Asp Phe Gly Lys
180 185 190

Ser Tyr Tyr Ala Ser Glu Gln His Leu Phe Ala Ala Arg Gln Arg Arg
195 200 205

Val Arg Gln His Ala Lys Ala Leu Lys Ile Arg Ala Lys His Ala Asn

210 215 220

Arg Ser Gly Ser Val Lys Arg Ala Leu Asp Arg Ser Arg Lys Gln Val
225 230 235 240
Thr Ala Leu Ala Arg Glu Phe Phe Lys Pro Ser Asp Val Val Arg Gly
245 250 255
Asp Ser Asp Ala Leu Ala His Val Val Gly Arg Asn Leu Gly Val Ser
260 265 270
Arg His Pro Ala Arg Glu Ile Pro Gln Thr Phe Thr Leu Pro Leu Cys
275 280 285

Ala Tyr Trp Glu Asp Val Asp Arg Val Ile Ser Cys Ser Ser Leu Leu
290 295 300
Ala Gly Glu Pro Phe Ala Arg Asp Gln Glu Ile Arg Ile Glu Gly Val
305 310 315 320
Ser Ser Ala Leu Gly Ser Leu Arg Leu Tyr Arg Gly Ala Ile Glu Trp
325 330 335
His Lys Pro Thr Ser Leu Tyr Ile Arg Cys Ser Asp Thr Arg Arg Lys
340 345 350

Phe Arg Pro Arg Gly Gly Leu Lys Lys Arg Trp Arg Gln Trp Ala Lys
355 360 365
Asp Leu Asp Arg Leu Val Glu Gln Arg Ala Cys Cys Ile Val Arg Ser
370 375 380
Leu Gln Ala Asp Val Glu Leu Leu Gln Thr Met Glu Arg Ala Gln Arg
385 390 395 400
Phe Tyr Asp Val His Asp Cys Ala Ala Thr His Val Gly Pro Val Ala
405 410 415

Val Arg Cys Ser Pro Cys Ala Gly Lys Gln Phe Asp Trp Asp Arg Tyr
420 425 430
Arg Leu Leu Ala Ala Leu Arg Gln Glu His Ala Leu Asn Tyr Leu Arg
435 440 445
Arg Leu Gln Arg Glu Trp Glu Ser Leu Glu Ala Gln Gln Val Lys Met
450 455 460

Pro Tyr Leu Arg Phe Lys Tyr Ala Arg Lys Leu Glu Val Ser Gly Pro
465 470 475 480

Leu Ile Gly Leu Glu Val Arg Arg Glu Pro Ser Met Gly Thr Ala Ile
485 490 495

Ala Glu Met

<210> 212

<211> 358

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 212

Ala Gly Thr Ala Gly Arg Arg His Gly Ser Leu Gly Ala Arg Arg Ser
1 5 10 15

Ile Asn Ile Ala Gly Val Thr Asp Arg His Gly Arg Trp Gly Cys Glu
20 25 30

Ser Cys Val Tyr Thr Arg Asp Gln Ala Gly Asn Arg Ala Arg Cys Ala
35 40 45

Pro Cys Asp Gln Ser Thr Tyr Ala Pro Asp Val Gln Glu Val Thr Ile
50 55 60

Gly Gln Arg Gln Ala Lys Tyr Thr Ile Phe Leu Thr Leu Gln Ser Phe
65 70 75 80

Ser Trp Thr Asn Thr Met Arg Asn Asn Lys Arg Ala Ala Ala Gly Arg
85 90 95

Ser Lys Arg Thr Thr Gly Lys Arg Ile Gly Gln Leu Ala Glu Ile Lys
100 105 110

Ile Thr Gly Val Gly Leu Ala His Ala His Asn Val Ile Gln Arg Ser
115 120 125

Leu Gln His Asn Ile Thr Lys Met Trp Arg Ala Glu Lys Gly Lys Ser
130 135 140

Lys Arg Val Ala Arg Leu Lys Lys Ala Lys Gln Leu Thr Lys Arg Arg

145 150 155 160

Ala Tyr Phe Arg Arg Arg Met Ser Arg Gln Ser Arg Gly Asn Gly Phe
165 170 175
Phe Arg Thr Gly Lys Gly Gly Ile His Ala Val Ala Pro Val Lys Ile
180 185 190
Gly Leu Asp Val Gly Met Ile Ala Ser Gly Ser Ser Glu Pro Ala Asp
195 200 205
Glu Gln Thr Val Thr Leu Asp Ala Ile Trp Lys Gly Arg Lys Lys Lys
210 215 220

Ile Arg Leu Ile Gly Ala Lys Gly Glu Leu Ala Val Ala Ala Cys Arg
225 230 235 240
Phe Arg Glu Gln Gln Thr Lys Gly Asp Lys Cys Ile Pro Leu Ile Leu
245 250 255
Gln Asp Gly Glu Val Arg Trp Asn Gln Asn Asn Trp Gln Cys His Pro
260 265 270
Lys Lys Leu Val Pro Leu Cys Gly Leu Glu Val Ser Arg Lys Phe Val
275 280 285

Ser Gln Ala Asp Arg Leu Ala Gln Asn Lys Val Ala Ser Pro Leu Ala
290 295 300
Ala Arg Phe Asp Lys Thr Ser Val Lys Gly Thr Leu Val Glu Ser Asp
305 310 315 320
Phe Ala Ala Val Leu Val Asn Val Thr Ser Ile Tyr Gln Gln Cys His
325 330 335
Ala Met Leu Leu Arg Ser Gln Glu Pro Thr Pro Ser Leu Arg Val Gln
340 345 350

Arg Thr Ile Thr Ser Met
355

<210> 213
<211> 369
<212> PRT
<213> Artificial sequence

<220><223> Synthetic sequence

<400> 213

Gly Val Arg Phe Ser Pro Ala Gln Ser Gln Val Phe Phe Arg Thr Val
1 5 10 15

Ile Pro Gln Ser Val Glu Ala Arg Phe Ala Ile Asn Met Ala Ala Ile
20 25 30

His Asp Ala Ala Gly Ala Phe Gly Cys Ser Val Cys Arg Phe Glu Asp
35 40 45

Arg Thr Pro Arg Asn Ala Lys Ala Val His Gly Cys Ser Pro Cys Thr
50 55 60

Arg Ser Thr Asn Arg Pro Asp Val Phe Val Leu Pro Val Gly Ala Ile
65 70 75 80

Lys Ala Lys Tyr Asp Val Phe Met Arg Leu Leu Gly Phe Asn Trp Thr
85 90 95

His Leu Asn Arg Arg Gln Ala Lys Arg Val Thr Val Arg Asp Arg Ile
100 105 110

Gly Gln Leu Asp Glu Leu Ala Ile Ser Met Leu Thr Gly Lys Ala Lys
115 120 125

Ala Val Leu Lys Lys Ser Ile Cys His Asn Val Asp Lys Ser Phe Lys
130 135 140

Ala Met Arg Gly Ser Leu Lys Lys Leu His Arg Lys Ala Ser Lys Thr
145 150 155 160

Gly Lys Ser Gln Leu Arg Ala Lys Leu Ser Asp Leu Arg Glu Arg Thr
165 170 175

Asn Thr Thr Gln Glu Gly Ser His Val Glu Gly Asp Ser Asp Val Ala
180 185 190

Leu Asn Lys Ile Gly Leu Asp Val Gly Leu Val Gly Lys Pro Asp Tyr
195 200 205

Pro Ser Glu Glu Ser Val Glu Val Val Val Cys Leu Tyr Phe Val Gly
210 215 220

Lys Val Leu Ile Leu Asp Ala Gln Gly Arg Ile Arg Asp Met Arg Ala

225 230 235 240

Lys Gln Tyr Asp Gly Phe Lys Ile Pro Ile Ile Gln Arg Gly Gln Leu
245 250 255
Thr Val Leu Ser Val Lys Asp Leu Gly Lys Trp Ser Leu Val Arg Gln
260 265 270
Asp Tyr Val Leu Ala Gly Asp Leu Arg Phe Glu Pro Lys Ile Ser Lys
275 280 285
Asp Arg Lys Tyr Ala Glu Cys Val Lys Arg Ile Ala Leu Ile Thr Leu
290 295 300

Gln Ala Ser Leu Gly Phe Lys Glu Arg Ile Pro Tyr Tyr Val Thr Lys
305 310 315 320
Gln Val Glu Ile Lys Asn Ala Ser His Ile Ala Phe Val Thr Glu Ala
325 330 335
Ile Gln Asn Cys Ala Glu Asn Phe Arg Glu Met Thr Glu Tyr Leu Met
340 345 350
Lys Tyr Gln Glu Lys Ser Pro Asp Leu Lys Val Leu Leu Thr Gln Leu
355 360 365

Met

<210> 214
<211> 486
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 214
Arg Ala Val Val Gly Lys Val Phe Leu Glu Gln Ala Arg Arg Ala Leu
1 5 10 15
Asn Leu Ala Thr Asn Phe Gly Thr Asn His Arg Thr Gly Cys Asn Gly
20 25 30
Cys Tyr Val Thr Pro Gly Lys Leu Ser Ile Pro Gln Asp Gly Glu Lys
35 40 45

```
Asn Ala Ala Gly Cys Thr Ser Cys Leu Met Lys Ala Thr Ala Ser Tyr
    50                  55                  60
Val Ser Tyr Pro Lys Pro Leu Gly Glu Lys Val Ala Lys Tyr Ser Thr
65                  70                  75                  80
Leu Asp Ala Leu Lys Gly Phe Pro Trp Tyr Ser Leu Arg Leu Asn Leu
                85                  90                  95
Arg Pro Asn Tyr Arg Gly Lys Pro Ile Asn Gly Val Gln Glu Val Ala
            100                 105                 110

Pro Val Ser Lys Phe Arg Leu Ala Glu Glu Val Ile Gln Ala Val Gln
        115                 120                 125
Arg Tyr His Phe Thr Glu Leu Glu Gln Ser Phe Pro Gly Gly Arg Arg
    130                 135                 140
Arg Leu Arg Glu Leu Arg Ala Phe Tyr Thr Lys Glu Tyr Arg Arg Ala
145                 150                 155                 160
Pro Glu Gln Arg Gln His Val Val Asn Gly Asp Arg Asn Ile Val Val
                165                 170                 175

Val Thr Val Leu His Glu Leu Gly Phe Ser Val Gly Met Phe Asn Glu
            180                 185                 190
Val Glu Leu Leu Pro Lys Thr Pro Ile Glu Cys Ala Val Asn Val Phe
        195                 200                 205
Ile Arg Gly Asn Arg Val Leu Leu Glu Val Arg Lys Pro Gln Phe Asp
    210                 215                 220
Lys Glu Arg Leu Leu Val Glu Ser Leu Trp Lys Lys Asp Ser Arg Arg
225                 230                 235                 240

His Thr Ala Lys Trp Thr Pro Pro Asn Asn Glu Gly Arg Ile Phe Thr
                245                 250                 255
Ala Glu Gly Trp Lys Asp Phe Gln Leu Pro Leu Leu Leu Gly Ser Thr
            260                 265                 270
Ser Arg Ser Leu Arg Ala Ile Glu Lys Glu Gly Phe Val Gln Leu Ala
        275                 280                 285
Pro Gly Arg Asp Pro Asp Tyr Asn Asn Thr Ile Asp Glu Gln His Ser
```

290 295 300

Gly Arg Pro Phe Leu Pro Leu Tyr Leu Tyr Leu Gln Gly Thr Ile Ser
305 310 315 320
Gln Glu Tyr Cys Val Phe Ala Gly Thr Trp Val Ile Pro Phe Gln Asp
325 330 335
Gly Ile Ser Pro Tyr Ser Thr Lys Asp Thr Phe Gln Pro Asp Leu Lys
340 345 350
Arg Lys Ala Tyr Ser Leu Leu Leu Asp Ala Val Lys His Arg Leu Gly
355 360 365

Asn Lys Val Ala Ser Gly Leu Gln Tyr Gly Arg Phe Pro Ala Ile Glu
370 375 380
Glu Leu Lys Arg Leu Val Arg Met His Gly Ala Thr Arg Lys Ile Pro
385 390 395 400
Arg Gly Glu Lys Asp Leu Leu Lys Lys Gly Asp Pro Asp Thr Pro Glu
405 410 415
Trp Trp Leu Leu Glu Gln Tyr Pro Glu Phe Trp Arg Leu Cys Asp Ala
420 425 430

Ala Ala Lys Arg Val Ser Gln Asn Val Gly Leu Leu Leu Ser Leu Lys
435 440 445
Lys Gln Pro Leu Trp Gln Arg Arg Trp Leu Glu Ser Arg Thr Arg Asn
450 455 460
Glu Pro Leu Asp Asn Leu Pro Leu Ser Met Ala Leu Thr Leu His Leu
465 470 475 480
Thr Asn Glu Glu Ala Leu
485

<210> 215
<211> 400
<212> PRT
<213> Artificial sequence

<220><223> Synthetic sequence
<400> 215

Ala Ala Val Tyr Ser Lys Phe Tyr Ile Glu Asn His Phe Lys Met Gly
1 5 10 15
Ile Pro Glu Thr Leu Ser Arg Ile Arg Gly Pro Ser Ile Ile Gln Gly
20 25 30
Phe Ser Val Asn Glu Asn Tyr Ile Asn Ile Ala Gly Val Gly Asp Arg
35 40 45
Asp Phe Ile Phe Gly Cys Lys Lys Cys Lys Tyr Thr Arg Gly Lys Pro
50 55 60

Ser Ser Lys Lys Ile Asn Lys Cys His Pro Cys Lys Arg Ser Thr Tyr
65 70 75 80
Pro Glu Pro Val Ile Asp Val Arg Gly Ser Ile Ser Glu Phe Lys Tyr
85 90 95
Lys Ile Tyr Asn Lys Leu Lys Gln Glu Pro Asn Gln Ser Ile Lys Gln
100 105 110
Asn Thr Lys Gly Arg Met Asn Pro Ser Asp His Thr Ser Ser Asn Asp
115 120 125

Gly Ile Ile Ile Asn Gly Ile Asp Asn Arg Ile Ala Tyr Asn Val Ile
130 135 140
Phe Ser Ser Tyr Lys His Leu Met Glu Lys Gln Ile Asn Leu Leu Arg
145 150 155 160
Asp Thr Thr Lys Arg Lys Ala Arg Gln Ile Lys Lys Tyr Asn Asn Ser
165 170 175
Gly Lys Lys Lys His Ser Leu Arg Ser Gln Thr Lys Gly Asn Leu Lys
180 185 190

Asn Arg Tyr His Met Leu Gly Met Phe Lys Lys Gly Ser Leu Thr Ile
195 200 205
Thr Asn Glu Gly Asp Phe Ile Thr Ala Val Arg Lys Val Gly Leu Asp
210 215 220
Ile Ser Leu Tyr Lys Asn Glu Ser Leu Asn Lys Gln Glu Val Glu Thr
225 230 235 240
Glu Leu Cys Leu Asn Ile Lys Trp Gly Arg Thr Lys Ser Tyr Thr Val

245 250 255

Ser Gly Tyr Ile Pro Leu Pro Ile Asn Ile Asp Trp Lys Leu Tyr Leu
260 265 270
Phe Glu Lys Glu Thr Gly Leu Thr Leu Arg Leu Phe Gly Asn Lys Tyr
275 280 285
Lys Ile Gln Ser Lys Lys Phe Leu Ile Ala Gln Leu Phe Lys Pro Lys
290 295 300
Arg Pro Pro Cys Ala Asp Pro Val Val Lys Lys Ala Gln Lys Trp Ser
305 310 315 320

Ala Leu Asn Ala His Val Gln Gln Met Ala Gly Leu Phe Ser Asp Ser
325 330 335
His Leu Leu Lys Arg Glu Leu Lys Asn Arg Met His Lys Gln Leu Asp
340 345 350
Phe Lys Ser Leu Trp Val Gly Thr Glu Asp Tyr Ile Lys Trp Phe Glu
355 360 365
Glu Leu Ser Arg Ser Tyr Val Glu Gly Ala Glu Lys Ser Leu Glu Phe
370 375 380

Phe Arg Gln Asp Tyr Phe Cys Phe Asn Tyr Thr Lys Gln Thr Thr Met
385 390 395 400

<210> 216
<211> 666
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 216

Pro Gln Gln Gln Arg Asp Leu Met Leu Met Ala Ala Asn Tyr Asp Gln
1 5 10 15
Asp Tyr Gly Asn Gly Cys Gly Pro Cys Thr Val Val Ala Ser Ala Ala
20 25 30
Tyr Arg Pro Asp Pro Gln Ala Gln His Gly Cys Lys Arg His Leu Arg

35 40 45

Thr Leu Gly Ala Ser Ala Val Thr His Val Gly Leu Gly Asp Arg Thr
50 55 60
Ala Thr Ile Thr Ala Leu His Arg Leu Arg Gly Pro Ala Ala Leu Ala
65 70 75 80
Ala Arg Ala Arg Ala Ala Gln Ala Ala Ser Ala Pro Met Thr Pro Asp
85 90 95
Thr Asp Ala Pro Asp Asp Arg Arg Arg Leu Glu Ala Ile Asp Ala Asp

100 105 110
Asp Val Val Leu Val Gly Ala His Arg Ala Leu Trp Ser Ala Val Arg
115 120 125
Arg Trp Ala Asp Asp Arg Arg Ala Ala Leu Arg Arg Arg Leu His Ser
130 135 140
Glu Arg Glu Trp Leu Leu Lys Asp Gln Ile Arg Trp Ala Glu Leu Tyr
145 150 155 160
Thr Leu Ile Glu Ala Ser Gly Thr Pro Pro Gln Gly Arg Trp Arg Asn

165 170 175
Thr Leu Gly Ala Leu Arg Gly Gln Ser Arg Trp Arg Arg Val Leu Ala
180 185 190
Pro Thr Met Arg Ala Thr Cys Ala Glu Thr His Ala Glu Leu Trp Asp
195 200 205
Ala Leu Ala Glu Leu Val Pro Glu Met Ala Lys Asp Arg Arg Gly Leu
210 215 220
Leu Arg Pro Pro Val Glu Ala Asp Ala Leu Trp Arg Ala Pro Met Ile

225 230 235 240
Val Glu Gly Trp Arg Gly Gly His Ser Val Val Val Asp Ala Val Ala
245 250 255
Pro Pro Leu Asp Leu Pro Gln Pro Cys Ala Trp Thr Ala Val Arg Leu
260 265 270
Ser Gly Asp Pro Arg Gln Arg Trp Gly Leu His Leu Ala Val Pro Pro
275 280 285
Leu Gly Gln Val Gln Pro Pro Asp Pro Leu Lys Ala Thr Leu Ala Val

290 295 300

Ser Met Arg His Arg Gly Gly Val Arg Val Arg Thr Leu Gln Ala Met

305 310 315 320

Ala Val Asp Ala Asp Ala Pro Met Gln Arg His Leu Gln Val Pro Leu

325 330 335

Thr Leu Gln Arg Gly Gly Gly Leu Gln Trp Gly Ile His Ser Arg Gly

340 345 350

Val Arg Arg Arg Glu Ala Arg Ser Met Ala Ser Trp Glu Gly Pro Pro

355 360 365

Ile Trp Thr Gly Leu Gln Leu Val Asn Arg Trp Lys Gly Gln Gly Ser

370 375 380

Ala Leu Leu Ala Pro Asp Arg Pro Pro Asp Thr Pro Pro Tyr Ala Pro

385 390 395 400

Asp Ala Ala Val Ala Pro Ala Gln Pro Asp Thr Lys Arg Ala Arg Arg

405 410 415

Thr Leu Lys Glu Ala Cys Thr Val Cys Arg Cys Ala Pro Gly His Met

420 425 430

Arg Gln Leu Gln Val Thr Leu Thr Gly Asp Gly Thr Trp Arg Arg Phe

435 440 445

Arg Leu Arg Ala Pro Gln Gly Ala Lys Arg Lys Ala Glu Val Leu Lys

450 455 460

Val Ala Thr Gln His Asp Glu Arg Ile Ala Asn Tyr Thr Ala Trp Tyr

465 470 475 480

Leu Lys Arg Pro Glu His Ala Ala Gly Cys Asp Thr Cys Asp Gly Asp

485 490 495

Ser Arg Leu Asp Gly Ala Cys Arg Gly Cys Arg Pro Leu Leu Val Gly

500 505 510

Asp Gln Cys Phe Arg Arg Tyr Leu Asp Lys Ile Glu Ala Asp Arg Asp

515 520 525

Asp Gly Leu Ala Gln Ile Lys Pro Lys Ala Gln Glu Ala Val Ala Ala

530 535 540

Met Ala Ala Lys Arg Asp Ala Arg Ala Gln Lys Val Ala Ala Arg Ala

545 550 555 560

Ala Lys Leu Ser Glu Ala Thr Gly Gln Arg Thr Ala Ala Thr Arg Asp

565 570 575

Ala Ser His Glu Ala Arg Ala Gln Lys Glu Leu Glu Ala Val Ala Thr

580 585 590

Glu Gly Thr Thr Val Arg His Asp Ala Ala Ala Val Ser Ala Phe Gly

595 600 605

Ser Trp Val Ala Arg Lys Gly Asp Glu Tyr Arg His Gln Val Gly Val

610 615 620

Leu Ala Asn Arg Leu Glu His Gly Leu Arg Leu Gln Glu Leu Met Ala

625 630 635 640

Pro Asp Ser Val Val Ala Asp Gln Gln Arg Ala Ser Gly His Ala Arg

645 650 655

Val Gly Tyr Arg Tyr Val Leu Thr Ala Met

660 665

<210> 217

<211> 560

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 217

Ala Val Ala His Pro Val Gly Arg Gly Asn Ala Gly Ser Pro Gly Ala

1 5 10 15

Arg Gly Pro Glu Glu Leu Pro Arg Gln Leu Val Asn Arg Ala Ser Asn

20 25 30

Val Thr Arg Pro Ala Thr Tyr Gly Cys Ala Pro Cys Arg His Val Arg

35 40 45

Leu Ser Ile Pro Lys Pro Val Leu Thr Gly Cys Arg Ala Cys Glu Gln

50 55 60

Thr Thr His Pro Ala Pro Lys Arg Ala Val Arg Gly Gly Ala Asp Ala

65 70 75 80

Ala Lys Tyr Asp Leu Ala Ala Phe Phe Ala Gly Trp Ala Ala Asp Leu

85 90 95

Glu Gly Arg Asn Arg Arg Arg Gln Val His Ala Pro Leu Asp Pro Gln

100 105 110

Pro Asp Pro Asn His Glu Pro Ala Val Thr Leu Gln Lys Ile Asp Leu

115 120 125

Ala Glu Val Ser Ile Glu Glu Phe Gln Arg Val Leu Ala Arg Ser Val

130 135 140

Lys His Arg His Asp Gly Arg Ala Ser Arg Glu Arg Glu Lys Ala Arg

145 150 155 160

Ala Tyr Ala Gln Val Ala Lys Lys Arg Arg Asn Ser His Ala His Gly

165 170 175

Ala Arg Thr Arg Arg Ala Val Arg Arg Gln Thr Arg Ala Val Arg Arg

180 185 190

Ala His Arg Met Gly Ala Asn Ser Gly Glu Ile Leu Val Ala Ser Gly

195 200 205

Ala Glu Asp Pro Val Pro Glu Ala Ile Asp His Ala Ala Gln Leu Arg

210 215 220

Arg Arg Ile Arg Ala Cys Ala Arg Asp Leu Glu Gly Leu Arg His Leu

225 230 235 240

Ser Arg Arg Tyr Leu Lys Thr Leu Glu Lys Pro Cys Arg Arg Pro Arg

245 250 255

Ala Pro Asp Leu Gly Arg Ala Arg Cys His Ala Leu Val Glu Ser Leu

260 265 270

Gln Ala Ala Glu Arg Glu Leu Glu Glu Leu Arg Arg Cys Asp Ser Pro

275 280 285

Asp Thr Ala Met Arg Arg Leu Asp Ala Val Leu Ala Ala Ala Ala Ser

290 295 300

Thr Asp Ala Thr Phe Ala Thr Gly Trp Thr Val Val Gly Met Asp Leu

305 310 315 320

Gly Val Ala Pro Arg Gly Ser Ala Ala Pro Glu Val Ser Pro Met Glu
325 330 335
Met Ala Ile Ser Val Phe Trp Arg Lys Gly Ser Arg Arg Val Ile Val
340 345 350
Ser Lys Pro Ile Ala Gly Met Pro Ile Arg Arg His Glu Leu Ile Arg
355 360 365
Leu Glu Gly Leu Gly Thr Leu Arg Leu Asp Gly Asn His Tyr Thr Gly
370 375 380

Ala Gly Val Thr Lys Gly Arg Gly Leu Ser Glu Gly Thr Glu Pro Asp
385 390 395 400
Phe Arg Glu Lys Ser Pro Ser Thr Leu Gly Phe Thr Leu Ser Asp Tyr
405 410 415
Arg His Glu Ser Arg Trp Arg Pro Tyr Gly Ala Lys Gln Gly Lys Thr
420 425 430
Ala Arg Gln Phe Phe Ala Ala Met Ser Arg Glu Leu Arg Ala Leu Val
435 440 445

Glu His Gln Val Leu Ala Pro Met Gly Pro Pro Leu Leu Glu Ala His
450 455 460
Glu Arg Arg Phe Glu Thr Leu Leu Lys Gly Gln Asp Asn Lys Ser Ile
465 470 475 480
His Ala Gly Gly Gly Gly Arg Tyr Val Trp Arg Gly Pro Pro Asp Ser
485 490 495
Lys Lys Arg Pro Ala Ala Asp Gly Asp Trp Phe Arg Phe Gly Arg Gly
500 505 510

His Ala Asp His Arg Gly Trp Ala Asn Lys Arg His Glu Leu Ala Ala
515 520 525
Asn Tyr Leu Gln Ser Ala Phe Arg Leu Trp Ser Thr Leu Ala Glu Ala
530 535 540
Gln Glu Pro Thr Pro Tyr Ala Arg Tyr Lys Tyr Thr Arg Val Thr Met
545 550 555 560
<210> 218
<211> 404

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 218

Trp Asp Phe Leu Thr Leu Gln Val Tyr Glu Arg His Thr Ser Pro Glu

1 5 10 15

Val Cys Val Ala Gly Asn Ser Thr Lys Cys Ala Ser Gly Thr Arg Lys

20 25 30

Ser Asp His Thr His Gly Val Gly Val Lys Leu Gly Ala Gln Glu Ile

35 40 45

Asn Val Ser Ala Asn Asp Asp Arg Asp His Glu Val Gly Cys Asn Ile

50 55 60

Cys Val Ile Ser Arg Val Ser Leu Asp Ile Lys Gly Trp Arg Tyr Gly

65 70 75 80

Cys Glu Ser Cys Val Gln Ser Thr Pro Glu Trp Arg Ser Ile Val Arg

85 90 95

Phe Asp Arg Asn His Lys Glu Ala Lys Gly Glu Cys Leu Ser Arg Phe

100 105 110

Glu Tyr Trp Gly Ala Gln Ser Ile Ala Arg Ser Leu Lys Arg Asn Lys

115 120 125

Leu Met Gly Gly Val Asn Leu Asp Glu Leu Ala Ile Val Gln Asn Glu

130 135 140

Asn Val Val Lys Thr Ser Leu Lys His Leu Phe Asp Lys Arg Lys Asp

145 150 155 160

Arg Ile Gln Ala Asn Leu Lys Ala Val Lys Val Arg Met Arg Glu Arg

165 170 175

Arg Lys Ser Gly Arg Gln Arg Lys Ala Leu Arg Arg Gln Cys Arg Lys

180 185 190

Leu Lys Arg Tyr Leu Arg Ser Tyr Asp Pro Ser Asp Ile Lys Glu Gly

195 200 205

Asn Ser Cys Ser Ala Phe Thr Lys Leu Gly Leu Asp Ile Gly Ile Ser

210 215 220

Pro Asn Lys Pro Pro Lys Ile Glu Pro Lys Val Glu Val Val Phe Ser

225 230 235 240

Leu Phe Tyr Gln Gly Ala Cys Asp Lys Ile Val Thr Val Ser Ser Pro

245 250 255

Glu Ser Pro Leu Pro Arg Ser Trp Lys Ile Lys Ile Asp Gly Ile Arg

260 265 270

Ala Leu Tyr Val Lys Ser Thr Lys Val Lys Phe Gly Gly Arg Thr Phe

275 280 285

Arg Ala Gly Gln Arg Asn Asn Arg Arg Lys Val Arg Pro Pro Asn Val

290 295 300

Lys Lys Gly Lys Arg Lys Gly Ser Arg Ser Gln Phe Phe Asn Lys Phe

305 310 315 320

Ala Val Gly Leu Asp Ala Val Ser Gln Gln Leu Pro Ile Ala Ser Val

325 330 335

Gln Gly Leu Trp Gly Arg Ala Glu Thr Lys Lys Ala Gln Thr Ile Cys

340 345 350

Leu Lys Gln Leu Glu Ser Asn Lys Pro Leu Lys Glu Ser Gln Arg Cys

355 360 365

Leu Phe Leu Ala Asp Asn Trp Val Val Arg Val Cys Gly Phe Leu Arg

370 375 380

Ala Leu Ser Gln Arg Gln Gly Pro Thr Pro Tyr Ile Arg Tyr Arg Tyr

385 390 395 400

Arg Cys Asn Met

<210> 219

<211> 392

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 219

Ala Arg Asn Val Gly Gln Arg Asn Ala Ser Arg Gln Ser Lys Arg Glu

1 5 10 15

Ser Ala Lys Ala Arg Ser Arg Arg Val Thr Gly Gly His Ala Ser Val

20 25 30

Thr Gln Gly Val Ala Leu Ile Asn Ala Ala Ala Asn Ala Asp Arg Asp

35 40 45

His Thr Thr Gly Cys Glu Pro Cys Thr Trp Glu Arg Val Asn Leu Pro

50 55 60

Leu Gln Glu Val Ile His Gly Cys Asp Ser Cys Thr Lys Ser Ser Pro

65 70 75 80

Phe Trp Arg Asp Ile Lys Val Val Asn Lys Gly Tyr Arg Glu Ala Lys

85 90 95

Glu Glu Ile Met Arg Ile Ala Ser Gly Ile Ser Ala Asp His Leu Ser

100 105 110

Arg Ala Leu Ser His Asn Lys Val Met Gly Arg Leu Asn Leu Asp Glu

115 120 125

Val Cys Ile Leu Asp Phe Arg Thr Val Leu Asp Thr Ser Leu Lys His

130 135 140

Leu Thr Asp Ser Arg Ser Asn Gly Ile Lys Glu His Ile Arg Ala Val

145 150 155 160

His Arg Lys Ile Arg Met Arg Arg Lys Ser Gly Lys Thr Ala Arg Ala

165 170 175

Leu Arg Lys Gln Tyr Phe Ala Leu Arg Arg Gln Trp Lys Ala Gly His

180 185 190

Lys Pro Asn Ser Ile Arg Glu Gly Asn Ser Leu Thr Ala Leu Arg Ala

195 200 205

Val Gly Phe Asp Val Gly Val Ser Glu Gly Thr Glu Pro Met Pro Ala

210 215 220

Pro Gln Thr Glu Val Val Leu Ser Val Phe Tyr Lys Gly Ser Ala Thr

225 230 235 240

Arg Ile Leu Arg Ile Ser Ser Pro His Pro Ile Ala Lys Arg Ser Trp

245 250 255

Lys Val Lys Ile Ala Gly Ile Lys Ala Leu Lys Leu Ile Arg Arg Glu
260 265 270
His Asp Phe Ser Phe Gly Arg Glu Thr Tyr Asn Ala Ser Gln Arg Ala
275 280 285
Glu Lys Arg Lys Phe Ser Pro His Ala Ala Arg Lys Asp Phe Phe Asn

290 295 300
Ser Phe Ala Val Gln Leu Asp Arg Leu Ala Gln Gln Leu Cys Val Ser
305 310 315 320
Ser Val Glu Asn Leu Trp Val Thr Glu Pro Gln Gln Lys Leu Leu Thr
325 330 335
Leu Ala Lys Asp Thr Ala Pro Tyr Gly Ile Arg Glu Gly Ala Arg Phe
340 345 350
Ala Asp Thr Arg Ala Arg Leu Ala Trp Asn Trp Val Phe Arg Val Cys

355 360 365
Gly Phe Thr Arg Ala Leu His Gln Glu Gln Glu Pro Thr Pro Tyr Cys
370 375 380
Arg Phe Thr Trp Arg Ser Lys Met
385 390

<210> 220
<211> 59
<212> DNA
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 220
ctacgccgat tatcttctga caactttcgc aagcggtgta aggtaaaaaa tgcgggcac 59

<210> 221
<211> 59
<212> DNA
<213> Artificial sequence

<220><223> Synthetic sequence
<400> 221
gtgcccgcat tttttacctt acaccgcttg cgaaagttgt cagaagataa tcggcgtag 59

<210> 222

<211> 90

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 222

tttatatgtt tctcctggag ataacgcaat cgtgacaact ttcgcaagcg gtgtaaggta 60

gcaggcttcc gaattccgcg tttttacggc 90

<210> 223

<211> 90

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 223

gccgtaaaaa cgcggaattc ggaagcctgc taccttacac cgcttgcgaa agttgtcacg 60

attgcgttat ctccaggaga aacatataaa 90

<210> 224

<211> 56

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 224

gatcttcagc tatacattat tgcaccaaca ctaaggcaga gtatgtttac ctggac 56

<210> 225

<211> 56

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 225

gatcttcagc tttgtattac tggaaggatg cttgcttgag gtgtaaaaac ctggac 56

<210> 226

<211> 5

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 226

ttttt 5

<210> 227

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 227

tttttt 6

<210> 228

<211> 7

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 228

ttttttt 7

<210> 229

<211> 8

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 229

tttttttt 8

<210> 230

<211> 9

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 230

ttttttttt 9

<210> 231

<211> 10

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 231

tttttttttt 10

<210> 232

<211> 11

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 232

tttttttttt t 11

<210> 233

<211> 12

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 233

tttttttttt tt 12

<210> 234

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 234

gccggggtgg tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagc 55

<210> 235

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 235

gccggggtgg tgcccatcct ggtcgagctg gacggcgacg tgctcggcca caagc 55

<210> 236

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 236

gccggggtgg tgcccatcct ggtcgagctg gacggcgagc taaacggcca caagc 55

<210> 237

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 237

gccggggtgg tgcccatcct ggtcgagctg gacggcctcg taaacggcca caagc 55

<210> 238

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 238

gccggggtgg tgcccatcct ggtcgagctg gacgcggacg taaacggcca caagc 55

<210> 239

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 239

gccggggtgg tgcccatcct ggtcgagctg gagcgcgacg taaacggcca caagc 55

<210> 240

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 240

gccggggtgg tgcccatcct ggtcgagctg ctcggcgacg taaacggcca caagc 55

<210> 241

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 241

gccggggtgg tgcccatcct ggtcgagcac gacggcgacg taaacggcca caagc 55

<210> 242

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 242

gccggggtgg tgcccatcct ggtcgacgtg gacggcgacg taaacggcca caagc 55

<210> 243

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 243

gccggggtgg tgcccatcct ggtcctgctg gacggcgacg taaacggcca caagc 55

<210> 244

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 244

gccggggtgg tgcccatcct ggaggagctg gacggcgacg taaacggcca caagc 55

<210> 245

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 245

gccggggtgg tgcccatcct cctcgagctg gacggcgacg taaacggcca caagc 55

<210> 246

<211> 34

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 246

gtgttaatac aaaggtacag gaacaaagaa tttg 34

<210> 247

<211> 18

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 247

caaagagaag cctcggcc 18

<210> 248

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 248

tttattcaag gcaatcacta tcagctgtgg aacacccagg taaactaaca caact 55

<210> 249

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 249

tttattcaag gcaatcacta tcagctgtgg aacacccagg tgctctaaca caact 55

<210> 250

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 250

tttattcaag gcaatcacta tcagctgtgg aacacccacc taaactaaca caact 55

<210> 251

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 251

tttattcaag gcaatcacta tcagctgtgg aacaccgtgg taaactaaca caact 55

<210> 252

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 252

tttattcaag gcaatcacta tcagctgtgg aacaggcagg taaactaaca caact 55

<210> 253

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 253

tttattcaag gcaatcacta tcagctgtgg aagtcccagg taaactaaca caact 55

<210> 254

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 254

tttattcaag gcaatcacta tcagctgtgg ttcacccagg taaactaaca caact 55

<210> 255

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 255

tttattcaag gcaatcacta tcagctgtcc aacacccagg taaactaaca caact 55

<210> 256

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 256

tttattcaag gcaatcacta tcagctcagg aacacccagg taaactaaca caact 55

<210> 257

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 257

tttattcaag gcaatcacta tcaggagtgg aacacccagg taaactaaca caact 55

<210> 258

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 258

tttattcaag gcaatcacta tctcctgtgg aacacccagg taaactaaca caact 55

<210> 259

<211> 55

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 259

tttattcaag gcaatcacta agagctgtgg aacacccagg taaactaaca caact 55

<210> 260

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 260

ctacgccgat tatcttctga caactttcgc aagcggtgta aggcgaaaaa tgcgggcac 59

<210> 261

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 261

ctacgccgat tatcttctga caactttcgc aagcggtgta aaataaaaaa tgcgggcac 59

<210> 262

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 262

ctacgccgat tatcttctga caactttcgc aagcggtgtg gggtaaaaaa tgcgggcac 59

<210> 263

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 263

ctacgccgat tatcttctga caactttcgc aagcggtaca aggtaaaaaa tgcgggcac 59

<210> 264

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 264

ctacgccgat tatcttctga caactttcgc aagcgacgta aggtaaaaaa tgcgggcac 59

<210> 265

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 265

ctacgccgat tatcttctga caactttcgc aagtagtgta aggtaaaaaa tgcgggcac 59

<210> 266

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 266

ctacgccgat tatcttctga caactttcgc agacggtgta aggtaaaaaa tgcgggcac 59

<210> 267

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 267

ctacgccgat tatcttctga caactttcgt gagcggtgta aggtaaaaaa tgcgggcac 59

<210> 268

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 268

ctacgccgat tatcttctga caacttttac aagcggtgta aggtaaaaaa tgcgggcac 59

<210> 269

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 269

ctacgccgat tatcttctga caactcccgc aagcggtgta aggtaaaaaa tgcgggcac 59

<210> 270

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 270

ctacgccgat tatcttctga caatcttcgc aagcggtgta aggtaaaaaa tgcgggcac 59

<210> 271

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 271

ctacgccgat tatcttctga cggctttcgc aagcggtgta aggtaaaaaa tgcgggcac 59

<210> 272

<211> 59

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 272

ctacgccgat tatcttctgg taactttcgc aagcggtgta aggtaaaaaa tgcgggcac 59

<210> 273

<211> 49

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 273

tatcttctga caactttcgc aagcggtgta aggtaaaaaa tgcgggcac 49

<210> 274

<211> 44

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 274

tctgacaact ttcgcaagcg gtgtaaggta aaaaatgcgg gcac 44

<210> 275

<211> 39

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 275

caactttcgc aagcggtgta aggtaaaaaa tgcgggcac 39

<210> 276

<211> 34

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 276

ttcgcaagcg gtgtaaggta aaaaatgcgg gcac 34

<210> 277

<211> 54

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 277

ctacgccgat tatcttctga caactttcgc aagcggtgta aggtaaaaaa tgcg 54

<210> 278

<211> 49

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 278

ctacgccgat tatcttctga caactttcgc aagcggtgta aggtaaaaa 49

<210> 279

<211> 40

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 279

ctacgccgat tatcttctga caactttcgc aagcggtgta 40

<210> 280

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 280

ctacgccgat tatcttctga caactttcgc aagcg 35

<210> 281

<211> 30

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 281

ctacgccgat tatcttctga caactttcgc 30

<210> 282

<211> 25

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 282

ctacgccgat tatcttctga caact 25

<210> 283

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 283

ctacgccgat tatcttctga 20

<210> 284

<211> 34

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 284

caactttcgc aagcggtgta aggtaaaaaa tgcg 34

<210> 285

<211> 63

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 285

atggaatgtg gcgaacgctt tcaacgaaac aactttcgca agcggtgtaa ggtaaaaaat 60

gcg 63

<210> 286

<211> 63

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 286

atggaatgtg gcgaacgctt agttggaaac aactttcgca agcggtgtaa ggtaaaaaat 60

gcg 63

<210> 287

<211> 63

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 287

atggaatgtg gcgaagcgaa agttggaaac aactttcgca agcggtgtaa ggtaaaaaat 60

gcg 63

<210> 288

<211> 63

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 288

atggaatgtg cgcttgcgaa agttggaaac aactttcgca agcggtgtaa ggtaaaaaat 60

gcg 63

<210> 289

<211> 63

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 289

atggatacac cgcttgcgaa agttggaaac aactttcgca agcggtgtaa ggtaaaaaat 60

gcg 63

<210> 290

<211> 63

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 290

taccttacac cgcttgcgaa agttggaaac aactttcgca agcggtgtaa ggtaaaaaat 60

gcg 63

<210> 291

<211> 40

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 291

gttttatctt ctgctggtgg ttcgttcggt atttttaatg 40

<210> 292

<211> 40

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 292

cattaaaaat accgaacgaa ccaccagcag aagataaaac 40

<210> 293

<211> 43

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 293

gaccatttgc gaaatgtatc taatggtcaa actaaatcta ctc 43

<210> 294

<211> 43

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 294

gagtagattt agtttgacca ttagatacat ttcgcaaatg gtc 43

<210> 295

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 295

cttgcagaac ccggatagac gaatgaagga atgcaac 37

<210> 296

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 296

cttgcaggcc ttgaatagag gagttaagga atgcaac 37

<210> 297

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 297

gttgcacagt gctaattaga gaaactagga atgcaac 37

<210> 298

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 298

ctagcatatt cagaacaaag ggattaagga atgcaac 37

<210> 299

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 299

ctttcatatt cagaaactag gggttaagga ctgcaac 37

<210> 300

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 300

gttgcatccc tacgtcgtga gcaccggtga gtgcaac 37

<210> 301

<211> 31

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 301

ggaaaggaat cccctgaagg aaacgagggg g 31

<210> 302

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 302

gtgtccatca atcagatttg cgttggccgg tgcaat 36

<210> 303

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 303

gtttcagcgc acgaattaac gagatgagag atgcaac 37

<210> 304

<211> 500

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 304

Lys Glu Pro Leu Asn Ile Gly Lys Thr Ala Lys Ala Val Phe Lys Glu
1 5 10 15

Ile Asp Pro Thr Ser Leu Asn Arg Ala Ala Asn Tyr Asp Ala Ser Ile
20 25 30

Glu Leu Asn Cys Lys Glu Cys Lys Phe Lys Pro Phe Lys Asn Val Lys
35 40 45

Arg Tyr Glu Phe Asn Phe Tyr Asn Asn Trp Tyr Arg Cys Asn Pro Asn
50 55 60

Ser Cys Leu Gln Ser Thr Tyr Lys Ala Gln Val Arg Lys Val Glu Ile
65 70 75 80

Gly Tyr Glu Lys Leu Lys Asn Glu Ile Leu Thr Gln Met Gln Tyr Tyr

85 90 95

Pro Trp Phe Gly Arg Leu Tyr Gln Asn Phe Phe His Asp Glu Arg Asp

100 105 110

Lys Met Thr Ser Leu Asp Glu Ile Gln Val Ile Gly Val Gln Asn Lys

115 120 125

Val Phe Phe Asn Thr Val Glu Lys Ala Trp Arg Glu Ile Ile Lys Lys

130 135 140

Arg Phe Lys Asp Asn Lys Glu Thr Met Glu Thr Ile Pro Glu Leu Lys

145 150 155 160

His Ala Ala Gly His Gly Lys Arg Lys Leu Ser Asn Lys Ser Leu Leu

165 170 175

Arg Arg Arg Phe Ala Phe Val Gln Lys Ser Phe Lys Phe Val Asp Asn

180 185 190

Ser Asp Val Ser Tyr Arg Ser Phe Ser Asn Asn Ile Ala Cys Val Leu

195 200 205

Pro Ser Arg Ile Gly Val Asp Leu Gly Gly Val Ile Ser Arg Asn Pro

210 215 220

Lys Arg Glu Tyr Ile Pro Gln Glu Ile Ser Phe Asn Ala Phe Trp Lys

225 230 235 240

Gln His Glu Gly Leu Lys Lys Gly Arg Asn Ile Glu Ile Gln Ser Val

245 250 255

Gln Tyr Lys Gly Glu Thr Val Lys Arg Ile Glu Ala Asp Thr Gly Glu

260 265 270

Asp Lys Ala Trp Gly Lys Asn Arg Gln Arg Arg Phe Thr Ser Leu Ile

275 280 285

Leu Lys Leu Val Pro Lys Gln Gly Gly Lys Lys Val Trp Lys Tyr Pro

290 295 300

Glu Lys Arg Asn Glu Gly Asn Tyr Glu Tyr Phe Pro Ile Pro Ile Glu

305 310 315 320

Phe Ile Leu Asp Ser Gly Glu Thr Ser Ile Arg Phe Gly Gly Asp Glu

325 330 335

Gly Glu Ala Gly Lys Gln Lys His Leu Val Ile Pro Phe Asn Asp Ser
340 345 350
Lys Ala Thr Pro Leu Ala Ser Gln Gln Thr Leu Leu Glu Asn Ser Arg
355 360 365

Phe Asn Ala Glu Val Lys Ser Cys Ile Gly Leu Ala Ile Tyr Ala Asn
370 375 380
Tyr Phe Tyr Gly Tyr Ala Arg Asn Tyr Val Ile Ser Ser Ile Tyr His
385 390 395 400
Lys Asn Ser Lys Asn Gly Gln Ala Ile Thr Ala Ile Tyr Leu Glu Ser
405 410 415
Ile Ala His Asn Tyr Val Lys Ala Ile Glu Arg Gln Leu Gln Asn Leu
420 425 430

Leu Leu Asn Leu Arg Asp Phe Ser Phe Met Glu Ser His Lys Lys Glu
435 440 445
Leu Lys Lys Tyr Phe Gly Gly Asp Leu Glu Gly Thr Gly Gly Ala Gln
450 455 460
Lys Arg Arg Glu Lys Glu Glu Lys Ile Glu Lys Glu Ile Glu Gln Ser
465 470 475 480
Tyr Leu Pro Arg Leu Ile Arg Leu Ser Leu Thr Lys Met Val Thr Lys
485 490 495

Gln Val Glu Met
500

<210> 305
<211> 507
<212> PRT
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 305
Glu Leu Ile Val Asn Glu Asn Lys Asp Pro Leu Asn Ile Gly Lys Thr
1 5 10 15
Ala Lys Ala Val Phe Lys Glu Ile Asp Pro Thr Ser Ile Asn Arg Ala
20 25 30

Ala Asn Tyr Asp Ala Ser Ile Glu Leu Ala Cys Lys Glu Cys Lys Phe
35 40 45

Lys Pro Phe Asn Asn Thr Lys Arg His Asp Phe Ser Phe Tyr Ser Asn
50 55 60
Trp His Arg Cys Ser Pro Asn Ser Cys Leu Gln Ser Thr Tyr Arg Ala
65 70 75 80
Lys Ile Arg Lys Thr Glu Ile Gly Tyr Glu Lys Leu Lys Asn Glu Ile
85 90 95
Leu Asn Gln Met Gln Tyr Tyr Pro Trp Phe Gly Arg Leu Tyr Gln Asn
100 105 110

Phe Phe Asn Asp Gln Arg Asp Lys Met Thr Ser Leu Asp Glu Ile Gln
115 120 125
Val Thr Gly Val Gln Asn Lys Ile Phe Phe Asn Thr Val Glu Lys Ala
130 135 140
Trp Arg Glu Ile Ile Lys Lys Arg Phe Arg Asp Asn Lys Glu Thr Met
145 150 155 160
Arg Thr Ile Pro Asp Leu Lys Asn Lys Ser Gly His Gly Ser Arg Lys
165 170 175

Leu Ser Asn Lys Ser Leu Leu Arg Arg Arg Phe Ala Phe Ala Gln Lys
180 185 190
Ser Phe Lys Leu Val Asp Asn Ser Asp Val Ser Tyr Arg Ala Phe Ser
195 200 205
Asn Asn Val Ala Cys Val Leu Pro Ser Lys Ile Gly Val Asp Ile Gly
210 215 220
Gly Ile Ile Asn Lys Asp Leu Lys Arg Glu Tyr Ile Pro Gln Glu Ile
225 230 235 240

Thr Phe Asn Val Phe Trp Lys Gln His Asp Gly Leu Lys Lys Gly Arg
245 250 255
Asn Ile Glu Ile His Ser Val Gln Tyr Lys Gly Glu Ile Val Lys Arg
260 265 270
Ile Glu Ala Asp Thr Gly Glu Asp Lys Ala Trp Gly Lys Asn Arg Gln

275 280 285

Arg Arg Phe Thr Ser Leu Ile Leu Lys Ile Thr Pro Lys Gln Gly Gly

290 295 300

Lys Lys Ile Trp Lys Phe Pro Glu Lys Lys Asn Ala Ser Asp Tyr Glu

305 310 315 320

Tyr Phe Pro Ile Pro Ile Glu Phe Ile Leu Asp Asn Gly Asp Ala Ser

325 330 335

Ile Lys Phe Gly Gly Glu Glu Gly Glu Val Gly Lys Gln Lys His Leu

340 345 350

Leu Ile Pro Phe Asn Asp Ser Lys Ala Thr Pro Leu Ser Ser Lys Gln

355 360 365

Met Leu Leu Glu Thr Ser Arg Phe Asn Ala Glu Val Lys Ser Thr Ile

370 375 380

Gly Leu Ala Leu Tyr Ala Asn Tyr Phe Val Ser Tyr Ala Arg Asn Tyr

385 390 395 400

Val Ile Lys Ser Thr Tyr His Lys Asn Ser Lys Lys Gly Gln Ile Val

405 410 415

Thr Glu Ile Tyr Leu Glu Ser Ile Ser Gln Asn Phe Val Arg Ala Ile

420 425 430

Gln Arg Gln Leu Gln Ser Leu Met Leu Asn Leu Lys Asp Trp Gly Phe

435 440 445

Met Gln Thr His Lys Lys Glu Leu Lys Lys Tyr Phe Gly Ser Asp Leu

450 455 460

Glu Gly Ser Lys Gly Gly Gln Lys Arg Arg Glu Lys Glu Glu Lys Ile

465 470 475 480

Glu Lys Glu Ile Glu Ala Ser Tyr Leu Pro Arg Leu Ile Arg Leu Ser

485 490 495

Leu Thr Lys Ser Val Thr Lys Ala Glu Glu Met

500 505

<210> 306

<211> 529

<212> PRT

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 306

Pro Glu Glu Lys Thr Ser Lys Leu Lys Pro Asn Ser Ile Asn Leu Ala
1 5 10 15

Ala Asn Tyr Asp Ala Asn Glu Lys Phe Asn Cys Lys Glu Cys Lys Phe
20 25 30

His Pro Phe Lys Asn Lys Lys Arg Tyr Glu Phe Asn Phe Tyr Asn Asn
35 40 45

Leu His Gly Cys Lys Ser Cys Thr Lys Ser Thr Asn Asn Pro Ala Val
50 55 60

Lys Arg Ile Glu Ile Gly Tyr Gln Lys Leu Lys Phe Glu Ile Lys Asn
65 70 75 80

Gln Met Glu Ala Tyr Pro Trp Phe Gly Arg Leu Arg Ile Asn Phe Tyr
85 90 95

Ser Asp Glu Lys Arg Lys Met Ser Glu Leu Asn Glu Met Gln Val Thr
100 105 110

Gly Val Lys Asn Lys Ile Phe Phe Asp Ala Ile Glu Cys Ala Trp Arg
115 120 125

Glu Ile Leu Lys Lys Arg Phe Arg Glu Ser Lys Glu Thr Leu Ile Thr
130 135 140

Ile Pro Lys Leu Lys Asn Lys Ala Gly His Gly Ala Arg Lys His Arg
145 150 155 160

Asn Lys Lys Leu Leu Ile Arg Arg Arg Ala Phe Met Lys Lys Asn Phe
165 170 175

His Phe Leu Asp Asn Asp Ser Ile Ser Tyr Arg Ser Phe Ala Asn Asn
180 185 190

Ile Ala Cys Val Leu Pro Ser Lys Val Gly Val Asp Ile Gly Gly Ile
195 200 205

Ile Ser Pro Asp Val Gly Lys Asp Ile Lys Pro Val Asp Ile Ser Leu
210 215 220

Asn Leu Met Trp Ala Ser Lys Glu Gly Ile Lys Ser Gly Arg Lys Val

225 230 235 240

Glu Ile Tyr Ser Thr Gln Tyr Asp Gly Asn Met Val Lys Lys Ile Glu

245 250 255

Ala Glu Thr Gly Glu Asp Lys Ser Trp Gly Lys Asn Arg Lys Arg Arg

260 265 270

Gln Thr Ser Leu Leu Leu Ser Ile Pro Lys Pro Ser Lys Gln Val Gln

275 280 285

Glu Phe Asp Phe Lys Glu Trp Pro Arg Tyr Lys Asp Ile Glu Lys Lys

290 295 300

Val Gln Trp Arg Gly Phe Pro Ile Lys Ile Ile Phe Asp Ser Asn His

305 310 315 320

Asn Ser Ile Glu Phe Gly Thr Tyr Gln Gly Gly Lys Gln Lys Val Leu

325 330 335

Pro Ile Pro Phe Asn Asp Ser Lys Thr Thr Pro Leu Gly Ser Lys Met

340 345 350

Asn Lys Leu Glu Lys Leu Arg Phe Asn Ser Lys Ile Lys Ser Arg Leu

355 360 365

Gly Ser Ala Ile Ala Ala Asn Lys Phe Leu Glu Ala Ala Arg Thr Tyr

370 375 380

Cys Val Asp Ser Leu Tyr His Glu Val Ser Ser Ala Asn Ala Ile Gly

385 390 395 400

Lys Gly Lys Ile Phe Ile Glu Tyr Tyr Leu Glu Ile Leu Ser Gln Asn

405 410 415

Tyr Ile Glu Ala Ala Gln Lys Gln Leu Gln Arg Phe Ile Glu Ser Ile

420 425 430

Glu Gln Trp Phe Val Ala Asp Pro Phe Gln Gly Arg Leu Lys Gln Tyr

435 440 445

Phe Lys Asp Asp Leu Lys Arg Ala Lys Cys Phe Leu Cys Ala Asn Arg

450 455 460

Glu Val Gln Thr Thr Cys Tyr Ala Ala Val Lys Leu His Lys Ser Cys

465 470 475 480

Ala Glu Lys Val Lys Asp Lys Asn Lys Glu Leu Ala Ile Lys Glu Arg

485 490 495

Asn Asn Lys Glu Asp Ala Val Ile Lys Glu Val Glu Ala Ser Asn Tyr

500 505 510

Pro Arg Val Ile Arg Leu Lys Leu Thr Lys Thr Ile Thr Asn Lys Ala

515 520 525

Met

<210> 307

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 307

cttttagaca gtttaaattc taaagggtat aaaac 35

<210> 308

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 308

gtcgaaatgc ccgcgcgggg gcgtcgtacc cgcgac 36

<210> 309

<211> 37

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 309

gttgcagcgg ccgacggagc gcgagcgtgg atgccac 37

<210> 310

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 310

ctttagactt ctccggaagt cgaattaatg gaaac 35

<210> 311

<211> 28

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 311

gggcgccccg cgcgagcggg ggttgaag 28

<210> 312

<211> 136

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 312

gaaggatgct tgcttgaggt gtagttgtca ttccttcatt cgtctattcg ggttctgcaa 60

ctatacatta ttgcaccaac actaaggcag agtatggttg cattccttca ttcgtctatt 120

cgggttctgc aacggg 136

<210> 313

<211> 102

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 313

agaatgcttg cttgaggtgt agttgcattc cttcattcgt ctattcgggt aattgcacca 60

acactaaggc agagtatggt tgcattcctt cattcgtcta gg 102

<210> 314

<211> 75

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 314

cgatgcttgc ttgaggtgta gttgcattgc accaacacta aggcagagta tggttgcatt 60

ccttcattcg tctag 75

<210> 315

<211> 90

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 315

gatgcttgct tgaggtgtag ttgcattcct tcattctgca ccaacactaa ggcagagtat 60

ggttgcattc cttttcgggt tctgcaacgg 90

<210> 316

<211> 73

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 316

gcttgcttga ggtgtagttg cattccttca ttccacctac actaaggcag agtatggttg 60

cattccttca ttc 73

<210> 317

<211> 73

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 317

gcttgcttga ggtgtagttg cattccttca ttccacctac actaaggcag agtatggttg 60

cattccttca ttc 73

<210> 318

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 318

gcttgcttga ggtgtagttg cattcgcaac actaaggcag agtatggttg cattccttca 60

ttc 63

<210> 319

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 319

cttgcttgag gtgtagttgc cgacactaag gcagagtatg gttgcattat tcgggttctg 60

caacgg 66

<210> 320

<211> 69

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 320

cttgcttgag gtgtagttgc attccttcat tcaaacacta aggcagagta tggttgcatt 60

ccttcattc 69

<210> 321

<211> 73

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 321

ttgcttgagg tgtagttgca ttccttcatt ctaacactaa ggcagagtat ggttgcattc 60

cttcattcgt cta 73

<210> 322

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 322

ttgcttgagg tgtagttgca tccacactaa ggcagagtat ggttgcattc cttcattc 58

<210> 323

<211> 67

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 323

tgcttaaggt gtagttgcat tccttcattc caacactaag gcagagtatg gttgcattcc 60

ttcattc 67

<210> 324

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 324

tgcttgaggt gtagttgcat tccttcattc cacactaagg cagagtatgg ttgcattcct 60

tcattc 66

<210> 325

<211> 72

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 325

tgcttgaggt gtagttgcat tccttcattc aacactaagg cagagtatgg ttgcattcct 60

tcattcgtct at 72

<210> 326

<211> 64

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 326

cttgaggtgt agttgcattc cttcattcaa cactaaggca gagtatggtt gcattccttc 60

attc 64

<210> 327

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 327

cttgaggtgt agttgcattc cttaacacta aggcagagta tggttgcatt ccttcattc 59

<210> 328

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 328

ttgaggtgta gttgcattcc ttcattcaac actaaggcag agtatggttg cattccttca 60

ttc 63

<210> 329

<211> 62

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 329

tgaggtgtag ttgcattcct tcattcaaca ctaaggcaga gtatggttgc attccttcat 60

tc 62

<210> 330

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 330

tgaggtgtag ttgcattcct tcgttcacac taaggcagag tatggttgca ttccttcatt 60

c 61

<210> 331

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 331

tgaggtgtag ttgcattcct tcattcacac taaggcagag tatggttgca ttccttcatt 60

c 61

<210> 332

<211> 61

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 332

ctacgccgat tatcttctga caactttcgc aagcggtgta aggtaaaaaa tgcgggcacc 60

c 61

<210> 333

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 333

ggaatgcaac taccttacac cgcttgcgaa 30

<210> 334

<211> 140

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 334

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 60

gagtgaaggt gggctgcttg catcagccta atgtcgagaa gtgctttctt cggaaagtaa 120

ccctcgaaac aaattcattt 140

<210> 335

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 335

gacgaatgaa ggaatgcaac taccttacac cgcttgcgaa 40

<210> 336

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 336

gacgaatgaa ggaatgcaac ccttacaccg cttgcgaaag 40

<210> 337

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 337

gacgaatgaa ggaatgcaac ttacaccgct tgcgaaagtt 40

<210> 338

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 338

gacgaatgaa ggaatgcaac acaccgcttg cgaaagttgt 40

<210> 339

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 339

gacgaatgaa ggaatgcaac cgtcgccgtc agctcgacc a 41

<210> 340

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 340

gacgaatgaa ggaatgcaac gatcgttacg ctaactatga 40

<210> 341

<211> 180

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 341

ttcactgata aagtggagaa ccgcttcacc aaaagctgtc ccttagggga ttagaacttg 60

agtgaaggtg ggctgcttgc atcagcctaa tgtcgagaag tgctttcttc ggaaagtaac 120

cctcgaaaca aattcatttg aaagaatgaa ggaatgcaac acttgacact taatgctcaa 180

<210> 342

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 342

gggtaatttc tactaagtgt agatacttga cacttaatgc tcaa 44

<210> 343

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 343

gacgaatgaa ggaatgcaac taccttacac cgcttgcgaa agttg 45

<210> 344

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 344

gacgaatgaa ggaatgcaac taccttacac cgcttgcg 38

<210> 345

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 345

gacgaatgaa ggaatgcaac taccttacac cgcttg 36

<210> 346

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 346

gacgaatgaa ggaatgcaac taccttacac cgct 34

<210> 347

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 347

gacgaatgaa ggaatgcaac taccttacac cg 32

<210> 348

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 348

gacgaatgaa ggaatgcaac taccttacac 30

<210> 349

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 349

gttgcagaac ccgaatagac gaatgaagga atgcaactac cttacaccgc ttgcgaa 57

<210> 350

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 350

gaatgaagga atgcaactac cttacaccgc ttgcgaa 37

<210> 351

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 351

atgaaggaat gcaactacct tacaccgctt gcgaa 35

<210> 352

<211> 179

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 352

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 60

gagtgaaggt gggctgcttg catcagccta atgtcgagaa gtgctttctt cggaaagtaa 120

ccctcgaaac aaattcattt ttcctctcca attctgcaca aaaaaaggtg agtccttat 179

<210> 353

<211> 110

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 353

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 60

gagtgaaggt gggctgcttg catcagccta atgtcgagaa gtgctttctt 110

<210> 354

<211> 72

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 354

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 60

gagtgaaggt gg 72

<210> 355

<211> 136

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 355

ttcactgata aagtggagaa ccgcttcacc aaaagctgtc ccttagggga ttagaacttg 60

agtgaaggtg ggctgcttgc atcagcctaa tgtcgagaag tgctttcttc ggaaagtaac 120

cctcgaaaca aattca 136

<

210> 356

<211> 234

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 356

ttcacacttc actgataaag tggagaaccg cttcaccaaa agctgtccct taggggatta 60

gaacttgagt gaaggtgggc tgcttgcatc agcctaatgt cgagaagtgc tttcttcgga 120

aagtaaccct cgaaacaaat tcatttttcc tctccaattc tgcacaaaaa aaggtgagtc 180

cttataaacc ggcgtgcaga acgccggctc acctttttttc ttcattcgat ttta 234

<210> 357

<211> 181

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 357

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 60

gagtgaaggt gggctgcttg catcagccta atgtcgagaa gtgctttctt cggaaagtaa 120

ccctcgaaac aaattcattt gaaagaatga aggaatgcaa ctaccttaca ccgcttgcga 180

a 181

<210> 358

<211> 222

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 358

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 60

gagtgaaggt gggctgcttg catcagccta atgtcgagaa gtgctttctt cggaaagtaa 120

ccctcgaaac aaattcattt ttcctctcca attctgcaca agaaagttgc agaacccgaa 180

tagacgaatg aaggaatgca actaccttac accgcttgcg aa 222

<210> 359

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 359

gacgaatgaa ggaatgcaac taccgaacga accaccagca gaaga 45

<210> 360

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 360

gacgaatgaa ggaatgcaac tcttctgctg gtggttcgtt cggta 45

<210> 361

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 361

gacgaatgaa ggaatgcaac gtttgaccat tagatacatt tcg 43

<210> 362

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 362

gacgaatgaa ggaatgcaac cgaaatgtat ctaatggtca aac 43

<210> 363

<211> 7580

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 363

gcaaccgcac ctgtggcgcc ggtgatgccg gccacgatgc gtccggcgta gaggatcgag 60

atctcgatcc cgcgaaatta atacgactca ctatagggag accacaacgg tttccctcta 120

gtgccggctc cggagagctc tttaattaag cggccgccct gcaggactcg agttctagaa 180

ataattttgt ttaactttaa gaaggagata tacatatgaa atcttctcac catcaccatc 240

accatcacca tcaccatggt tcttctatga aaatcgaaga aggtaaactg gtaatctgga 300

ttaacggcga taaaggctat aacggtctcg ctgaagtcgg taagaaattc gagaaagata 360

ccggaattaa agtcaccgtt gagcatccgg ataaactgga agagaaattc ccacaggttg 420

cggcaactgg cgatggccct gacattatct tctgggcaca cgaccgcttt ggtggctacg 480

ctcaatctgg cctgttggct gaaatcaccc cggacaaagc gttccaggac aagctgtatc 540

cgtttacctg ggatgccgta cgttacaacg gcaagctgat tgcttacccg atcgctgttg 600

aagcgttatc gctgatttat aacaaagatc tgctgccgaa cccgccaaaa acctgggaag 660

agatcccggc gctggataaa gaactgaaag cgaaaggtaa gagcgcgctg atgttcaacc 720

tgcaagaacc gtacttcacc tggccgctga ttgctgctga cgggggttat gcgttcaagt 780

atgaaaacgg caagtacgac attaaagacg tgggcgtgga taacgctggc gcgaaagcgg 840

gtctgacctt cctggttgac ctgattaaaa acaaacacat gaatgcagac accgattact 900

ccatcgcaga agctgccttt aataaaggcg aaacagcgat gaccatcaac ggcccgtggg 960
catggtccaa catcgacacc agcaaagtga attatggtgt aacggtactg ccgaccttca 1020
agggtcaacc atccaaaccg ttcgttggcg tgctgagcgc aggtattaac gccgccagtc 1080
cgaacaaaga gctggcaaaa gagttcctcg aaaactatct gctgactgat gaaggtctgg 1140
aagcggttaa taaagacaaa ccgctgggtg ccgtagcgct gaagtcttac gaggaagagt 1200
tggcgaaaga tccacgtatt gccgccacta tggaaaacgc ccagaaaggt gaaatcatgc 1260
cgaacatccc gcagatgtcc gctttctggt atgccgtgcg tactgcggtg atcaacgccg 1320

ccagcggtcg tcagactgtc gatgaagccc tgaaagacgc gcagactaat tcgagctcga 1380
acaacaacaa caataacaat aacaacaacc tcgggatcga ggaaaacctg tacttccaat 1440
ccaatgcaat ggaagaaagc attattaccg gtgtgaaatt caaactgcgc atcgataaag 1500
aaaccaccaa aaaactgaac gagtacttcg atgaatatgg caaagcaatt aacttcgccg 1560
tgaagatcat tcagaaagaa ctggcagatg atcgttttgc aggtaaagca aaactggacc 1620
agaataaaaa cccgatcctg gatgaaaacg gcaaaaaaat ctatgaattc ccggatgaat 1680
tttgcagctg tggtaaacag gttaacaagt acgttaacaa caaaccgttt tgccaagagt 1740

gctataaaat ccgctttacc gaaaatggta ttcgcaaacg tatgtatagc gccaaaggtc 1800
gtaaagccga acataaaatc aatatcctga acagcaccaa caagatcagc aaaacccatt 1860
ttaactatgc cattcgcgaa gccttcattc tggataaaag catcaaaaag cagcgcaaaa 1920
aacgtaatga acgtctgcgt gaaagtaaaa aacgtctgca gcagtttatc gatatgcgtg 1980
atggtaaacg tgaaatttgc ccgaccatta aaggtcagaa agtggatcgt tttattcatc 2040
cgagctggat caccaaagat aaaaagctgg aagattttcg cggttatacc ctgagcatta 2100
tcaacagcaa aattaagatt ctggatcgca acatcaaacg cgaagaaaaa agcctgaaag 2160

aaaaaggcca gatcatcttt aaagccaaac gtctgatgct ggataaatcc attcgttttg 2220
ttggtgatcg caaagtgctg tttacaatta gtaaaaccct gccgaaagag tatgaactgg 2280
atctgccgag caaagaaaaa cggctgaatt ggctgaaaga gaagatcgag attatcaaga 2340
accagaaacc gaaatatgcc tatctgctgc gcaaaaacat tgagagcgaa aaaaaaccga 2400
actatgagta ctatctgcag tacaccctgg aaattaaacc ggaactgaaa gatttttatg 2460
atggtgccat tggtattgac cgtggcatta atcatattgc cgtttgcacc tttattagca 2520
acgatggtaa agttacccct ccgaaatttt tcagcagcgg tgaaattctg cgtctgaaaa 2580

atctgcagaa agagcgtgat cgctttctgc tgcgtaaaca caacaaaaat cgcaaaaaag 2640

gcaacatgcg cgtgatcgaa aacaaaatca atctgatcct gcaccgttat agcaagcaga 2700

ttgttgatat ggccaaaaag ctgaatgcca gcattgtttt tgaagaactg ggtcgtattg 2760

gtaaaagccg caccaaaatg aaaaaaagcc agcgttataa actgagcctg ttcatcttca 2820

agaaactgag cgatctggtt gattacaaaa gccgtcgtga aggtattcgt gttacctatg 2880

ttccgcctga atataccagc aaagaatgta gccattgcgg tgaaaaagtt aatacccagc 2940

gtccgtttaa tggcaactat agcctgttta aatgcaacaa atgtggcatc cagctgaaca 3000

gcgattataa tgcaagcatc aacattgcga aaaagggcct gaaaattccg aatagcacct 3060

aataacattg gaagtggata acggatccgc gatcgcggcg cgccacctgg tggccggccg 3120

gtaccacgcg tgcgcgctga tccggctgct aacaaagccc gaaaggaagc tgagttggct 3180

gctgccaccg ctgagcaata actagcataa ccccttgggg cctctaaacg ggtcttgagg 3240

ggtttttgc tgaaaggagg aactatatcc ggatatccac aggacgggtg tggtcgccat 3300

gatcgcgtag tcgatagtgg ctccaagtag cgaagcgagc aggactgggc ggcggccaaa 3360

gcggtcggac agtgctccga gaacgggtgc gcatagaaat tgcatcaacg catatagcgc 3420

tagcagcacg ccatagtgac tggcgatgct gtcggaatgg acgatatccc gcaagaggcc 3480

cggcagtacc ggcataacca agcctatgcc tacagcatcc agggtgacgg tgccgaggat 3540

gacgatgagc gcattgttag atttcataca cggtgcctga ctgcgttagc aatttaactg 3600

tgataaacta ccgcattaaa gcttatcgat gataagctgt caaacatgag aattcttgaa 3660

gacgaaaggg cctcgtgata cgcctatttt tataggttaa tgtcatgata ataatggttt 3720

cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt 3780

tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat 3840

aacattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt 3900

ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 3960

ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 4020

tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc 4080

tatgtggcgc ggtattatcc cgtgttgacg ccgggcaaga gcaactcggt cgccgcatac 4140

actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 4200

gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca 4260

acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg 4320

gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg 4380

acgagcgtga caccacgatg cctgcagcaa tggcaacaac gttgcgcaaa ctattaactg 4440

gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 4500

ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 4560

gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 4620

cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 4680

agatcgctga gataggtgcc tcactgatta agcattggta actgtcagac caagtttact 4740

catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga 4800

tccttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt 4860

cagaccccgt agaaaagatc aaaggatctt cttgagatcc ttttttttctg cgcgtaatct 4920

gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc 4980

taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 5040

ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc 5100

tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 5160

ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggtt 5220

cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 5280

agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg 5340

gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt 5400

atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 5460

gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc ctggccttttt 5520

gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta 5580

ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt 5640

cagtgagcga ggaagcggaa gagcgcctga tgcggtattt tctccttacg catctgtgcg 5700

gtatttcaca ccgcaatggt gcactctcag tacaatctgc tctgatgccg catagttaag 5760

ccagtataca ctccgctatc gctacgtgac tgggtcatgg ctgcgccccg acacccgcca 5820

acacccgctg acgcgccctg acgggcttgt ctgctcccgg catccgctta cagacaagct 5880

gtgaccgtct ccgggagctg catgtgtcag aggttttcac cgtcatcacc gaaacgcgcg 5940

aggcagctgc ggtaaagctc atcagcgtgg tcgtgaagcg attcacagat gtctgcctgt 6000

tcatccgcgt ccagctcgtt gagtttctcc agaagcgtta atgtctggct tctgataaag 6060

cgggccatgt taagggcggt tttttcctgt ttggtcactg atgcctccgt gtaaggggga 6120

tttctgttca tgggggtaat gataccgatg aaacgagaga ggatgctcac gatacgggtt 6180

actgatgatg aacatgcccg gttactggaa cgttgtgagg gtaaacaact ggcggtatgg 6240

atgcggcggg accagagaaa aatcactcag ggtcaatgcc agcgcttcgt taatacagat 6300

gtaggtgttc cacagggtag ccagcagcat cctgcgatgc agatccggaa cataatggtg 6360

cagggcgctg acttccgcgt ttccagactt tacgaaacac ggaaaccgaa gaccattcat 6420
gttgttgctc aggtcgcaga cgttttgcag cagcagtcgc ttcacgttcg ctcgcgtatc 6480
ggtgattcat tctgctaacc agtaaggcaa ccccgccagc ctagccgggt cctcaacgac 6540
aggagcacga tcatgcgcac ccgtggccag gacccaacgc tgcccgagat gcgccgcgtg 6600
cggctgctgg agatggcgga cgcgatggat atgttctgcc aagggttggt ttgcgcattc 6660
acagttctcc gcaagaattg attggctcca attcttggag tggtgaatcc gttagcgagg 6720
tgccgccggc ttccattcag gtcgaggtgg cccggctcca tgcaccgcga cgcaacgcgg 6780

ggaggcagac aaggtatagg gcggcgccta caatccatgc caacccgttc catgtgctcg 6840
ccgaggcggc ataaatcgcc gtgacgatca gcggtccaat gatcgaagtt aggctggtaa 6900
gagccgcgag cgatccttga agctgtccct gatggtcgtc atctacctgc ctggacagca 6960
tggcctgcaa cgcgggcatc ccgatgccgc cggaagcgag aagaatcata atggggaagg 7020
ccatccagcc tcgcgtcgcg aacgccagca agacgtagcc cagcgcgtcg gccgccatgc 7080
cggcgataat ggcctgcttc tcgccgaaac gtttggtggc gggaccagtg acgaaggctt 7140
gagcgagggc gtgcaagatt ccgaataccg caagcgacag gccgatcatc gtcgcgctcc 7200

agcgaaagcg gtcctcgccg aaaatgaccc agagcgctgc cggcacctgt cctacgagtt 7260
gcatgataaa gaagacagtc ataagtgcgg cgacgatagt catgccccgc gcccaccgga 7320
aggagctgac tgggttgaag gctctcaagg gcatcggtcg acgctctccc ttatgcgact 7380
cctgcattag gaagcagccc agtagtaggt tgaggccgtt gagcaccgcc gccgcaagga 7440
atggtgcatg caaggagatg gcgcccaaca gtcccccggc cacggggcct gccaccatac 7500
ccacgccgaa acaagcgctc atgagcccga agtggcgagc ccgatcttcc ccatcggtga 7560
tgtcggcgat ataggcgcca 7580

<210> 364
<211> 4715
<212> DNA
<213> Artificial sequence
<220><223> Synthetic sequence
<400> 364
atcgttgata gagttatttt accactccct atcagtgata gagaaaagaa ttcaaaagat 60
ctaaagagga gaaaggatct atggaagaaa gcattattac cggtgtgaaa ttcaaactgc 120
gcatcgataa agaaaccacc aaaaaactga acgagtactt cgatgaatat ggcaaagcaa 180

ttaacttcgc cgtgaagatc attcagaaag aactggcaga tgatcgtttt gcaggtaaag 240

caaaactgga ccagaataaa aacccgatcc tggatgaaaa cggcaaaaaa atctatgaat 300

tcccggatga attttgcagc tgtggtaaac aggttaacaa gtacgttaac aacaaaccgt 360

tttgccaaga gtgctataaa atccgcttta ccgaaaatgg tattcgcaaa cgtatgtata 420

gcgccaaagg tcgtaaagcc gaacataaaa tcaatatcct gaacagcacc aacaagatca 480

gcaaaaccca ttttaactat gccattcgcg aagccttcat tctggataaa agcatcaaaa 540

agcagcgcaa aaaacgtaat gaacgtctgc gtgaaagtaa aaaacgtctg cagcagttta 600

tcgatatgcg tgatggtaaa cgtgaaattt gcccgaccat taaaggtcag aaagtggatc 660

gttttattca tccgagctgg atcaccaaag ataaaaagct ggaagatttt cgcggttata 720

ccctgagcat tatcaacagc aaaattaaga ttctggatcg caacatcaaa cgcgaagaaa 780

aaagcctgaa agaaaaaggc cagatcatct ttaaagccaa acgtctgatg ctggataaat 840

ccattcgttt tgttggtgat cgcaaagtgc tgtttacaat tagtaaaacc ctgccgaaag 900

agtatgaact ggatctgccg agcaaagaaa aacggctgaa ttggctgaaa gagaagatcg 960

agattatcaa gaaccagaaa ccgaaatatg cctatctgct gcgcaaaaac attgagagcg 1020

aaaaaaaacc gaactatgag tactatctgc agtacaccct ggaaattaaa ccggaactga 1080

aagattttta tgatggtgcc attggtattg accgtggcat taatcatatt gccgtttgca 1140

cctttattag caacgatggt aaagttaccc ctccgaaatt tttcagcagc ggtgaaattc 1200

tgcgtctgaa aaatctgcag aaagagcgtg atcgctttct gctgcgtaaa cacaacaaaa 1260

atcgcaaaaa aggcaacatg cgcgtgatcg aaaacaaaat caatctgatc ctgcaccgtt 1320

atagcaagca gattgttgat atggccaaaa agctgaatgc cagcattgtt tttgaagaac 1380

tgggtcgtat tggtaaaagc cgcaccaaaa tgaaaaaaag ccagcgttat aaactgagcc 1440

tgttcatctt caagaaactg agcgatctgg ttgattacaa aagccgtcgt gaaggtattc 1500

gtgttaccta tgttccgcct gaatatacca gcaaagaatg tagccattgc ggtgaaaaag 1560

ttaataccca gcgtccgttt aatggcaact atagcctgtt taaatgcaac aaatgtggca 1620

tccagctgaa cagcgattat aatgcaagca tcaacattgc gaaaaagggc ctgaaaattc 1680

cgaatagcac ctaataatgt tggttaagcc acaatatgga atattgttct tatggacagt 1740

attgacttaa attaataatc ttcgaaggct atatgcggaa gatttggcgt tgttgtaacg 1800

caataagggg taaccctgaa aaggtttgaa atcatataaa cctagtttta tttgagttta 1860

ggctcagata aaatgaacag accaatcttt aattccgttc tgatttaaaa aatcagaatc 1920

tctttataaa tagtattaca aaaagtgtac attccaaaat ccgaaagcag aattgacctt 1980

tttaagccta aaaaagccaa atttcaaggc tctttcatac tcagaacaaa gggattaagg 2040

aatgcaacta ccttacaccg cttgcgaaag ttgtcagaag ataatctttc atactcagaa 2100
caaagggatt aaggaatgca actatcttat ccatttcttg acatcaaatt ttcttgcagc 2160
atctgaattg cttaattgct ttccttgctt cagcaggaaa tagccaagat tttccagttc 2220
tgctggcgtt attgcaagca cggggatttt gtgcttgctg tagattttgt tcttcaggac 2280
cttctttttc catgctagag tcacactggc tcaccttcgg gtgggccttt ctgcgtttat 2340
acctagggat atattccgct tcctcgctca ctgactcgct acgctcggtc gttcgactgc 2400
ggcgagcgga aatggcttac gaacggggcg gagatttcct ggaagatgcc aggaagatac 2460

ttaacaggga agtgagaggg ccgcggcaaa gccgtttttc cataggctcc gccccccctga 2520
caagcatcac gaaatctgac gctcaaatca gtggtggcga aacccgacag gactataaag 2580
ataccaggcg tttccccctg gcggctccct cgtgcgctct cctgttcctg cctttcggtt 2640
taccggtgtc attccgctgt tatggccgcg tttgtctcat tccacgcctg acactcagtt 2700
ccgggtaggc agttcgctcc aagctggact gtatgcacga acccccccgtt cagtccgacc 2760
gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggaaagacat gcaaaagcac 2820
cactggcagc agccactggt aattgattta gaggagttag tcttgaagtc atgcgccggt 2880

taaggctaaa ctgaaaggac aagttttggt gactgcgctc ctccaagcca gttacctcgg 2940
ttcaaagagt tggtagctca gagaaccttc gaaaaaccgc cctgcaaggc ggtttttttcg 3000
ttttcagagc aagagattac gcgcagacca aaacgatctc aagaagatca tcttattaat 3060
cagataaaat atttctagat ttcagtgcaa tttatctctt caaatgtagc acctgaagtc 3120
agccccatac gatataagtt gttactagtg cttggattct caccaataaa aaacgcccgg 3180
cggcaaccga gcgttctgaa caaatccaga tggagttctg aggtcattac tggatctatc 3240
aacaggagtc caagcgagct cgatatcaaa ttacgccccg ccctgccact catcgcagta 3300

ctgttgtaat tcattaagca ttctgccgac atggaagcca tcacaaacgg catgatgaac 3360
ctgaatcgcc agcggcatca gcaccttgtc gccttgcgta taatatttgc ccatggtgaa 3420
aacgggggcg aagaagttgt ccatattggc cacgtttaaa tcaaaactgg tgaaactcac 3480
ccagggattg gctgagacga aaaacatatt ctcaataaac cctttaggga aataggccag 3540
gttttcaccg taacacgcca catcttgcga atatatgtgt agaaactgcc ggaaatcgtc 3600
gtggtattca ctccagagcg atgaaaacgt ttcagtttgc tcatggaaaa cggtgtaaca 3660
agggtgaaca ctatcccata tcaccagctc accgtctttc attgccatac gaaattccgg 3720

atgagcattc atcaggcggg caagaatgtg aataaaggcc ggataaaact tgtgcttatt 3780

tttctttacg gtctttaaaa aggccgtaat atccagctga acggtctggt tataggtaca 3840

ttgagcaact gactgaaatg cctcaaaatg ttctttacga tgccattggg atatatcaac 3900

ggtggtatat ccagtgattt ttttctccat tttagcttcc ttagctcctg aaaatctcga 3960

taactcaaaa aatacgcccg gtagtgatct tatttcatta tggtgaaagt tggaacctct 4020

tacgtgccga tcaacgtctc attttcgcca gatatcgacg tcttaagacc cactttcaca 4080

tttaagttgt ttttctaatc cgcatatgat caattcaagg ccgaataaga aggctggctc 4140

tgcaccttgg tgatcaaata attcgatagc ttgtcgtaat aatggcggca tactatcagt 4200

agtaggtgtt tccctttctt ctttagcgac ttgatgctct tgatcttcca atacgcaacc 4260

taaagtaaaa tgccccacag cgctgagtgc atataatgca ttctctagtg aaaaaccttg 4320

ttggcataaa aaggctaatt gattttcgag agtttcatac tgtttttctg taggccgtgt 4380

acctaaatgt acttttgctc catcgcgatg acttagtaaa gcacatctaa aacttttagc 4440

gttattacgt aaaaaatctt gccagctttc cccttctaaa gggcaaaagt gagtatggtg 4500

cctatctaac atctcaatgg ctaaggcgtc gagcaaagcc cgcttatttt ttacatgcca 4560

atacaatgta ggctgctcta cacctagctt ctgggcgagt ttacgggttg ttaaaccttc 4620

gattccgacc tcattaagca gctctaatgc gctgttaatc actttacttt tatctaatct 4680

agacatcatt aattcctaat ttttgttgac actct 4715

<210> 365

<211> 4670

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 365

taattcctaa tttttgttga cactctatcg ttgatagagt tattttacca ctccctatca 60

gtgatagaga aaagaattca aaagatctaa agaggagaaa ggatctatga aatctcacca 120

tcaccatcac catgaaaacc tgtacttcca atccaatatt ggaagtggaa tggccaaaaa 180

caccattacc aaaacactga aactgcgtat tgtgcgtccg tataatagcg cagaagtgga 240

aaaaattgtt gccgacgaaa aaaacaaccg cgaaaaaatc gcactggaaa agaacaaaga 300

caaagtgaaa gaagcctgca gcaaacatct gaaagttgca gcatattgta ccacacaggt 360

tgaacgtaat gcatgcctgt tttgtaaagc acgtaaactg gatgacaaat tctaccaaaa 420

actgcgtggt cagtttccgg atgcagtttt ttggcaagaa atcagcgaaa tttttcgcca 480

gctgcagaaa caggcagcag aaatctataa tcagagcctg atcgaactgt actacgagat 540

ttttatcaaa ggcaaaggta ttgcaaatgc cagcagcgtt gaacattatc tgagtgatgt 600

ttgttatacc cgtgcagcag aactgtttaa aaacgcagca attgcaagcg gtctgcgtag 660
caaaatcaaa agcaattttc gtctgaaaga actgaaaaac atgaaaagtg gtctgccgac 720
caccaaaagc gataattttc cgattccgct ggttaaacag aaaggtggtc agtataccgg 780
ttttgaaatt agcaatcata atagcgactt catcatcaag attccgtttg gtcgttggca 840
ggtcaaaaaa gagattgata aatatcgtcc gtgggagaaa tttgactttg aacaggttca 900
gaaaagcccg aaaccgatta gcctgctgct gagcacccag cgtcgtaaac gtaataaagg 960
ttggagcaaa gatgaaggca ccgaagccga aatcaaaaaa gttatgaatg gcgattatca 1020

gaccagctac attgaagtta aacgtggcag caaaatcggt gaaaaaagcg catggatgct 1080
gaatctgagc attgatgttc cgaaaattga taaaggtgtg gatccgagca ttattggtgg 1140
tattgatgtt ggtgttaaat caccgctggt ttgcgcaatt aacaatgcat ttagccgtta 1200
tagcatcagc gataacgacc tgtttcactt caacaagaaa atgtttgcac gtcgtcgtat 1260
cctgctgaaa aaaaaccgtc ataaacgtgc aggtcatggt gcaaaaaaca aactgaaacc 1320
gatcaccatt ctgaccgaaa aaagtgaacg ttttcgcaaa aagctgattg aacgttgggc 1380
atgtgaaatc gcggatttct tcattaaaaa caaagttggc accgtgcaga tggaaaatct 1440

ggaaagcatg aaacgtaaag aggacagcta ttttaacatt cgcctgcgtg gcttttggcc 1500
gtatgcagaa atgcagaaca aaatcgaatt caaactgaag cagtatggca tcgaaattcg 1560
taaagttgca ccgaataata ccagcaaaac ctgtagcaaa tgtggccatc tgaacaacta 1620
tttcaacttc gagtaccgca agaaaaacaa attcccgcac tttaaatgcg aaaaatgcaa 1680
cttcaaagaa aacgccgatt ataatgcagc cctgaatatt tcaaacccga aactgaaaag 1740
caccaaagag gaaccgtaaa tatttatact ttattatcct tcattgacaa aaatgagaat 1800
gttatcccag ataacatttg atgtacacag attcacactt cactgataaa gtggagaacc 1860

gcttcaccaa aagctgtccc ttaggggatt agaacttgag tgaaggtggg ctgcttgcat 1920
cagcctaatg tcgagaagtg ctttcttcgg aaagtaaccc tcgaaacaaa ttcatttttc 1980
ctctccaatt ctgcacaaaa aaaggtgagt ccttataaac cggcgtgcag aacgccggct 2040
cacctttttt cttcattcga ttttatgctt aaaagccgta aaaacgcgga attcggcgcc 2100
gttgcagaac ccgaatagac gaatgaagga atgcaactac cttacaccgc ttgcgaaagt 2160
tgtcagaaga taatcttgca gaacccgaat agacgaatga aggaatgcaa tcttgacaga 2220
gcccgattgc gttatctcca ggagaaacat ataaaagcat caaccgctga tcggactaga 2280

gtcacactgg ctcaccttcg ggtgggcctt tctgcgttta tacctaggga tatattccgc 2340

ttcctcgctc actgactcgc tacgctcggt cgttcgactg cggcgagcgg aaatggctta 2400

cgaacggggc ggagatttcc tggaagatgc caggaagata cttaacaggg aagtgagagg 2460

gccgcggcaa agccgttttt ccataggctc cgccccccctg acaagcatca cgaaatctga 2520

cgctcaaatc agtggtggcg aaacccgaca ggactataaa gataccaggc gtttcccccct 2580

ggcggctccc tcgtgcgctc tcctgttcct gcctttcggt ttaccggtgt cattccgctg 2640

ttatggccgc gtttgtctca ttccacgcct gacactcagt tccgggtagg cagttcgctc 2700

caagctggac tgtatgcacg aacccccccgt tcagtccgac cgctgcgcct tatccggtaa 2760

ctatcgtctt gagtccaacc cggaaagaca tgcaaaagca ccactggcag cagccactgg 2820

taattgattt agaggagtta gtcttgaagt catgcgccgg ttaaggctaa actgaaagga 2880

caagttttgg tgactgcgct cctccaagcc agttacctcg gttcaaagag ttggtagctc 2940

agagaacctt cgaaaaaccg ccctgcaagg cggttttttc gttttcagag caagagatta 3000

cgcgcagacc aaaacgatct caagaagatc atcttattaa tcagataaaa tatttctaga 3060

tttcagtgca atttatctct tcaaatgtag cacctgaagt cagccccata cgatataagt 3120

tgttactagt gcttggattc tcaccaataa aaaacgcccg gcggcaaccg agcgttctga 3180

acaaatccag atggagttct gaggtcatta ctggatctat caacaggagt ccaagcgagc 3240

tcgatatcaa attacgcccc gccctgccac tcatcgcagt actgttgtaa ttcattaagc 3300

attctgccga catggaagcc atcacaaacg gcatgatgaa cctgaatcgc cagcggcatc 3360

agcaccttgt cgccttgcgt ataatatttg cccatggtga aaacgggggc gaagaagttg 3420

tccatattgg ccacgtttaa atcaaaactg gtgaaactca cccagggatt ggctgagacg 3480

aaaaacatat tctcaataaa ccctttaggg aaataggcca ggttttcacc gtaacacgcc 3540

acatcttgcg aatatatgtg tagaaactgc cggaaatcgt cgtggtattc actccagagc 3600

gatgaaaacg tttcagtttg ctcatggaaa acggtgtaac aagggtgaac actatcccat 3660

atcaccagct caccgtcttt cattgccata cgaaattccg gatgagcatt catcaggcgg 3720

gcaagaatgt gaataaaggc cggataaaac ttgtgcttat ttttctttac ggtctttaaa 3780

aaggccgtaa tatccagctg aacggtctgg ttataggtac attgagcaac tgactgaaat 3840

gcctcaaaat gttctttacg atgccattgg gatatatcaa cggtggtata tccagtgatt 3900

tttttctcca ttttagcttc cttagctcct gaaaatctcg ataactcaaa aaatacgccc 3960

ggtagtgatc ttatttcatt atggtgaaag ttggaacctc ttacgtgccg atcaacgtct 4020

cattttcgcc agatatcgac gtcttaagac ccactttcac atttaagttg tttttctaat 4080

ccgcatatga tcaattcaag gccgaataag aaggctggct ctgcaccttg gtgatcaaat 4140

aattcgatag cttgtcgtaa taatggcggc atactatcag tagtaggtgt ttccctttct 4200

tctttagcga cttgatgctc ttgatcttcc aatacgcaac ctaaagtaaa atgccccaca 4260

gcgctgagtg catataatgc attctctagt gaaaaacctt gttggcataa aaaggctaat 4320

tgattttcga gagtttcata ctgtttttct gtaggccgtg tacctaaatg tacttttgat 4380

ccatcgcgat gacttagtaa agcacatcta aaacttttag cgttattacg taaaaaatct 4440

tgccagcttt ccccttctaa agggcaaaag tgagtatggt gcctatctaa catctcaatg 4500

gctaaggcgt cgagcaaagc ccgcttattt tttacatgcc aatacaatgt aggctgctct 4560

acacctagct tctgggcgag tttacgggtt gttaaacctt cgattccgac ctcattaagc 4620

agctctaatg cgctgttaat cactttactt ttatctaatc tagacatcat 4670

<210> 366

<211> 4607

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 366

ttacttttat ctaatctaga catcattaat tcctaatttt tgttgacact ctatcgttga 60

tagagttatt ttaccactcc ctatcagtga tagagaaaag aattcaaaag atctaaagag 120

gagaaaggat ctatggccaa aaacaccatt accaaaacac tgaaactgcg tattgtgcgt 180

ccgtataata gcgcagaagt ggaaaaaatt gttgccgacg aaaaaaacaa ccgcgaaaaa 240

atcgcactgg aaaagaacaa agacaaagtg aaagaagcct gcagcaaaca tctgaaagtt 300

gcagcatatt gtaccacaca ggttgaacgt aatgcatgcc tgttttgtaa agcacgtaaa 360

ctggatgaca aattctacca aaaactgcgt ggtcagtttc cggatgcagt tttttggcaa 420

gaaatcagcg aaatttttcg ccagctgcag aaacaggcag cagaaatcta taatcagagc 480

ctgatcgaac tgtactacga gatttttatc aaaggcaaag gtattgcaaa tgccagcagc 540

gttgaacatt atctgagtga tgtttgttat acccgtgcag cagaactgtt taaaaacgca 600

gcaattgcaa gcggtctgcg tagcaaaatc aaaagcaatt ttcgtctgaa agaactgaaa 660

aacatgaaaa gtggtctgcc gaccaccaaa agcgataatt ttccgattcc gctggttaaa 720

cagaaaggtg gtcagtatac cggttttgaa attagcaatc ataatagcga cttcatcatc 780

aagattccgt ttggtcgttg gcaggtcaaa aaagagattg ataaatatcg tccgtgggag 840

aaatttgact ttgaacaggt tcagaaaagc ccgaaaccga ttagcctgct gctgagcacc 900

cagcgtcgta aacgtaataa aggttggagc aaagatgaag gcaccgaagc cgaaatcaaa 960

aaagttatga atggcgatta tcagaccagc tacattgaag ttaaacgtgg cagcaaaatc 1020

ggtgaaaaaa gcgcatggat gctgaatctg agcattgatg ttccgaaaat tgataaaggt 1080

gtggatccga gcattattgg tggtattgat gttggtgtta aatcaccgct ggtttgcgca 1140

attaacaatg catttagccg ttatagcatc agcgataacg acctgtttca cttcaacaag 1200

aaaatgtttg cacgtcgtcg tatcctgctg aaaaaaaacc gtcataaacg tgcaggtcat 1260

ggtgcaaaaa acaaactgaa accgatcacc attctgaccg aaaaaagtga acgttttcgc 1320

aaaaagctga ttgaacgttg ggcatgtgaa atcgcggatt tcttcattaa aaacaaagtt 1380

ggcaccgtgc agatggaaaa tctggaaagc atgaaacgta aagaggacag ctattttaac 1440

attcgcctgc gtggcttttg gccgtatgca gaaatgcaga acaaaatcga attcaaactg 1500

aagcagtatg gcatcgaaat tcgtaaagtt gcaccgaata ataccagcaa aacctgtagc 1560

aaatgtggcc atctgaacaa ctatttcaac ttcgagtacc gcaagaaaaa caaattcccg 1620

cactttaaat gcgaaaaatg caacttcaaa gaaaacgccg attataatgc agccctgaat 1680

atttcaaacc cgaaactgaa aagcaccaaa gaggaaccgt aaatatttat actttattat 1740

ccttcattga caaaaatgag aatgttatcc cagataacat ttgatgtaca cagattcaca 1800

cttcactgat aaagtggaga accgcttcac caaaagctgt cccttagggg attagaactt 1860

gagtgaaggt gggctgcttg catcagccta atgtcgagaa gtgctttctt cggaaagtaa 1920

ccctcgaaac aaattcattt ttcctctcca attctgcaca aaaaaaggtg agtccttata 1980

aaccggcgtg cagaacgccg gctcaccttt tttcttcatt cgattttatg cttaaaagcc 2040

gtaaaaacgc ggaattcggc gccgttgcag aacccgaata gacgaatgaa ggaatgcaac 2100

taccttacac cgcttgcgaa agttgtcaga agataatctt gcagaacccg aatagacgaa 2160

tgaaggaatg caatcttgac agagcccgat tgcgttatct ccaggagaaa catataaaag 2220

catcaaccgc tgatcggact agagtcacac tggctcacct tcgggtgggc ctttctgcgt 2280

ttatacctag ggatatattc cgcttcctcg ctcactgact cgctacgctc ggtcgttcga 2340

ctgcggcgag cggaaatggc ttacgaacgg ggcggagatt tcctggaaga tgccaggaag 2400

atacttaaca gggaagtgag agggccgcgg caaagccgtt tttccatagg ctccgccccc 2460

ctgacaagca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 2520

aaagatacca ggcgtttccc cctggcggct ccctcgtgcg ctctcctgtt cctgcctttc 2580

ggtttaccgg tgtcattccg ctgttatggc cgcgtttgtc tcattccacg cctgacactc 2640

agttccgggt aggcagttcg ctccaagctg gactgtatgc acgaaccccc cgttcagtcc 2700

gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggaaag acatgcaaaa 2760

gcaccactgg cagcagccac tggtaattga tttagaggag ttagtcttga agtcatgcgc 2820

cggttaaggc taaactgaaa ggacaagttt tggtgactgc gctcctccaa gccagttacc 2880

tcggttcaaa gagttggtag ctcagagaac cttcgaaaaa ccgccctgca aggcggtttt 2940

ttcgttttca gagcaagaga ttacgcgcag accaaaacga tctcaagaag atcatcttat 3000

taatcagata aaatatttct agatttcagt gcaatttatc tcttcaaatg tagcacctga 3060

agtcagcccc atacgatata agttgttact agtgcttgga ttctcaccaa taaaaaacgc 3120

ccggcggcaa ccgagcgttc tgaacaaatc cagatggagt tctgaggtca ttactggatc 3180

tatcaacagg agtccaagcg agctcgatat caaattacgc cccgccctgc cactcatcgc 3240

agtactgttg taattcatta agcattctgc cgacatggaa gccatcacaa acggcatgat 3300

gaacctgaat cgccagcggc atcagcacct tgtcgccttg cgtataatat ttgcccatgg 3360

tgaaaacggg ggcgaagaag ttgtccatat tggccacgtt taaatcaaaa ctggtgaaac 3420

tcacccaggg attggctgag acgaaaaaca tattctcaat aaaacccttta gggaaatagg 3480

ccaggttttc accgtaacac gccacatctt gcgaatatat gtgtagaaac tgccggaaat 3540

cgtcgtggta ttcactccag agcgatgaaa acgtttcagt ttgctcatgg aaaacggtgt 3600

aacaagggtg aacactatcc catatcacca gctcaccgtc tttcattgcc atacgaaatt 3660

ccggatgagc attcatcagg cgggcaagaa tgtgaataaa ggccggataa aacttgtgct 3720

tatttttctt tacggtcttt aaaaaggccg taatatccag ctgaacggtc tggttatagg 3780

tacattgagc aactgactga aatgcctcaa aatgttcttt acgatgccat tgggatatat 3840

caacggtggt atatccagtg attttttttct ccattttagc ttccttagct cctgaaaatc 3900

tcgataactc aaaaaatacg cccggtagtg atcttatttc attatggtga aagttggaac 3960

ctcttacgtg ccgatcaacg tctcattttc gccagatatc gacgtcttaa gacccacttt 4020

cacatttaag ttgtttttct aatccgcata tgatcaattc aaggccgaat aagaaggctg 4080

gctctgcacc ttggtgatca aataattcga tagcttgtcg taataatggc ggcatactat 4140

cagtagtagg tgtttccctt tcttctttag cgacttgatg ctcttgatct tccaatacgc 4200

aacctaaagt aaaatgcccc acagcgctga gtgcatataa tgcattctct agtgaaaaac 4260

cttgttggca taaaaaggct aattgatttt cgagagtttc atactgtttt tctgtaggcc 4320

gtgtacctaa atgtactttt gctccatcgc gatgacttag taaagcacat ctaaaacttt 4380

tagcgttatt acgtaaaaaa tcttgccagc tttccccttc taaagggcaa aagtgagtat 4440

ggtgcctatc taacatctca atggctaagg cgtcgagcaa agcccgctta ttttttacat 4500

gccaatacaa tgtaggctgc tctacaccta gcttctgggc gagtttacgg gttgttaaac 4560

cttcgattcc gacctcatta agcagctcta atgcgctgtt aatcact 4607

<210> 367

<211> 7556

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 367

acccacgccg aaacaagcgc tcatgagccc gaagtggcga gcccgatctt ccccatcggt 60

gatgtcggcg atataggcgc cagcaaccgc acctgtggcg ccggtgatgc cggccacgat 120

gcgtccggcg tagaggatcg agatctcgat cccgcgaaat taatacgact cactataggg 180

agaccacaac ggtttccctc tagtgccggc tccggagagc tctttaatta agcggccgcc 240

ctgcaggact cgagttctag aaataatttt gtttaacttt aagaaggaga tatacatatg 300

aaatcttctc accatcacca tcaccatcac catcaccatg gttcttctat gaaaatcgaa 360

gaaggtaaac tggtaatctg gattaacggc gataaaggct ataacggtct cgctgaagtc 420

ggtaagaaat tcgagaaaga taccggaatt aaagtcaccg ttgagcatcc ggataaactg 480

gaagagaaat tcccacaggt tgcggcaact ggcgatggcc ctgacattat cttctgggca 540

cacgaccgct ttggtggcta cgctcaatct ggcctgttgg ctgaaatcac cccggacaaa 600

gcgttccagg acaagctgta tccgtttacc tgggatgccg tacgttacaa cggcaagctg 660

attgcttacc cgatcgctgt tgaagcgtta tcgctgattt ataacaaaga tctgctgccg 720

aacccgccaa aaacctggga agagatcccg gcgctggata aagaactgaa agcgaaaggt 780

aagagcgcgc tgatgttcaa cctgcaagaa ccgtacttca cctggccgct gattgctgct 840

gacgggggtt atgcgttcaa gtatgaaaac ggcaagtacg acattaaaga cgtgggcgtg 900

gataacgctg gcgcgaaagc gggtctgacc ttcctggttg acctgattaa aaacaaacac 960

atgaatgcag acaccgatta ctccatcgca gaagctgcct ttaataaagg cgaaacagcg 1020

atgaccatca acggcccgtg ggcatggtcc aacatcgaca ccagcaaagt gaattatggt 1080

gtaacggtac tgccgacctt caagggtcaa ccatccaaac cgttcgttgg cgtgctgagc 1140

gcaggtatta acgccgccag tccgaacaaa gagctggcaa aagagttcct cgaaaactat 1200

ctgctgactg atgaaggtct ggaagcggtt aataaagaca aaccgctggg tgccgtagcg 1260

ctgaagtctt acgaggaaga gttggcgaaa gatccacgta ttgccgccac tatggaaaac 1320

gcccagaaag gtgaaatcat gccgaacatc ccgcagatgt ccgctttctg gtatgccgtg 1380

cgtactgcgg tgatcaacgc cgccagcggt cgtcagactg tcgatgaagc cctgaaagac 1440

gcgcagacta attcgagctc gaacaacaac aacaataaca ataacaacaa cctcgggatc 1500

gaggaaaacc tgtacttcca atccaatgca atggccaaaa acaccattac caaaacactg 1560

aaactgcgta ttgtgcgtcc gtataatagc gcagaagtgg aaaaaattgt tgccgacgaa 1620

aaaaacaacc gcgaaaaaat cgcactggaa aagaacaaag acaaagtgaa agaagcctgc 1680

agcaaacatc tgaaagttgc agcatattgt accacacagg ttgaacgtaa tgcatgcctg 1740

ttttgtaaag cacgtaaact ggatgacaaa ttctaccaaa aactgcgtgg tcagtttccg 1800

gatgcagttt tttggcaaga aatcagcgaa atttttcgcc agctgcagaa acaggcagca 1860

gaaatctata atcagagcct gatcgaactg tactacgaga tttttatcaa aggcaaaggt 1920

attgcaaatg ccagcagcgt tgaacattat ctgagtgatg tttgttatac ccgtgcagca 1980

gaactgttta aaaacgcagc aattgcaagc ggtctgcgta gcaaaatcaa aagcaatttt 2040

cgtctgaaag aactgaaaaa catgaaaagt ggtctgccga ccaccaaaag cgataatttt 2100

ccgattccgc tggttaaaca gaaaggtggt cagtataccg gttttgaaat tagcaatcat 2160

aatagcgact tcatcatcaa gattccgttt ggtcgttggc aggtcaaaaa agagattgat 2220

aaatatcgtc cgtgggagaa atttgacttt gaacaggttc agaaaagccc gaaaccgatt 2280

agcctgctgc tgagcaccca gcgtcgtaaa cgtaataaag gttggagcaa agatgaaggc 2340

accgaagccg aaatcaaaaa agttatgaat ggcgattatc agaccagcta cattgaagtt 2400

aaacgtggca gcaaaatctg tgaaaaaagc gcatggatgc tgaatctgag cattgatgtt 2460

ccgaaaattg ataaaggtgt ggatccgagc attattggtg gtattgatgt tggtgttaaa 2520

tcaccgctgg tttgcgcaat taacaatgca tttagccgtt atagcatcag cgataacgac 2580

ctgtttcact tcaacaagaa aatgtttgca cgtcgtcgta tcctgctgaa aaaaaaccgt 2640

cataaacgtg caggtcatgg tgcaaaaaac aaactgaaac cgatcaccat tctgaccgaa 2700

aaaagtgaac gttttcgcaa aaagctgatt gaacgttggg catgtgaaat cgcggatttc 2760

ttcattaaaa acaaagttgg caccgtgcag atggaaaatc tggaaagcat gaaacgtaaa 2820

gaggacagct attttaacat tcgcctgcgt ggcttttggc cgtatgcaga aatgcagaac 2880

aaaatcgaat tcaaactgaa gcagtatggc atcgaaattc gtaaagttgc accgaataat 2940

accagcaaaa cctgtagcaa atgtggccat ctgaacaact atttcaactt cgagtaccgc 3000

aagaaaaaca aattcccgca ctttaaatgc gaaaaatgca acttcaaaga aaacgccgat 3060

tataatgcag ccctgaatat ttcaaacccg aaactgaaaa gcaccaaaga ggaaccgtaa 3120

taacattgga agtggataac ggatccgcga tcgcggcgcg ccacctggtg gccggccggt 3180

accacgcgtg cgcgctgatc cggctgctaa caaagcccga aaggaagctg agttggctgc 3240

tgccaccgct gagcaataac tagcataacc ccttggggcc tctaaacggg tcttgagggg 3300

tttttgctg aaaggaggaa ctatatccgg atatccacag gacgggtgtg gtcgccatga 3360

tcgcgtagtc gatagtggct ccaagtagcg aagcgagcag gactgggcgg cggccaaagc 3420
ggtcggacag tgctccgaga acgggtgcgc atagaaattg catcaacgca tatagcgcta 3480
gcagcacgcc atagtgactg gcgatgctgt cggaatggac gatatcccgc aagaggcccg 3540
gcagtaccgg cataaccaag cctatgccta cagcatccag ggtgacggtg ccgaggatga 3600
cgatgagcgc attgttagat ttcatacacg gtgcctgact gcgttagcaa tttaactgtg 3660
ataaactacc gcattaaagc ttatcgatga taagctgtca aacatgagaa ttcttgaaga 3720
cgaaagggcc tcgtgatacg cctattttta taggttaatg tcatgataat aatggtttct 3780

tagacgtcag gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc 3840
taaatacatt caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa 3900
cattgaaaaa ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt 3960
gcggcatttt gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct 4020
gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc 4080
cttgagagtt ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta 4140
tgtggcgcgg tattatcccg tgttgacgcc gggcaagagc aactcggtcg ccgcatacac 4200

tattctcaga atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc 4260
atgacagtaa gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac 4320
ttacttctga caacgatcgg aggaccgaag gagctaaccg ctttttgca caacatgggg 4380
gatcatgtaa ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac 4440
gagcgtgaca ccacgatgcc tgcagcaatg gcaacaacgt tgcgcaaact attaactggc 4500
gaactactta ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt 4560
gcaggaccac ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga 4620

gccggtgagc gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc 4680
cgtatcgtag ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag 4740
atcgctgaga taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca 4800
tatatacttt agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc 4860
ctttttgata atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca 4920
gaccccgtag aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc 4980
tgcttgcaaa caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta 5040

ccaactcttt ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt 5100

ctagtgtagc cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc 5160

gctctgctaa tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg 5220

ttggactcaa gacgatagtt accggataag gcgcagcggt cgggctgaac gggggttcg 5280

tgcacacagc ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag 5340

ctatgagaaa gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc 5400

agggtcggaa caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat 5460

agtcctgtcg ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg 5520

gggcggagcc tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc 5580

tggccttttg ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt 5640

accgcctttg agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca 5700

gtgagcgagg aagcggaaga gcgcctgatg cggtattttc tccttacgca tctgtgcggt 5760

atttcacacc gcaatggtgc actctcagta caatctgctc tgatgccgca tagttaagcc 5820

agtatacact ccgctatcgc tacgtgactg ggtcatggct gcgccccgac acccgccaac 5880

acccgctgac gcgccctgac gggcttgtct gctcccggca tccgcttaca gacaagctgt 5940

gaccgtctcc gggagctgca tgtgtcagag gttttcaccg tcatcaccga aacgcgcgag 6000

gcagctgcgg taaagctcat cagcgtggtc gtgaagcgat tcacagatgt ctgcctgttc 6060

atccgcgtcc agctcgttga gtttctccag aagcgttaat gtctggcttc tgataaagcg 6120

ggccatgtta agggcggttt tttcctgttt ggtcactgat gcctccgtgt aaggggggatt 6180

tctgttcatg ggggtaatga taccgatgaa acgagagagg atgctcacga tacgggttac 6240

tgatgatgaa catgcccggt tactggaacg ttgtgagggt aaacaactgg cggtatggat 6300

gcggcgggac cagagaaaaa tcactcaggg tcaatgccag cgcttcgtta atacagatgt 6360

aggtgttcca cagggtagcc agcagcatcc tgcgatgcag atccggaaca taatggtgca 6420

gggcgctgac ttccgcgttt ccagacttta cgaaacacgg aaaccgaaga ccattcatgt 6480

tgttgctcag gtcgcagacg ttttgcagca gcagtcgctt cacgttcgct cgcgtatcgg 6540

tgattcattc tgctaaccag taaggcaacc ccgccagcct agccgggtcc tcaacgacag 6600

gagcacgatc atgcgcaccc gtggccagga cccaacgctg cccgagatgc gccgcgtgcg 6660

gctgctggag atggcggacg cgatggatat gttctgccaa gggttggttt gcgcattcac 6720

agttctccgc aagaattgat tggctccaat tcttggagtg gtgaatccgt tagcgaggtg 6780

ccgccggctt ccattcaggt cgaggtggcc cggctccatg caccgcgacg caacgcgggg 6840

aggcagacaa ggtatagggc ggcgcctaca atccatgcca acccgttcca tgtgctcgcc 6900

gaggcggcat aaatcgccgt gacgatcagc ggtccaatga tcgaagttag gctggtaaga 6960

gccgcgagcg atccttgaag ctgtccctga tggtcgtcat ctacctgcct ggacagcatg 7020

gcctgcaacg cgggcatccc gatgccgccg gaagcgagaa gaatcataat ggggaaggcc 7080

atccagcctc gcgtcgcgaa cgccagcaag acgtagccca gcgcgtcggc cgccatgccg 7140

gcgataatgg cctgcttctc gccgaaacgt ttggtggcgg gaccagtgac gaaggcttga 7200

gcgagggcgt gcaagattcc gaataccgca agcgacaggc cgatcatcgt cgcgctccag 7260

cgaaagcggt cctcgccgaa aatgacccag agcgctgccg gcacctgtcc tacgagttgc 7320

atgataaaga agacagtcat aagtgcggcg acgatagtca tgccccgcgc ccaccggaag 7380

gagctgactg ggttgaaggc tctcaagggc atcggtcgac gctctccctt atgcgactcc 7440

tgcattagga agcagcccag tagtaggttg aggccgttga gcaccgccgc cgcaaggaat 7500

ggtgcatgca aggagatggc gcccaacagt cccccggcca cggggcctgc caccat 7556

<210> 368

<211> 30

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 368

ttcatttgag cattaaatgt caagttctgc 30

<210> 369

<211> 30

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 369

ttcatttgag cattaagtgt caagttctgc 30

<210> 370

<211> 21

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 370

aaactcgtaa ttcacagttc a 21

<210> 371

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 371

tatatatata 10

<210> 372

<211> 6

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 372

tatata 6

<210> 373

<211> 5

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 373

tatat 5