(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0090512
(43) 공개일자 2022년06월29일

(51) 국제특허분류(Int. Cl.)
*A61K 39/00* (2006.01) *A61P 35/00* (2006.01)
*A61P 35/02* (2006.01) *C07K 14/705* (2006.01)
*C07K 14/725* (2006.01) *C12N 15/113* (2010.01)
*C12N 5/0783* (2010.01)
(52) CPC특허분류
*A61K 39/001102* (2021.08)
*A61K 39/001129* (2018.08)
(21) 출원번호 10-2022-7013872
(22) 출원일자(국제) 2020년09월25일
심사청구일자 없음
(85) 번역문제출일자 2022년04월25일
(86) 국제출원번호 PCT/US2020/052822
(87) 국제공개번호 WO 2021/062227
국제공개일자 2021년04월01일
(30) 우선권주장
62/907,254 2019년09월27일 미국(US)

(71) 출원인
**빔 테라퓨틱스, 인크.**
미국 02139 매사추세츠주 케임브리지 랜즈다운 스트리트 26 세컨드 플로어
**유씨엘 비즈니스 리미티드**
영국, 더블유씨1이 6비티 런던, 고워 스트리트, 유니버시티 칼리지 런던
(72) 발명자
**카심 와심**
영국 블룸스버리 그레이터 런던 더블유1티 4티피 97 토트넘 코트 로드
**게르케 제이슨 마이클**
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트 빔 테라퓨틱스, 인크. 내
*(뒷면에 계속)*
(74) 대리인
**문두현**

전체 청구항 수 : 총 137 항

(54) 발명의 명칭 **액체암의 치료를 위한 조성물 및 방법**

## (57) 요 약

아래에 기재된 바와 같이, 본 발명은 향상된 항-신생물 활성, 면역 억제에 대한 저항성 및 이식편 대 숙주 반응을 유발할 위험의 감소 또는 이들의 조합을 갖는 유전적으로 변형된 면역 세포를 특징으로 한다. 본 발명은 또한 이러한 변형된 면역 효과기 세포를 생산하고 사용하는 방법을 특징으로 한다.

(52) CPC특허분류
***A61P 35/00*** (2018.01)
***A61P 35/02*** (2018.01)
***C07K 14/7051*** (2013.01)
***C07K 14/70596*** (2013.01)
***C12N 15/1138*** (2013.01)
***C12N 5/0636*** (2013.01)
*A61K 2039/5156* (2013.01)
*A61K 2039/804* (2018.08)
*C12N 2310/20* (2017.05)
(72) 발명자
**에드워즈 아론 디.**
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트 빔 테라퓨틱스, 인크. 내
**머레이 라이언**
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트 빔 테라퓨틱스, 인크. 내

# 명 세 서

## 청구범위

### 청구항 1

2종 이상의 면역 세포를 포함하는 조성물로서, 각 면역 세포는,

a) CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 상이한 키메라 항원 수용체(여기서, 상기 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 포함함); 및

b) TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 면역원성 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이

를 포함하는, 조성물.

### 청구항 2

제1항에 있어서, 상기 면역 세포 중 하나는 CD5를 표적으로 하는 키메라 항원 수용체를 포함하고, 또 다른 면역 세포는 CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 키메라 항원 수용체를 포함하는, 조성물.

### 청구항 3

제2항에 있어서, 상기 조성물은 각각 상이한 항원을 표적으로 하는 키메라 항원 수용체를 포함하는 적어도 3종의 면역 세포를 포함하되, 상기 표적화된 항원은 CD3, CD5 및 CD7인, 조성물.

### 청구항 4

제1항에 있어서, 상기 면역 세포 중 하나는 CD7을 표적으로 하는 키메라 항원 수용체를 발현하고, 또 다른 면역 세포는 CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 키메라 항원 수용체를 발현하는, 조성물.

### 청구항 5

제1항에 있어서, 상기 면역 세포 중 하나는 CD3을 표적으로 하는 키메라 항원 수용체를 발현하고, 또 다른 면역 세포는 CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 키메라 항원 수용체를 발현하는, 조성물.

### 청구항 6

제1항에 있어서, 상기 면역 세포 중 하나는 CD33을 표적으로 하는 키메라 항원 수용체를 발현하고, 또 다른 면역 세포는 CD123을 표적으로 하는 키메라 항원 수용체를 발현하는, 조성물.

### 청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 하나의 면역 세포는 2개, 3개, 4개 이상의 상이한 키메라 항원 수용체를 발현하는, 조성물.

### 청구항 8

제7항에 있어서, 상기 키메라 항원 수용체 중 하나는 CD5를 표적으로 하고, 또 다른 키메라 항원 수용체는 CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는, 조성물.

### 청구항 9

제7항에 있어서, 상기 키메라 항원 수용체 중 하나는 CD7을 표적으로 하고, 또 다른 키메라 항원 수용체는 CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는, 조성물.

**청구항 10**

제7항에 있어서, 상기 키메라 항원 수용체 중 하나는 CD3을 표적으로 하고, 또 다른 키메라 항원 수용체는 CD33, CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는, 조성물.

**청구항 11**

제7항에 있어서, 상기 키메라 항원 수용체 중 하나는 CD33을 표적으로 하고, 또 다른 키메라 항원 수용체는 CD123을 표적으로 하는, 조성물.

**청구항 12**

적어도 3종의 면역 세포를 포함하는 조성물로서, 각각은 상이한 키메라 항원 수용체를 포함하되, 하나의 키메라 항원 수용체는 CD3을 표적으로 하고, 하나는 CD5를 표적으로 하며, 세 번째는 CD7을 표적으로 하고, 상기 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하며; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 추가로 포함하는, 조성물.

**청구항 13**

적어도 2종의 면역 세포를 포함하는 조성물로서, 각각은 상이한 키메라 항원 수용체를 포함하되, 하나의 키메라 항원 수용체는 CD3을 표적으로 하고, 또 다른 하나는 CD7을 표적으로 하며, 상기 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 추가로 포함하는, 조성물.

**청구항 14**

적어도 2종의 면역 세포를 포함하는 조성물로서, 각각은 상이한 키메라 항원 수용체를 포함하되, 하나의 키메라 항원 수용체는 CD5를 표적으로 하고, 또 다른 하나는 CD7을 표적으로 하며, 상기 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 추가로 포함하는, 조성물.

**청구항 15**

적어도 2종의 면역 세포를 포함하는 조성물로서, 각각은 상이한 키메라 항원 수용체를 포함하되, 하나의 키메라 항원 수용체는 CD3을 표적으로 하고, 또 다른 하나는 CD5를 표적으로 하며, 상기 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 추가로 포함하는, 조성물.

**청구항 16**

적어도 2종의 면역 세포를 포함하는 조성물로서, 각각은 상이한 키메라 항원 수용체를 포함하되, 하나의 키메라 항원 수용체는 CD33을 표적으로 하고, 또 다른 하나는 CD123을 표적으로 하며, 상기 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 추가로 포함하는, 조성물.

**청구항 17**

제1항 내지 제16항 중 어느 한 항에 있어서, 상기 돌연변이는 유전자를 침묵시키거나 또는 종결 코돈을 유전자에 도입하는 C에서 T 또는 A에서 G로의 돌연변이인, 조성물.

**청구항 18**

제1항 내지 제17항 중 어느 한 항에 있어서, 상기 돌연변이는 조기 종결 코돈을 도입하거나 또는 스플라이스 공

여체 또는 수용체 부위를 변경하는, 조성물.

**청구항 19**

제18항에 있어서, 상기 돌연변이는 데아미네이스 도메인을 포함하는 염기 편집기에 의해 생성되는, 조성물.

**청구항 20**

제19항에 있어서, 상기 데아미네이스는 사이티딘 또는 아데노신 데아미네이스인, 조성물.

**청구항 21**

제20항에 있어서, 상기 염기 편집기는 BE4인, 조성물.

**청구항 22**

제1항 내지 제21항 중 어느 한 항에 있어서, 상기 돌연변이는 돌연변이가 없는 상응하는 대조군 세포에 비해 암호화된 폴리펩타이드의 발현을 약 50% 이상 감소시키는, 조성물.

**청구항 23**

제1항 내지 제22항 중 어느 한 항에 있어서, 상기 조성물은 면역 세포의 집단을 포함하는, 조성물.

**청구항 24**

제23항에 있어서, 상기 집단의 적어도 50%는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및/또는 면역원성 폴리펩타이드를 포함하는, 조성물.

**청구항 25**

제1항 내지 제24항 중 어느 한 항에 있어서, 상기 면역 세포는 동족 살해 저항성인, 조성물.

**청구항 26**

제1항 내지 제25항 중 어느 한 항에 있어서, 상기 면역 세포는 항-신생물 활성이 증가된, 조성물.

**청구항 27**

제1항 내지 제26항 중 어느 한 항에 있어서, 상기 면역 세포는 검출 가능한 전위를 포함하지 않는, 조성물.

**청구항 28**

제1항 내지 제27항 중 어느 한 항에 있어서, 상기 면역 세포는 1% 미만의 indel을 포함하는, 조성물.

**청구항 29**

제1항 내지 제28항 중 어느 한 항에 있어서, 상기 면역 세포는 포유동물 세포인, 조성물.

**청구항 30**

제29항에 있어서, 상기 포유동물 세포는 인간 또는 설치류 세포인, 조성물.

**청구항 31**

제1항 내지 제30항 중 어느 한 항에 있어서, 상기 면역 세포는 세포독성 T 세포, 조절 T 세포, T 헬퍼 세포, 수지상 세포, B 세포 또는 NK 세포 또는 이들의 선조세포인, 조성물.

**청구항 32**

제31항에 있어서, 상기 선조세포는 조혈 줄기 세포인, 조성물.

**청구항 33**

약제학적 조성물로서, 유효량의 제1항 내지 제32항 중 어느 한 항에 따른 조성물 및 약제학적으로 허용 가능한

부형제를 포함하는, 약제학적 조성물.

**청구항 34**

염기 편집 시스템으로서, 핵산 프로그램 가능 DNA 결합 단백질(nucleic acid programmable DNA binding protein: napDNAbp)과 데아미네이스 도메인을 포함하는 융합 단백질 및 각각 상이한 항원을 표적으로 하는 적어도 2개의 가이드 폴리뉴클레오타이드를 포함하되, 상기 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는, 염기 편집 시스템.

**청구항 35**

제34항에 있어서, 상기 하나의 가이드 폴리뉴클레오타이드는 CD5를 표적으로 하고, 또 다른 하나는 CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는, 염기 편집 시스템.

**청구항 36**

제34항에 있어서, 상기 시스템은 각각이 항원 CD3, CD5 및 CD7 중 하나를 표적으로 하는 3개의 가이드 폴리뉴클레오타이드를 포함하는, 염기 편집 시스템.

**청구항 37**

제34항에 있어서, 상기 하나의 가이드 폴리뉴클레오타이드는 CD7을 표적으로 하고, 또 다른 하나는 CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는, 염기 편집 시스템.

**청구항 38**

제34항에 있어서, 상기 하나의 가이드 폴리뉴클레오타이드는 CD3을 표적으로 하고, 또 다른 하나는 CD33, CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는, 염기 편집 시스템.

**청구항 39**

제34항에 있어서, 상기 하나의 가이드 폴리뉴클레오타이드는 CD33을 표적으로 하고, 또 다른 하나는 CD123을 표적으로 하는, 염기 편집 시스템.

**청구항 40**

제34항에 있어서, 상기 가이드 폴리뉴클레오타이드는 각각 표 26으로부터 선택되는 핵산 서열을 포함하는, 염기 편집 시스템.

**청구항 41**

제40항에 있어서, 상기 가이드 폴리뉴클레오타이드는 AGCGACUGCAGAAAGAAGAG 또는 CAUACCAGCUGAGCCGUCCG로부터 선택되는 핵산 서열을 포함하는, 염기 편집 시스템.

**청구항 42**

제34항 내지 제41항 중 어느 한 항에 있어서, 상기 융합 단백질은 하나 이상의 우라실 글리코실레이스 저해제(uracil glycosylase inhibitor: UGI) 및/또는 하나 이상의 핵 국부화 서열(nuclear localization sequence: NLS)을 추가로 포함하는, 염기 편집 시스템.

**청구항 43**

제34항 내지 제42항 중 어느 한 항에 있어서, 상기 napDNAbp는 Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 또는 Cas12j/CasΦ 폴리펩타이드 또는 이들의 일부를 포함하는, 염기 편집 시스템.

**청구항 44**

제43항에 있어서, 상기 napDNAbp는 Cas12 폴리펩타이드 또는 이의 단편을 포함하는, 염기 편집 시스템.

**청구항 45**

제43항에 있어서, 상기 napDNAbp는 Cas9 폴리펩타이드 또는 이의 단편을 포함하는, 염기 편집 시스템.

**청구항 46**

제45항에 있어서, 상기 Cas9는 활성이 없는 Cas9(dead Cas9: dCas9) 또는 Cas9 닉케이스(Cas9 nickase: nCas9)인, 염기 편집 시스템.

**청구항 47**

제45항에 있어서, 상기 Cas9는 변형된 스타필로코커스 아우레우스 Cas9(*Staphylococcus aureus* Cas9: SaCas9), 스트렙토코커스 써모필루스 1 Cas9(*Streptococcus thermophilus* 1 Cas9: St1Cas9) 또는 변형된 스트렙토코커스 피오게네스 Cas9(*Streptococcus pyogenes* Cas9: SpCas9)인, 염기 편집 시스템.

**청구항 48**

제45항에 있어서, 상기 Cas9는 변경된 프로토스페이서-인접 모티프(protospacer-adjacent motif: PAM) 특이성을 포함하는, 염기 편집 시스템.

**청구항 49**

제48항에 있어서, 상기 변경된 PAM은 핵산 서열 5'-NGC-3'에 대한 특이성을 갖는, 염기 편집 시스템.

**청구항 50**

제34항 내지 제49항 중 어느 한 항에 있어서, 상기 데아미네이스 도메인은 사이티딘 또는 아데노신을 탈아미노화할 수 있는, 염기 편집 시스템.

**청구항 51**

제50항에 있어서, 상기 데아미네이스 도메인은 사이티딘 또는 아데노신 데아미네이스 도메인인, 염기 편집 시스템.

**청구항 52**

제51항에 있어서, 상기 사이티딘 데아미네이스는 APOBEC 데아미네이스인, 염기 편집 시스템.

**청구항 53**

제51항에 있어서, 상기 아데노신 데아미네이스는 TadA 변이체인, 염기 편집 시스템.

**청구항 54**

제51항에 있어서, 상기 TadA 변이체는 TadA*8 변이체인, 염기 편집 시스템.

**청구항 55**

제54항에 있어서, 상기 TadA*8 변이체는 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23 또는 TadA*8.24인, 염기 편집 시스템.

**청구항 56**

제53항에 있어서, 상기 TadA 변이체는 TadA*9 변이체인, 염기 편집 시스템.

**청구항 57**

약제학적 조성물로서, 제34항 내지 제56항 중 어느 한 항에 따른 염기 편집 시스템을 포함하는, 약제학적 조성

물.

**청구항 58**

제34항 내지 제56항 중 어느 한 항에 따른 염기 편집 시스템을 암호화하는 폴리뉴클레오타이드 및 가이드 폴리뉴클레오타이드.

**청구항 59**

벡터로서, 제58항의 폴리뉴클레오타이드를 포함하는, 벡터.

**청구항 60**

제59항에 있어서, 상기 벡터는 바이러스 벡터인, 벡터.

**청구항 61**

제60항에 있어서, 상기 바이러스 벡터는 레트로바이러스 벡터, 아데노바이러스 벡터, 렌티바이러스 벡터, 헤르페스바이러스 벡터 또는 아데노-관련 바이러스 벡터(AAV)인, 벡터.

**청구항 62**

약제학적 조성물로서, 제58항의 폴리뉴클레오타이드 또는 제59항 내지 제61항 중 어느 한 항에 따른 벡터를 포함하는, 약제학적 조성물.

**청구항 63**

면역원성이 감소된 CAR-발현 면역 세포를 생산하는 방법으로서, CAR-발현 면역 세포에서 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)과 데아미네이스 도메인을 포함하는 융합 단백질 및 각각 상이한 항원을 암호화하는 폴리뉴클레오타이드를 표적화하는 2개의 가이드 폴리뉴클레오타이드를 포함하는 염기 편집 시스템을 발현시키는 단계를 포함하되, 상기 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되어 면역원성이 감소된 CAR-발현 면역 세포를 생성하는, 방법.

**청구항 64**

제63항에 있어서, 상기 면역 세포는 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 표적으로 하는 가이드 폴리뉴클레오타이드를 발현하거나 또는 이와 접촉되는, 방법.

**청구항 65**

면역원성이 감소된 CAR-발현 면역 세포를 생산하는 방법으로서,

(a) CAR-발현 면역 세포에서 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)과 데아미네이스 도메인을 포함하는 융합 단백질을 포함하는 염기 편집 시스템을 발현시키는 단계; 및

(b) 상기 CAR-발현 면역 세포를 적어도 각각 상이한 항원을 암호화하는 폴리뉴클레오타이드를 표적화하는 2개의 가이드 폴리뉴클레오타이드와 접촉시키는 단계로서, 상기 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되어 면역원성이 감소된 CAR-발현 면역 세포를 생성하는, 상기 접촉시키는 단계

를 포함하는, 방법.

**청구항 66**

제65항에 있어서, 상기 방법은 상기 CAR-발현 면역 세포를 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 표적으로 하는 가이드 폴리뉴클레오타이드와 접촉시키는 단계를 더 포함하는, 방법.

**청구항 67**

제63항 또는 제64항에 있어서, 상기 면역 세포는 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택

되는 항원을 표적으로 하는 키메라 항원 수용체(CAR)를 발현하는, 방법.

**청구항 68**

제63항 또는 제64항에 있어서, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 돌연변이를 상기 면역 세포에 도입시켜 면역원성이 감소된 CAR-발현 면역 세포의 집단을 생성하는 단계를 더 포함하는, 방법.

**청구항 69**

면역원성이 감소된 CAR-발현 면역 세포의 집단을 생산하는 방법으로서,

a) CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 하나의 면역 세포에 도입하고, 상기 항원 중 하나의 상이한 돌연변이를 제2 면역 세포에 도입하는 단계; 및

b) TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 돌연변이를 상기 면역 세포에 도입하여 면역원성이 감소된 CAR-발현 면역 세포의 집단을 생성하는 단계

를 포함하는, 방법.

**청구항 70**

제69항에 있어서, 상기 면역 세포에 의해 발현된 키메라 항원 수용체는 CD3, CD5, CD7, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는, 방법.

**청구항 71**

제65항 내지 제68항 중 어느 한 항에 있어서, 상기 방법에 의해 생성된 면역 세포는 CD3, CD5 및/또는 CD7 항원을 표적으로 하고, CD3, CD5 및/또는 CD7 항원을 발현하지 못하거나 감소된 수준의 이들을 발현하는 키메라 항원 수용체를 발현하는, 방법.

**청구항 72**

제65항 내지 제68항 중 어느 한 항에 있어서, 상기 방법에 의해 생성된 면역 세포는 CD33 및 CD123 항원을 표적으로 하고, CD33 및 CD123 항원을 발현하지 못하거나 감소된 수준의 이들을 발현하는 키메라 항원 수용체를 발현하는, 방법.

**청구항 73**

제71항에 있어서, 상기 CAR은 CD5 키메라 항원 수용체(CAR)인, 방법.

**청구항 74**

제73항에 있어서, 상기 CD5 CAR은 표 28에 제시된 CD5 CAR 작제물에 의해 암호화되는, 방법.

**청구항 75**

면역원성이 감소된 면역 세포를 생산하는 방법으로서,

a) 내인성 CD5 유전자 서열 또는 CD5의 발현을 감소시키거나 또는 제거하는 조절 요소에 돌연변이를 도입하는 단계; 및

b) 표 28에 제시된 CD5 CAR 작제물을 세포에서 발현시키는 단계

를 포함하는, 방법.

**청구항 76**

제74항 또는 제75항에 있어서, 상기 CD5 CAR 작제물은 하기로부터 선택되는 아미노산 서열을 포함하거나 또는 이로 구성되는 CD5 CAR 폴리펩타이드를 암호화하는, 방법:

a)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL 300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ 360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKFWVLVVVGGVLACYSLLVTVAFIIFWV 420
421 RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQ 480
481 LYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGE 540
541 RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR 573;
```

b)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA 300
301 AGGAVHTRGLDFACDFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPR 360
361 RPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDK 420
421 RRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTAT 480
481 KDTYDALHMQALPPR 495;
```

c)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAIEVMYPPPYLDNEKSNGTIIHVKGKHLCPS 300
301 PLFPGPSKPFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTR 360
361 KHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDP 420
421 EMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDA 480
481 LHMQALPPR 489;
```

d)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL 300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ 360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKPTTTPAPRPPTPAPTIASQPLSLRPEA 420
421 CRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMR 480
481 PVQTTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVL 540
541 DKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLST 600
601 ATKDTYDALHMQALPPR 617; 또는
```

e)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA 300
301 AGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQT 360
361 TQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRR 420
421 DPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKD 480
481 DALHMQALPPRX 494.
```

**청구항 77**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 방법에 의해 생성된 면역 세포는 상응하는 대조군 세포와 비교하여 동족 살해 저항성 및/또는 증가된 항-신생물 활성을 나타내는, 방법.

**청구항 78**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 방법은 생체내 또는 생체외에서 수행되는, 방법.

**청구항 79**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 방법에 의해 생성된 면역 세포는 검출 가능한 전위를 포함하지 않는, 방법.

**청구항 80**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 방법에 의해 생성된 면역 세포는 1% 미만의 indel을 포함하는, 방법.

**청구항 81**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 방법에 의해 생성된 면역 세포는 5% 미만의 비-표적 편집을 포함하는, 방법.

**청구항 82**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 방법에 의해 생성된 면역 세포는 5% 미만의 표적외 편집을 포함하는, 방법.

**청구항 83**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 돌연변이는 핵염기 변형에 의해 생성되는, 방법.

**청구항 84**

제83항에 있어서, 상기 돌연변이는 엑손에 있는, 방법.

**청구항 85**

제83항 또는 제84항에 있어서, 상기 돌연변이는 단백질 발현을 감소시키거나 또는 제거하는 조기 종결 코돈을 생성하는, 방법.

**청구항 86**

제83항 또는 제84항에 있어서, 상기 돌연변이는 스플라이스 공여체 부위 또는 스플라이스 수용체 부위에 있는, 방법.

**청구항 87**

제63항 내지 제76항 중 어느 한 항에 있어서, 상기 하나 이상의 돌연변이는 표적 폴리뉴클레오타이드를 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp), 데아미네이스 및 하나 이상의 가이드 폴리뉴클레오타이드를 포함하는 융합 단백질을 포함하는 염기 편집 시스템과 접촉시킴으로써 생성되는, 방법.

**청구항 88**

제87항에 있어서, 상기 데아미네이스는 아데노신 또는 사이티딘 데아미네이스인, 방법.

**청구항 89**

제88항에 있어서, 상기 사이티딘 데아미네이스는 BE4인, 방법.

**청구항 90**

제87항에 있어서, 상기 돌연변이는 돌연변이가 없는 상응하는 대조군 세포에 비해 암호화된 폴리펩타이드의 발현을 적어도 약 50% 이상 감소시키는, 방법.

**청구항 91**

제87항에 있어서, 상기 가이드 폴리뉴클레오타이드는 표 26에 제공되는 것들로부터 선택되는 서열을 포함하는, 방법.

**청구항 92**

제91항에 있어서, 상기 가이드 폴리뉴클레오타이드는 AGCGACUGCAGAAAGAAGAG 또는 CAUACCAGCUGAGCCGUCCG로부터 선택되는 핵산 서열을 포함하는, 방법.

**청구항 93**

제87항에 있어서, 상기 하나 이상의 가이드 핵산 서열 각각은 *CD5, FAS, LAG-3, CD52, TRAC, B2M, CIITA, TRBC1, TRBC2* 및/또는 P*DC1/PD-1* 유전자 또는 조절 요소에 대해 napDNAbp를 표적화하는, 방법.

**청구항 94**

제87항에 있어서, 상기 염기 편집기 및 하나 이상의 가이드 핵산 서열은 전기천공법, 뉴클레오펙션, 양이온성 지질-매개성 방법, 바이러스 형질도입 또는 이들의 조합을 통해 상기 면역 세포에 도입되는, 방법.

**청구항 95**

제63항 내지 제94항 중 어느 한 항에 있어서, 면역 세포의 집단을 생성하기 위해 배양에서 상기 면역 세포를 확장시키는 단계를 더 포함하는, 방법.

**청구항 96**

제95항에 있어서, 상기 항원 또는 폴리펩타이드의 발현은 상기 면역 세포의 집단의 적어도 약 50%에서 감소되는, 방법.

**청구항 97**

제95항에 있어서, 상기 변형된 면역 세포의 집단으로부터 TCRα/β+ 세포를 고갈시키는 단계를 더 포함하는, 방법.

**청구항 98**

제63항 내지 제97항 중 어느 한 항에 따른 방법에 의해 생성된 면역원성이 감소된 CAR-발현 면역 세포.

**청구항 99**

약제학적 조성물로서, 제98항의 면역 세포를 포함하는, 약제학적 조성물.

**청구항 100**

제99항에 있어서, 약제학적으로 허용 가능한 부형제를 더 포함하는, 약제학적 조성물.

**청구항 101**

신생물 세포를 죽이는 방법으로서, CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 발현하는 신생물 세포를 2종 이상의 면역 세포와 접촉시키는 단계를 포함하되, 각각의 면역 세포는 상기 세포에 의해 발현되는 상기 항원 중 2개를 표적으로 하는 상이한 키메라 항원 수용체를 발현하고, 상기 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하는, 방법.

**청구항 102**

제101항에 있어서, 상기 방법은 시험관내 또는 생체내에서 수행되는, 방법.

**청구항 103**

제101항에 있어서, 상기 신생물 세포는 신생물로부터 유래되는, 방법.

**청구항 104**

대상체에서 신생물을 치료하는 방법으로서, 2종 이상의 면역 세포를 상기 대상체에게 투여하는 단계를 포함하되, 각각의 면역 세포는 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 상기 대상체의 신생물 세포에 의해 발현된 항원을 표적으로 하는 상이한 키메라 항원 수용체를 발현하고, 상기 면역 세포는 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하며, 상기 면역 세포 각각은 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 더 포함하는, 방법.

**청구항 105**

제103항 또는 제104항에 있어서, 상기 신생물은 혈액암인, 방법.

**청구항 106**

제105항에 있어서, 상기 신생물은 액체암인, 방법.

**청구항 107**

제105항에 있어서, 상기 혈액암은 백혈병, 골수종 및/또는 림프종인, 방법.

**청구항 108**

제107항에 있어서, 상기 혈액암은 B 세포암인, 방법.

**청구항 109**

제105항에 있어서, 상기 혈액암은 다음 T-세포 급성 림프아구성 백혈병(T-cell acute lymphoblastic leukemia: T-ALL), 용상 진균증(mycosis fungoides: MF), 세자리 증후군(Sézary syndrome: SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 $ALK^+$, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 $CD30^+$ 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종 중 적어도 하나로부터 선택되는, 방법.

**청구항 110**

제109항에 있어서, 상기 혈액암은 T-세포 급성 림프아구성 백혈병(T-ALL) 세포인, 방법.

**청구항 111**

제109항에 있어서, 상기 혈액암은 급성 골수성 백혈병(AML)인, 방법.

**청구항 112**

선택된 대상체에서 신생물을 치료하는 방법으로서, 2종 이상의 면역 세포를 상기 선택된 대상체에게 투여하는 단계를 포함하되, 각각의 면역 세포는 상기 대상체의 신생물 세포에 의해 발현된 항원을 표적으로 하는 상이한 키메라 항원 수용체를 발현하고, 상기 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되고, 상기 면역 세포는 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하며, 상기 면역 세포 각각은 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 추가로 포함하고, 상기 대상체는 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 발현하는 신생물을 갖는 대상체로부터 선택되는, 방법.

**청구항 113**

제109항에 있어서, 상기 혈액암은 T-세포 급성 림프아구성 백혈병(T-ALL) 세포인, 방법.

**청구항 114**

제109항에 있어서, 상기 혈액암은 급성 골수성 백혈병(AML)인, 방법.

**청구항 115**

제101항 내지 제112항 중 어느 한 항에 있어서, 상기 상이한 키메라 항원 수용체를 발현하는 2종 이상의 면역 세포는 순차적으로 투여되는, 방법.

**청구항 116**

제101항 내지 제112항 중 어느 한 항에 있어서, 상기 상이한 키메라 항원 수용체를 발현하는 2종 이상의 면역 세포는 동시에 투여되는, 방법.

**청구항 117**

대상체에서 신생물 세포를 항원-의존적으로 죽이는 방법으로서, 2종 이상의 면역 세포를 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 발현하는 신생물을 갖는 대상체에게 투여하는 단계를 포함하되, 각각의 면역 세포는 상기 항원 중 하나를 표적으로 하는 상이한 키메라 항원 수용체를 발현하고, 상기 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하는, 방법.

**청구항 118**

대상체에서 급성 골수성 백혈병(AML) 세포를 항원-의존적으로 죽이는 방법으로서, 2종 이상의 면역 세포를 CD33 및 CD123 항원을 발현하는 AML을 갖는 대상체에게 투여하는 단계를 포함하되, 각각의 면역 세포는 상기 항원 중 하나를 표적으로 하는 상이한 키메라 항원 수용체를 발현하고, 상기 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하는, 방법.

**청구항 119**

대상체에서 T-세포 급성 림프아구성 백혈병(T-ALL) 세포를 항원-의존적으로 죽이는 방법으로서, 적어도 3종의 면역 세포를 CD3, CD5 및 CD7 항원을 발현하는 T-ALL을 갖는 대상체에게 투여하는 단계를 포함하되, 각각의 면역 세포는 상기 항원 중 하나를 표적으로 하는 상이한 키메라 항원 수용체를 발현하고, 상기 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이, 및 TRAC, LAG-3, FAS, CIITA, TRBC1,

TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하는, 방법.

**청구항 120**

선택된 대상체에서 암을 치료하는 방법으로서, 적어도 2종의 면역 세포를 상기 대상체에게 투여하는 단계를 포함하되, 각각의 면역 세포는 CD33 또는 CD123 항원을 표적으로 하는 키메라 항원 수용체를 발현하고, 상기 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하고, 상기 대상체는 상기 암을 CD33 및 CD123 항원을 발현하는 것으로 특성화함으로써 선택되는, 방법.

**청구항 121**

제120항에 있어서, 상기 CD33 및 CD123 항원을 발현하는 암은 AML인, 방법.

**청구항 122**

선택된 대상체에서 암을 치료하는 방법으로서, 3종 이상의 면역 세포를 상기 대상체에게 투여하는 단계를 포함하되, 각각의 면역 세포는 CD3, CD5 및 CD7 항원을 표적으로 하는 상이한 키메라 항원 수용체를 발현하고, 상기 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하고, 상기 대상체는 상기 암을 CD3, CD5 및 CD7 항원을 발현하는 것으로 특성화함으로써 선택되는, 방법.

**청구항 123**

제122항에 있어서, 상기 CD3, CD5 및 CD7 항원을 발현하는 암은 T-ALL인, 방법.

**청구항 124**

제101항 내지 제123항 중 어느 한 항에 있어서, 상기 면역 세포는 세포독성 T 세포, 조절 T 세포, T 헬퍼 세포, 수지상 세포, B 세포 또는 NK 세포인, 방법.

**청구항 125**

제101항 내지 제123항 중 어느 한 항에 있어서, 상기 대상체는 이전에 림프제거로 치료된, 방법.

**청구항 126**

제125항에 있어서, 상기 림프제거는 사이클로포스파마이드(cyclophosphamide), 플루다라빈(fludarabine) 및/또는 알렘투주맙(alemtuzumab)(Cy/Flu/캄파스(Campath))의 투여를 포함하는, 방법.

**청구항 127**

제101항 내지 제126항 중 어느 한 항에 있어서, 상기 대상체는 화학요법에 대해 불응성이거나 또는 높은 종양 부담을 갖는, 방법.

**청구항 128**

제101항 내지 제127항 중 어느 한 항에 있어서, 상기 대상체는 후속적으로 동종이계 조혈 줄기 세포 이식(allogeneic hematopoietic stem cell transplantation: allo-HSCT)으로 치료되는, 방법.

**청구항 129**

제1항 내지 제32항, 제63항 내지 제97항 및 제101항 내지 제111항 중 어느 한 항에 있어서, 상기 면역 세포는 단일 인간 공여체로부터 유래되는, 조성물 또는 방법.

**청구항 130**

제101항 내지 제121항 중 어느 한 항에 있어서, 상기 면역 세포는 대상체에 대해 자가 조직인, 방법.

**청구항 131**

제101항 내지 제121항 중 어느 한 항에 있어서, 상기 면역 세포는 대상체에 대해 동종이계인, 방법.

**청구항 132**

제101항 내지 제121항 중 어느 한 항에 있어서, 상기 대상체는 포유동물 대상체인, 방법.

**청구항 133**

제101항 내지 제121항 중 어느 한 항에 있어서, 상기 대상체는 인간 또는 설치류 대상체인, 방법.

**청구항 134**

제101항 내지 제121항 중 어느 한 항에 있어서, 상기 대상체는 소아 인간 대상체인, 방법.

**청구항 135**

키트로서, 암의 치료에 사용하기 위한 제1항 내지 제32항 중 어느 한 항에 따른 조성물을 포함하는, 키트.

**청구항 136**

키트로서, 면역원성이 감소된 CAR-발현 면역 세포를 생성하는데 사용하기 위한 제34항 내지 제56항 중 어느 한 항에 따른 염기 편집 시스템을 포함하는, 키트.

**청구항 137**

제135항 또는 제136항에 있어서, 키트의 사용을 위한 서면 설명서를 추가로 포함하는, 키트.

## *발명의 설명*

### *기 술 분 야*

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 출원은 2019년 9월 27일자로 출원된 미국 특허 가출원 제62/907,254호의 우선권 및 이익을 주장하는 국제 PCT출원이며, 그 전체 내용은 전문이 참조에 의해 본 명세서에 원용된다.

### *배 경 기 술*

[0003] 자가 및 동종이계 면역요법은 키메라 항원 수용체를 발현하는 면역 세포가 대상체에게 투여되는 신생물 치료 접근법이다. 키메라 항원 수용체(chimera antigen receptor: CAR)를 발현하는 면역 세포를 생성하기 위해, 면역 세포는 먼저 대상체(자가) 또는 치료를 받는 대상체로부터 분리한 공여체(동종이계)로부터 수집되고, 키메라 항원 수용체를 발현하도록 유전적으로 변형된다. 그 결과 생성된 세포는 이의 세포(예를 들어, CAR T-세포) 표면에서 키메라 항원 수용체를 발현하며, 대상체에게 투여 시, 키메라 항원 수용체는 신생물 세포에 의해 발현된 마커에 결합한다. 신생물 마커와의 이러한 상호작용은 CAR-T 세포를 활성화한 다음, 세포는 신생물 세포를 죽인다. 그러나, 자가 또는 동종이계 세포 요법이 효과적이고 효율적이기 위해서는, T 세포 신호전달 저해와 같은 중요한 상태 및 세포 반응이 극복되거나 회피되어야 한다. 동종이계 세포 요법의 경우, 이식편 대 숙주 질환 및 CAR-T 세포의 숙주 거부가 추가적인 문제를 제공할 수 있다. 이러한 과정에 관여하는 유전자를 편집하면 CAR-T 세포 기능 및 면역억제 또는 저해에 대한 저항성을 향상시킬 수 있지만, 이러한 편집을 위한 현재 방법론은 CAR-T 세포에서 대규모의 게놈 재배열을 유도하여 이의 효능에 부정적인 영향을 미칠 가능성이 있다. 따라서, 면역 세포, 특히 CAR-T 세포를 보다 정확하게 변형시키기 위한 기법에 대한 상당한 요구 사항이 존재한다. 본 출원은 이것 및 기타 중요한 요구 사항에 관한 것이다.

### *발명의 내용*

[0004] 아래 기술되는 바와 같이, 본 발명은 향상된 항-신생물 활성, 면역 억제에 대한 저항성 및 숙주 CD8+ T 세포가 이식편을 자기가 아닌 것으로 인식하는(예를 들어, 이식 수용체(transplant recipient)가 이식된 기관에 대해

면역 반응을 생성하는 경우) 이식편 대 숙주 반응 또는 숙주 대 이식편 반응을 유발의 감소된 위험 또는 이들의 조합을 갖는 유전적으로 변형된 면역 세포를 특징으로 한다. 일 실시형태에서, 이식편 대 숙주 질환(graft versus host disease: GVHD)이 있거나 이를 발병하는 경향이 있는 대상체에는 기능적 TRAC가 결여되어 있거나 또는 이의 수준이 감소된 CAR-T 세포가 투여된다. 일 실시형태에서, 숙주 대 이식편 질환(host versus graft disease: HVGD)이 있거나 이를 발병하는 경향이 있는 대상체에는 기능적 베타2 마이크로글로불린(beta2 microglobulin: B2M)이 결여되어 있거나 또는 이의 수준이 감소된 CAR-T 세포가 투여된다. 본 발명은 또한 이러한 변형된 면역 세포를 생산하고 사용하는 방법을 특징으로 한다.

[0005] 일 양태에서, 본 발명은 2종 이상의 면역 세포를 포함하는 조성물을 제공하며, 각 면역 세포는 a) CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 상이한 키메라 항원 수용체(여기서, 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 포함함); 및 b) TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 면역원성 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함한다. 일부 실시형태에서, 면역 세포 중 하나는 CD5를 표적으로 하는 키메라 항원 수용체를 포함하고, 또 다른 면역 세포는 CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 키메라 항원 수용체를 포함한다. 일부 실시형태에서, 조성물은 상이한 항원을 표적으로 하는 키메라 항원 수용체를 각각 포함하는 적어도 3종의 면역 세포를 포함하되, 표적화된 항원은 CD3, CD5 및 CD7이다. 일부 실시형태에서, 면역 세포 중 하나는 CD7을 표적으로 하는 키메라 항원 수용체를 발현하고, 또 다른 면역 세포는 CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 키메라 항원 수용체를 발현한다. 일부 실시형태에서, 면역 세포 중 하나는 CD3을 표적으로 하는 키메라 항원 수용체를 발현하고, 또 다른 면역 세포는 CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 키메라 항원 수용체를 발현한다. 일부 실시형태에서, 면역 세포 중 하나는 CD33을 표적으로 하는 키메라 항원 수용체를 발현하고, 또 다른 면역 세포는 CD123을 표적으로 하는 키메라 항원 수용체를 발현한다. 일부 실시형태에서, 하나의 면역 세포는 2개, 3개, 4개 이상의 상이한 키메라 항원 수용체를 발현한다. 일부 실시형태에서, 키메라 항원 수용체 중 하나는 CD5를 표적으로 하고, 또 다른 키메라 항원 수용체는 CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 한다. 일부 실시형태에서, 키메라 항원 수용체 중 하나는 CD7을 표적으로 하고, 또 다른 키메라 항원 수용체는 CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 한다. 일부 실시형태에서, 키메라 항원 수용체 중 하나는 CD3을 표적으로 하고, 또 다른 키메라 항원 수용체는 CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 한다. 일부 실시형태에서, 키메라 항원 수용체 중 하나는 CD33을 표적으로 하고, 또 다른 키메라 항원 수용체는 CD123을 표적으로 한다.

[0006] 또 다른 양태에서, 본 발명은 상이한 키메라 항원 수용체를 각각 포함하는 적어도 3종의 면역 세포를 포함하는 조성물을 제공하되, 하나의 키메라 항원 수용체는 CD3을 표적으로 하고, 다른 하나는 CD5를 표적으로 하며, 세 번째는 CD7을 표적으로 하고, 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 더 포함한다.

[0007] 또 다른 양태에서, 본 발명은 상이한 키메라 항원 수용체를 각각 포함하는 적어도 2종의 면역 세포를 포함하는 조성물을 제공하되, 하나의 키메라 항원 수용체는 CD3을 표적으로 하고, 또 다른 하나는 CD7을 표적으로 하며, 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 더 포함한다.

[0008] 일 양태에서, 본 발명은 상이한 키메라 항원 수용체를 각각 포함하는 적어도 2종의 면역 세포를 포함하는 조성물을 제공하되, 하나의 키메라 항원 수용체는 CD5를 표적으로 하고, 또 다른 CD7을 표적으로 하고, 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 더 포함한다.

[0009] 또 다른 양태에서, 본 발명은 상이한 키메라 항원 수용체를 각각 포함하는 적어도 2종의 면역 세포를 포함하는 조성물을 제공하되, 하나의 키메라 항원 수용체는 CD3을 표적으로 하고, 또 다른 하나는 CD5를 표적으로 하며, 세포는 표적화된 항원의 발현을 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연

변이를 더 포함한다.

[0010] 또 다른 양태에서, 본 발명은 상이한 키메라 항원 수용체를 각각 포함하는 적어도 2종의 면역 세포를 포함하는 조성물을 제공하되, 하나의 키메라 항원 수용체는 CD33을 표적으로 하고, 또 다른 하나는 CD123을 표적으로 하며, 세포는 표적화된 항원의 발현을 돌연변이를 더 포함하고; TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 더 포함한다.

[0011] 일부 실시형태에서, 돌연변이는 유전자를 침묵시키거나 또는 종결 코돈을 유전자에 도입하는 C에서 T로 또는 A에서 G로의 돌연변이이다. 일부 실시형태에서, 돌연변이는 조기 종결 코돈을 도입하거나 또는 스플라이스 공여체 또는 수용체 부위(splice donor or acceptor site)를 변경한다. 일부 실시형태에서, 돌연변이는 데아미네이스 도메인을 포함하는 염기 편집기(염기 편집기)에 의해 생성된다. 일부 실시형태에서, 데아미네이스는 사이티딘 또는 아데노신 데아미네이스이다. 일부 실시형태에서, 염기 편집기는 BE4이다. 일부 실시형태에서, 돌연변이는 돌연변이가 없는 상응하는 대조군 세포에 비해 암호화된 폴리펩타이드의 발현을 약 50% 이상 감소시킨다. 일부 실시형태에서, 조성물은 면역 세포의 집단을 포함한다. 일부 실시형태에서, 집단의 적어도 50%는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및/또는 면역원성 폴리펩타이드를 포함한다. 일부 실시형태에서, 면역 세포는 동족 살해 저항성(fratricide resistant)이다. 일부 실시형태에서, 면역 세포는 증가된 항-신생물 활성을 갖는다. 일부 실시형태에서, 면역 세포는 검출 가능한 전위를 포함하지 않는다. 일부 실시형태에서, 면역 세포는 1% 미만의 indel을 포함한다. 일부 실시형태에서, 면역 세포는 포유동물 세포이다. 일부 실시형태에서, 포유동물 세포는 인간 또는 설치류 세포이다. 일부 실시형태에서, 면역 세포는 세포독성 T 세포, 조절 T 세포, T 헬퍼 세포, 수지상 세포, B 세포 또는 NK 세포 또는 이들의 선조세포이다. 일부 실시형태에서, 선조세포는 조혈 줄기 세포이다.

[0012] 일 양태에서, 본 발명은 유효량의 본 명세서에 제공된 임의의 조성물 및 약제학적으로 허용 가능한 부형제를 포함하는 약제학적 조성물을 제공한다.

[0013] 또 다른 양태에서, 본 발명은 핵산 프로그램 가능 DNA 결합 단백질(nucleic acid programmable DNA binding protein: napDNAbp)과 데아미네이스 도메인을 포함하는 융합 단백질 및 상이한 항원을 각각 표적화하는 적어도 2개의 가이드 폴리뉴클레오타이드를 포함하는 염기 편집 시스템을 제공하되, 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택된다. 일부 실시형태에서, 하나의 가이드 폴리뉴클레오타이드는 CD5를 표적으로 하고, 또 다른 하나는 CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 한다. 일부 실시형태에서, 각각이 항원 CD3, CD5 및 CD7 중 하나를 표적으로 하는 3개의 가이드 폴리뉴클레오타이드를 포함하는 시스템. 일부 실시형태에서, 하나의 가이드 폴리뉴클레오타이드는 CD7을 표적으로 하고, 또 다른 하나는 CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 한다. 일부 실시형태에서, 하나의 가이드 폴리뉴클레오타이드는 CD3을 표적으로 하고, 또 다른 하나는 CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 한다. 일부 실시형태에서, 하나의 가이드 폴리뉴클레오타이드는 CD33을 표적으로 하고, 또 다른 하나는 CD123을 표적으로 한다.

[0014] 일부 실시형태에서, 가이드 폴리뉴클레오타이드 각각은 표 26으로부터 선택되는 핵산 서열을 포함한다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 AGCGACUGCAGAAAGAAGAG 또는 CAUACCAGCUGAGCCGUCCG로부터 선택되는 핵산 서열을 포함한다. 일부 실시형태에서, 융합 단백질은 하나 이상의 우라실 글리코실레이스 저해제(uracil glycosylase inhibitor: UGI) 및/또는 하나 이상의 핵 국부화 서열(nuclear localization sequence: NLS)을 더 포함한다. 일부 실시형태에서, napDNAbp는 Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 또는 Cas12j/CasΦ 폴리펩타이드 또는 이들의 일부를 포함한다. 일부 실시형태에서, napDNAbp는 Cas12 폴리펩타이드 또는 이의 단편을 포함한다. 일부 실시형태에서, napDNAbp는 Cas9 폴리펩타이드 또는 이의 단편을 포함한다. 일부 실시형태에서, Cas9는 활성이 없는 Cas9(dead Cas9: dCas9) 또는 Cas9 닉케이스(Cas9 nickase: nCas9)이다. 일부 실시형태에서, Cas9는 변형된 스타필로코커스 아우레우스(*Staphylococcus aureus*) Cas9(SaCas9), 스트렙토코커스 써모필루스(*Streptococcus thermophilus*) 1 Cas9(St1Cas9) 또는 변형된 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9(SpCas9)이다. 일부 실시형태에서, Cas9는 변경된 프로토스페이서-인접 모티프(protospacer-adjacent motif: PAM) 특이성을 포함한다. 일부 실시형태에서, 변경된 PAM은 핵산 서열 5'-NGC-3'에 대해 특이성을 갖는다.

[0015] 일부 실시형태에서, 데아미네이스 도메인은 사이티딘 또는 아데노신을 탈아미노화할 수 있다. 일부 실시형태에서, 데아미네이스 도메인은 사이티딘 또는 아데노신 데아미네이스 도메인이다. 일부 실시형태에서, 사이티딘 데

아미네이스는 APOBEC 데아미네이스이다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 변이체이다. 일부 실시형태에서, TadA 변이체는 TadA*8 변이체이다. 일부 실시형태에서, TadA*8 변이체는 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23 또는 TadA*8.24이다. 일부 실시형태에서, TadA 변이체는 TadA*9 변이체이다.

[0016] 일 양태에서, 본 발명은 본 명세서에 제공된 바와 같은 임의의 염기 편집 시스템을 포함하는 약제학적 조성물을 제공한다.

[0017] 또 다른 양태에서, 본 발명은 본 명세서에 제공된 바와 같은 임의의 염기 편집 시스템을 암호화하는 폴리뉴클레오타이드 및 가이드 폴리뉴클레오타이드를 제공한다.

[0018] 또 다른 양태에서, 본 발명은 본 명세서에 제공된 바와 같은 임의의 폴리뉴클레오타이드를 포함하는 벡터를 제공한다. 일부 실시형태에서, 벡터는 바이러스 벡터이다. 일부 실시형태에서, 바이러스 벡터는 레트로바이러스 벡터, 아데노바이러스 벡터, 렌티바이러스 벡터, 헤르페스바이러스 벡터 또는 아데노-관련 바이러스 벡터(adeno-associated viral vector: AAV)이다.

[0019] 일 양태에서, 본 발명은 본 명세서에 제공된 바와 같은 임의의 폴리뉴클레오타이드 또는 임의의 벡터를 포함하는 약제학적 조성물을 제공한다.

[0020] 일 양태에서, 본 발명은 면역원성이 감소된 CAR-발현 면역 세포를 생산하는 방법을 제공한다. 일부 실시형태에서, 방법은 CAR-발현 면역 세포를 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)과 데아미네이스 도메인을 포함하는 융합 단백질 및 각각 상이한 항원을 암호화하는 폴리뉴클레오타이드를 표적화하는 2개의 가이드 폴리뉴클레오타이드를 포함하는 염기 편집 시스템에서 발현시키는 단계를 포함하되, 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되어 면역원성이 감소된 CAR-발현 면역 세포를 생성한다. 일부 실시형태에서, 면역 세포는 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 표적으로 하는 가이드 폴리뉴클레오타이드를 발현하거나 또는 이와 접촉된다.

[0021] 일부 실시형태에서, 방법은 (a) CAR-발현 면역 세포를 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)과 데아미네이스 도메인을 포함하는 융합 단백질을 포함하는 염기 편집 시스템에서 발현시키는 단계; 및 (b) CAR-발현 면역 세포를 적어도 각각 상이한 항원을 암호화하는 폴리뉴클레오타이드를 표적화하는 2개의 가이드 폴리뉴클레오타이드와 접촉시키는 단계를 포함하되, 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되어 면역원성이 감소된 CAR-발현 면역 세포를 생성한다.

[0022] 일부 실시형태에서, 방법은 CAR-발현 면역 세포를 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 표적으로 하는 가이드 폴리뉴클레오타이드와 접촉시키는 단계를 더 포함한다. 일부 실시형태에서, 면역 세포는 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 하는 키메라 항원 수용체(CAR)를 발현한다. 일부 실시형태에서, 방법은 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 돌연변이를 면역 세포에 도입하여 면역원성이 감소된 CAR-발현 면역 세포의 집단을 생성하는 단계를 더 포함한다.

[0023] 일 양태에서, 본 발명은 면역원성이 감소된 CAR-발현 면역 세포의 집단을 생성하는 방법을 제공한다. 일부 실시형태에서, 방법은 a) CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원의 발현을 감소시키거나 또는 제거하는 돌연변이를 하나의 면역 세포에 도입하고, 상기 항원 중 하나의 상이한 돌연변이를 제2 면역 세포에 도입하는 단계; 및 b) TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 돌연변이를 면역 세포에 도입하여 면역원성이 감소된 CAR-발현 면역 세포의 집단을 생성하는 단계를 포함한다. 일부 실시형태에서, 면역 세포에 의해 발현된 키메라 항원 수용체는 CD3, CD5, CD7, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 표적으로 한다. 일부 실시형태에서, 방법에 의해 생성된 면역 세포는 CD3, CD5 및/또는 CD7 항원을 표적으로 하고, CD3, CD5 및/또는 CD7 항원을 발현하지 않거나 감소된 수준의 이들을 발현하는 키메라 항원 수용체를 발현한다. 일부 실시형태에서, 방법에 의해 생성된 면역 세포는 CD33 및 CD123 항원을 표적으로 하고, CD33 및 CD123 항원을 발현하지 않거나 감소된 수준의 이들을 발현하는 키메라 항원 수용체를 발현한다. 일부 실시형태

에서, CAR은 CD5 키메라 항원 수용체(CAR)이다. 일부 실시형태에서, CD5 CAR은 표 28에 제시된 CD5 CAR 작제물에 의해 암호화된다.

[0024] 또 다른 양태에서, 본 발명은 면역원성이 감소된 면역 세포를 생산하는 방법을 제공한다. 일부 실시형태에서, 방법은 a) CD5의 발현을 감소시키거나 또는 제거하는 내인성 CD5 유전자 서열 또는 조절 요소에 돌연변이를 도입하는 단계; 및 b) 세포에서 표 28에 제시된 CD5 CAR 작제물을 발현시키는 단계를 포함한다. 일부 실시형태에서, CD5 CAR 작제물은 다음으로부터 선택되는 아미노산 서열을 포함하거나 또는 이로 구성되는 CD5 CAR 폴리펩타이드를 암호화한다:

a)

```
 1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL 300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ 360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKFWVLVVVGGVLACYSLLVTVAFIIFWV 420
421 RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQ 480
481 LYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGE 540
541 RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR 573;
```

[0025]

b)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA 300
301 AGGAVHTRGLDFACDFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPR 360
361 RPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDK 420
421 RRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTAT 480
481 KDTYDALHMQALPPR 495;
```

[0026]

c)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAIEVMYPPPYLDNEKSNGTIIHVKGKHLCPS 300
301 PLFPGPSKPFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTR 360
361 KHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDP 420
421 EMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDA 480
481 LHMQALPPR 489;
```

[0027]

d)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL 300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ 360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKPTTTPAPRPPTPAPTIASQPLSLRPEA 420
421 CRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMR 480
481 PVQTTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVL 540
541 DKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLST 600
601 ATKDTYDALHMQALPPR 617; 또는
```

e)

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA 300
301 AGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQT 360
361 TQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRR 420
421 DPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKD 480
481 DALHMQALPPRX 494.
```

[0028]

[0029] 일부 실시형태에서, 방법에 의해 생성된 면역 세포는 상응하는 대조군 세포와 비교하여 동족 살해 저항성 및/또는 증가된 항-신생물 활성을 나타낸다. 일부 실시형태에서, 방법은 생체내 또는 생체외에서 수행된다. 일부 실시형태에서, 방법에 의해 생성된 면역 세포는 검출 가능한 전위를 포함하지 않는다. 일부 실시형태에서, 방법에 의해 생성된 면역 세포는 1% 미만의 indel을 포함한다. 일부 실시형태에서, 방법에 의해 생성된 면역 세포는 5% 미만의 비-표적 편집을 포함한다. 일부 실시형태에서, 방법에 의해 생성된 면역 세포는 5% 미만의 표적외 편집을 포함한다. 일부 실시형태에서, 돌연변이는 핵염기 변형에 의해 생성된다. 일부 실시형태에서, 돌연변이는 엑손에 있다. 일부 실시형태에서, 돌연변이는 단백질 발현을 감소시키거나 또는 제거하는 조기 종결 코돈을 생성한다. 일부 실시형태에서, 돌연변이는 스플라이스 공여체 부위 또는 스플라이스 수용체 부위에 있다. 일부 실시형태에서, 하나 이상의 돌연변이는 표적 폴리뉴클레오타이드를 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp), 데아미네이스를 포함하는 융합 단백질 및 하나 이상의 가이드 폴리뉴클레오타이드를 포함하는 염기 편집 시스템과 접촉시킴으로써 생성된다.

[0030] 일부 실시형태에서, 데아미네이스는 아데노신 또는 사이티딘 데아미네이스이다. 일부 실시형태에서, 사이티딘 데아미네이스는 BE4이다. 일부 실시형태에서, 돌연변이는 돌연변이가 없는 상응하는 대조군 세포에 비해 적어도 약 50% 이상 암호화된 폴리펩타이드의 발현을 감소시킨다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 표 26에 제공된 것들로부터 선택되는 서열을 포함한다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 AGCGACUGCAGAAAGAAGAG 또는 CAUACCAGCUGAGCCGUCCG로부터 선택되는 핵산 서열을 포함한다. 일부 실시형태에서, 하나 이상의 가이드 핵산 서열 각각은 *CD5, FAS, LAG-3, CD52, TRAC, B2M, CIITA, TRBC1, TRBC2* 및/또는 *PDC1/PD-1* 유전자 또는 조절 요소에 대한 napDNAbp를 표적으로 한다. 일부 실시형태에서, 염기 편집기 및 하나 이상의 가이드 핵산 서열은 전기천공법, 뉴클레오펙션, 양이온성 지질-매개성 방법, 바이러스 형질도입 또는 이들의 조합을 통해 면역 세포에 도입된다. 일부 실시형태에서, 방법은 면역 세포의 집단을 생성하기 위해 배양에서 면역 세포를 확장시키는 단계를 더 포함한다. 일부 실시형태에서, 항원 또는 폴리펩타이드의 발현은 면역 세포의 집단의 적어도 약 50%에서 감소된다. 일부 실시형태에서, 방법은 변형된 면역 세포의 집단으로부터 TCRα/β+ 세포를 고갈시키는 단계를 더 포함한다.

[0031] 일 양태에서, 본 발명은 본 명세서에 제공된 바와 같은 임의의 방법에 의해 생성된 면역원성이 감소된 CAR-발현 면역 세포를 제공한다.

[0032] 또 다른 양태에서, 본 발명은 본 명세서에 제공된 바와 같은 임의의 면역 세포를 포함하는 약제학적 조성물을 제공한다. 일부 실시형태에서, 약제학적 조성물은 약제학적으로 허용 가능한 부형제를 포함한다.

[0033] 또 다른 양태에서, 본 발명은 신생물 세포를 죽이는 방법을 제공한다. 일부 실시형태에서, 방법은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 발현하는 신생물 세포를 2종 이상의 면역 세포와 접촉시키는 단계를 포함하며, 각 면역 세포는 세포에 의해 발현된 상기 항원 중 2개를 표적으로 하는 상이한 키메라 항원 수용체를 발현하되, 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함한다. 일부 실시형태에서, 방법은 시험관내 또는 생체내에서 수행된다. 일부 실시형태에서, 신생물 세포는 신생물로부터 유래된다.

[0034] 일 양태에서, 본 발명은 대상체에서 신생물을 치료하는 방법을 제공한다. 일부 실시형태에서, 방법은 2종 이상의 면역 세포를 대상체에게 투여하는 단계를 포함하며, 각 면역 세포는 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 대상체의 신생물 세포에 의해 발현된 항원을 표적으로 하는 상이한 키메라 항원 수용체를 발현하되, 면역 세포는 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하고, 면역 세포 각각은 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 더 포함한다.

[0035] 일부 실시형태에서, 신생물은 혈액암이다. 일부 실시형태에서, 신생물은 액체암이다. 일부 실시형태에서, 혈액암은 백혈병, 골수종 및/또는 림프종이다. 일부 실시형태에서, 혈액암은 B 세포암이다. 일부 실시형태에서, 혈액암은 다음의 T-세포 급성 림프아구성 백혈병(T-cell acute lymphocytic leukemia: T-ALL), 용상 진균증(mycosis fungoides: MF), 세자리 증후군(Sézary syndrome: SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 $ALK^+$, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 $CD30^+$ 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종 중 적어도 하나로부터 선택된다. 일부 실시형태에서, 혈액암은 T-세포 급성 림프아구성 백혈병(T-ALL) 세포이다. 일부 실시형태에서, 혈액암은 급성 골수성 백혈병(acute myelogenous leukemia: AML)이다.

[0036] 또 다른 양태에서, 본 발명은 선택된 대상체에서 신생물을 치료하는 방법을 제공한다. 일부 실시형태에서, 방법은 2종 이상의 면역 세포를 선택된 대상체에게 투여하는 단계를 포함하며, 각 면역 세포는 대상체의 신생물 세포에 의해 발현된 항원을 표적으로 하는 상이한 키메라 항원 수용체를 발현하되, 항원은 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되고, 면역 세포는 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하며, 면역 세포 각각은 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 더 포함하고, 대상체는 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 발현하는 신생물을 갖는 것으로 선택된다. 일부 실시형태에서, 혈액암은 T-세포 급성 림프아구성 백혈병(T-ALL) 세포이다. 일부 실시형태에서, 혈액암은 급성 골수성 백혈병(AML)이다. 일부 실시형태에서, 상이한 키메라 항원 수용체를 발현하는 2종 이상의 면역 세포는 순차적으로 투여된다. 일부 실시형태에서, 상이한 키메라 항원 수용체를 발현하는 2종 이상의 면역 세포는 동시에 투여된다.

[0037] 또 다른 양태에서, 본 발명은 대상체에서 신생물 세포의 항원-의존적 살해 방법을 제공한다. 일부 실시형태에서, 방법은 2종 이상의 면역 세포를 CD5, CD7, CD3, CD33 및 CD123으로 이루어진 군으로부터 선택되는 항원을 발현하는 신생물을 갖는 대상체에게 투여하는 단계를 포함하며, 각 면역 세포는 상기 항원 중 하나를 표적으로 하는 상이한 키메라 항원 수용체를 발현하되, 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함한다.

[0038] 일 양태에서, 본 발명은 대상체에서의 급성 골수성 백혈병(AML) 세포의 항원-의존적 살해 방법을 제공한다. 일부 실시형태에서, 방법은 2종 이상의 면역 세포를 CD33 및 CD123 항원을 발현하는 AML을 갖는 대상체에게 투여하는 단계를 포함하며, 각 면역 세포는 상기 항원 중 하나를 표적으로 하는 상이한 키메라 항원 수용체를 발현하되, 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시

키거나 또는 제거하는 하나 이상의 돌연변이를 포함한다.

[0039] 일 양태에서, 본 발명은 대상체에서의 T-세포 급성 림프아구성 백혈병(T-ALL) 세포의 항원-의존적 살해 방법을 제공한다. 일부 실시형태에서, 방법은 적어도 3종의 면역 세포를 CD3, CD5 및 CD7 항원을 발현하는 T-ALL을 갖는 대상체에게 투여하는 단계를 포함하며, 각 면역 세포는 상기 항원 중 하나를 표적으로 하는 상이한 키메라 항원 수용체를 발현하되, 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함한다.

[0040] 또 다른 양태에서, 본 발명은 선택된 대상체에서 암을 치료하는 방법을 제공한다. 일부 실시형태에서, 방법은 적어도 2종의 면역 세포를 대상체에게 투여하는 단계를 포함하며, 각 면역 세포는 CD33 또는 CD123 항원을 표적으로 하는 키메라 항원 수용체를 발현하되, 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하고, 대상체는 암이 CD33 및 CD123 항원을 발현하는 것으로 특성화함으로써 선택된다. 일부 실시형태에서, CD33 및 CD123 항원을 발현하는 암은 AML이다.

[0041] 또 다른 양태에서, 본 발명은 선택된 대상체에서 암을 치료하는 방법을 제공한다. 일부 실시형태에서, 방법은 3종 이상의 면역 세포를 대상체에게 투여하는 단계를 포함하며, 각 면역 세포는 CD3, CD5 및 CD7 항원을 표적으로 하는 상이한 키메라 항원 수용체를 발현하되, 면역 세포는 표적화된 항원의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1로 이루어진 군으로부터 선택되는 폴리펩타이드의 발현을 감소시키거나 또는 제거하는 하나 이상의 돌연변이를 포함하고, 대상체는 암이 CD3, CD5 및 CD7 항원을 발현하는 것으로 특성화함으로써 선택된다. 일부 실시형태에서, CD3, CD5 및 CD7 항원을 발현하는 암은 T-ALL이다.

[0042] 일부 실시형태에서, 면역 세포는 세포독성 T 세포, 조절 T 세포, T 헬퍼 세포, 수지상 세포, B 세포 또는 NK 세포이다. 일부 실시형태에서, 대상체는 이전에 림프제거(lymphodepletion)로 치료되었다. 일부 실시형태에서, 림프제거는 사이클로포스파마이드(cyclophosphamide), 플루다라빈(fludarabine) 및/또는 알렘투주맙(alemtuzumab)(Cy/Flu/캄파스(Campath))의 투여를 수반한다. 일부 실시형태에서, 대상체는 화학요법에 불응성(refractory)이거나 또는 높은 종양 부담을 갖는다. 일부 실시형태에서, 대상체는 후속적으로 동종이계 조혈 줄기 세포 이식(allogeneic hematopoietic stem cell transplantation: allo-HSCT)으로 치료된다. 일부 실시형태에서, 면역 세포는 단일 인간 공여체로부터 유래된다. 일부 실시형태에서, 면역 세포는 대상체에 대해 자가 조직이다. 일부 실시형태에서, 면역 세포는 대상체에 대해 동종이계이다. 일부 실시형태에서, 대상체는 포유동물 대상체이다. 일부 실시형태에서, 대상체는 인간 또는 설치류 대상체이다. 일부 실시형태에서, 대상체는 소아 인간 대상체이다.

[0043] 일 양태에서, 본 발명은 암의 치료에 사용하기 위한 본 명세서에 제공된 임의의 조성물을 포함하는 키트를 제공한다. 또 다른 양태에서, 본 발명은 면역원성이 감소된 CAR-발현 면역 세포를 생성하는데 사용하기 위한 본 명세서에 제공된 바와 같은 임의의 염기 편집 시스템을 포함하는 키트를 제공한다. 일부 실시형태에서, 본 명세서에 제공된 바와 같은 임의의 키트는 키트의 사용을 위한 서면 설명서를 포함한다.

[0044] 본 명세서의 설명 및 실시예는 본 개시내용의 실시형태를 상세히 예시한다. 본 개시내용은 본 명세서에 기술된 특정 실시형태로 제한되지 않으며 그에 따라 달라질 수 있음을 이해하여야 한다. 당업자는 본 개시내용의 다양한 변형 및 수정이 있음을 인식할 것이며, 이는 그 범주 내에 포함된다.

[0045] 본 명세서에 개시된 일부 실시형태의 실시는, 달리 지시되지 않는 한, 당업계의 기술 범위 내에 있는 면역학, 생화학, 화학, 분자 생물학, 미생물학, 세포 생물학, 게놈 및 재조합 DNA의 종래의 기법을 사용한다. 예를 들어, 문헌[Sambrook and Green, Molecular Cloning: A Laboratory Manual, 4th Edition (2012); the series Current Protocols in Molecular Biology (F. M. Ausubel, *et al.* eds.); the series Methods In Enzymology (Academic Press, Inc.), PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual, and Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition (R.I. Freshney, ed. (2010)] 참조.

[0046] 본 개시내용의 다양한 특징이 단일 실시형태의 맥락에서 기술될 수 있지만, 특징은 또한 별도로 또는 임의의 적합한 조합으로 제공될 수 있다. 반대로, 본 개시내용은 명료함을 위해 별도의 실시형태의 맥락에서 본 명세서에

기술될 수 있지만, 본 개시내용은 또한 단일 실시형태로 구현될 수 있다. 본 명세서에서 사용된 부문 제목은 구조적 목적만을 위한 것이며 기재된 주제를 제한하는 것으로 해석되어서는 안된다.

[0047] 본 개시내용의 특징은 첨부된 청구범위에 구체적으로 설명된다. 본 발명의 특징 및 이점에 대한 더 나은 이해는 본 개시내용의 원리가 이용되는 예시적인 실시형태를 설명하는 다음의 상세한 설명을 참조하고, 아래에서 설명되는 첨부 도면을 참조하여 얻어질 것이다.

[0048] **정의**

[0049] 다음 정의는 해당 기술 분야의 정의를 보완하며 현재 출원에 대한 것이며, 관련되거나 관련되지 않은 사례, 예를 들어, 임의의 공동 소유의 특허 또는 출원에 귀속되어서는 안된다. 본 명세서에 기재된 것과 유사하거나 또는 동등한 임의의 방법 및 물질이 본 개시내용의 테스트를 위한 실시에 사용될 수 있지만, 바람직한 물질 및 방법이 본 명세서에 기재되어 있다. 따라서, 본 명세서에서 사용되는 용어는 단지 특정 실시형태를 설명하기 위해 사용된 것으로 한정하려는 의도가 아니다.

[0050] 달리 정의되지 않는 한, 본 명세서에서 사용되는 모든 기술 및 과학 용어는 본 발명이 속하는 기술 분야에서 당업자에 의해 일반적으로 이해되는 의미를 갖는다. 다음 참고 문헌은 본 발명에 사용된 많은 용어에 대한 일반적인 정의를 당업자에게 제공한다: 문헌[Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); 및 Hale & Marham, The Harper Collins Dictionary of Biology (1991)].

[0051] 본 출원에서, 단수의 사용은 달리 특별히 언급하지 않는 한 복수를 포함한다. 본 명세서에서 사용된 바와 같이, 단수 형태는 문맥에서 달리 명백하게 지시하지 않는 한, 복수의 참조를 포함한다는 점에 유의하여야 한다. 본 출원에서, "또는"의 사용은 달리 명시되지 않는 한 "및/또는"을 의미하며 포괄적인 것으로 이해된다. 또한, "포함하는"뿐만 아니라 "포함하다(include)", "포함하다(includes)" 및 "포함된(included)"과 같은 다른 형태의 사용은 제한적이지 않다.

[0052] 본 명세서 및 청구항(들)에서 사용되는 단어 "포함하는(comprising)"(및 "포함하다(comprise)" 및 "포함하다(comprises)"와 같은 포함하는의 임의의 형태), "가지는(having)"(및 "가지다(have)" 및 "가지다(has)"와 같은 가지는의 임의의 형태), "포함하는(including)"(및 "포함하다(includes)" 및 "포함하다(include)"와 같은 포함하는의 임의의 형태) 또는 "함유하는(containing)"(및 "함유하다(contains)" 및 "함유하다(contain)"과 같은 함유하는의 임의의 형태)은 포괄적이거나 또는 개방형(open-ended)이며, 추가적인 인용되지 않은 요소 또는 방법 단계를 배제하지 않는다. 본 명세서에서 논의된 임의의 실시형태는 본 개시내용의 임의의 방법 또는 조성물과 관련하여 구현될 수 있으며 그 반대도 마찬가지인 것으로 상정된다. 또한, 본 개시내용의 조성물은 본 개시내용의 방법을 달성하기 위해 사용될 수 있다.

[0053] 용어 "약" 또는 "대략"은 당업자에 의해 결정된 바와 같은 특정 값에 대해 허용 가능한 오차 범위 내를 의미하며, 이는 값이 측정 또는 결정되는 방식, 즉, 측정 시스템의 한계에 부분적으로 의존할 것이다. 예를 들어, "약"은 당업계의 관행에 따라 하나(1) 또는 하나(1) 이상의 표준 편차 이내를 의미할 수 있다. 대안적으로, "약"은 주어진 값의 최대 20%, 최대 10%, 최대 5% 또는 최대 1%의 범위를 의미할 수 있다. 대안적으로, 특히 생물학적 시스템 또는 과정과 관련하여, 용어는 값의 5-배 이내 또는 2-배 이내와 같은 10배 이내를 의미할 수 있다. 특정 값이 출원 및 청구범위에 기술되어 있는 경우, 달리 언급되지 않는 한, 용어 "약"은 특정 값에 대해 허용 가능한 오차 범위 내를 의미하는 것으로 가정하여야 한다.

[0054] 본 명세서에 제공된 범위는 범위 내의 모든 값에 대한 약칭으로 이해된다. 예를 들어, 1 내지 50의 범위는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50으로 이루어진 군으로부터의 임의의 숫자, 숫자의 조합 또는 하위 범위를 포함하는 것으로 이해된다.

[0055] 명세서에서 "일부 실시형태", "실시형태", "하나의 실시형태" 또는 "다른 실시형태"에 대한 참조는 실시형태와 관련하여 기재된 특정 특징, 구조 또는 특성이 적어도 일부 실시형태에 포함된다는 것을 의미하지만, 본 개시내용의 모든 실시형태가 반드시 그런 것은 아니다.

[0056] "아데노신 데아미네이스"는 아데닌 또는 아데노신의 가수분해성 탈아미노화를 촉매할 수 있는 폴리펩타이드 또는 이의 단편을 의미한다. 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 아데노신에서 이노신으로 또는 데옥시아데노신에서 데옥시이노신으로의 가수분해성 탈아미노화를 촉매하는 아데노신 데아미네이스이

다. 일부 실시형태에서, 아데노신 데아미네이스는 데옥시리보핵산(DNA)에서 아데닌 또는 아데노신의 가수분해성 탈아미노화를 촉매한다. 본 명세서에 제공된 아데노신 데아미네이스(예를 들어, 조작된 아데노신 데아미네이스, 진화된 아데노신 데아미네이스)는 세균과 같은 임의의 유기체로부터 유래할 수 있다.

[0057] 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스와 같은 유기체로부터의 자연-발생적 데아미네이스의 변이체이다. 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 자연에서 발생하지 않는다. 예를 들어, 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 자연-발생적 데아미네이스와 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75% 적어도 80%, 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 적어도 99.1%, 적어도 99.2%, 적어도 99.3%, 적어도 99.4%, 적어도 99.5%, 적어도 99.6%, 적어도 99.7%, 적어도 99.8% 또는 적어도 99.9% 동일하다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균, S. 아우레우스, B. 서브틸리스(*B. subtilis*), S. 티피(*S. typhi*), S. 푸트레파시엔스(*S. putrefaciens*), H. 인플루엔자에(*H. influenzae*), C. 크레센투스(*C. crescentus*) 또는 G. 설퍼레듀센스(*G. sulfurreducens*)와 같은 세균으로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 데아미네이스이다. 일부 실시형태에서, TadA 데아미네이스는 대장균 TadA(ecTadA) 데아미네이스 또는 이의 단편이다.

[0058] 일부 실시형태에서, ecTadA 데아미네이스는 절단된 ecTadA이다. 예를 들어, 절단된 ecTadA는 전장 ecTadA에 비해 하나 이상의 N-말단 아미노산이 누락될 수 있다. 일부 실시형태에서, 절단된 ecTadA는 전장 ecTadA에 비해 1개, 2개, 3개, 4개, 5 개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개 또는 20개의 N-말단 아미노산 잔기가 누락될 수 있다. 일부 실시형태에서, 절단된 ecTadA는 전장 ecTadA에 비해 1개, 2개, 3개, 4개, 5 개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개 또는 20개의 C-말단 아미노산 잔기가 누락될 수 있다. 일부 실시형태에서, ecTadA 데아미네이스는 N-말단 메티오닌을 포함하지 않는다. 일부 실시형태에서, TadA 데아미네이스는 N-말단 절단된 TadA이다. 특정 실시형태에서, TadA는 그 전문이 참조에 의해 본 명세서에 원용되어 있는 PCT/US2017/045381에 기재된 TadA 중 어느 하나이다.

[0059] 일부 실시형태에서, TadA 데아미네이스는 TadA 변이체이다. 일부 실시형태에서, TadA 변이체는 TadA*7.10이다. 일부 실시형태에서, TadA 변이체는 TadA*8이다. 일부 실시형태에서, TadA*8은 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23 또는 TadA*8.24이다. 일부 실시형태에서, TadA*8은 TadA*8a, TadA*8b, TadA*8c, TadA*8d 또는 TadA*8e이다. 일부 실시형태에서, TadA*8은 TadA*8e이다. 일부 실시형태에서, TadA 변이체는 TadA*9이다.

[0060] "아데노신 데아미네이스 염기 편집기 8(ABE8) 폴리펩타이드" 또는 "ABE8"은 다음 참조 서열의 82번 및/또는 166번 아미노산 위치에 변경을 포함하는 아데노신 데아미네이스 변이체를 포함하는 본 명세서에 정의된 바와 같은 염기 편집기를 의미한다:

[0061]
```
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD
```

[0062] 일부 실시형태에서, ABE8은 참조 서열에 대해 본 명세서에 기재된 바와 같은 추가 변경을 포함한다.

[0063] "아데노신 데아미네이스 염기 편집기 8(ABE8) 폴리뉴클레오타이드"는 ABE8을 암호화하는 폴리뉴클레오타이드를 의미한다.

[0064] "아데노신 데아미네이스 염기 편집기 9(ABE9) 폴리펩타이드" 또는 "ABE9"는 다음 참조 서열에서 다음 변경 중 하나 이상을 포함하는 아데노신 데아미네이스 변이체를 포함하는 본 명세서에 정의된 바와 같은 염기 편집기를 의미한다: R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R 및 A158K:

```
        10         20         30         40         50
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
        60         70         80         90        100
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
       110        120        130        140        150
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
       160
MPRQVFNAQK KAQSSTD
```

[0065]

[0066] 참조 서열에서 변경된 관련 염기는 밑줄과 굵은 글꼴로 표시된다. 일부 실시형태에서, ABE9는 참조 서열에 대해 본 명세서에 기재된 바와 같은 추가 변경을 포함한다. ABE9 염기 편집기에 대한 세부사항은 그 전문이 참조에 의해 본 명세서에 원용되어 있는 PCT 출원 PCT/2020/049975에 기술되어 있다.

[0067] "아데노신 데아미네이스 염기 편집기 9(ABE9) 폴리뉴클레오타이드"는 ABE9를 암호화하는 폴리뉴클레오타이드를 의미한다.

[0068] "투여하는"은 본 명세서에 기재된 하나 이상의 조성물을 환자 또는 대상체에게 제공하는 것으로 본 명세서에서 지칭된다. 예로서 그리고 제한 없이, 조성물 투여, 예를 들어, 주사는 정맥내(i.v.) 주사, 피하(s.c.) 주사, 피내(i.d.) 주사, 복강내(i.p.) 주사 또는 근육내(i.m.) 주사에 의해 수행될 수 있다. 하나 이상의 이러한 경로가 이용될 수 있다. 비경구 투여는, 예를 들어, 볼루스 주사 또는 시간 경과에 따른 점진적 관류에 의한 것일 수 있다. 일부 실시형태에서, 비경구 투여는 혈관내로, 정맥내로, 근육내로, 동맥내로, 척추강내로, 종양내로, 피내로, 복강내로, 기관내로, 피하로, 표피하로, 관절내로, 피막하로, 지주막하로 및 흉골내로 주입 또는 주사하는 것을 포함한다. 대안적으로 또는 동시에, 투여는 경구 경로에 의한 것일 수 있다.

[0069] "작용제"는 임의의 소분자 화학 화합물, 항체, 핵산 분자 또는 폴리펩타이드 또는 이들의 단편을 의미한다

[0070] "변경"은 본 명세서에 기재된 것들과 같은 당업계의 표준 공지 방법에 의해 검출되는 바와 같은 유전자 또는 폴리펩타이드의 구조, 발현 수준 또는 활성의 변화(예를 들어, 증가 또는 감소)를 의미한다. 본 명세서에서 사용되는 바와 같이, 변경(예를 들어, 증가 또는 감소)은 폴리뉴클레오타이드 또는 폴리펩타이드 서열의 변화 또는 발현 수준의 변화, 예컨대, 10%의 변화, 25%의 변화, 40%의 변화 및 50%의 변화 또는 그 이상을 포함한다.

[0071] 본 명세서에서 사용되는 "동종이계"는 비교시 세포와 유전적으로 상이한 동일한 종의 세포를 지칭한다.

[0072] "개선하다"는 질환의 발달 및 진행을 감소, 억제, 감쇠, 약화, 저지 또는 안정화시키는 것을 의미한다.

[0073] "유사체"는 동일하지 않지만 유사한 기능적 또는 구조적 특징을 갖는 분자를 의미한다. 예를 들어, 폴리뉴클레오타이드 또는 폴리펩타이드 유사체는 상응하는 자연-발생적 폴리뉴클레오타이드 또는 폴리펩타이드의 생물학적 활성을 유지하면서 자연 발생적 폴리뉴클레오타이드 또는 폴리펩타이드에 비해 유사체의 기능을 향상시키는 소정의 변형을 갖는다. 이러한 변형은, 예를 들어, 리간드 결합을 변경하지 않고도 DNA에 대한 유사체의 친화성, 효율성, 특이성, 프로테이스 또는 뉴클레이스 저항성, 막 투과성 및/또는 반감기를 증가시킬 수 있다. 유사체는 비천연 뉴클레오타이드 또는 아미노산을 포함할 수 있다.

[0074] "항-신생물 활성"은 신생물의 성숙 및/또는 증식을 방지 또는 저해하는 것을 의미한다.

[0075] 본 명세서에서 사용되는 "자가"는 동일한 대상체로부터의 세포를 지칭한다.

[0076] 본 명세서에서 사용되는 용어 "항체"는 특정 항원에 특이적으로 결합하거나 또는 면역학적으로 반응성인 면역글로불린 분자를 지칭하고, 키메라 항체, 인간화된 항체, 이종접합체 항체(예를 들어, 이중-, 삼중- 및 사중-특이성 항체, 다이어바디, 트라이어바디 및 테트라바디) 및, 예를 들어, Fab', F(ab')2, Fab, Fv, rIgG 및 scFv 단편을 포함하는 항체의 항원 결합 단편을 포함하지만 이들로 제한되지 않는 다클론, 단일클론, 유전적으로 조작된 및 그렇지 않으면 변형된 형태의 항체를 포함한다. 달리 지시되지 않는 한, 용어 "단일클론 항체"(monoclonal antibody; mAb)는 온전한 분자뿐만 아니라 표적 단백질에 특이적으로 결합할 수 있는 항체 단편(예를 들어, Fab 및 F(ab')2 단편 포함)을 둘 다 포함하는 것을 의미한다. 본 명세서에서 사용되는 바와 같이, Fab 및 F(ab')2 단편은 온전한 항체의 Fc 단편이 결여되어 있는 항체 단편을 지칭한다. 이러한 항체 단편의 예는 본 명세서에 기재되어 있다.

[0077] "B 세포 성숙 항원(B cell maturation antigen: BCMA) 또는 종양 괴사 인자 수용체 패밀리 구성원 17 폴리펩타이드(tumor necrosis factor receptor superfamily member 17 polypeptide)"는 성숙 B 림프구에서 발현되는 NCBI 등록 번호 NP_001183 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. 예시적인 BCMA 폴리펩타이드 서열이 아래에 제공된다.

[0078] >NP_001183.2 종양 괴사 인자 수용체 패밀리 구성원 17 [호모 사피엔스]

[0079]
```
MLQMAGQCSQNEYFDSLLHACIPCQLRCSSNTPPLTCQRYCNASVTNSVKGTNAILWTCLGLSL
IISLAVFVLMFLLRKINSEPLKDEFKNTGSGLLGMANIDLEKSRTGDEIILPRGLEYTVEECTC
EDCIKSKPKVDSDHCFPLPAMEEGATILVTTKTNDYCKSLPAALSATEIEKSISAR
```

[0080] 이 항원은 재발성 또는 불응성 다발성 골수종 및 다른 혈액학적 신생물 요법에서 표적화될 수 있다.

[0081] "B 세포 성숙 항원 또는 종양 괴사 인자 수용체 패밀리 구성원 17(BCMA) 폴리뉴클레오타이드"는 BCMA 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. BCMA 유전자는 B 세포 활성화 인자를 인식하는 세포 표면 수용체를 암호화한다. 예시적인 B2M 폴리뉴클레오타이드 서열이 아래에 제공된다.

[0082] >NM_001192.2 호모 사피엔스 TNF 수용체 슈퍼패밀리 구성원 17(TNFRSF17), mRNA

[0083]
```
AAGACTCAAACTTAGAAACTTGAATTAGATGTGGTATTCAAATCCTTAGCTGCCGCGAAGACAC
AGACAGCCCCCGTAAGAACCCACGAAGCAGGCGAAGTTCATTGTTCTCAACATTCTAGCTGCTC
TTGCTGCATTTGCTCTGGAATTCTTGTAGAGATATTACTTGTCCTTCCAGGCTGTTCTTTCTGT
AGCTCCCTTGTTTTCTTTTTGTGATCATGTTGCAGATGGCTGGGCAGTGCTCCCAAAATGAATA
TTTTGACAGTTTGTTGCATGCTTGCATACCTTGTCAACTTCGATGTTCTTCTAATACTCCTCCT
CTAACATGTCAGCGTTATTGTAATGCAAGTGTGACCAATTCAGTGAAAGGAACGAATGCGATTC
TCTGGACCTGTTTGGGACTGAGCTTAATAATTTCTTTGGCAGTTTTCGTGCTAATGTTTTTGCT
AAGGAAGATAAACTCTGAACCATTAAAGGACGAGTTTAAAAACACAGGATCAGGTCTCCTGGGC
ATGGCTAACATTGACCTGGAAAAGAGCAGGACTGGTGATGAAATTATTCTTCCGAGAGGCCTCG
AGTACACGGTGGAAGAATGCACCTGTGAAGACTGCATCAAGAGCAAACCGAAGGTCGACTCTGA
CCATTGCTTTCCACTCCCAGCTATGGAGGAAGGCGCAACCATTCTTGTCACCACGAAAACGAAT
GACTATTGCAAGAGCCTGCCAGCTGCTTTGAGTGCTACGGAGATAGAGAAATCAATTTCTGCTA
GGTAATTAACCATTTCGACTCGAGCAGTGCCACTTTAAAAATCTTTTGTCAGAATAGATGATGT
GTCAGATCTCTTTAGGATGACTGTATTTTTCAGTTGCCGATACAGCTTTTTGTCCTCTAACTGT
GGAAACTCTTTATGTTAGATATATTTCTCTAGGTTACTGTTGGGAGCTTAATGGTAGAAACTTC
CTTGGTTTCATGATTAAACTCTTTTTTTTCCTGA
```

[0084]

[0085] "염기 편집기(base editor: BE)" 또는 "핵염기 편집기(nucleobase editor: NBE)"는 폴리뉴클레오타이드에 결합하며 핵염기 변형 활성을 갖는 작용제를 의미한다. 일 실시형태에서, 작용제는 핵산 프로그램 가능 DNA 결합 단백질을 사용하여 특정 서열에서 폴리뉴클레오타이드에 결합한다. 또 다른 실시형태에서, 염기 편집기는 핵산 분자(예를 들어, DNA) 내의 사이티딘 염기를 변형시킬 수 있는 효소이다. 일부 실시형태에서, 염기 편집기는 핵산 분자 내의 염기를 탈아미노화할 수 있다. 일부 실시형태에서, 염기 편집기는 DNA 분자 내의 염기를 탈아미노화할 수 있다. 일부 실시형태에서, 염기 편집기는 DNA의 사이티딘을 탈아미노화할 수 있다. 일부 실시형태에서, 염기 편집기는 사이티딘 데아미네이스 또는 아데노신 데아미네이스를 포함하는 융합 단백질이다. 일부 실시형태에서, 염기 편집기는 사이티딘 데아미네이스 또는 아데노신 데아미네이스에 융합된 Cas9 단백질이다. 일부 실시형태에서, 염기 편집기는 사이티딘 데아미네이스 또는 아데노신 데아미네이스에 융합된 Cas9 닉케이스(nCas9)이다. 일부 실시형태에서, 염기 편집기는 염기 절단 복구(base excision repair), 예를 들어, UGI 도메인의 저해제에 융합된다. 일부 실시형태에서, 융합 단백질은 데아미네이스에 융합된 Cas9 닉케이스 및 UGI 도메인과 같은 염기 절단 복구 저해제를 포함한다.

[0086] 일부 실시형태에서, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 핵염기 편집기 폴리펩타이드는 다음 도메인 A-B: $NH_2$-[A-B]-COOH를 포함하되, A는 사이티딘 데아미네이스 도메인, 아데노신 데아미네이스 도메인 또는 이의 활성 단편을 포함하고, B는 핵산 서열 특이적 결합 활성을 갖는 하나 이상의 도메인을 포함한다. 일 실시

형태에서, 이전 양태의 사이티딘 또는 아데노신 데아미네이스 핵염기 편집기 폴리펩타이드는 $NH_2$-[$A_n$-$B_o$]-COOH를 포함하되, A는 사이티딘 데아미네이스 도메인, 아데노신 데아미네이스 도메인 또는 이의 활성 단편을 포함하고, n은 1, 2, 3, 4 또는 5의 정수이며; B는 핵산 서열 특이적 결합 활성을 갖는 도메인을 포함하고; o는 1, 2, 3, 4 또는 5의 정수이다. 일 실시형태에서, 폴리펩타이드는 하나 이상의 핵 국부화 서열을 포함한다. 일 실시형태에서, 폴리펩타이드는 N-말단 또는 C-말단에 있는 상기 핵 국부화 서열 중 적어도 하나를 포함한다. 일 실시형태에서, 폴리펩타이드는 이분(bipartite) 핵 국부화 신호인 핵 국부화 신호를 포함한다. 일 실시형태에서, 폴리펩타이드는 링커에 의해 연결되는 하나 이상의 도메인을 포함한다.

[0087] 일부 실시형태에서, 염기 편집기는 사이티딘 염기 편집기(cytidine base editor: CBE)이다. 일부 실시형태에서, 염기 편집기는 아데노신 염기 편집기(adenosine base editor: ABE)이다. 일부 실시형태에서, 염기 편집기는 아데노신 염기 편집기(ABE)와 사이티딘 염기 편집기(CBE)이다. 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스에 융합된 뉴클레이스-불활성 Cas9(dCas9)이다. 일부 실시형태에서, Cas9는 원형 순열(circular permutant) Cas9(예를 들어, spCas9 또는 saCas9)이다. 원형 순열 Cas9는 당업계에 공지되어 있고, 예를 들어, 문헌[Oakes *et al.*, Cell 176, 254-267, 2019]에 기술되어 있다. 일부 실시형태에서, 염기 편집기는 염기 절단 복구 저해제, 예를 들어, UGI 도메인 또는 dISN 도메인에 융합된다. 일부 실시형태에서, 융합 단백질은 데아미네이스 및 UGI 또는 dISN 도메인과 같은 염기 절단 복구 저해제에 융합된 Cas9 닉케이스를 포함한다. 다른 실시형태에서, 염기 편집기는 무염기성 염기 편집기이다.

[0088] 일부 실시형태에서, 아데노신 데아미네이스는 TadA로부터 진화된다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인은 CRISPR 관련(예를 들어, Cas 또는 Cpf1) 효소이다. 일부 실시형태에서, 염기 편집기는 데아미네이스 도메인에 융합된 촉매적으로 활성이 없는 Cas9(dCas9)이다. 일부 실시형태에서, 염기 편집기는 데아미네이스 도메인에 융합된 Cas9 닉케이스(nCas9)이다. 일부 실시형태에서, 염기 편집기는 염기 절단 복구 저해제(BER)에 융합된다. 일부 실시형태에서, 염기 절단 복구 저해제는 우라실 DNA 글리코실레이스 저해제(uracil DNA glycosylase inhibitor: UGI)이다. 일부 실시형태에서, 염기 절단 복구 저해제는 이노신 염기 절단 복구 저해제이다.

[0089] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 변이체(예를 들어, TadA*7.10)를 원형 순열 Cas9(예를 들어, spCAS9) 및 이분된 핵 국부화 서열을 포함하는 스캐폴드로 클로닝함으로써 생성된다. 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 변이체(예를 들어, TadA*8)를 원형 순열 Cas9(예를 들어, spCAS9 또는 saCAS9) 및 이분된 핵 국부화 서열을 포함하는 스캐폴드로 클로닝함으로써 생성된다(예를 들어, ABE8). 원형 순열 Cas9는 당업계에 공지되어 있고, 예를 들어, 문헌[Oakes *et al.*, Cell 176, 254-267, 2019]에 기술되어 있다.

[0090] 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인은 CRISPR 관련(예를 들어, Cas 또는 Cpf1) 효소이다. 일부 실시형태에서, 염기 편집기는 데아미네이스 도메인에 융합된 촉매적으로 활성이 없는 Cas9(dCas9)이다. 일부 실시형태에서, 염기 편집기는 데아미네이스 도메인에 융합된 Cas9 닉케이스(nCas9)이다. 일부 실시형태에서, 염기 편집기는 염기 절단 복구 저해제(BER)에 융합된다. 일부 실시형태에서, 염기 절단 복구 저해제는 우라실 DNA 글리코실레이스 저해제(UGI)이다. 일부 실시형태에서, 염기 절단 복구 저해제는 이노신 염기 절단 복구 저해제이다.

[0091] 염기 편집기에 대한 세부사항은 PCT 특허 PCT/2017/045381(WO 2018/027078) 및 PCT/US2016/058344(WO 2017/070632)에 기술되어 있으며, 이들 각각은 그 전문이 참조에 의해 본 명세서에 원용된다. 또한, 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), 및 Rees, H.A., *et al.*, "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1]을 참조한다.

[0092] 예로서, 본 명세서에 기재된 염기 편집 조성물, 시스템 및 방법에 사용되는 바와 같은 아데닌 염기 편집기(ABE)는 아래 제공되는 바와 같은 핵산 서열(8877개의 염기쌍)을 갖는다(애드진(Addgene), 매사추세츠주 워터타운

소재; 문헌[Gaudelli NM, *et al.*, Nature. 2017 Nov 23;551(7681):464-471. doi: 10.1038/nature24644; Koblan LW, *et al.*, Nat Biotechnol. 2018 Oct;36(9):843-846. doi: 10.1038/nbt.4172.]). ABE 핵산 서열에 대해 적어도 95% 이상의 동일성을 갖는 폴리뉴클레오타이드 서열이 또한 포함된다.

[0093]

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT
GACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG
TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCC
ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGT
CAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACA
GCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGT
ATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGT
GCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCA
CACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCC
TGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGT
GTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATG
AACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTA
GAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGG
AGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCC
GGAGGATCCTCTGAGGTGGAGTTTTCCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGG
CACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTG
GAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTG

GTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAGCCTTGCGTGATGTGCGCCG
GCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTTTGGCGTGAGGAACGCAAAAACCGGCGCCGCAGG
CTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCA
GATGAATGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGG
CCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGA
GAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGCCTGGCC
ATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGA
AACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGT
CCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGC
CTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGAC
CTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACC
TGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGA
GGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCC
TGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAG
CAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTT
CTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCA
AGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGC
TCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGG
ACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAA
CGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTAC
CCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGC
TGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACAT
ACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGA
AGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCC
GGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGG
CTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAA
GCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTA
AGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGAC

[0094]

[0095]

TCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAG
AGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTT
CGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACG
ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCG
GAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAAC
GCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACA
AGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTT
CTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGC
GGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAA
AGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAG
TACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA
TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAA
ACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGG
CTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATC
GAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCT
ACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAA
TCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA
GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
AGCTGGGAGGTGACTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAG
GAAAGTCTAACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTT
CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCAC
TGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT
CTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTA
ATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGA
AGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG
TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT
CCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
AGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGAT
ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC
GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTA
TCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA

GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAACTC
ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
CACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCT
CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGT
TGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCC
CAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGA
TCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
TGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
ATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA
AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGA
GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG
TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGA
TCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAA
GCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT
GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCAT
TAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCC
CAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCAT
TGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

[0096]

[0097] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 염기 편집기 8(ABE8)이다. 일부 실시형태에서, ABE8은 아래 **표 13, 표 14** 또는 **표 16**의 염기 편집기로부터 선택된다. 일부 실시형태에서, ABE8은 TadA로부터 진화된 아데노신 데아미네이스 변이체를 포함한다. 일부 실시형태에서, ABE8의 아데노신 데아미네이스 변이체는 아래 **표 11, 표 13** 또는 **표 14**에 기재되어 있는 바와 같은 TadA*8 변이체이다. 일부 실시형태에서, 아데노신 데아미네이스 변이체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R 및/또는 Q154R로 이루어진 군으로부터 선택된 변경 중 하나 이상을 포함하는 TadA*7.10 변이체(예를 들어, TadA*8)이다. 다양한 실시형태에서, ABE8은 다음으로 이루어진 군으로부터 선택되는 변경의 조합을 갖는 TadA*7.10 변이체(예를 들어, TadA*8)를 포함한다: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R. 일부 실시형태에서, ABE8은 단량체 작제물이다. 일부 실시형태에서, ABE8은 이종이량체 작제물이다. 일부 실시형태에서, ABE8 염기 편집기는 다음 서열을 포함한다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0098]

[0099] 예로서, 본 명세서에 기재된 염기 편집 조성물, 시스템 및 방법에 사용되는 바와 같은 사이티딘 염기 편집기(CBE)는 아래 제공되는 바와 같은 다음 핵산 서열(8877개의 염기쌍)을 갖는다(애드진, 매사추세츠주 워터타운 소재; 문헌[Komor AC, *et al.*, 2017, Sci Adv., 30;3(8):eaao4774. doi: 10.1126/sciadv.aao4774]). BE4 핵산 서열에 대해 적어도 95% 이상의 동일성을 갖는 폴리뉴클레오타이드 서열이 또한 포함된다.

```
   1 ATATGCCAAG TACGCCCCCT ATTGACGTCA ATGACGGTAA ATGGCCCGCC TGGCATTATG
  61 CCCAGTACAT GACCTTATGG GACTTTCCTA CTTGGCAGTA CATCTACGTA TTAGTCATCG
 121 CTATTACCAT GGTGATGCGG TTTTGGCAGT ACATCAATGG GCGTGGATAG CGGTTTGACT
 181 CACGGGGATT TCCAAGTCTC CACCCCATTG ACGTCAATGG GAGTTTGTTT TGGCACCAAA
 241 ATCAACGGGA CTTTCCAAAA TGTCGTAACA ACTCCGCCCC ATTGACGCAA ATGGGCGGTA
 301 GGCGTGTACG GTGGGAGGTC TATATAAGCA GAGCTGGTTT AGTGAACCGT CAGATCCGCT
 361 AGAGATCCGC GGCCGCTAAT ACGACTCACT ATAGGGAGAG CCGCCACCAT GAGCTCAGAG
 421 ACTGGCCCAG TGGCTGTGGA CCCCACATTG AGACGGCGGA TCGAGCCCCA TGAGTTTGAG
 481 GTATTCTTCG ATCCGAGAGA GCTCCGCAAG GAGACCTGCC TGCTTTACGA AATTAATTGG
 541 GGGGGCCGGC ACTCCATTTG GCGACATACA TCACAGAACA CTAACAAGCA CGTCGAAGTC
 601 AACTTCATCG AGAAGTTCAC GACAGAAAGA TATTTCTGTC CGAACACAAG GTGCAGCATT
 661 ACCTGGTTTC TCAGCTGGAG CCCATGCGGC GAATGTAGTA GGGCCATCAC TGAATTCCTG
 721 TCAAGGTATC CCCACGTCAC TCTGTTTATT TACATCGCAA GGCTGTACCA CCACGCTGAC
 781 CCCCGCAATC GACAAGGCCT GCGGGATTTG ATCTCTTCAG GTGTGACTAT CCAAATTATG
 841 ACTGAGCAGG AGTCAGGATA CTGCTGGAGA AACTTTGTGA ATTATAGCCC GAGTAATGAA
 901 GCCCACTGGC CTAGGTATCC CCATCTGTGG GTACGACTGT ACGTTCTTGA ACTGTACTGC
 961 ATCATACTGG GCCTGCCTCC TTGTCTCAAC ATTCTGAGAA GGAAGCAGCC ACAGCTGACA
1021 TTCTTTACCA TCGCTCTTCA GTCTTGTCAT TACCAGCGAC TGCCCCCACA CATTCTCTGG
1081 GCCACCGGGT TGAAATCTGG TGGTTCTTCT GGTGGTTCTA GCGGCAGCGA GACTCCCGGG
1141 ACCTCAGAGT CCGCCACACC CGAAAGTTCT GGTGGTTCTT CTGGTGGTTC TGATAAAAAG
1201 TATTCTATTG GTTTAGCCAT CGGCACTAAT TCCGTTGGAT GGGCTGTCAT AACCGATGAA
1261 TACAAAGTAC CTTCAAAGAA ATTTAAGGTG TTGGGGAACA CAGACCGTCA TTCGATTAAA
1321 AAGAATCTTA TCGGTGCCCT CCTATTCGAT AGTGGCGAAA CGGCAGAGGC GACTCGCCTG
1381 AAACGAACCG CTCGGAGAAG GTATACACGT CGCAAGAACC GAATATGTTA CTTACAAGAA
1441 ATTTTTAGCA ATGAGATGGC CAAAGTTGAC GATTCTTTCT TTCACCGTTT GGAAGAGTCC
1501 TTCCTTGTCG AAGAGGACAA GAAACATGAA CGGCACCCCA TCTTTGGAAA CATAGTAGAT
1561 GAGGTGGCAT ATCATGAAAA GTACCCAACG ATTTATCACC TCAGAAAAAA GCTAGTTGAC
1621 TCAACTGATA AAGCGGACCT GAGGTTAATC TACTTGGCTC TTGCCCATAT GATAAAGTTC
1681 CGTGGGCACT TTCTCATTGA GGGTGATCTA AATCCGGACA ACTCGGATGT CGACAAACTG
1741 TTCATCCAGT TAGTACAAAC CTATAATCAG TTGTTTGAAG AGAACCCTAT AAATGCAAGT
1801 GGCGTGGATG CGAAGGCTAT TCTTAGCGCC CGCCTCTCTA AATCCCGACG GCTAGAAAAC
1861 CTGATCGCAC AATTACCCGG AGAGAAGAAA AATGGGTTGT TCGGTAACCT TATAGCGCTC
1921 TCACTAGGCC TGACACCAAA TTTTAAGTCG AACTTCGACT TAGCTGAAGA TGCCAAATTG
1981 CAGCTTAGTA AGGACACGTA CGATGACGAT CTCGACAATC TACTGGCACA AATTGGAGAT
2041 CAGTATGCGG ACTTATTTTT GGCTGCCAAA AACCTTAGCG ATGCAATCCT CCTATCTGAC
2101 ATACTGAGAG TTAATACTGA GATTACCAAG GCGCCGTTAT CCGCTTCAAT GATCAAAAGG
2161 TACGATGAAC ATCACCAAGA CTTGACACTT CTCAAGGCCC TAGTCCGTCA GCAACTGCCT
2221 GAGAAATATA AGGAAATATT CTTTGATCAG TCGAAAAACG GGTACGCAGG TTATATTGAC
2281 GGCGGAGCGA GTCAAGAGGA ATTCTACAAG TTTATCAAAC CCATATTAGA GAAGATGGAT
2341 GGGACGGAAG AGTTGCTTGT AAAACTCAAT CGCGAAGATC TACTGCGAAA GCAGCGGACT
```

[0100]

2401 TTCGACAACG GTAGCATTCC ACATCAAATC CACTTAGGCG AATTGCATGC TATACTTAGA
2461 AGGCAGGAGG ATTTTTATCC GTTCCTCAAA GACAATCGTG AAAAGATTGA GAAAATCCTA
2521 ACCTTTCGCA TACCTTACTA TGTGGGACCC CTGGCCCGAG GGAACTCTCG GTTCGCATGG
2581 ATGACAAGAA AGTCCGAAGA AACGATTACT CCATGGAATT TTGAGGAAGT TGTCGATAAA
2641 GGTGCGTCAG CTCAATCGTT CATCGAGAGG ATGACCAACT TTGACAAGAA TTTACCGAAC
2701 GAAAAAGTAT TGCCTAAGCA CAGTTTACTT TACGAGTATT TCACAGTGTA CAATGAACTC
2761 ACGAAAGTTA AGTATGTCAC TGAGGGCATG CGTAAACCCG CCTTTCTAAG CGGAGAACAG
2821 AAGAAAGCAA TAGTAGATCT GTTATTCAAG ACCAACCGCA AAGTGACAGT TAAGCAATTG
2881 AAAGAGGACT ACTTTAAGAA AATTGAATGC TTCGATTCTG TCGAGATCTC CGGGGTAGAA
2941 GATCGATTTA ATGCGTCACT TGGTACGTAT CATGACCTCC TAAAGATAAT TAAAGATAAG
3001 GACTTCCTGG ATAACGAAGA GAATGAAGAT ATCTTAGAAG ATATAGTGTT GACTCTTACC
3061 CTCTTTGAAG ATCGGGAAAT GATTGAGGAA AGACTAAAAA CATACGCTCA CCTGTTCGAC
3121 GATAAGGTTA TGAAACAGTT AAAGAGGCGT CGCTATACGG GCTGGGGACG ATTGTCGCGG
3181 AAACTTATCA ACGGGATAAG AGACAAGCAA AGTGGTAAAA CTATTCTCGA TTTTCTAAAG
3241 AGCGACGGCT TCGCCAATAG GAACTTTATG CAGCTGATCC ATGATGACTC TTTAACCTTC
3301 AAAGAGGATA TACAAAAGGC ACAGGTTTCC GGACAAGGGG ACTCATTGCA CGAACATATT
3361 GCGAATCTTG CTGGTTCGCC AGCCATCAAA AAGGGCATAC TCCAGACAGT CAAAGTAGTG
3421 GATGAGCTAG TTAAGGTCAT GGGACGTCAC AAACCGGAAA ACATTGTAAT CGAGATGGCA
3481 CGCGAAAATC AAACGACTCA GAAGGGGCAA AAAAACAGTC GAGAGCGGAT GAAGAGAATA
3541 GAAGAGGGTA TTAAAGAACT GGGCAGCCAG ATCTTAAAGG AGCATCCTGT GGAAAATACC
3601 CAATTGCAGA ACGAGAAACT TTACCTCTAT TACCTACAAA ATGGAAGGGA CATGTATGTT
3661 GATCAGGAAC TGGACATAAA CCGTTTATCT GATTACGACG TCGATCACAT TGTACCCCAA
3721 TCCTTTTTGA AGGACGATTC AATCGACAAT AAAGTGCTTA CACGCTCGGA TAAGAACCGA
3781 GGGAAAAGTG ACAATGTTCC AAGCGAGGAA GTCGTAAAGA AAATGAAGAA CTATTGGCGG
3841 CAGCTCCTAA ATGCGAAACT GATAACGCAA AGAAAGTTCG ATAACTTAAC TAAAGCTGAG
3901 AGGGGTGGCT TGTCTGAACT TGACAAGGCC GGATTTATTA AACGTCAGCT CGTGGAAACC
3961 CGCCAAATCA CAAAGCATGT TGCACAGATA CTAGATTCCC GAATGAATAC GAAATACGAC
4021 GAGAACGATA AGCTGATTCG GGAAGTCAAA GTAATCACTT TAAAGTCAAA ATTGGTGTCG
4081 GACTTCAGAA AGGATTTTCA ATTCTATAAA GTTAGGGAGA TAAATAACTA CCACCATGCG
4141 CACGACGCTT ATCTTAATGC CGTCGTAGGG ACCGCACTCA TTAAGAAATA CCCGAAGCTA
4201 GAAAGTGAGT TTGTGTATGG TGATTACAAA GTTTATGACG TCCGTAAGAT GATCGCGAAA
4261 AGCGAACAGG AGATAGGCAA GGCTACAGCC AAATACTTCT TTTATTCTAA CATTATGAAT
4321 TTCTTTAAGA CGGAAATCAC TCTGGCAAAC GGAGAGATAC GCAAACGACC TTTAATTGAA
4381 ACCAATGGGG AGACAGGTGA AATCGTATGG GATAAGGGCC GGGACTTCGC GACGGTGAGA
4441 AAAGTTTTGT CCATGCCCCA AGTCAACATA GTAAAGAAAA CTGAGGTGCA GACCGGAGGG
4501 TTTTCAAAGG AATCGATTCT TCCAAAAAGG AATAGTGATA AGCTCATCGC TCGTAAAAAG
4561 GACTGGGACC CGAAAAAGTA CGGTGGCTTC GATAGCCCTA CAGTTGCCTA TTCTGTCCTA
4621 GTAGTGGCAA AAGTTGAGAA GGGAAAATCC AAGAAACTGA AGTCAGTCAA AGAATTATTG
4681 GGGATAACGA TTATGGAGCG CTCGTCTTTT GAAAAGAACC CCATCGACTT CCTTGAGGCG
4741 AAAGGTTACA AGGAAGTAAA AAAGGATCTC ATAATTAAAC TACCAAAGTA TAGTCTGTTT
4801 GAGTTAGAAA ATGGCCGAAA ACGGATGTTG GCTAGCGCCG GAGAGCTTCA AAAGGGGAAC
4861 GAACTCGCAC TACCGTCTAA ATACGTGAAT TTCCTGTATT TAGCGTCCCA TTACGAGAAG

[0101]

4921 TTGAAAGGTT CACCTGAAGA TAACGAACAG AAGCAACTTT TTGTTGAGCA GCACAAACAT
4981 TATCTCGACG AAATCATAGA GCAAATTTCG GAATTCAGTA AGAGAGTCAT CCTAGCTGAT
5041 GCCAATCTGG ACAAAGTATT AAGCGCATAC AACAAGCACA GGGATAAACC CATACGTGAG
5101 CAGGCGGAAA ATATTATCCA TTTGTTTACT CTTACCAACC TCGGCGCTCC AGCCGCATTC
5161 AAGTATTTTG ACACAACGAT AGATCGCAAA CGATACACTT CTACCAAGGA GGTGCTAGAC
5221 GCGACACTGA TTCACCAATC CATCACGGGA TTATATGAAA CTCGGATAGA TTTGTCACAG
5281 CTTGGGGGTG ACTCTGGTGG TTCTGGAGGA TCTGGTGGTT CTACTAATCT GTCAGATATT
5341 ATTGAAAAGG AGACCGGTAA GCAACTGGTT ATCCAGGAAT CCATCCTCAT GCTCCCAGAG
5401 GAGGTGGAAG AAGTCATTGG GAACAAGCCG GAAAGCGATA TACTCGTGCA CACCGCCTAC
5461 GACGAGAGCA CCGACGAGAA TGTCATGCTT CTGACTAGCG ACGCCCCTGA ATACAAGCCT
5521 TGGGCTCTGG TCATACAGGA TAGCAACGGT GAGAACAAGA TTAAGATGCT CTCTGGTGGT
5581 TCTGGAGGAT CTGGTGGTTC TACTAATCTG TCAGATATTA TTGAAAAGGA GACCGGTAAG
5641 CAACTGGTTA TCCAGGAATC CATCCTCATG CTCCCAGAGG AGGTGGAAGA AGTCATTGGG
5701 AACAAGCCGG AAAGCGATAT ACTCGTGCAC ACCGCCTACG ACGAGAGCAC CGACGAGAAT
5761 GTCATGCTTC TGACTAGCGA CGCCCCTGAA TACAAGCCTT GGGCTCTGGT CATACAGGAT
5821 AGCAACGGTG AGAACAAGAT TAAGATGCTC TCTGGTGGTT CTCCCAAGAA GAAGAGGAAA
5881 GTCTAACCGG TCATCATCAC CATCACCATT GAGTTTAAAC CCGCTGATCA GCCTCGACTG
5941 TGCCTTCTAG TTGCCAGCCA TCTGTTGTTT GCCCCTCCCC CGTGCCTTCC TTGACCCTGG
6001 AAGGTGCCAC TCCCACTGTC CTTTCCTAAT AAAATGAGGA AATTGCATCG CATTGTCTGA
6061 GTAGGTGTCA TTCTATTCTG GGGGGTGGGG TGGGGCAGGA CAGCAAGGGG GAGGATTGGG
6121 AAGACAATAG CAGGCATGCT GGGGATGCGG TGGGCTCTAT GGCTTCTGAG GCGGAAAGAA
6181 CCAGCTGGGG CTCGATACCG TCGACCTCTA GCTAGAGCTT GGCGTAATCA TGGTCATAGC
6241 TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
6301 TAAAGTGTAA AGCCTAGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT
6361 CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC
6421 GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC
6481 TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT
6541 TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG
6601 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG
6661 AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT
6721 ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA
6781 CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT
6841 GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC
6901 CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC AACCCGGTAA
6961 GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG
7021 TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGAACAG
7081 TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT
7141 GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
7201 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC
7261 AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA
7321 CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA
[0102] 7381 CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT

```
7441 TTCGTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT
7501 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT
7561 TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT
7621 CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA
7681 ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG
7741 GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
7801 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG
7861 CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG
7921 TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC
7981 GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA
8041 CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
8101 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT
8161 TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG
8221 GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA
8281 GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA
8341 AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC GACGGATCGG
8401 GAGATCGATC TCCCGATCCC CTAGGGTCGA CTCTCAGTAC AATCTGCTCT GATGCCGCAT
8461 AGTTAAGCCA GTATCTGCTC CCTGCTTGTG TGTTGGAGGT CGCTGAGTAG TGCGCGAGCA
8521 AAATTTAAGC TACAACAAGG CAAGGCTTGA CCGACAATTG CATGAAGAAT CTGCTTAGGG
8581 TTAGGCGTTT TGCGCTGCTT CGCGATGTAC GGGCCAGATA TACGCGTTGA CATTGATTAT
8641 TGACTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT
8701 TCCGCGTTAC ATAACTTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCCGCC
8761 CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC
8821 GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATC
```

[0103]

[0104] 일부 실시형태에서, 사이티딘 염기 편집기는 다음 중 하나로부터 선택되는 핵산 서열을 갖는 BE4이다:

[0105] 원래의 BE4 핵산 서열:

```
ATGagctcagagactggcccagtggctgtggaccccacattgagacggcggatcgagccccatgagtttg
aggtattcttcgatccgagagagctccgcaaggagacctgcctgctttacgaaattaattgggggggccg
gcactccatttggcgacatacatcacagaacactaacaagcacgtcgaagtcaacttcatcgagaagttc
acgacagaaagatatttctgtccgaacacaaggtgcagcattacctggtttctcagctggagccgcgaat
gtagtagggccatcactgaattcctgtcaaggtatccccacgtcactctgtttatttacatcgcaaggct
gtaccaccacgctgacccccgcaatcgacaaggcctgcgggatttgatctcttcaggtgtgactatccaa
attatgactgagcaggagtcaggatactgctggagaaactttgtgaattatagcccgagtaatgaagccc
actggcctaggtatccccatctgtgggtacgactgtacgttcttgaactgtactgcatcatactgggcct
gcctccttgtctcaacattctgagaaggaagcagccacagctgacattctttaccatcgctcttcagtct
tgtcattaccagcgactgcccccacacattctctgggccaccgggttgaaatctggtggttcttctggtg
gttctagcggcagcgagactcccgggacctcagagtccgccacacccgaaagttctggtggttcttctgg
tggttctgataaaaagtattctattggtttagccatcggcactaattccgttggatgggctgtcataacc
gatgaatacaaagtaccttcaaagaaatttaaggtgttggggaacacagaccgtcattcgattaaaaaga
```

[0106]

atcttatcggtgccctcctattcgatagtggcgaaacggcagaggcgactcgcctgaaacgaaccgctcg
gagaaggtatacacgtcgcaagaaccgaatatgttacttacaagaaatttttagcaatgagatggccaaa
gttgacgattctttctttcaccgtttggaagagtccttccttgtcgaagaggacaagaaacatgaacggc
accccatctttggaaacatagtagatgaggtggcatatcatgaaaagtacccaacgatttatcacctcag
aaaaaagctagttgactcaactgataaagcggacctgaggttaatctacttggctcttgcccatatgata
aagttccgtgggcactttctcattgagggtgatctaaatccggacaactcggatgtcgacaaactgttca
tccagttagtacaaacctataatcagttgtttgaagagaaccctataaatgcaagtggcgtggatgcgaa
ggctattcttagcgcccgcctctctaaatcccgacggctagaaaacctgatcgcacaattacccggagag
aagaaaaatgggttgttcggtaaccttatagcgctctcactaggcctgacaccaaattttaagtcgaact
tcgacttagctgaagatgccaaattgcagcttagtaaggacacgtacgatgacgatctcgacaatctact
ggcacaaattggagatcagtatgcggacttatttttggctgccaaaaaccttagcgatgcaatcctccta
tctgacatactgagagttaatactgagattaccaaggcgccgttatccgcttcaatgatcaaaaggtacg
atgaacatcaccaagacttgacacttctcaaggccctagtccgtcagcaactgcctgagaaatataagga
aatattctttgatcagtcgaaaaacgggtacgcaggttatattgacggcggagcgagtcaagaggaattc
tacaagtttatcaaacccatattagagaagatggatgggacggaagagttgcttgtaaaactcaatcgcg
aagatctactgcgaaagcagcggactttcgacaacggtagcattccacatcaaatccacttaggcgaatt
gcatgctatacttagaaggcaggaggatttttatccgttcctcaaagacaatcgtgaaaagattgagaaa
atcctaacctttcgcataccttactatgtgggacccctggcccgagggaactctcggttcgcatggatga
caagaaagtccgaagaaacgattactccatggaattttgaggaagttgtcgataaaggtgcgtcagctca
atcgttcatcgagaggatgaccaactttgacaagaatttaccgaacgaaaaagtattgcctaagcacagt
ttactttacgagtatttcacagtgtacaatgaactcacgaaagttaagtatgtcactgagggcatgcgta
aacccgcctttctaagcggagaacagaagaaagcaatagtagatctgttattcaagaccaaccgcaaagt
gacagttaagcaattgaaagaggactactttaagaaaattgaatgcttcgattctgtcgagatctccggg
gtagaagatcgatttaatgcgtcacttggtacgtatcatgacctcctaaagataattaaagataaggact
tcctggataacgaagagaatgaagatatcttagaagatatagtgttgactcttaccctctttgaagatcg
ggaaatgattgaggaaagactaaaaacatacgctcacctgttcgacgataaggttatgaaacagttaaag
aggcgtcgctatacgggctggggacgattgtcgcggaaacttatcaacgggataagagacaagcaaagtg
gtaaaactattctcgattttctaaagagcgacggcttcgccaataggaactttatgcagctgatccatga
tgactctttaaccttcaaagaggatatacaaaaggcacaggtttccggacaaggggactcattgcacgaa
catattgcgaatcttgctggttcgccagccatcaaaaagggcatactccagacagtcaaagtagtggatg
agctagttaaggtcatgggacgtcacaaaccggaaaacattgtaatcgagatggcacgcgaaaatcaaac
gactcagaaggggcaaaaaaacagtcgagagcggatgaagagaatagaagagggtattaaagaactgggc
agccagatcttaaaggagcatcctgtggaaaatacccaattgcagaacgagaaactttacctctattacc
tacaaaatggaagggacatgtatgttgatcaggaactggacataaaccgtttatctgattacgacgtcga
tcacattgtaccccaatccttttttgaaggacgattcaatcgacaataaagtgcttacacgctcggataag
aaccgagggaaaagtgacaatgttccaagcgaggaagtcgtaaagaaaatgaagaactattggcggcagc
tcctaaatgcgaaactgataacgcaaagaaagttcgataacttaactaaagctgagaggggtggcttgtc
tgaacttgacaaggccggatttattaaacgtcagctcgtggaaacccgccaaatcacaaagcatgttgca

[0107]

cagatactagattcccgaatgaatacgaaatacgacgagaacgataagctgattcgggaagtcaaagtaa
tcactttaaagtcaaaattggtgtcggacttcagaaaggattttcaattctataaagttagggagataaa
taactaccaccatgcgcacgacgcttatcttaatgccgtcgtagggaccgcactcattaagaaatacccg
aagctagaaagtgagtttgtgtatggtgattacaaagtttatgacgtccgtaagatgatcgcgaaaagcg
aacaggagataggcaaggctacagccaaatacttcttttattctaacattatgaatttctttaagacgga
aatcactctggcaaacggagagatacgcaaacgacctttaattgaaaccaatggggagacaggtgaaatc
gtatgggataagggccgggacttcgcgacggtgagaaaagttttgtccatgccccaagtcaacatagtaa
agaaaactgaggtgcagaccggagggttttcaaaggaatcgattcttccaaaaaggaatagtgataagct
catcgctcgtaaaaaggactgggacccgaaaaagtacggtggcttcgatagccctacagttgcctattct
gtcctagtagtggcaaaagttgagaagggaaaatccaagaaactgaagtcagtcaaagaattattgggga
taacgattatggagcgctcgtcttttgaaaagaaccccatcgacttccttgaggcgaaaggttacaagga
agtaaaaaaggatctcataattaaactaccaaagtatagtctgtttgagttagaaaatggccgaaaacgg
atgttggctagcgccggagagcttcaaaaggggaacgaactcgcactaccgtctaaatacgtgaatttcc
tgtatttagcgtcccattacgagaagttgaaaggttcacctgaagataacgaacagaagcaacttttttgt
tgagcagcacaaacattatctcgacgaaatcatagagcaaatttcggaattcagtaagagagtcatcctа
gctgatgccaatctggacaaagtattaagcgcatacaacaagcacagggataaacccatacgtgagcagg
cggaaaatattatccatttgtttactcttaccaacctcggcgctccagccgcattcaagtattttgacac
aacgatagatcgcaaacgatacacttctaccaaggaggtgctagacgcgacactgattcaccaatccatc
acgggattatatgaaactcggatagatttgtcacagcttgggggtgactctggtggttctggaggatctg
gtggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggttatccaggaatccat
cctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgatatactcgtgcacacc
gcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgcccctgaatacaagccttggg
ctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtggttctggaggatctgg
tggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggttatccaggaatccatc
ctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgatatactcgtgcacaccg
cctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgcccctgaatacaagccttgggc
tctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtggttctAAAAGGACGGCG
[0108] GACGGATCAGAGTTCGAGAGTCCGAAAAAAAAACGAAAGGTCGAAtaa

[0109] BE4 코돈 최적화 1 핵산 서열:

ATGTCATCCGAAACCGGGCCAGTGGCCGTAGACCCAACACTCAGGAGGCGGATAGAACCCCATGAGTTTG
AAGTGTTCTTCGACCCCAGAGAGCTGCGCAAAGAGACTTGCCTCCTGTATGAAATAAATTGGGGGGGTCG
CCATTCAATTTGGAGGCACACTAGCCAGAATACTAACAAACACGTGGAGGTAAATTTTATCGAGAAGTTT
ACCACCGAAAGATACTTTTGCCCCAATACACGGTGTTCAATTACCTGGTTTCTGTCATGGAGTCCATGTG
GAGAATGTAGTAGAGCGATAACTGAGTTCCTGTCTCGATATCCTCACGTCACGTTGTTTATATACATCGC
TCGGCTTTATCACCATGCGGACCCGCGGAACAGGCAAGGTCTTCGGGACCTCATATCCTCTGGGGTGACC
ATCCAGATAATGACGGAGCAAGAGAGCGGATACTGCTGGCGAAACTTTGTTAACTACAGCCCAAGCAATG
AGGCACACTGGCCTAGATATCCGCATCTCTGGGTTCGACTGTATGTCCTTGAACTGTACTGCATAATTCT
[0110] GGGACTTCCGCCATGCTTGAACATTCTGCGGCGGAAACAACCACAGCTGACCTTTTTCACGATTGCTCTC

CAAAGTTGTCACTACCAGCGATTGCCACCCCACATCTTGTGGGCTACTGGACTCAAGTCTGGAGGAAGTT
CAGGCGGAAGCAGCGGGTCTGAAACGCCCGGAACCTCAGAGAGCGCAACGCCCGAAAGCTCTGGAGGGTC
AAGTGGTGGTAGTGATAAGAAATACTCCATCGGCCTCGCCATCGGTACGAATTCTGTCGGTTGGGCCGTT
ATCACCGATGAGTACAAGGTCCCTTCTAAGAAATTCAAGGTTTTGGGCAATACAGACCGCCATTCTATAA
AAAAAAACCTGATCGGCGCCCTTTTGTTTGACAGTGGTGAGACTGCTGAAGCGACTCGCCTGAAGCGAAC
TGCCAGGAGGCGGTATACGAGGCGAAAAAACCGAATTTGTTACCTCCAGGAGATTTTCTCAAATGAAATG
GCCAAGGTAGATGATAGTTTTTTTCACCGCTTGGAAGAAAGTTTTCTCGTTGAGGAGGACAAAAAGCACG
AGAGGCACCCAATCTTTGGCAACATAGTCGATGAGGTCGCATACCATGAGAAATATCCTACGATCTATCA
TCTCCGCAAGAAGCTGGTCGATAGCACGGATAAAGCTGACCTCCGGCTGATCTACCTTGCTCTTGCTCAC
ATGATTAAATTCAGGGGCCATTTCCTGATAGAAGGAGACCTCAATCCCGACAATTCTGATGTCGACAAAC
TGTTTATTCAGCTCGTTCAGACCTATAATCAACTCTTTGAGGAGAACCCCATCAATGCTTCAGGGGTGGA
CGCAAAGGCCATTTTGTCCGCGCGCTTGAGTAAATCACGACGCCTCGAGAATTTGATAGCTCAACTGCCG
GGTGAGAAGAAAAACGGGTTGTTTGGGAATCTCATAGCGTTGAGTTTGGGACTTACGCCAAACTTTAAGT
CTAACTTTGATTTGGCCGAAGATGCCAAATTGCAGCTGTCCAAAGATACCTATGATGACGACTTGGATAA
CCTTCTTGCGCAGATTGGTGACCAATACGCGGATCTGTTTCTTGCCGCAAAAAATCTGTCCGACGCCATA
CTCTTGTCCGATATACTGCGCGTCAATACTGAGATAACTAAGGCTCCCCTCAGCGCGTCCATGATTAAAA
GATACGATGAGCACCACCAAGATCTCACTCTGTTGAAAGCCCTGGTTCGCCAGCAGCTTCCAGAGAAGTA
TAAGGAGATATTTTTCGACCAATCTAAAAACGGCTATGCGGGTTACATTGACGGTGGCGCCTCTCAAGAA
GAATTCTACAAGTTTATAAAGCCGATACTTGAGAAAATGGACGGTACAGAGGAATTGTTGGTTAAGCTCA
ATCGCGAGGACTTGTTGAGAAAGCAGCGCACATTTGACAATGGTAGTATTCCACACCAGATTCATCTGGG
CGAGTTGCATGCCATTCTTAGAAGACAAGAAGATTTTTATCCGTTTCTGAAAGATAACAGAGAAAAGATT
GAAAAGATACTTACCTTTCGCATACCGTATTATGTAGGTCCCCTGGCTAGAGGGAACAGTCGCTTCGCTT
GGATGACTCGAAAATCAGAAGAAACAATAACCCCCTGGAATTTTGAAGAAGTGGTAGATAAAGGTGCGAG
TGCCCAATCTTTTATTGAGCGGATGACAAATTTTGACAAGAATCTGCCTAACGAAAAGGTGCTTCCCAAG
CATTCCCTTTTGTATGAATACTTTACAGTATATAATGAACTGACTAAAGTGAAGTACGTTACCGAGGGGA
TGCGAAAGCCAGCTTTTCTCAGTGGCGAGCAGAAAAAAGCAATAGTTGACCTGCTGTTCAAGACGAATAG
GAAGGTTACCGTCAAACAGCTCAAAGAAGATTACTTTAAAAAGATCGAATGTTTTGATTCAGTTGAGATA
AGCGGAGTAGAGGATAGATTTAACGCAAGTCTTGGAACTTATCATGACCTTTTGAAGATCATCAAGGATA
AAGATTTTTTGGACAACGAGGAGAATGAAGATATCCTGGAAGATATAGTACTTACCTTGACGCTTTTTGA
AGATCGAGAGATGATCGAGGAGCGACTTAAGACGTACGCACATCTCTTTGACGATAAGGTTATGAAACAA
TTGAAACGCCGGCGGTATACTGGCTGGGGCAGGCTTTCTCGAAAGCTGATTAATGGTATCCGCGATAAGC
AGTCTGGAAAGACAATCCTTGACTTTCTGAAAAGTGATGGATTTGCAAATAGAAACTTTATGCAGCTTAT
ACATGATGACTCTTTGACGTTCAAGGAAGACATCCAGAAGGCACAGGTATCCGGCCAAGGGGATAGCCTC
CATGAACACATAGCCAACCTGGCCGGCTCACCAGCTATTAAAAAGGGAATATTGCAAACCGTTAAGGTTG
TTGACGAACTCGTTAAGGTTATGGGCCGACACAAACCAGAGAATATCGTGATTGAGATGGCTAGGGAGAA
TCAGACCACTCAAAAAGGTCAGAAAAATTCTCGCGAAAGGATGAAGCGAATTGAAGAGGGAATCAAAGAA
CTTGGCTCTCAAATTTTGAAAGAGCACCCGGTAGAAAACACTCAGCTGCAGAATGAAAAGCTGTATCTGT
ATTATCTGCAGAATGGTCGAGATATGTACGTTGATCAGGAGCTGGATATCAATAGGCTCAGTGACTACGA

[0111]

TGTCGACCACATCGTTCCTCAATCTTTCCTGAAAGATGACTCTATCGACAACAAAGTGTTGACGCGATCA
GATAAGAACCGGGGAAAATCCGACAATGTACCCTCAGAAGAAGTTGTCAAGAAGATGAAAAACTATTGGA
GACAATTGCTGAACGCCAAGCTCATAACACAACGCAAGTTCGATAACTTGACGAAAGCCGAAAGAGGTGG
GTTGTCAGAATTGGACAAAGCTGGCTTTATTAAGCGCCAATTGGTGGAGACCCGGCAGATTACGAAACAC
GTAGCACAAATTTTGGATTCACGAATGAATACCAAATACGACGAAAACGACAAATTGATACGCGAGGTGA
AAGTGATTACGCTTAAGAGTAAGTTGGTTTCCGATTTCAGGAAGGATTTTCAGTTTTACAAAGTAAGAGA
AATAAACAACTACCACCACGCCCATGATGCTTACCTCAACGCGGTAGTTGGCACAGCTCTTATCAAAAAA
TATCCAAAGCTGGAAAGCGAGTTCGTTTACGGTGACTATAAAGTATACGACGTTCGGAAGATGATAGCCA
AATCAGAGCAGGAAATTGGGAAGGCAACCGCAAAATACTTCTTCTATTCAAACATCATGAACTTCTTTAA
GACGGAGATTACGCTCGCGAACGGCGAAATACGCAAGAGGCCCCTCATAGAGACTAACGGCGAAACCGGG
GAGATCGTATGGGACAAAGGACGGGACTTTGCGACCGTTAGAAAAGTACTTTCAATGCCACAAGTGAATA
TTGTTAAAAAGACAGAAGTACAAACAGGGGGGTTCAGTAAGGAATCCATTTTGCCCAAGCGGAACAGTGA
TAAATTGATAGCAAGGAAAAAAGATTGGGACCCTAAGAAGTACGGTGGTTTCGACTCTCCTACCGTTGCA
TATTCAGTCCTTGTAGTTGCGAAAGTGGAAAAGGGGAAAAGTAAGAAGCTTAAGAGTGTTAAAGAGCTTC
TGGGCATAACCATAATGGAACGGTCTAGCTTCGAGAAAAATCCAATTGACTTTCTCGAGGCTAAAGGTTA
CAAGGAGGTAAAAAAGGACCTGATAATTAAACTCCCAAAGTACAGTCTCTTCGAGTTGGAGAATGGGAGG
AAGAGAATGTTGGCATCTGCAGGGGAGCTCCAAAAGGGGAACGAGCTGGCTCTGCCTTCAAAATACGTGA
ACTTTCTGTACCTGGCCAGCCACTACGAGAAACTCAAGGGTTCTCCTGAGGATAACGAGCAGAAACAGCT
GTTTGTAGAGCAGCACAAGCATTACCTGGACGAGATAATTGAGCAAATTAGTGAGTTCTCAAAAAGAGTA
ATCCTTGCAGACGCGAATCTGGATAAAGTTCTTTCCGCCTATAATAAGCACCGGGACAAGCCTATACGAG
AACAAGCCGAGAACATCATTCACCTCTTTACCCTTACTAATCTGGGCGCGCCGGCCGCCTTCAAATACTT
CGACACCACGATAGACAGGAAAAGGTATACGAGTACCAAAGAAGTACTTGACGCCACTCTCATCCACCAG
TCTATAACAGGGTTGTACGAAACGAGGATAGATTTGTCCCAGCTCGGCGGCGACTCAGGAGGGTCAGGCG
GCTCCGGTGGATCAACGAATCTTTCCGACATAATCGAGAAAGAAACCGGCAAACAGTTGGTGATCCAAGA
ATCAATCCTGATGCTGCCTGAAGAAGTAGAAGAGGTGATTGGCAACAAACCTGAGTCTGACATTCTTGTC
CACACCGCGTATGACGAGAGCACGGACGAGAACGTTATGCTTCTCACTAGCGACGCCCCTGAGTATAAAC
CATGGGCGCTGGTCATCCAAGATTCCAATGGGGAAAACAAGATTAAGATGCTTAGTGGTGGGTCTGGAGG
GAGCGGTGGGTCCACGAACCTCAGCGACATTATTGAAAAAGAGACTGGTAAACAACTTGTAATACAAGAG
TCTATTCTGATGTTGCCTGAAGAGGTGGAGGAGGTGATTGGGAACAAACCGGAGTCTGATATACTTGTTC
ATACCGCCTATGACGAATCTACTGATGAGAATGTGATGCTTTTaACGTCAGACGCTCCCGAGTACAAACC
CTGGGCTCTGGTGATTCAGGACAGCAATGGTGAGAATAAGATTAAAATGTTGAGTGGGGGCTCAAAGCGC

[0112] ACGGCTGACGGTAGCGAATTTGAGAGCCCCAAAAAAAAACGAAAGGTCGAAtaa

[0113] BE4 코돈 최적화 2 핵산 서열:

ATGAGCAGCGAGACAGGCCCTGTGGCTGTGGATCCTACACTGCGGAGAAGAATCGAGCCCCACGAGTTCG
AGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAG
ACACTCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTT
ACGACCGAGCGGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCG

[0114] GCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCTATATCGC

CCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGACTGCGCGACCTGATCAGCAGCGGAGTGACC
ATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACTTCGTGAACTACAGCCCCAGCAACG
AAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTACTGCATCATCCT
GGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCACAATCGCCCTG
CAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGAGCGGAGGATCTA
GCGGCGGCTCTAGCGGATCTGAGACACCTGGCACAAGCGAGTCTGCCACACCTGAGAGTAGCGGCGGATC
TTCTGGCGGCTCCGACAAGAAGTACTCTATCGGACTGGCCATCGGCACCAACTCTGTTGGATGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCA
AGAAGAATCTGATCGGCGCCCTGCTGTTCGACTCTGGCGAAACAGCCGAAGCCACCAGACTGAAGAGAAC
CGCCAGGCGGAGATACACCCGGCGGAAGAACCGGATCTGCTACCTGCAAGAGATCTTCAGCAACGAGATG
GCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGACAAGAAGCACG
AGCGGCACCCCATCTTCGGCAACATCGTGGATGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCA
CCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGAGACTGATCTACCTGGCTCTGGCCCAC
ATGATCAAGTTCCGGGGCCACTTTCTGATCGAGGGCGATCTGAACCCCGACAACAGCGACGTGGACAAGC
TGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCTCTGGCGTGGA
CGCCAAGGCTATCCTGTCTGCCAGACTGAGCAAGAGCAGAAGGCTGGAAAACCTGATCGCCCAGCTGCCT
GGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGACTGACCCCTAACTTCAAGA
GCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAA
TCTGCTGGCCCAGATCGGCGATCAGTACGCCGACTTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATC
CTGCTGAGCGATATCCTGAGAGTGAACACCGAGATCACAAAGGCCCCTCTGAGCGCCTCTATGATCAAGA
GATACGACGAGCACCACCAGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTA
CAAAGAGATTTTCTTCGATCAGTCCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAG
GAATTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTGGTCAAGCTGA
ACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAATGGCTCTATCCCTCACCAGATCCACCTGGG
AGAGCTGCACGCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATC
GAGAAGATCCTGACCTTCAGGATCCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTCGCCT
GGATGACCAGAAAGAGCGAGGAAACCATCACACCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCCAG
CGCTCAGTCCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCCAAG
CACTCCCTGCTGTATGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAA
TGAGAAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATTGTGGATCTGCTGTTCAAGACCAACCG
GAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTGGAAATC
AGCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGACCTGCTGAAAATTATCAAGGACA
AGGACTTCCTGGACAACGAAGAGAACGAGGACATTCTCGAGGACATCGTGCTGACCCTGACACTGTTTGA
GGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCCACCTGTTCGACGACAAAGTGATGAAGCAA
CTGAAGCGGAGGCGGTACACAGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCATCCGGGATAAGC
AGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGAT
CCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAAGGCGATTCTCTG
CACGAGCACATTGCCAACCTGGCCGGATCTCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGG

[0115]

TGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAA
CCAGACCACACAGAAGGGCCAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAG
CTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTGGACATCAACCGGCTGAGCGACTACGA
TGTGGACCATATCGTGCCCCAGAGCTTTCTGAAGGACGACTCCATCGATAACAAGGTCCTGACCAGAAGC
GACAAGAACCGGGGCAAGAGCGATAACGTGCCCTCCGAAGAGGTGGTCAAGAAGATGAAGAACTACTGGC
GACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAGTTCGATAACCTGACCAAGGCCGAGAGAGGCGG
CCTGAGCGAACTTGATAAGGCCGGCTTCATTAAGCGGCAGCTGGTGGAAACCCGGCAGATCACCAAACAC
GTGGCACAGATTCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGA
AAGTCATCACCCTGAAGTCTAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGGGA
AATCAACAACTACCATCACGCCCACGACGCCTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAGAAG
TATCCCAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCA
AGAGCGAACAAGAGATCGGCAAGGCTACCGCCAAGTACTTTTTCTACAGCAACATCATGAACTTTTTCAA
GACAGAGATCACCCTGGCCAACGGCGAGATCCGGAAAAGACCCCTGATCGAGACAAACGGCGAAACCGGG
GAGATCGTGTGGGATAAGGGCAGAGATTTTGCCACAGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATA
TCGTGAAGAAAACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACAGCGA
TAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGATAGCCCTACCGTGGCC
TATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAAAAGCTCAAGAGCGTGAAAGAGCTGC
TGGGGATCACCATCATGGAAAGAAGCAGCTTTGAGAAGAACCCGATCGACTTTCTGGAAGCCAAGGGCTA
CAAAGAAGTCAAGAAGGACCTCATCATCAAGCTCCCCAAGTACAGCCTGTTCGAGCTGGAAAATGGCCGG
AAGCGGATGCTGGCCTCAGCAGGCGAACTGCAGAAAGGCAATGAACTGGCCCTGCCTAGCAAATACGTCA
ACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCT
GTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
ATCCTGGCCGACGCTAACCTGGATAAGGTGCTGTCTGCCTATAACAAGCACCGGGACAAGCCTATCAGAG
AGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCTGCCGCCTTCAAGTACTT
CGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACACTGATCCACCAG
TCTATCACCGGCCTGTACGAAACCCGGATCGACCTGTCTCAGCTCGGCGGCGATTCTGGTGGTTCTGGCG
GAAGTGGCGGATCCACCAATCTGAGCGACATCATCGAAAAAGAGACAGGCAAGCAGCTCGTGATCCAAGA
ATCCATCCTGATGCTGCCTGAAGAGGTTGAGGAAGTGATCGGCAACAAGCCTGAGTCCGACATCCTGGTG
CACACCGCCTACGATGAGAGCACCGATGAGAACGTCATGCTGCTGACAAGCGACGCCCCTGAGTACAAGC
CTTGGGCTCTCGTGATTCAGGACAGCAATGGGGAGAACAAGATCAAGATGCTGAGCGGAGGTAGCGGAGG
CAGTGGCGGAAGCACAAACCTGTCTGATATCATTGAAAAAGAAACCGGGAAGCAACTGGTCATTCAAGAG
TCCATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAACCCGAGAGCGATATTCTGGTCC
ACACAGCCTATGACGAGTCTACAGACGAAAACGTGATGCTCCTGACCTCTGACGCTCCCGAGTATAAGCC
CTGGGCACTTGTTATCCAGGACTCTAACGGGGAAAACAAAATCAAAATGTTGTCCGGCGGCAGCAAGCGG

[0116] ACAGCCGATGGATCTGAGTTCGAGAGCCCCAAGAAGAAACGGAAGGTgGAGtaa

[0117] "염기 편집 활성"은 폴리뉴클레오타이드 내의 염기를 화학적으로 변경하는 작용을 하는 것을 의미한다. 일 실시형태에서, 제1 염기는 제2 염기로 전환된다. 일 실시형태에서, 염기 편집 활성은, 예를 들어, 표적 C·G를 T·A로 전환하는 사이티딘 데아미네이스 활성이다. 또 다른 실시형태에서, 염기 편집 활성은, 예를 들어, A·T를 G·C로 전환하는 아데노신 또는 아데닌 데아미네이스 활성이다. 또 다른 실시형태에서, 염기 편집 활성은, 예를 들어, 표적 C·G를 T·A로 전환하는 사이티딘 데아미네이스 활성 및, 예를 들어, A·T를 G·C로 전환하는 아데노신 또는 아데닌 데아미네이스 활성이다.

[0118] 일부 실시형태에서, 염기 편집 활성은 편집의 효율성에 의해 평가된다. 염기 편집 효율은 임의의 적합한 수단에 의해, 예를 들어, 생어 시퀀싱 또는 차세대 시퀀싱에 의해 측정될 수 있다. 일부 실시형태에서, 염기 편집 효율은 염기 편집기에 의해 영향을 받은 핵염기 전환이 있는 총 시퀀싱 판독물의 백분율, 예를 들어, 표적 A·T 염기쌍이 G·C 염기쌍으로 전환되거나 또는 표적 C·G 염기쌍이 T·A 염기쌍으로 전환된 총 시퀀싱 판독물의 백분율에 의해 측정된다. 일부 실시형태에서, 염기 편집 효율은 염기 편집이 세포의 집단에서 수행될 때 염기 편집기에 의해 영향을 받은 핵염기 전환이 있는 총 세포의 백분율에 의해 측정된다.

[0119] 용어 "염기 편집 시스템"은 표적 뉴클레오타이드 서열의 핵염기를 편집하는 시스템을 지칭한다. 다양한 실시형태에서, 염기 편집 시스템은 (1) 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인(예를 들어, Cas9); (2) 상기 핵염기를 탈아미노화하기 위한 데아미네이스 도메인(예를 들어, 아데노신 데아미네이스 또는

사이티딘 데아미네이스); 및 (3) 하나 이상의 가이드 폴리뉴클레오타이드(예를 들어, 가이드 RNA)를 포함한다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인이다. 일부 실시형태에서, 염기 편집기는 사이티딘 염기 편집기(CBE)이다. 일부 실시형태에서, 염기 편집기는 아데닌 또는 아데노신 염기 편집기(ABE)이다. 일부 실시형태에서, 염기 편집 시스템은 아데노신 데아미네이스 염기 편집기 8(ABE8)이다.

[0120] 일부 실시형태에서, ABE8은 단량체 작제물이다. 일부 실시형태에서, ABE8은 ABE8.1-m, ABE8.2-m, ABE8.3-m, ABE8.4-m, ABE8.5-m, ABE8.6-m, ABE8.7-m, ABE8.8-m, ABE8.9-m, ABE8.10-m, ABE8.11-m, ABE8.12-m, ABE8.13-m, ABE8.14-m, ABE8.15-m, ABE8.16-m, ABE8.17-m, ABE8.18-m, ABE8.19-m, ABE8.20-m, ABE8.21-m, ABE8.22-m, ABE8.23-m, ABE8.24-m이다. 일부 실시형태에서, ABE8은 이종체(heteromeric) 작제물이다. 일부 실시형태에서, ABE8은 ABE8.1-d, ABE8.2-d, ABE8.3-d, ABE8.4-d, ABE8.5-d, ABE8.6-d, ABE8.7-d, ABE8.8-d, ABE8.9-d, ABE8.10-d, ABE8.11-d, ABE8.12-d, ABE8.13-d, ABE8.14-d, ABE8.15-d, ABE8.16-d, ABE8.17-d, ABE8.18-d, ABE8.19-d, ABE8.20-d, ABE8.21-d, ABE8.22-d, ABE8.23-d 또는 ABE8.24-d이다.

[0121] 일부 실시형태에서, 염기 편집 시스템은 하나 이상의 염기 편집 성분을 포함할 수 있다. 예를 들어, 염기 편집 시스템은 하나 이상의 데아미네이스를 포함할 수 있다. 일부 실시형태에서, 염기 편집 시스템은 하나 이상의 사이티딘 데아미네이스를 포함할 수 있다. 일부 실시형태에서, 염기 편집 시스템은 하나 이상의 아데노신 데아미네이스를 포함할 수 있다. 일부 실시형태에서, 단일 가이드 폴리뉴클레오타이드는 표적 핵산 서열에 대해 상이한 데아미네이스를 표적화하기 위해 사용될 수 있다. 일부 실시형태에서, 단일 쌍의 가이드 폴리뉴클레오타이드는 표적 핵산 서열에 대해 상이한 데아미네이스를 표적화하기 위해 사용될 수 있다.

[0122] 염기 편집 시스템의 핵염기 성분 및 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 성분은 공유적으로 또는 비공유적으로 서로 회합되거나 또는 이들의 회합 및 상호작용의 임의의 조합일 수 있다. 예를 들어, 일부 실시형태에서, 데아미네이스 도메인은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미네이스 도메인에 융합 또는 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미네이스 도메인과 비공유적으로 상호작용하거나 또는 이와 회합함으로써 데아미네이스 도메인을 표적 뉴클레오타이드 서열로 표적화할 수 있다. 예를 들어, 일부 실시형태에서, 핵염기 편집 성분, 예를 들어, 데아미네이스 도메인은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 일부인 추가적인 이종성 부분 또는 도메인과 상호작용하거나, 이와 회합할 수 있거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인을 포함할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드에 결합하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드에 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가적인 이종성 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머레이스 Ku 결합 모티프 및 Ku 단백질, 텔로머레이스 Sm7 결합 모티프 및 Sm7 단백질 또는 RNA 인식 모티프일 수 있다.

[0123] 염기 편집 시스템은 가이드 폴리뉴클레오타이드 성분을 더 포함할 수 있다. 염기 편집 시스템의 성분은 공유 결합, 비공유 상호작용 또는 이들의 회합 및 상호작용의 임의의 조합을 통해 서로 회합될 수 있음을 이해하여야 한다. 일부 실시형태에서, 데아미네이스 도메인은 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다. 예를 들어, 일부 실시형태에서, 염기 편집 시스템의 핵염기 편집 성분, 예를 들어, 데아미네이스 도메인은 가이드 폴리뉴클레오타이드의 부분 또는 분절(예를 들어, 폴리뉴클레오타이드 모티프)과 상호작용하거나, 이와 회합할 수 있거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)은 데아미네이스 도메인에 융합 또는 연결될 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드에 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드에 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부

실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가적인 이종성 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머레이스 Ku 결합 모티프 및 Ku 단백질, 텔로머레이스 Sm7 결합 모티프 및 Sm7 단백질 또는 RNA 인식 모티프일 수 있다.

[0124] 일부 실시형태에서, 염기 편집 시스템은 염기 절단 복구 저해제(BER) 성분을 더 포함할 수 있다. 염기 편집 시스템의 성분은 공유 결합, 비공유 상호작용 또는 이들의 회합 및 상호작용의 임의의 조합을 통해 서로 회합될 수 있음을 이해하여야 한다. BER 성분의 저해제는 BER 저해제를 포함할 수 있다. 일부 실시형태에서, BER의 저해제는 우라실 DNA 글리코실레이스 저해제(UGI)일 수 있다. 일부 실시형태에서, BER의 저해제는 이노신 BER 저해제일 수 있다. 일부 실시형태에서, BER의 저해제는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 BER의 저해제에 융합 또는 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미네이스 도메인 및 BER의 저해제에 융합 또는 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 BER의 저해제와 비공유적으로 상호작용하거나 또는 이와 회합함으로써 BER의 저해제를 표적 뉴클레오타이드 서열로 표적화할 수 있다. 예를 들어, 일부 실시형태에서, BER 성분의 저해제는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 일부인 추가적인 이종성 부분 또는 도메인과 상호작용하고, 이와 회합하거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인을 포함할 수 있다.

[0125] 일부 실시형태에서, BER의 저해제는 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다. 예를 들어, 일부 실시형태에서, BER의 저해제는 가이드 폴리뉴클레오타이드의 부분 또는 분절(예를 들어, 폴리뉴클레오타이드 모티프)과 상호작용할 수 있거나, 이와 회합할 수 있거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드의 추가적인 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)은 BER의 저해제에 융합 또는 연결될 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드에 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가적인 이종성 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머레이스 Ku 결합 모티프 및 Ku 단백질, 텔로머레이스 Sm7 결합 모티프 및 Sm7 단백질 또는 RNA 인식 모티프일 수 있다.

[0126] "베타-2 마이크로글로불린(B2M) 폴리펩타이드"는 UniProt 등록 번호 P61769 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖고 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 B2M 폴리펩타이드 서열이 아래에 제공된다.

[0127] >sp|P61769|B2MG_인간 베타-2-마이크로글로불린 OS=호모 사피엔스 OX=9606 GN=B2M PE=1 SV=1

[0128] MSRSVALAVLALLSLSGLEAIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLL
KNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDM

[0129] "베타-2-마이크로글로불린(B2M) 폴리뉴클레오타이드"는 B2M 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. 베타-2-마이크로글로불린 유전자는 주요 조직적합성 복합체와 관련된 혈청 단백질을 암호화한다. B2M은 숙주 CD8+ T 세포에 의한 비-자기 인식에 관여한다. 예시적인 B2M 폴리뉴클레오타이드 서열이 아래에 제공된다.

[0130] >DQ217933.1 호모 사피엔스 베타-2-마이크로글로빈(B2M) 유전자, 완전 cds

CATGTCATAAATGGTAAGTCCAAGAAAAATACAGGTATTCCCCCCCAAAGAAAACTGTAAAATC
GACTTTTTTCTATCTGTACTGTTTTTTATTGGTTTTTAAATTGGTTTTCCAAGTGAGTAAATCA
GAATCTATCTGTAATGGATTTTAAATTTAGTGTTTCTCTGTGATGTAGTAAACAAGAAACTAGA
GGCAAAAATAGCCCTGTCCCTTGCTAAACTTCTAAGGCACTTTTCTAGTACAACTCAACACTAA
CATTTCAGGCCTTTAGTGCCTTATATGAGTTTTTAAAAGGGGGAAAAGGGAGGGAGCAAGAGTG
TCTTAACTCATACATTTAGGCATAACAATTATTCTCATATTTTAGTTATTGAGAGGGCTGGTAG
AAAAACTAGGTAAATAATATTAATAATTATAGCGCTTATTAAACACTACAGAACACTTACTATG
TACCAGGCATTGTGGGAGGCTCTCTCTTGTGCATTATCTCATTTCATTAGGTCCATGGAGAGTA
TTGCATTTTCTTAGTTTAGGCATGGCCTCCACAATAAAGATTATCAAAAGCCTAAAAATATGTA
AAAGAAACCTAGAAGTTATTTGTTGTGCTCCTTGGGGAAGCTAGGCAAATCCTTTCAACTGAAA
ACCATGGTGACTTCCAAGATCTCTGCCCCTCCCCATCGCCATGGTCCACTTCCTCTTCTCACTG
TTCCTCTTAGAAAAGATCTGTGGACTCCACCACCACGAAATGGCGGCACCTTATTTATGGTCAC
TTTAGAGGGTAGGTTTTCTTAATGGGTCTGCCTGTCATGTTTAACGTCCTTGGCTGGGTCCAAG
GCAGATGCAGTCCAAACTCTCACTAAAATTGCCGAGCCCTTTGTCTTCCAGTGTCTAAAATATT
AATGTCAATGGAATCAGGCCAGAGTTTGAATTCTAGTCTCTTAGCCTTTGTTTCCCCTGTCCAT
AAAATGAATGGGGGTAATTCTTTCCTCCTACAGTTTATTTATATATTCACTAATTCATTCATTC
ATCCATCCATTCGTTCATTCGGTTTACTGAGTACCTACTATGTGCCAGCCCCTGTTCTAGGGTG
GAAACTAAGAGAATGATGTACCTAGAGGGCGCTGGAAGCTCTAAAGCCCTAGCAGTTACTGCTT
TTACTATTAGTGGTCGTTTTTTTCTCCCCCCCGCCCCCCGACAAATCAACAGAACAAAGAAAAT
TACCTAAACAGCAAGGACATAGGGAGGAACTTCTTGGCACAGAACTTTCCAAACACTTTTTCCT
GAAGGGATACAAGAAGCAAGAAAGGTACTCTTTCACTAGGACCTTCTCTGAGCTGTCCTCAGGA
TGCTTTTGGGACTATTTTTCTTACCCAGAGAATGGAGAAACCCTGCAGGGAATTCCCAAGCTGT
AGTTATAAACAGAAGTTCTCCTTCTGCTAGGTAGCATTCAAAGATCTTAATCTTCTGGGTTTCC
GTTTTCTCGAATGAAAAATGCAGGTCCGAGCAGTTAACTGGCTGGGGCACCATTAGCAAGTCAC
TTAGCATCTCTGGGGCCAGTCTGCAAAGCGAGGGGGCAGCCTTAATGTGCCTCCAGCCTGAAGT
CCTAGAATGAGCGCCCGGTGTCCCAAGCTGGGGCGCGCACCCCAGATCGGAGGGCGCCGATGTA
CAGACAGCAAACTCACCCAGTCTAGTGCATGCCTTCTTAAACATCACGAGACTCTAAGAAAAGG
AAACTGAAAACGGGAAAGTCCCTCTCTCTAACCTGGCACTGCGTCGCTGGCTTGGAGACAGGTG
ACGGTCCCTGCGGGCCTTGTCCTGATTGGCTGGGCACGCGTTTAATATAAGTGGAGGCGTCGCG
CTGGCGGGCATTCCTGAAGCTGACAGCATTCGGGCCGAGATGTCTCGCTCCGTGGCCTTAGCTG
TGCTCGCGCTACTCTCTCTTTCTGGCCTGGAGGCTATCCAGCGTGAGTCTCTCCTACCCTCCCG
CTCTGGTCCTTCCTCTCCCGCTCTGCACCCTCTGTGGCCCTCGCTGTGCTCTCTCGCTCCGTGA
CTTCCCTTCTCCAAGTTCTCCTTGGTGGCCCGCCGTGGGGCTAGTCCAGGGCTGGATCTCGGGG
AAGCGGCGGGGTGGCCTGGGAGTGGGGAAGGGGGTGCGCACCCGGGACGCGCGCTACTTGCCCC
[0131] TTTCGGCGGGGAGCAGGGGAGACCTTTGGCCTACGGCGACGGGAGGGTCGGGACAAAGTTTAGG

GCGTCGATAAGCGTCAGAGCGCCGAGGTTGGGGGAGGGTTTCTCTTCCGCTCTTTCGCGGGGCC
TCTGGCTCCCCCAGCGCAGCTGGAGTGGGGGACGGGTAGGCTCGTCCCAAAGGCGCGGCGCTGA
GGTTTGTGAACGCGTGGAGGGGCGCTTGGGGTCTGGGGGAGGCGTCGCCCGGGTAAGCCTGTCT
GCTGCGGCTCTGCTTCCCTTAGACTGGAGAGCTGTGGACTTCGTCTAGGCGCCCGCTAAGTTCG
CATGTCCTAGCACCTCTGGGTCTATGTGGGGCCACACCGTGGGGAGGAAACAGCACGCGACGTT
TGTAGAATGCTTGGCTGTGATACAAAGCGGTTTCGAATAATTAACTTATTTGTTCCCATCACAT
GTCACTTTTAAAAAATTATAAGAACTACCCGTTATTGACATCTTTCTGTGTGCCAAGGACTTTA
TGTGCTTTGCGTCATTTAATTTTGAAAACAGTTATCTTCCGCCATAGATAACTACTATGGTTAT
CTTCTGCCTCTCACAGATGAAGAAACTAAGGCACCGAGATTTTAAGAAACTTAATTACACAGGG
GATAAATGGCAGCAATCGAGATTGAAGTCAAGCCTAACCAGGGCTTTTGCGGGAGCGCATGCCT
TTTGGCTGTAATTCGTGCATTTTTTTTTAAGAAAAACGCCTGCCTTCTGCGTGAGATTCTCCAG
AGCAAACTGGGCGGCATGGGCCCTGTGGTCTTTTCGTACAGAGGGCTTCCTCTTTGGCTCTTTG
CCTGGTTGTTTCCAAGATGTACTGTGCCTCTTACTTTCGGTTTTGAAAACATGAGGGGGTTGGG
CGTGGTAGCTTACGCCTGTAATCCCAGCACTTAGGGAGGCCGAGGCGGGAGGATGGCTTGAGGT
CCGTAGTTGAGACCAGCCTGGCCAACATGGTGAAGCCTGGTCTCTACAAAAAATAATAACAAAA
ATTAGCCGGGTGTGGTGGCTCGTGCCTGTGGTCCCAGCTGCTCCGGTGGCTGAGGCGGGAGGAT
CTCTTGAGCTTAGGCTTTTGAGCTATCATGGCGCCAGTGCACTCCAGCGTGGGCAACAGAGCGA
GACCCTGTCTCTCAAAAAAGAAAAAAAAAAAAAAAGAAAGAGAAAAGAAAAGAAAGAAAGAAGT
GAAGGTTTGTCAGTCAGGGGAGCTGTAAAACCATTAATAAAGATAATCCAAGATGGTTACCAAG
ACTGTTGAGGACGCCAGAGATCTTGAGCACTTTCTAAGTACCTGGCAATACACTAAGCGCGCTC
ACCTTTTCCTCTGGCAAAACATGATCGAAAGCAGAATGTTTTGATCATGAGAAAATTGCATTTA
ATTTGAATACAATTTATTTACAACATAAAGGATAATGTATATATCACCACCATTACTGGTATTT
GCTGGTTATGTTAGATGTCATTTTAAAAAATAACAATCTGATATTTAAAAAAAAATCTTATTTT
GAAAATTTCCAAAGTAATACATGCCATGCATAGACCATTTCTGGAAGATACCACAAGAAACATG
TAATGATGATTGCCTCTGAAGGTCTATTTTCCTCCTCTGACCTGTGTGTGGGTTTTGTTTTTGT
TTTACTGTGGGCATAAATTAATTTTTCAGTTAAGTTTTGGAAGCTTAAATAACTCTCCAAAAGT
CATAAAGCCAGTAACTGGTTGAGCCCAAATTCAAACCCAGCCTGTCTGATACTTGTCCTCTTCT
TAGAAAAGATTACAGTGATGCTCTCACAAAATCTTGCCGCCTTCCCTCAAACAGAGAGTTCCAG
GCAGGATGAATCTGTGCTCTGATCCCTGAGGCATTTAATATGTTCTTATTATTAGAAGCTCAGA
TGCAAAGAGCTCTCTTAGCTTTTAATGTTATGAAAAAAATCAGGTCTTCATTAGATTCCCCAAT
CCACCTCTTGATGGGGCTAGTAGCCTTTCCTTAATGATAGGGTGTTTCTAGAGAGATATATCTG
GTCAAGGTGGCCTGGTACTCCTCCTTCTCCCCACAGCCTCCCAGACAAGGAGGAGTAGCTGCCT
TTTAGTGATCATGTACCCTGAATATAAGTGTATTTAAAAGAATTTTATACACATATATTTAGTG
TCAATCTGTATATTTAGTAGCACTAACACTTCTCTTCATTTTCAATGAAAAATATAGAGTTTAT
[0132] AATATTTTCTTCCCACTTCCCCATGGATGGTCTAGTCATGCCTCTCATTTTGGAAAGTACTGTT

TCTGAAACATTAGGCAATATATTCCCAACCTGGCTAGTTTACAGCAATCACCTGTGGATGCTAA
TTAAAACGCAAATCCCACTGTCACATGCATTACTCCATTTGATCATAATGGAAAGTATGTTCTG
TCCCATTTGCCATAGTCCTCACCTATCCCTGTTGTATTTTATCGGGTCCAACTCAACCATTTAA
GGTATTTGCCAGCTCTTGTATGCATTTAGGTTTTGTTTCTTTGTTTTTTAGCTCATGAAATTAG
GTACAAAGTCAGAGAGGGGTCTGGCATATAAAACCTCAGCAGAAATAAAGAGGTTTTGTTGTTT
GGTAAGAACATACCTTGGGTTGGTTGGGCACGGTGGCTCGTGCCTGTAATCCCAACACTTTGGG
AGGCCAAGGCAGGCTGATCACTTGAAGTTGGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAA
TCCCGTCTCTACTGAAAATACAAAAATTAACCAGGCATGGTGGTGTGTGCCTGTAGTCCCAGGA
ATCACTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATCTCACCACTGCACACTGCACTC
CAGCCTGGGCAATGGAATGAGATTCCATCCCAAAAAATAAAAAAATAAAAAAATAAAGAACATA
CCTTGGGTTGATCCACTTAGGAACCTCAGATAATAACATCTGCCACGTATAGAGCAATTGCTAT
GTCCCAGGCACTCTACTAGACACTTCATACAGTTTAGAAAATCAGATGGGTGTAGATCAAGGCA
GGAGCAGGAACCAAAAAGAAAGGCATAAACATAAGAAAAAAAATGGAAGGGGTGGAAACAGAGT
ACAATAACATGAGTAATTTGATGGGGGCTATTATGAACTGAGAAATGAACTTTGAAAAGTATCT
TGGGGCCAAATCATGTAGACTCTTGAGTGATGTGTTAAGGAATGCTATGAGTGCTGAGAGGGCA
TCAGAAGTCCTTGAGAGCCTCCAGAGAAAGGCTCTTAAAAATGCAGCGCAATCTCCAGTGACAG
AAGATACTGCTAGAAATCTGCTAGAAAAAAAACAAAAAAGGCATGTATAGAGGAATTATGAGGG
AAAGATACCAAGTCACGGTTTATTCTTCAAAATGGAGGTGGCTTGTTGGGAAGGTGGAAGCTCA
TTTGGCCAGAGTGGAAATGGAATTGGGAGAAATCGATGACCAAATGTAAACACTTGGTGCCTGA
TATAGCTTGACACCAAGTTAGCCCCAAGTGAAATACCCTGGCAATATTAATGTGTCTTTTCCCG
ATATTCCTCAGGTACTCCAAAGATTCAGGTTTACTCACGTCATCCAGCAGAGAATGGAAAGTCA
AATTTCCTGAATTGCTATGTGTCTGGGTTTCATCCATCCGACATTGAAGTTGACTTACTGAAGA
ATGGAGAGAGAATTGAAAAAGTGGAGCATTCAGACTTGTCTTTCAGCAAGGACTGGTCTTTCTA
TCTCTTGTACTACACTGAATTCACCCCCACTGAAAAAGATGAGTATGCCTGCCGTGTGAACCAT
GTGACTTTGTCACAGCCCAAGATAGTTAAGTGGGGTAAGTCTTACATTCTTTTGTAAGCTGCTG
AAAGTTGTGTATGAGTAGTCATATCATAAAGCTGCTTTGATATAAAAAAGGTCTATGGCCATAC
TACCCTGAATGAGTCCCATCCCATCTGATATAAACAATCTGCATATTGGGATTGTCAGGGAATG
TTCTTAAAGATCAGATTAGTGGCACCTGCTGAGATACTGATGCACAGCATGGTTTCTGAACCAG
TAGTTTCCCTGCAGTTGAGCAGGGAGCAGCAGCAGCACTTGCACAAATACATATACACTCTTAA
CACTTCTTACCTACTGGCTTCCTCTAGCTTTTGTGGCAGCTTCAGGTATATTTAGCACTGAACG
AACATCTCAAGAAGGTATAGGCCTTTGTTTGTAAGTCCTGCTGTCCTAGCATCCTATAATCCTG
GACTTCTCCAGTACTTTCTGGCTGGATTGGTATCTGAGGCTAGTAGGAAGGGCTTGTTCCTGCT
GGGTAGCTCTAAACAATGTATTCATGGGTAGGAACAGCAGCCTATTCTGCCAGCCTTATTTCTA
ACCATTTTAGACATTTGTTAGTACATGGTATTTTAAAAGTAAAACTTAATGTCTTCCTTTTTTT
[0133] TCTCCACTGTCTTTTTCATAGATCGAGACATGTAAGCAGCATCATGGAGGTAAGTTTTTGACCT

TGAGAAAATGTTTTTGTTTCACTGTCCTGAGGACTATTTATAGACAGCTCTAACATGATAACCC
TCACTATGTGGAGAACATTGACAGAGTAACATTTTAGCAGGGAAAGAAGAATCCTACAGGGTCA
TGTTCCCTTCTCCTGTGGAGTGGCATGAAGAAGGTGTATGGCCCCAGGTATGGCCATATTACTG
ACCCTCTACAGAGAGGGCAAAGGAACTGCCAGTATGGTATTGCAGGATAAAGGCAGGTGGTTAC
CCACATTACCTGCAAGGCTTTGATCTTTCTTCTGCCATTTCCACATTGGACATCTCTGCTGAGG
AGAGAAAATGAACCACTCTTTTCCTTTGTATAATGTTGTTTTATTCTTCAGACAGAAGAGAGGA
GTTATACAGCTCTGCAGACATCCCATTCCTGTATGGGGACTGTGTTTGCCTCTTAGAGGTTCCC
AGGCCACTAGAGGAGATAAAGGGAAACAGATTGTTATAACTTGATATAATGATACTATAATAGA
TGTAACTACAAGGAGCTCCAGAAGCAAGAGAGAGGGAGGAACTTGGACTTCTCTGCATCTTTAG
TTGGAGTCCAAAGGCTTTTCAATGAAATTCTACTGCCCAGGGTACATTGATGCTGAAACCCCAT
TCAAATCTCCTGTTATATTCTAGAACAGGGAATTGATTTGGGAGAGCATCAGGAAGGTGGATGA
TCTGCCCAGTCACACTGTTAGTAAATTGTAGAGCCAGGACCTGAACTCTAATATAGTCATGTGT
TACTTAATGACGGGGACATGTTCTGAGAAATGCTTACACAAACCTAGGTGTTGTAGCCTACTAC
ACGCATAGGCTACATGGTATAGCCTATTGCTCCTAGACTACAAACCTGTACAGCCTGTTACTGT
ACTGAATACTGTGGGCAGTTGTAACACAATGGTAAGTATTTGTGTATCTAAACATAGAAGTTGC
AGTAAAAATATGCTATTTTAATCTTATGAGACCACTGTCATATATACAGTCCATCATTGACCAA
AACATCATATCAGCATTTTTTCTTCTAAGATTTTGGGAGCACCAAAGGGATACACTAACAGGAT
ATACTCTTTATAATGGGTTTGGAGAACTGTCTGCAGCTACTTCTTTTAAAAAGGTGATCTACAC
AGTAGAAATTAGACAAGTTTGGTAATGAGATCTGCAATCCAAATAAAATAAATTCATTGCTAAC
CTTTTTCTTTTCTTTTCAGGTTTGAAGATGCCGCATTTGGATTGGATGAATTCCAAATTCTGCT
TGCTTGCTTTTTAATATTGATATGCTTATACACTTACACTTTATGCACAAAATGTAGGGTTATA
ATAATGTTAACATGGACATGATCTTCTTTATAATTCTACTTTGAGTGCTGTCTCCATGTTTGAT
GTATCTGAGCAGGTTGCTCCACAGGTAGCTCTAGGAGGGCTGGCAACTTAGAGGTGGGGAGCAG
AGAATTCTCTTATCCAACATCAACATCTTGGTCAGATTTGAACTCTTCAATCTCTTGCACTCAA
AGCTTGTTAAGATAGTTAAGCGTGCATAAGTTAACTTCCAATTTACATACTCTGCTTAGAATTT
GGGGGAAAATTTAGAAATATAATTGACAGGATTATTGGAAATTTGTTATAATGAATGAAACATT
TTGTCATATAAGATTCATATTTACTTCTTATACATTTGATAAAGTAAGGCATGGTTGTGGTTAA
TCTGGTTTATTTTTGTTCCACAAGTTAAATAAATCATAAAACTTGATGTGTTATCTCTTATATC
TCACTCCCACTATTACCCCTTTATTTTCAAACAGGGAAACAGTCTTCAAGTTCCACTTGGTAAA
AAATGTGAACCCCTTGTATATAGAGTTTGGCTCACAGTGTAAAGGGCCTCAGTGATTCACATTT
TCCAGATTAGGAATCTGATGCTCAAAGAAGTTAAATGGCATAGTTGGGGTGACACAGCTGTCTA
GTGGGAGGCCAGCCTTCTATATTTTAGCCAGCGTTCTTTCCTGCGGGCCAGGTCATGAGGAGTA
TGCAGACTCTAAGAGGGAGCAAAAGTATCTGAAGGATTTAATATTTTAGCAAGGAATAGATATA
CAATCATCCCTTGGTCTCCCTGGGGGATTGGTTTCAGGACCCCTTCTTGGACACCAAATCTATG
[0134] GATATTTAAGTCCCTTCTATAAAATGGTATAGTATTTGCATATAACCTATCCACATCCTCCTGT

ATACTTTAAATCATTTCTAGATTACTTGTAATACCTAATACAATGTAAATGCTATGCAAATAGT
TGTTATTGTTTAAGGAATAATGACAAGAAAAAAAAGTCTGTACATGCTCAGTAAAGACACAACC
ATCCCTTTTTTTCCCCAGTGTTTTTGATCCATGGTTTGCTGAATCCACAGATGTGGAGCCCCTG
[0135] GATACGGAAGGCCCGCTGTACTTTGAATGACAAATAACAGATTTAAA

[0136] 용어 "Cas9" 또는 "Cas9 도메인"은 Cas9 단백질 또는 이의 단편(예를 들어, Cas9의 활성, 불활성 또는 부분적으로 활성인 DNA 절단 도메인 및/또는 Cas9의 gRNA 결합 도메인을 포함하는 단백질)을 포함하는 RNA-가이드된 뉴클레이스를 지칭한다. Cas9 뉴클레이스는 또한 때때로 Casn1 뉴클레이스 또는 CRISPR("클러스터링된 규칙적 간격의 짧은 회문 반복부(clustered regularly interspaced short palindromic repeat)")-관련 뉴클레이스로도 지칭된다. CRISPR은 이동성 유전 요소(바이러스, 전이성 요소 및 접합성 플라스미드)에 대한 보호를 제공하는 적응 면역 시스템이다. CRISPR 클러스터는 스페이서, 앞선 이동성 요소에 상보적인 서열 및 표적 침입 핵산을 포함한다. CRISPR 클러스터는 전사되어 CRISPR RNA(crRNA)로 처리된다. 유형 II CRISPR 시스템에서, pre-crRNA의 올바른 처리에는 트랜스-암호화된 작은 RNA(tracrRNA), 내인성 리보뉴클레이스 3(rnc) 및 Cas9 단백질이 필요하다. tracrRNA는 pre-crRNA의 리보뉴클레이스 3-보조 처리를 위한 가이드 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 원형 dsDNA 표적을 핵내 분해적으로

(endonucleolytically) 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 핵내 분해적으로 절단된 다음, 3'-5'로 트리밍된다. 사실상, DNA-결합 및 절단은 전형적으로 단백질과 두 RNA가 모두 필요하다. 단일 가이드 RNA("sgRNA" 또는 간단히 "gRNA")는 두 crRNA 및 tracrRNA의 양태를 단일 RNA 종에 통합하도록 조작될 수 있다. 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Jinek M., *et al.*, Charpentier E. *Science* 337:816-821(2012)]을 참조한다. Cas9는 CRISPR 반복 서열에서 짧은 모티프(PAM 또는 프로토스페이서 인접 모티프)를 인식하여 자기 대 비-자기를 구별하는데 도움이 된다. Cas9 뉴클레이스 서열 및 구조는 당업자에게 잘 알려져 있다(예를 들어, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌["Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., *et al.*, Nature 471:602-607(2011); 및 "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., *et al. Science* 337:816-821(2012)] 참조). Cas9 오쏘로그는 S. 피오게네스 및 S. 써모필루스를 포함하지만 이들로 제한되지 않는 다양한 종에서 설명되었다. 추가적인 적합한 Cas9 뉴클레이스 및 서열은 본 개시내용에 기초하여 당업자에게 명백할 것이며, 이러한 Cas9 뉴클레이스 및 서열은 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737]에 개시된 유기체로부터의 서열 및 유전자좌로부터의 Cas9 서열을 포함한다.

[0137] 뉴클레이스-불활성화 Cas9 단백질은 "dCas9" 단백질(뉴클레이스-"활성이 없는" Cas9의 경우) 또는 촉매적으로 불활성인 Cas9로 상호교환적으로 지칭될 수 있다. 불활성 DNA 절단 도메인을 갖는 Cas9 단백질(또는 이의 단편)을 생성하는 방법은 공지되어 있다(예를 들어, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Jinek *et al., Science.* 337:816-821(2012); Qi *et al.,* "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28;152(5):1173-83] 참조). 예를 들어, Cas9의 DNA 절단 도메인은 HNH 뉴클레이스 하위도메인 및 RuvC1 하위도메인의 2개의 하위도메인을 포함하는 것으로 알려져 있다. HNH 하위도메인은 gRNA에 상보적인 가닥을 절단하는 반면, RuvC1 하위도메인은 비-상보적 가닥을 절단한다. 이들 하위도메인 내의 돌연변이는 Cas9의 뉴클레이스 활성을 침묵시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 S. 피오게네스 Cas9의 뉴클레이스 활성을 완전히 불활성화한다(문헌[Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5):1173-83 (2013)]). 일부 실시형태에서, dCas9는 Cas9 뉴클레이스 활성을 불활성화하는 하나 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 상응하거나 또는 이를 부분적으로 또는 전체적으로 포함한다. 일부 실시형태에서, dCas9 도메인은 또 다른 Cas9에서 D10A 및 H840A 돌연변이 또는 상응하는 돌연변이를 포함한다. 일부 실시형태에서, Cas9 뉴클레이스는 불활성(예를 들어, 불활성화된) DNA 절단 도메인을 가지며, 즉, Cas9는 "nCas9" 단백질("닉케이스" Cas9의 경우)로도 지칭되는 닉케이스이다. 이의 변이체 및 상동체를 포함하는 추가적인 Cas9 단백질(예를 들어, 뉴클레이스 활성이 없는 Cas9(dCas9), Cas9 닉케이스(nCas9) 또는 뉴클레이스 활성이 있는 Cas9)은 본 개시내용의 범위 내에 있음이 이해되어야 한다. 예시적인 Cas9 단백질은 제한 없이 본 명세서에 제공된 것들을 포함한다. 일부 실시형태에서, Cas9 단백질은 뉴클레이스 활성이 없는 Cas9(dCas9)이다. 일부 실시형태에서, Cas9 단백질은 Cas9 닉케이스(nCas9)이다. 일부 실시형태에서, Cas9 단백질은 뉴클레이스 활성이 있는 Cas9이다.

[0138] 일부 실시형태에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시형태에서, 단백질은 다음 2개의 Cas9 도메인 중 하나를 포함한다: (1) Cas9의 gRNA 결합 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 실시형태에서, Cas9 또는 이의 단편을 포함하는 단백질은 "Cas9 변이체"로 지칭된다. Cas9 변이체는 Cas9 또는 이의 단편과 상동성을 공유한다. 예를 들어, Cas9 변이체는 야생형 Cas9와 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일하다. 일부 실시형태에서, Cas9 변이체는 야생형 Cas9와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 아미노산 변화를 가질 수 있다. 일부 실시형태에서, Cas9 변이체는 단편이 야생형 Cas9의 상응하는 단편과 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일한 Cas9의 단편(예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함한다. 일부 실시형태에서, 단편은 상응하는 야생형 Cas9의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적

어도 90%, 적어도 95% 동일, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5%이다.

[0139] 일부 실시형태에서, 단편은 적어도 100개의 아미노산 길이이다. 일부 실시형태에서, 단편은 적어도 100개, 150개, 200개, 250개, 300개, 350개, 400개, 450개, 500개, 550개, 600개, 650개, 700개, 750개, 800개, 850개, 900개, 950개, 1000개, 1050개, 1100개, 1150개, 1200개, 1250개 또는 적어도 1300개의 아미노산 길이이다.

[0140] 일부 실시형태에서, Cas9는 코리네박테리움 울서란스(*Corynebacterium ulcerans*)(NCBI 참조: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아(*Corynebacterium diphtheria*)(NCBI 참조: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라(*Spiroplasma syrphidicola*)(NCBI 참조: NC_021284.1); 프레보텔라 인테르메디아(*Prevotella intermedia*)(NCBI 참조: NC_017861.1); 스피로플라스마 타이와넨세(*Spiroplasma taiwanense*)(NCBI 참조: NC_021846.1); 스트렙토코커스 이니아에(*Streptococcus iniae*)(NCBI 참조: NC_021314.1); 벨리엘라 발티카(*Belliella baltica*)(NCBI 참조: NC_018010.1); 싸이크로플렉서스 토르퀴스 I(*Psychroflexus torquisI*)(NCBI 참조: NC_018721.1); 스트렙토코커스 써모필루스(NCBI 참조: YP_820832.1), 리스테리아 이노쿠아(*Listeria innocua*)(NCBI 참조: NP_472073.1), 캄필로박터 제주니(*Campylobacter jejuni*)(NCBI 참조: YP_002344900.1) 또는 네이세리아 메닌지티디스(*Neisseria meningitidis*)(NCBI 참조: YP_002342100.1)로부터의 Cas9 또는 임의의 다른 유기체로부터의 Cas9를 지칭한다.

[0141] 일부 실시형태에서, Cas9는 네이세리아 메닌지티디스(Nme)로부터 유래한다. 일부 실시형태에서, Cas9는 Nme1, Nme2 또는 Nme3이다. 일부 실시형태에서, Nme1, Nme2 또는 Nme3에 대한 PAM-상호작용 도메인은 각각 $N_4$GAT, $N_4$CC 및 $N_4$CAAA이다(예를 들어, 문헌[Edraki, A., *et al.*, A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for *In Vivo* Genome Editing, Molecular Cell (2018)] 참조).

[0142] 일부 실시형태에서, 본 명세서에 제공된 바와 같은 Cas9 융합 단백질은 Cas9 단백질의 전장 아미노산 서열, 예를 들어, 본 명세서에 제공된 Cas9 서열 중 하나를 포함한다. 그러나, 다른 실시형태에서, 본 명세서에 제공된 바와 같은 융합 단백질은 전장 Cas9 서열을 포함하지 않고, 이의 하나 이상의 단편만을 포함한다. 예를 들어, 일부 실시형태에서, 본 명세서에 제공된 Cas9 융합 단백질은 Cas9 단편을 포함하되, 단편은 crRNA 및 tracrRNA 또는 sgRNA에 결합하지만 기능적 뉴클레이스 도메인은 포함하지 않으며, 예를 들어, 이는 뉴클레이스 도메인의 절단된 버전만을 포함하거나 또는 뉴클레이스 도메인을 전혀 포함하지 않는다.

[0143] 적합한 Cas9 도메인 및 Cas9 단편의 예시적인 아미노산 서열이 본 명세서에 제공되며, Cas9 도메인 및 단편의 추가적인 적합한 서열은 당업자에게 명백할 것이다.

[0144] 일부 실시형태에서, Cas9는 단일-세포 원핵 미생물계 및 이의 도메인을 구성하는 고세균(예를 들어, 나노고세균)으로부터의 Cas9를 지칭한다. 일부 실시형태에서, Cas9는 CasX 또는 CasY로 지칭되며, 이는, 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Burstein *et al.*, "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21]에 기술되어 있다. 게놈-분해 메타게노믹스(genome-resolved metagenomics)를 사용하여, 고대 생명체의 도메인에서 처음 보고된 Cas9를 포함하여 다수의 CRISPR-Cas 시스템이 확인되었다. 이러한 다른 Cas9 단백질이 활성 CRISPR-Cas 시스템의 일부로서 거의 연구되지 않은 나노고세균에서 발견되었다. 세균에서, 이전에 알려지지 않은 두 시스템인 CRISPR-CasX 및 CRISPR-CasY가 발견되었으며, 이들은 아직 발견되지 않은 가장 컴팩트한 시스템 중 하나이다. 일부 실시형태에서, Cas9는 CasX 또는 CasX의 변이체를 지칭한다. 일부 실시형태에서, Cas9는 CasY 또는 CasY의 변이체를 지칭한다. 다른 RNA-가이드된 DNA 결합 단백질이 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)로서 사용될 수 있으며, 본 개시내용의 범위 내에 있음을 이해하여야 한다.

[0145] 특정 실시형태에서, 본 발명의 방법에 유용한 napDNAbp는 당업계에 공지되어 있으며, 예를 들어, 문헌[Oakes *et al.*, Cell 176, 254-267, 2019]에 기술되어 있는 원형 순열을 포함한다.

[0146] 염기 편집기에 혼입될 수 있는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 비제한적인 예는 CRISPR 단백질-유래 도메인, 제한 뉴클레이스, 메가뉴클레이스, TAL 뉴클레이스(TAL nuclease: TALEN) 및 아연 핑거 뉴클레이스(아연 핑거 nuclease: ZFN)를 포함한다.

[0147] 일부 실시형태에서, 본 명세서에 제공된 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 또는 임의의 융합 단백질은 CasX 또는 CasY 단백질일 수 있다. 일부 실시형태에서, napDNAbp는 CasX 단백질이다. 일부 실시형태에서, napDNAbp는 CasY 단백질이다. 일부 실시형태에서, napDNAbp는 자연-발생적 CasX 또는 CasY 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적

어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, napDNAbp는 자연-발생적 CasX 또는 CasY 단백질이다. 일부 실시형태에서, napDNAbp는 본 명세서에 기재된 임의의 CasX 또는 CasY 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종으로부터의 CasX 및 CasY가 또한 본 개시내용에 따라 사용될 수 있음을 이해하여야 한다.

[0148] 용어 "Cas12b" 또는 "Cas12b 도메인"은 Cas12b/C2c1 단백질 또는 이의 단편(예를 들어, Cas12b의 활성, 불활성 또는 부분적으로 활성인 DNA 절단 도메인 및/또는 Cas12b의 gRNA 결합 도메인을 포함하는 단백질, 이들 각각의 내용은 참조에 의해 본 명세서에 원용됨)을 포함하는 RNA-가이드된 뉴클레이스를 지칭한다. Cas12b 오쏘로그는 알리사이클로바실러스 아시도테레스트리스(*Alicyclobacillus acidoterrestris*), 알리사이클로바실러스 아시도필루스(*Alicyclobacillus acidophilus*)(문헌[Teng *et al.*, Cell Discov. 2018 Nov 27;4:63]), 바실러스 히사시(*Bacillus hisashi*) 및 바실러스 종 V3-13을 포함하지만 이들로 제한되지 않는 다양한 종에서 기술되어 있다. 추가적인 적합한 Cas12b 뉴클레이스 및 서열은 본 개시내용에 기초하여 당업자에게 명백할 것이다.

[0149] 일부 실시형태에서, Cas12b 또는 이들의 단편을 포함하는 단백질은 "Cas12b 변이체"로 지칭된다. Cas12b 변이체는 Cas12b 또는 이의 단편과 상동성을 공유한다. 예를 들어, Cas12b 변이체는 야생형 Cas12b와 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일하다. 일부 실시형태에서, Cas12b 변이체는 야생형 Cas12b와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 아미노산 변화를 가질 수 있다. 일부 실시형태에서, Cas12b 변이체는 단편이 야생형 Cas12b의 상응하는 단편과 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일한 Cas12b의 단편(예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함한다. 일부 실시형태에서, 단편은 상응하는 야생형 Cas12b의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 동일, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 이다. 예시적인 Cas12b 폴리펩타이드는 본 명세서에 열거되어 있다.

[0150] "Cbl 프로토-종양 유전자 B(Cbl proto-oncogene B: CBLB) 폴리펩타이드"는 면역 반응의 조절에 관여하는 GenBank 등록 번호 ABC86700.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. 예시적인 CBLB 폴리펩타이드 서열이 아래에 제공된다.

[0151] >ABC86700.1 CBL-B [호모 사피엔스]

MANSMNGRNPGGRGGNPRKGRILGIIDAIQDAVGPPKQAAADRRTVEKTWKLMDKVVRLCQNPK
LQLKNSPPYILDILPDTYQHLRLILSKYDDNQKLAQLSENEYFKIYIDSLMKKSKRAIRLFKEG
KERMYEEQSQDRRNLTKLSLIFSHMLAEIKAIFPNGQFQGDNFRITKADAAEFWRKFFGDKTIV
PWKVFRQCLHEVHQISSGLEAMALKSTIDLTCNDYISVFEFDIFTRLFQPWGSILRNWNFLAVT
HPGYMAFLTYDEVKARLQKYSTKPGSYIFRLSCTRLGQWAIGYVTGDGNILQTIPHNKPLFQAL
IDGSREGFYLYPDGRSYNPDLTGLCEPTPHDHIKVTQEQYELYCEMGSTFQLCKICAENDKDVK
IEPCGHLMCTSCLTAWQESDGQGCPFCRCEIKGTEPIIVDPFDPRDEGSRCCSIIDPFGMPMLD
LDDDDDREESLMMNRLANVRKCTDRQNSPVTSPGSSPLAQRRKPQPDPLQIPHLSLPPVPPRLD
LIQKGIVRSPCGSPTGSPKSSPCMVRKQDKPLPAPPPPLRDPPPPPPERPPPIPPDNRLSRHIH
HVESVPSRDPPMPLEAWCPRDVFGTNQLVGCRLLGEGSPKPGITASSNVNGRHSRVGSDPVLMR
KHRRHDLPLEGAKVFSNGHLGSEEYDVPPRLSPPPPVTTLLPSIKCTGPLANSLSEKTRDPVEE
DDDEYKIPSSHPVSLNSQPSHCHNVKPPVRSCDNGHCMLNGTHGPSSEKKSNIPDLSIYLKGDV
FDSASDPVPLPPARPPTRDNPKHGSSLNRTPSDYDLLIPPLGEDAFDALPPSLPPPPPPARHSL
IEHSKPPGSSSRPSSGQDLFLLPSDPFVDLASGQVPLPPARRLPGENVKTNRTSQDYDQLPSCS
DGSQAPARPPKPRPRRTAPEIHHRKPHGPEAALENVDAKIAKLMGEGYAFEEVKRALEIAQNNV
EVARSILREFAFPPPVSPRLNL

[0152]

[0153] "Cbl 프로토-종양 유전자 B(CBLB) 폴리뉴클레오타이드"는 CBLB 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. CBLB 유전자는 E3 유비퀴틴 라이게이스를 암호화한다. 예시적인 CBLB 핵산 서열이 아래에 제공된다.

[0154] >DQ349203.1 호모 사피엔스 CBL-B mRNA, 완전 cds

ATGGCAAACTCAATGAATGGCAGAAACCCTGGTGGTCGAGGAGGAAATCCCCGAAAAGGTCGAA
TTTTGGGTATTATTGATGCTATTCAGGATGCAGTTGGACCCCCTAAGCAAGCTGCCGCAGATCG
CAGGACCGTGGAGAAGACTTGGAAGCTCATGGACAAAGTGGTAAGACTGTGCCAAAATCCCAAA
CTTCAGTTGAAAAATAGCCCACCATATATACTTGATATTTTGCCTGATACATATCAGCATTTAC
GACTTATATTGAGTAAATATGATGACAACCAGAAACTTGCCCAACTCAGTGAGAATGAGTACTT
TAAAATCTACATTGATAGCCTTATGAAAAAGTCAAAACGGGCAATAAGACTCTTTAAAGAAGGC
AAGGAGAGAATGTATGAAGAACAGTCACAGGACAGACGAAATCTCACAAAACTGTCCCTTATCT
TCAGTCACATGCTGGCAGAAATCAAAGCAATCTTTCCCAATGGTCAATTCCAGGGAGATAACTT
TCGTATCACAAAAGCAGATGCTGCTGAATTCTGGAGAAAGTTTTTTGGAGACAAAACTATCGTA
CCATGGAAAGTATTCAGACAGTGCCTTCATGAGGTCCACCAGATTAGCTCTGGCCTGGAAGCAA
TGGCTCTAAAATCAACAATTGATTTAACTTGCAATGATTACATTTCAGTTTTTGAATTTGATAT
TTTTACCAGGCTGTTTCAGCCTTGGGGCTCTATTTTGCGGAATTGGAATTTCTTAGCTGTGACA
CATCCAGGTTACATGGCATTTCTCACATATGATGAAGTTAAAGCACGACTACAGAAATATAGCA
CCAAACCCGGAAGCTATATTTTCCGGTTAAGTTGCACTCGATTGGGACAGTGGGCCATTGGCTA
TGTGACTGGGGATGGGAATATCTTACAGACCATACCTCATAACAAGCCCTTATTTCAAGCCCTG
ATTGATGGCAGCAGGGAAGGATTTTATCTTTATCCTGATGGGAGGAGTTATAATCCTGATTTAA
CTGGATTATGTGAACCTACACCTCATGACCATATAAAAGTTACACAGGAACAATATGAATTATA
TTGTGAAATGGGCTCCACTTTTCAGCTCTGTAAGATTTGTGCAGAGAATGACAAAGATGTCAAG
ATTGAGCCTTGTGGGCATTTGATGTGCACCTCTTGCCTTACGGCATGGCAGGAGTCGGATGGTC
AGGGCTGCCCTTTCTGTCGTTGTGAAATAAAAGGAACTGAGCCCATAATCGTGGACCCCTTTGA
TCCAAGAGATGAAGGCTCCAGGTGTTGCAGCATCATTGACCCCTTTGGCATGCCGATGCTAGAC
TTGGACGACGATGATGATCGTGAGGAGTCCTTGATGATGAATCGGTTGGCAAACGTCCGAAAGT
GCACTGACAGGCAGAACTCACCAGTCACATCACCAGGATCCTCTCCCCTTGCCCAGAGAAGAAA
GCCACAGCCTGACCCACTCCAGATCCCACATCTAAGCCTGCCACCCGTGCCTCCTCGCCTGGAT
CTAATTCAGAAAGGCATAGTTAGATCTCCCTGTGGCAGCCCAACGGGTTCACCAAAGTCTTCTC
CTTGCATGGTGAGAAAACAAGATAAACCACTCCCAGCACCACCTCCTCCCTTAAGAGATCCTCC
TCCACCGCCACCTGAAAGACCTCCACCAATCCCACCAGACAATAGACTGAGTAGACACATCCAT

[0155] CATGTGGAAAGCGTGCCTTCCAGAGACCCGCCAATGCCTCTTGAAGCATGGTGCCCTCGGGATG

TGTTTGGGACTAATCAGCTTGTGGGATGTCGACTCCTAGGGGAGGGCTCTCCAAAACCTGGAAT
CACAGCGAGTTCAAATGTCAATGGAAGGCACAGTAGAGTGGGCTCTGACCCAGTGCTTATGCGG
AAACACAGACGCCATGATTTGCCTTTAGAAGGAGCTAAGGTCTTTTCCAATGGTCACCTTGGAA
GTGAAGAATATGATGTTCCTCCCCGGCTTTCTCCTCCTCCTCCAGTTACCACCCTCCTCCCTAG
CATAAAGTGTACTGGTCCGTTAGCAAATTCTCTTTCAGAGAAAACAAGAGACCCAGTAGAGGAA
GATGATGATGAATACAAGATTCCTTCATCCCACCCTGTTTCCCTGAATTCACAACCATCTCATT
GTCATAATGTAAAACCTCCTGTTCGGTCTTGTGATAATGGTCACTGTATGCTGAATGGAACACA
TGGTCCATCTTCAGAGAAGAAATCAAACATCCCTGACTTAAGCATATATTTAAAGGGAGATGTT
TTTGATTCAGCCTCTGATCCCGTGCCATTACCACCTGCCAGGCCTCCAACTCGGGACAATCCAA
AGCATGGTTCTTCACTCAACAGGACGCCCTCTGATTATGATCTTCTCATCCCTCCATTAGGTGA
AGATGCTTTTGATGCCCTCCCTCCATCTCTCCCACCTCCCCCACCTCCTGCAAGGCATAGTCTC
ATTGAACATTCAAAACCTCCTGGCTCCAGTAGCCGGCCATCCTCAGGACAGGATCTTTTTCTTC
TTCCTTCAGATCCCTTTGTTGATCTAGCAAGTGGCCAAGTTCCTTTGCCTCCTGCTAGAAGGTT
ACCAGGTGAAAATGTCAAAACTAACAGAACATCACAGGACTATGATCAGCTTCCTTCATGTTCA
GATGGTTCACAGGCACCAGCCAGACCCCCTAAACCACGACCGCGCAGGACTGCACCAGAAATTC
ACCACAGAAAACCCCATGGGCCTGAGGCGGCATTGGAAAATGTCGATGCAAAAATTGCAAAACT
CATGGGAGAGGGTTATGCCTTTGAAGAGGTGAAGAGAGCCTTAGAGATAGCCCAGAATAATGTC
GAAGTTGCCCGGAGCATCCTCCGAGAATTTGCCTTCCCTCCTCCAGTATCCCCACGTCTAAATC

[0156] TATAG

[0157] "키메라 항원 수용체" 또는 "CAR"은 항원에 대한 특이성을 면역 효과기 세포에 부여하는 하나 이상의 세포내 신호전달 도메인(예를 들어, T 세포 신호전달 도메인)에 연결된 세포외 항원 결합 도메인을 포함하는 합성 또는 조작된 수용체를 의미한다. 일부 실시형태에서, CAR은 막관통 도메인을 포함한다.

[0158] "키메라 항원 수용체 T 세포" 또는 "CAR-T 세포"는 CAR의 항체-유래 표적화 도메인에 의해 결정되는 항원 특이성을 갖는 CAR을 발현하는 T 세포를 의미한다. 본 명세서에서 사용되는 "CAR-T 세포"는 T 세포 또는 NK 세포를 포함한다. 본 명세서에서 사용되는 "CAR-T 세포"는 CAR 또는 T 세포 수용체(TCR)를 발현하도록 조작된 세포를 포함한다. 일부 실시형태에서, CAR-T 세포는 선택적으로 정의된 비율의 T 헬퍼 CD4+ 및/또는 T 효과기 CD8+ 세포일 수 있다. CAR(예를 들어, 암의 치료용)을 제조하는 방법은 공개적으로 이용 가능하다(예를 들어, 문헌 [Park *et al.*, Trends Biotechnol., 29:550-557, 2011; Grupp *et al.*, N Engl J Med., 368:1509-1518, 2013; Han *et al.*, J. Hematol Oncol. 6:47, 2013; Haso *et al.*, (2013) Blood, 121, 1165-1174]; PCT 공개 WO2012/079000, WO2013/059593; 및 미국 공개 제2012/0213783호 참조, 이들 각각은 그 전문이 참조에 의해 본 명세서에 원용됨).

[0159] "클래스 II, 주요 조직적합성 복합체, 전사촉진인자(class II, major histocompatibility complex, transactivator: CIITA)"는 NCBI 등록 번호 NP_001273331.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0160] >NP_001273331.1 MHC 클래스 II 전사촉진인자 동형단백질 1 [호모 사피엔스]

MRCLAPRPAGSYLSEPQGSSQCATMELGPLEGGYLELLNSDADPLCLYHFYDQMDLAGEEEIEL
YSEPDTDTINCDQFSRLLCDMEGDEETREAYANIAELDQYVFQDSQLEGLSKDIFIEHIGPDEV
IGESMEMPAEVGQKSQKRPFPEELPADLKHWKPAEPPTVVTGSLLVGPVSDCSTLPCLPLPALF
NQEPASGQMRLEKTDQIPMPFSSSSLSCLNLPEGPIQFVPTISTLPHGLWQISEAGTGVSSIFI
YHGEVPQASQVPPPSGFTVHGLPTSPDRPGSTSPFAPSATDLPSMPEPALTSRANMTEHKTSPT
QCPAAGEVSNKLPKWPEPVEQFYRSLQDTYGAEPAGPDGILVEVDLVQARLERSSSKSLERELA
TPDWAERQLAQGGLAEVLLAAKEHRRPRETRVIAVLGKAGQGKSYWAGAVSRAWACGRLPQYDF
VFSVPCHCLNRPGDAYGLQDLLFSLGPQPLVAADEVFSHILKRPDRVLLILDGFEELEAQDGFL
HSTCGPAPAEPCSLRGLLAGLFQKKLLRGCTLLLTARPRGRLVQSLSKADALFELSGFSMEQAQ
AYVMRYFESSGMTEHQDRALTLLRDRPLLLSHSHSPTLCRAVCQLSEALLELGEDAKLPSTLTG
LYVGLLGRAALDSPPGALAELAKLAWELGRRHQSTLQEDQFPSADVRTWAMAKGLVQHPPRAAE
SELAFPSFLLQCFLGALWLALSGEIKDKELPQYLALTPRKKRPYDNWLEGVPRFLAGLIFQPPA
RCLGALLGPSAAASVDRKQKVLARYLKRLQPGTLRARQLLELLHCAHEAEEAGIWQHVVQELPG
RLSFLGTRLTPPDAHVLGKALEAAGQDFSLDLRSTGICPSGLGSLVGLSCVTRFRAALSDTVAL
WESLQQHGETKLLQAAEEKFTIEPFKAKSLKDVEDLGKLVQTQRTRSSSEDTAGELPAVRDLKK
LEFALGPVSGPQAFPKLVRILTAFSSLQHLDLDALSENKIGDEGVSQLSATFPQLKSLETLNLS
QNNITDLGAYKLAEALPSLAASLLRLSLYNNCICDVGAESLARVLPDMVSLRVMDVQYNKFTAA
GAQQLAASLRRCPHVETLAMWTPTIPFSVQEHLQQQDSRISLR

[0161]

[0162] "클래스 II, 주요 조직적합성 복합체, 전사촉진인자(CIITA)"는 CIITA 폴리펩타이드를 암호화하는 핵산을 의미한다. 예시적인 CIIT핵산 서열이 아래에 제공된다.

[0163] >NM_001286402.1 호모 사피엔스 클래스 II 주요 조직적합성 복합체 전사촉진인자(CIITA), 전사체 변이체 1, mRNA

GGTTAGTGATGAGGCTAGTGATGAGGCTGTGTGCTTCTGAGCTGGGCATCCGAAGGCATCCTTG
GGGAAGCTGAGGGCACGAGGAGGGGCTGCCAGACTCCGGGAGCTGCTGCCTGGCTGGGATTCCT
ACACAATGCGTTGCCTGGCTCCACGCCCTGCTGGGTCCTACCTGTCAGAGCCCCAAGGCAGCTC
ACAGTGTGCCACCATGGAGTTGGGGCCCCTAGAAGGTGGCTACCTGGAGCTTCTTAACAGCGAT
GCTGACCCCCTGTGCCTCTACCACTTCTATGACCAGATGGACCTGGCTGGAGAAGAAGAGATTG
AGCTCTACTCAGAACCCGACACAGACACCATCAACTGCGACCAGTTCAGCAGGCTGTTGTGTGA

[0164]

CATGGAAGGTGATGAAGAGACCAGGGAGGCTTATGCCAATATCGCGGAACTGGACCAGTATGTC
TTCCAGGACTCCCAGCTGGAGGGCCTGAGCAAGGACATTTTCATAGAGCACATAGGACCAGATG
AAGTGATCGGTGAGAGTATGGAGATGCCAGCAGAAGTTGGGCAGAAAAGTCAGAAAAGACCCTT
CCCAGAGGAGCTTCCGGCAGACCTGAAGCACTGGAAGCCAGCTGAGCCCCCCACTGTGGTGACT
GGCAGTCTCCTAGTGGGACCAGTGAGCGACTGCTCCACCCTGCCCTGCCTGCCACTGCCTGCGC
TGTTCAACCAGGAGCCAGCCTCCGGCCAGATGCGCCTGGAGAAAACCGACCAGATTCCCATGCC
TTTCTCCAGTTCCTCGTTGAGCTGCCTGAATCTCCCTGAGGGACCCATCCAGTTTGTCCCCACC
ATCTCCACTCTGCCCCATGGGCTCTGGCAAATCTCTGAGGCTGGAACAGGGGTCTCCAGTATAT
TCATCTACCATGGTGAGGTGCCCCAGGCCAGCCAAGTACCCCCTCCCAGTGGATTCACTGTCCA
CGGCCTCCCAACATCTCCAGACCGGCCAGGCTCCACCAGCCCCTTCGCTCCATCAGCCACTGAC
CTGCCCAGCATGCCTGAACCTGCCCTGACCTCCCGAGCAAACATGACAGAGCACAAGACGTCCC
CCACCCAATGCCCGGCAGCTGGAGAGGTCTCCAACAAGCTTCCAAAATGGCCTGAGCCGGTGGA
GCAGTTCTACCGCTCACTGCAGGACACGTATGGTGCCGAGCCCGCAGGCCCGGATGGCATCCTA
GTGGAGGTGGATCTGGTGCAGGCCAGGCTGGAGAGGAGCAGCAGCAAGAGCCTGGAGCGGGAAC
TGGCCACCCCGGACTGGGCAGAACGGCAGCTGGCCCAAGGAGGCCTGGCTGAGGTGCTGTTGGC
TGCCAAGGAGCACCGGCGGCCGCGTGAGACACGAGTGATTGCTGTGCTGGGCAAAGCTGGTCAG
GGCAAGAGCTATTGGGCTGGGGCAGTGAGCCGGGCCTGGGCTTGTGGCCGGCTTCCCCAGTACG
ACTTTGTCTTCTCTGTCCCCTGCCATTGCTTGAACCGTCCGGGGGATGCCTATGGCCTGCAGGA
TCTGCTCTTCTCCCTGGGCCCACAGCCACTCGTGGCGGCCGATGAGGTTTTCAGCCACATCTTG
AAGAGACCTGACCGCGTTCTGCTCATCCTAGACGGCTTCGAGGAGCTGGAAGCGCAAGATGGCT
TCCTGCACAGCACGTGCGGACCGGCACCGGCGGAGCCCTGCTCCCTCCGGGGGCTGCTGGCCGG
CCTTTTCCAGAAGAAGCTGCTCCGAGGTTGCACCCTCCTCCTCACAGCCCGGCCCCGGGGCCGC
CTGGTCCAGAGCCTGAGCAAGGCCGACGCCCTATTTGAGCTGTCCGGCTTCTCCATGGAGCAGG
CCCAGGCATACGTGATGCGCTACTTTGAGAGCTCAGGGATGACAGAGCACCAAGACAGAGCCCT
GACGCTCCTCCGGGACCGGCCACTTCTTCTCAGTCACAGCCACAGCCCTACTTTGTGCCGGGCA
GTGTGCCAGCTCTCAGAGGCCCTGCTGGAGCTTGGGGAGGACGCCAAGCTGCCCTCCACGCTCA
CGGGACTCTATGTCGGCCTGCTGGGCCGTGCAGCCCTCGACAGCCCCCCCGGGGCCCTGGCAGA
GCTGGCCAAGCTGGCCTGGGAGCTGGGCCGCAGACATCAAAGTACCCTACAGGAGGACCAGTTC
CCATCCGCAGACGTGAGGACCTGGGCGATGGCCAAAGGCTTAGTCCAACACCCACCGCGGGCCG
CAGAGTCCGAGCTGGCCTTCCCCAGCTTCCTCCTGCAATGCTTCCTGGGGGCCCTGTGGCTGGC
TCTGAGTGGCGAAATCAAGGACAAGGAGCTCCCGCAGTACCTAGCATTGACCCCAAGGAAGAAG
AGGCCCTATGACAACTGGCTGGAGGGCGTGCCACGCTTTCTGGCTGGGCTGATCTTCCAGCCTC
CCGCCCGCTGCCTGGGAGCCCTACTCGGGCCATCGGCGGCTGCCTCGGTGGACAGGAAGCAGAA
GGTGCTTGCGAGGTACCTGAAGCGGCTGCAGCCGGGGACACTGCGGGCGCGGCAGCTGCTGGAG
CTGCTGCACTGCGCCCACGAGGCCGAGGAGGCTGGAATTTGGCAGCACGTGGTACAGGAGCTCC

[0165]

CCGGCCGCCTCTCTTTTCTGGGCACCCGCCTCACGCCTCCTGATGCACATGTACTGGGCAAGGC
CTTGGAGGCGGCGGGCCAAGACTTCTCCCTGGACCTCCGCAGCACTGGCATTTGCCCCTCTGGA
TTGGGGAGCCTCGTGGGACTCAGCTGTGTCACCCGTTTCAGGGCTGCCTTGAGCGACACGGTGG
CGCTGTGGGAGTCCCTGCAGCAGCATGGGGAGACCAAGCTACTTCAGGCAGCAGAGGAGAAGTT
CACCATCGAGCCTTTCAAAGCCAAGTCCCTGAAGGATGTGGAAGACCTGGGAAAGCTTGTGCAG
ACTCAGAGGACGAGAAGTTCCTCGGAAGACACAGCTGGGGAGCTCCCTGCTGTTCGGGACCTAA
AGAAACTGGAGTTTGCGCTGGGCCCTGTCTCAGGCCCCCAGGCTTTCCCCAAACTGGTGCGGAT
CCTCACGGCCTTTTCCTCCCTGCAGCATCTGGACCTGGATGCGCTGAGTGAGAACAAGATCGGG
GACGAGGGTGTCTCGCAGCTCTCAGCCACCTTCCCCCAGCTGAAGTCCTTGGAAACCCTCAATC
TGTCCCAGAACAACATCACTGACCTGGGTGCCTACAAACTCGCCGAGGCCCTGCCTTCGCTCGC
TGCATCCCTGCTCAGGCTAAGCTTGTACAATAACTGCATCTGCGACGTGGGAGCCGAGAGCTTG
GCTCGTGTGCTTCCGGACATGGTGTCCCTCCGGGTGATGGACGTCCAGTACAACAAGTTCACGG
CTGCCGGGGCCCAGCAGCTCGCTGCCAGCCTTCGGAGGTGTCCTCATGTGGAGACGCTGGCGAT
GTGGACGCCCACCATCCCATTCAGTGTCCAGGAACACCTGCAACAACAGGATTCACGGATCAGC
CTGAGATGATCCCAGCTGTGCTCTGGACAGGCATGTTCTCTGAGGACACTAACCACGCTGGACC
TTGAACTGGGTACTTGTGGACACAGCTCTTCTCCAGGCTGTATCCCATGAGCCTCAGCATCCTG
GCACCCGGCCCCTGCTGGTTCAGGGTTGGCCCCTGCCCGGCTGCGGAATGAACCACATCTTGCT
CTGCTGACAGACACAGGCCCGGCTCCAGGCTCCTTTAGCGCCCAGTTGGGTGGATGCCTGGTGG
CAGCTGCGGTCCACCCAGGAGCCCCGAGGCCTTCTCTGAAGGACATTGCGGACAGCCACGGCCA
GGCCAGAGGGAGTGACAGAGGCAGCCCCATTCTGCCTGCCCAGGCCCCTGCCACCCTGGGGAGA
AAGTACTTCTTTTTTTTTATTTTTAGACAGAGTCTCACTGTTGCCCAGGCTGGCGTGCAGTGGT
GCGATCTGGGTTCACTGCAACCTCCGCCTCTTGGGTTCAAGCGATTCTTCTGCTTCAGCCTCCC
GAGTAGCTGGGACTACAGGCACCCACCATCATGTCTGGCTAATTTTTCATTTTTAGTAGAGACA
GGGTTTTGCCATGTTGGCCAGGCTGGTCTCAAACTCTTGACCTCAGGTGATCCACCCACCTCAG
CCTCCCAAAGTGCTGGGATTACAAGCGTGAGCCACTGCACCGGGCCACAGAGAAAGTACTTCTC
CACCCTGCTCTCCGACCAGACACCTTGACAGGGCACACCGGGCACTCAGAAGACACTGATGGGC
AACCCCCAGCCTGCTAATTCCCCAGATTGCAACAGGCTGGGCTTCAGTGGCAGCTGCTTTTGTC
TATGGGACTCAATGCACTGACATTGTTGGCCAAAGCCAAAGCTAGGCCTGGCCAGATGCACCAG
CCCTTAGCAGGGAAACAGCTAATGGGACACTAATGGGGCGGTGAGAGGGGAACAGACTGGAAGC
ACAGCTTCATTTCCTGTGTCTTTTTTCACTACATTATAAATGTCTCTTTAATGTCACAGGCAGG
TCCAGGGTTTGAGTTCATACCCTGTTACCATTTTGGGGTACCCACTGCTCTGGTTATCTAATAT
[0166] GTAACAAGCCACCCCAAATCATAGTGGCTTAAAACAACACTCACATTTA

[0167] 본 개시내용에서, "포함하다(comprises)", "포함하는(comprising)", "포함하는(containing)" 및 "갖는(having)" 등은 미국 특허법에서 부여된 의미를 가질 수 있으며, "포함하다(includes)", "포함하는(including)" 등을 의미할 수 있고; "본질적으로 구성되는(consisting essentially of)" 또는 "본질적으로 구성되는(consists essentially)"도 마찬가지로 미국 특허법에 정의된 의미를 가지며, 용어는 개방형(open-ended)이며, 인용된 것 이상의 존재에 의해 인용된 것의 기본 또는 신규한 특성이 변경되지 않는 한 인용된 것보다 더 많은 존재를 허용하지만 선행 기술의 실시형태는 배제한다.

[0168] "세포독성 T-림프구 관련 단백질 4(cytotoxic T-lymphocyte associated protein 4: CTLA-4) 폴리펩타이드"는 NCBI 등록 번호 EAW70354.1 또는 이의 단편에 대해 적어도 약 85%의 서열 동일성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다:

[0169] >EAW70354.1 세포독성 T-림프구-관련 단백질 4 [호모 사피엔스]

MACLGFQRHKAQLNLATRTWPCTLLFFLLFIPVFCKAMHVAQPAVVLASSRGIASFVCEYASPG
KATEVRVTVLRQADSQVTEVCAATYMMGNELTFLDDSICTGTSSGNQVNLTIQGLRAMDTGLYI
CKVELMYPPPYYLGIGNGTQIYVIDPEPCPDSDFLLWILAAVSSGLFFYSFLLTAVSLSKMLKK
[0170] RSPLTTGVYVKMPPTEPECEKQFQPYFIPIN

[0171] "세포독성 T-림프구 관련 단백질 4(CTLA-4) 폴리뉴클레오타이드"는 CTLA-4 폴리펩타이드를 암호화하는 핵산 분

자를 의미한다. CTLA-4 유전자는 면역글로불린 패밀리를 암호화하고, 저해 신호를 T 세포에 전달하는 단백질을 암호화한다. 예시적인 CTLA-4 핵산 서열이 아래에 제공된다.

[0172] >BC074842.2 호모 사피엔스 세포독성 T-림프구-관련 단백질 4, mRNA(cDNA 클론 MGC:104099 IMAGE:30915552), 완전 cds

GACCTGAACACCGCTCCCATAAAGCCATGGCTTGCCTTGGATTTCAGCGGCACAAGGCTCAGCT
GAACCTGGCTACCAGGACCTGGCCCTGCACTCTCCTGTTTTTTCTTCTCTTCATCCCTGTCTTC
TGCAAAGCAATGCACGTGGCCCAGCCTGCTGTGGTACTGGCCAGCAGCCGAGGCATCGCCAGCT
TTGTGTGTGAGTATGCATCTCCAGGCAAAGCCACTGAGGTCCGGGTGACAGTGCTTCGGCAGGC
TGACAGCCAGGTGACTGAAGTCTGTGCGGCAACCTACATGATGGGGAATGAGTTGACCTTCCTA
GATGATTCCATCTGCACGGGCACCTCCAGTGGAAATCAAGTGAACCTCACTATCCAAGGACTGA
GGGCCATGGACACGGGACTCTACATCTGCAAGGTGGAGCTCATGTACCCACCGCCATACTACCT
GGGCATAGGCAACGGAACCCAGATTTATGTAATTGATCCAGAACCGTGCCCAGATTCTGACTTC
CTCCTCTGGATCCTTGCAGCAGTTAGTTCGGGGTTGTTTTTTTATAGCTTTCTCCTCACAGCTG
TTTCTTTGAGCAAAATGCTAAAGAAAAGAAGCCCTCTTACAACAGGGGTCTATGTGAAAATGCC
CCCAACAGAGCCAGAATGTGAAAAGCAATTTCAGCCTTATTTTATTCCCATCAATTGAGAAACC

[0173] ATTATGAAGAAGAGAGTCCATATTTCAATTTCCAAGAGCTGAGG

[0174] "분화 클러스터 2(CD2) 폴리펩타이드"는 NCBI 등록 번호 NP_001758.2 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0175] >NP_001758.2 T-세포 표면 항원 CD2 동형단백질 2 전구체 [호모 사피엔스]

```
  1 MSFPCKFVAS FLLIFNVSSK GAVSKEITNA LETWGALGQD INLDIPSFQM SDDIDDIKWE
 61 KTSDKKKIAQ FRKEKETFKE KDTYKLFKNG TLKIKHLKTD DQDIYKVSIY DTKGKNVLEK
121 IFDLKIQERV SKPKISWTCI NTTLTCEVMN GTDPELNLYQ DGKHLKLSQR VITHKWTTSL
181 SAKFKCTAGN KVSKESSVEP VSCPEKGLDI YLIIGICGGG SLLMVFVALL VFYITKRKKQ
241 RSRRNDEELE TRAHRVATEE RGRKPHQIPA STPQNPATSQ HPPPPPGHRS QAPSHRPPPP
301 GHRVQHQPQK RPPAPSGTQV HQQKGPPLPR PRVQPKPPHG AAENSLSPSS N
```

[0176]

[0177] CD2 세포질 도메인(235번 내지 351번 아미노산 잔기)은 굵은 글씨로 표시된다. 호모 사피엔스로부터의 예시적인 CD2 폴리펩타이드의 구조는 **도 4**에 나타나 있다.

[0178] "분화 클러스터 2(CD2) 폴리뉴클레오타이드"는 CD2 폴리펩타이드를 암호화하는 핵산을 의미한다. 예시적인 CD2 핵산 서열이 아래에 제공된다.

[0179] > NM_001767.5 호모 사피엔스 CD2 분자(CD2), 전사체 변이체 2, mRNA

```
   1 agtctcactt cagttccttt tgcatgaaga gctcagaatc aaaagaggaa accaacccct
  61 aagatgagct ttccatgtaa atttgtagcc agcttccttc tgattttcaa tgtttcttcc
 121 aaaggtgcag tctccaaaga gattacgaat gccttggaaa cctggggtgc cttgggtcag
 181 gacatcaact tggacattcc tagttttcaa atgagtgatg atattgacga tataaaatgg
 241 gaaaaaactt cagacaagaa aaagattgca caattcagaa aagagaaaga gactttcaag
 301 gaaaaagata catataagct atttaaaaat ggaactctga aaattaagca tctgaagacc
 361 gatgatcagg atatctacaa ggtatcaata tatgatacaa aaggaaaaaa tgtgttggaa
 421 aaaatatttg atttgaagat tcaagagagg gtctcaaaac caaagatctc ctggacttgt
 481 atcaacacaa ccctgacctg tgaggtaatg aatggaactg accccgaatt aaacctgtat
 541 caagatggga aacatctaaa actttctcag agggtcatca cacacaagtg gaccaccagc
 601 ctgagtgcaa aattcaagtg cacagcaggg aacaaagtca gcaaggaatc cagtgtcgag
 661 cctgtcagct gtccagagaa aggtctggac atctatctca tcattggcat atgtggagga
 721 ggcagcctct tgatggtctt tgtggcactg ctcgttttct atatcaccaa aaggaaaaaa
 781 cagaggagtc ggagaaatga tgaggagctg gagacaagag cccacagagt agctactgaa
 841 gaaaggggcc ggaagcccca ccaaattcca gcttcaaccc ctcagaatcc agcaacttcc
 901 caacatcctc ctccaccacc tggtcatcgt tcccaggcac ctagtcatcg tcccccgcct
 961 cctggacacc gtgttcagca ccagcctcag aagaggcctc ctgctccgtc gggcacacaa
1021 gttcaccagc agaaaggccc gcccctcccc agacctcgag ttcagccaaa acctccccat
1081 ggggcagcag aaaactcatt gtccccttcc tctaattaaa aaagatagaa actgtctttt
1141 tcaataaaaa gcactgtgga tttctgccct cctgatgtgc atatccgtac ttccatgagg
1201 tgttttctgt gtgcagaaca ttgtcacctc ctgaggctgt gggccacagc cacctctgca
1261 tcttcgaact cagccatgtg gtcaacatct ggagtttttg gtctcctcag agagctccat
1321 cacaccagta aggagaagca atataagtgt gattgcaaga atggtagagg accgagcaca
1381 gaaatcttag agatttcttg tcccctctca ggtcatgtgt agatgcgata aatcaagtga
1441 ttggtgtgcc tgggtctcac tacaagcagc ctatctgctt aagagactct ggagtttctt
1501 atgtgccctg gtggacactt gcccaccatc ctgtgagtaa aagtgaaata aaagctttga
1561 ctaga
```

[0180]

[0181] "분화 클러스터 5(CD5) 폴리펩타이드"는 NCBI 등록 번호 NP_001333385.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0182] >NP_001333385.1 T-세포 표면 당단백질 CD5 동형단백질 2 [호모 사피엔스]

```
MVCSQSWGRSSKQWEDPSQASKVCQRLNCGVPLSLGPFLVTYTPQSSIICYGQLGSFSNCSHSR
NDMCHSLGLTCLEPQKTTPPTTRPPPTTTPEPTAPPRLQLVAQSGGQHCAGVVEFYSGSLGGTI
SYEAQDKTQDLENFLCNNLQCGSFLKHLPETEAGRAQDPGEPREHQPLPIQWKIQNSSCTSLEH
CFRKIKPQKSGRVLALLCSGFQPKVQSRLVGGSSICEGTVEVRQGAQWAALCDSSSARSSLRWE
EVCREQQCGSVNSYRVLDAGDPTSRGLFCPHQKLSQCHELWERNSYCKKVFVTCQDPNPAGLAA
GTVASIILALVLLVVLLVVCGPLAYKKLVKKFRQKKQRQWIGPTGMNQNMSFHRNHTATVRSHA
ENPTASHVDNEYSQPPRNSHLSAYPALEGALHRSSMQPDNSSDSDYDLHGAQRL
```

[0183]

[0184] "분화 클러스터 5(CD5) 폴리뉴클레오타이드"는 CD5 폴리펩타이드를 암호화하는 핵산을 의미한다. 예시적인 CD5 핵산 서열이 아래에 제공된다.

[0185] >NM_001346456.1 호모 사피엔스 CD5 분자(CD5), 전사체 변이체 2, mRNA

```
   1 gagtcttgct gatgctcccg gctgaataaa ccccttcctt ctttaacttg gtgtctgagg
  61 ggttttgtct gtggcttgtc ctgctacatt tcttggttcc ctgaccagga agcaaagtga
 121 ttaacggaca gttgaggcag ccccttaggc agcttaggcc tgccttgtgg agcatccccg
 181 cggggaactc tggccagctt gagcgacacg gatcctcaga gcgctcccag gtaggcaatt
 241 gccccagtgg aatgcctcgt cagagcagtg catggcaggc ccctgtggag gatcaacgca
 301 gtggctgaac acagggaagg aactggcact tggagtccgg acaactgaaa cttgtcgctt
 361 cctgcctcgg acggctcagc tggtatgacc cagatttcca ggcaaggctc acccgttcca
 421 actcgaagtg ccagggccag ctggaggtct acctcaagga cggatggcac atggtttgca
 481 gccagagctg gggccggagc tccaagcagt gggaggaccc cagtcaagcg tcaaaagtct
 541 gccagcggct gaactgtggg gtgcccttaa gccttggccc cttccttgtc acctacacac
 601 ctcagagctc aatcatctgc tacggacaac tgggctcctt ctccaactgc agccacagca
 661 gaaatgacat gtgtcactct ctgggcctga cctgcttaga accccagaag acaacacctc
 721 caacgacaag gcccccgccc accacaactc cagagcccac agctcctccc aggctgcagc
 781 tggtggcaca gtctggcggc cagcactgtg ccggcgtggt ggagttctac agcggcagcc
 841 tgggggggtac catcagctat gaggcccagg acaagaccca ggacctggag aacttcctct
 901 gcaacaacct ccagtgtggc tccttcttga agcatctgcc agagactgag gcaggcagag
 961 cccaagaccc aggggagcca cgggaacacc agcccttgcc aatccaatgg aagatccaga
1021 actcaagctg tacctccctg gagcattgct tcaggaaaat caagccccag aaaagtggcc
1081 gagttcttgc cctcctttgc tcaggtttcc agcccaaggt gcagagccgt ctggtggggg
1141 gcagcagcat ctgtgaaggc accgtggagg tgcgccaggg ggctcagtgg gcagccctgt
1201 gtgacagctc ttcagccagg agctcgctgc ggtgggagga ggtgtgccgg gagcagcagt
1261 gtggcagcgt caactcctat cgagtgctgg acgctggtga cccaacatcc cgggggctct
1321 tctgtcccca tcagaagctg tcccagtgcc acgaactttg ggagagaaat tcctactgca
1381 agaaggtgtt tgtcacatgc caggatccaa accccgcagg cctggccgca ggcacggtgg
1441 caagcatcat cctggccctg gtgctcctgg tggtgctgct ggtcgtgtgc ggccccttg
1501 cctacaagaa gctagtgaag aaattccgcc agaagaagca gcgccagtgg attggcccaa
```

[0186]

```
1561 cgggaatgaa ccaaaacatg tctttccatc gcaaccacac ggcaaccgtc cgatcccatg
1621 ctgagaaccc cacagcctcc cacgtggata acgaatacag ccaacctccc aggaactccc
1681 acctgtcagc ttatccagct ctggaagggg ctctgcatcg ctcctccatg cagcctgaca
1741 actcctccga cagtgactat gatctgcatg gggctcagag gctgtaaaga actgggatcc
1801 atgagcaaaa agccgagagc cagacctgtt tgtcctgaga aaactgtccg ctcttcactt
1861 gaaatcatgt ccctatttct accccggcca gaacatggac agaggccaga agccttccgg
1921 acaggcgctg ctgccccgag tggcaggcca gctcacactc tgctgcacaa cagctcggcc
1981 gcccctccac ttgtggaagc tgtggtgggc agagccccaa aacaagcagc cttccaacta
2041 gagactcggg ggtgtctgaa gggggccccc tttccctgcc cgctggggag cggcgtctca
2101 gtgaaatcgg ctttctcctc agactctgtc cctggtaagg agtgacaagg aagctcacag
2161 ctgggcgagt gcattttgaa tagttttttg taagtagtgc ttttcctcct tcctgacaaa
2221 tcgagcgctt tggcctcttc tgtgcagcat ccacccctgc ggatcctct ggggaggaca
2281 ggaaggggac tcccggagac ctctgcagcc gtggtggtca gaggctgctc acctgagcac
2341 aaagacagct ctgcacattc accgcagctg ccagccaggg gtctgggtgg gcaccaccct
2401 gacccacagc gtcaccccac tccctctgtc ttatgactcc cctccccaac cccctcatct
2461 aaagacacct tcctttccac tggctgtcaa gcccacaggg caccagtgcc acccagggcc
2521 cggcacaaag gggcgcctag taaaccttaa ccaacttggt tttttgcttc acccagcaat
2581 taaaagtccc aagctgaggt agtttcagtc catcacagtt catcttctaa cccaagagtc
2641 agagatgggg ctggtcatgt tcctttggtt tgaataactc ccttgacgaa aacagactcc
2701 tctagtactt ggagatcttg gacgtacacc taatcccatg gggcctcggc ttccttaact
2761 gcaagtgaga agaggaggtc tacccaggag cctcgggtct gatcaaggga gaggccaggc
2821 gcagctcact gcggcggctc cctaagaagg tgaagcaaca tgggaacaca tcctaagaca
2881 ggtccttttct ccacgccatt tgatgctgta tctcctggga gcacaggcat caatggtcca
2941 agccgcataa taagtctgga agagcaaaag ggagttacta ggatatgggg tgggctgctc
3001 ccagaatctg ctcagctttc tgcccccacc aacaccctcc aaccaggcct tgccttctga
3061 gagcccccgt ggccaagccc aggtcacaga tcttcccccg accatgctgg gaatccagaa
3121 acagggaccc catttgtctt cccatatctg gtggaggtga gggggctcct caaaagggaa
3181 ctgagaggct gctcttaggg agggcaaagg ttcgggggca gccagtgtct cccatcagtg
3241 cctttttaa taaaagctct ttcatctata gtttggccac catacagtgg cctcaaagca
3301 accatggcct acttaaaaac caaaccaaaa ataaagagtt tagttgagga gaaaaaaaaa
3361 aaaaaaaaaa aaaaaa
```

[0187]

[0188] "분화 클러스터 7(CD7) 폴리펩타이드"는 NCBI 참조 서열: NP_006128.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0189] >NP_006128.1 T-세포 항원 CD7 전구체 [호모 사피엔스]

```
  1 MAGPPRLLLL PLLLALARGL PGALAAQEVQ QSPHCTTVPV GASVNITCST SGGLRGIYLR
 61 QLGPQPQDII YYEDGVVPTT DRRFRGRIDF SGSQDNLTIT MHRLQLSDTG TYTCQAITEV
121 NVYGSGTLVL VTEEQSQGWH RCSDAPPRAS ALPAPPTGSA LPDPQTASAL PDPPAASALP
181 AALAVISFLL GLGLGVACVL ARTQIKKLCS WRDKNSAACV VYEDMSHSRC NTLSSPNQYQ
```

[0190]

[0191] "분화 클러스터 7(CD7) 폴리뉴클레오타이드"는 CD7 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. 예시적인 CD7 핵산 서열이 아래에 제공된다.

[0192] >NM_006137.7 호모 사피엔스 CD7 분자(CD7), mRNA

```
   1 ctctctgagc tctgagcgcc tgcggtctcc tgtgtgctgc tctctgtggg gtcctgtaga
  61 cccagagagg ctcagctgca ctcgcccggc tgggagagct gggtgtgggg aacatggccg
 121 ggcctccgag gctcctgctg ctgcccctgc ttctggcgct ggctcgcggc ctgcctgggg
 181 ccctggctgc ccaagaggtg cagcagtctc cccactgcac gactgtcccc gtgggagcct
 241 ccgtcaacat cacctgctcc accagcgggg gcctgcgtgg gatctacctg aggcagctcg
 301 ggccacagcc ccaagacatc atttactacg aggacggggt ggtgcccact acggacagac
 361 ggttccgggg ccgcatcgac ttctcagggt cccaggacaa cctgactatc accatgcacc
 421 gcctgcagct gtcggacact ggcacctaca cctgccaggc catcacggag gtcaatgtct
 481 acggctccgg caccctggtc ctggtgacag aggaacagtc ccaaggatgg cacagatgct
 541 cggacgcccc accaagggcc tctgccctcc ctgccccacc gacaggctcc gccctccctg
 601 acccgcagac agcctctgcc ctccctgacc cgccagcagc ctctgccctc cctgcggccc
 661 tggcggtgat ctccttcctc ctcgggctgg gcctgggggt ggcgtgtgtg ctggcgagga
 721 cacagataaa gaaactgtgc tcgtggcggg ataagaattc ggcggcatgt gtggtgtacg
 781 aggacatgtc gcacagccgc tgcaacacgc tgtcctcccc caaccagtac cagtgaccca
 841 gtgggccccc gcacgtcccg cctgtggtcc ccccagcacc ttccctgccc caccatgccc
 901 cccaccctgc cacacccctc accctgctgt cctcccacgg ctgcagcaga gtttgaaggg
 961 cccagccgtg cccagctcca agcagacaca caggcagtgg ccaggcccca cggtgcttct
1021 cagtggacaa tgatgcctcc tccgggaagc cttccctgcc cagcccacgc cgccaccggg
1081 aggaagcctg actgtccttt ggctgcatct cccgaccatg gccaaggagg gcttttctgt
1141 gggatgggcc tgggcacgcg gccctctcct gtcagtgccg gcccacccac cagcaggccc
1201 ccaaccccca ggcagcccgg cagaggacgg gaggagacca gtcccccacc cagccgtacc
1261 agaaataaag gcttctgtgc ttcc
```

[0193]

[0194] "분화 클러스터 33(CD33) 폴리펩타이드"는 NCBI 참조 서열: NP_001763.3 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. CD33은 또한 Siglec-3으로도 알려져 있다. 예시적인 아미노산 서열이 아래에 제공된다.

[0195] > NP_001763.3 골수 세포 표면 항원 CD33 동형단백질 1 전구체 [호모 사피엔스]

```
  1 MPLLLLLPLL WAGALAMDPN FWLQVQESVT VQEGLCVLVP CTFFHPIPYY DKNSPVHGYW
 61 FREGAIISRD SPVATNKLDQ EVQEETQGRF RLLGDPSRNN CSLSIVDARR RDNGSYFFRM
121 ERGSTKYSYK SPQLSVHVTD LTHRPKILIP GTLEPGHSKN LTCSVSWACE QGTPPIFSWL
181 SAAPTSLGPR TTHSSVLIIT PRPQDHGTNL TCQVKFAGAG VTTERTIQLN VTYVPQNPTT
241 GIFPGDGSGK QETRAGVVHG AIGGAGVTAL LALCLCLIFF IVKTHRRKAA RTAVGRNDTH
301 PTTGSASPKH QKKSKLHGPT ETSSCSGAAP TVEMDEELHY ASLNFHGMNP SKDTSTEYSE
361 VRTQ
```

[0196]

[0197] "분화 클러스터 33(CD33) 폴리뉴클레오타이드"는 CD33 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. 예시적인 CD33 핵산 서열이 아래에 제공된다.

[0198] > NM_001772.4 호모 사피엔스 CD33 분자(CD33), 전사체 변이체 1, mRNA

```
   1 ctgctcacac aggaagccct ggaagctgct tcctcagaca tgccgctgct gctactgctg
  61 cccctgctgt gggcaggggc cctggctatg gatccaaatt tctggctgca agtgcaggag
 121 tcagtgacgg tacaggaggg tttgtgcgtc ctcgtgccct gcactttctt ccatcccata
 181 ccctactacg acaagaactc cccagttcat ggttactggt tccgggaagg agccattata
 241 tccagggact ctccagtggc cacaaacaag ctagatcaag aagtacagga ggagactcag
 301 ggcagattcc gcctccttgg ggatcccagt aggaacaact gctccctgag catcgtagac
 361 gccaggagga gggataatgg ttcatacttc tttcggatgg agagaggaag taccaaatac
 421 agttacaaat ctccccagct ctctgtgcat gtgacagact tgacccacag gcccaaaatc
 481 ctcatccctg gcactctaga acccggccac tccaaaaacc tgacctgctc tgtgtcctgg
 541 gcctgtgagc agggaacacc cccgatcttc tcctggttgt cagctgcccc cacctccctg
 601 ggccccagga ctactcactc ctcggtgctc ataatcaccc cacggcccca ggaccacggc
 661 accaacctga cctgtcaggt gaagttcgct ggagctggtg tgactacgga gagaaccatc
 721 cagctcaacg tcacctatgt tccacagaac ccaacaactg gtatctttcc aggagatggc
 781 tcagggaaac aagagaccag agcaggagtg gttcatgggg ccattggagg agctggtgtt
 841 acagccctgc tcgctctttg tctctgcctc atcttcttca tagtgaagac ccacaggagg
 901 aaagcagcca ggacagcagt gggcaggaat gacacccacc ctaccacagg gtcagcctcc
 961 ccgaaacacc agaagaagtc caagttacat ggccccactg aaacctcaag ctgttcaggt
```

[0199]

```
1021 gccgccccta ctgtggagat ggatgaggag ctgcattatg cttccctcaa ctttcatggg
1081 atgaatcctt ccaaggacac ctccaccgaa tactcagagg tcaggaccca gtgaggaacc
1141 cacaagagca tcaggctcag ctagaagatc cacatcctct acaggtcggg gaccaaaggc
1201 tgattcttgg agatttaaca ccccacaggc aatgggttta tagacattat gtgagtttcc
1261 tgctatatta acatcatctt agactttgca agcagagagt cgtggaatca aatctgtgct
1321 ctttcatttg ctaagtgtat gatgtcacac aagctcctta accttccatg tctccatttt
1381 cttctctgtg aagtaggtat aagaagtcct atctcatagg gatgctgtga gcattaaata
1441 aaggtacaca tggaaaacac ca
```

[0200]

[0201] "분화 클러스터 52(CD52) 폴리펩타이드"는 NCBI 참조 서열: NP_001794.2 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. CD52는 CAMPATH-1로도 알려져 있다. 예시적인 아미노산 서열이 아래에 제공된다.

[0202] > NP_001794.2 CAMPATH-1 항원 전구체 [호모 사피엔스]

```
  1 MKRFLFLLLT ISLLVMVQIQ TGLSGQNDTS QTSSPSASSN ISGGIFLFFV ANAIIHLFCF
 61 S
```

[0203]

[0204] "분화 클러스터 52(CD52) 폴리뉴클레오타이드"는 CD52 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. 예시적인 CD52 핵산 서열이 아래에 제공된다.

[0205] > NM_001803.3 호모 사피엔스 CD52 분자(CD52), mRNA

```
  1 agacagccct gagatcacct aaaaagctgc taccaagaca gccacgaaga tcctaccaaa
 61 atgaagcgct tcctcttcct cctactcacc atcagcctcc tggttatggt acagatacaa
121 actggactct caggacaaaa cgacaccagc caaaccagca gcccctcagc atccagcaac
181 ataagcggag gcattttcct tttcttcgtg gccaatgcca taatccacct cttctgcttc
241 agttgaggtg acacgtctca gccttagccc tgtgcccccc gaaacagctg ccaccatcac
301 tcgcaagaga atcccctcca tctttgggag gggttgatgc cagacatcac caggttgtag
361 aagttgacag gcagtgccat gggggcaaca gccaaaatag gggggtaatg atgtaggggc
421 caagcagtgc ccagctgggg gtcaataaag ttacccttgt acttgca
```

[0206]

[0207] "분화 클러스터 123(CD123) 폴리펩타이드"는 NCBI 참조 서열: NP_002174.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. CD123은 또한 인터류킨-3 수용체로도 알려져 있다. 예시적인 아미노산 서열이 아래에 제공된다.

[0208] > NP_002174.1 인터류킨-3 수용체 서브유닛 알파 동형단백질 1 전구체 [호모 사피엔스]

```
  1 MVLLWLTLLL IALPCLLQTK EDPNPPITNL RMKAKAQQLT WDLNRNVTDI ECVKDADYSM
 61 PAVNNSYCQF GAISLCEVTN YTVRVANPPF STWILFPENS GKPWAGAENL TCWIHDVDFL
121 SCSWAVGPGA PADVQYDLYL NVANRRQQYE CLHYKTDAQG TRIGCRFDDI SRLSSGSQSS
181 HILVRGRSAA FGIPCTDKFV VFSQIEILTP PNMTAKCNKT HSFMHWKMRS HFNRKFRYEL
241 QIQKRMQPVI TEQVRDRTSF QLLNPGTYTV QIRARERVYE FLSAWSTPQR FECDQEEGAN
301 TRAWRTSLLI ALGTLLALVC VFVICRRYLV MQRLFPRIPH MKDPIGDSFQ NDKLVVWEAG
361 KAGLEECLVT EVQVVQKT
```

[0209]

[0210] "분화 클러스터 123(CD123) 폴리뉴클레오타이드"는 CD123 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. 예시적인 CD123 핵산 서열이 아래에 제공된다.

[0211] > NM_002183.4 호모 사피엔스 인터류킨 3 수용체 서브유닛 알파(IL3RA), 전사체 변이체 1, mRNA

```
   1 cttcggtttc tcttcgggga aagctgcttt cagcgcacac gggaagatat cagaaacatc
  61 ctaggatcag gacaccccag atcttctcaa ctggaaccac gaaggctgtt tcttccacac
 121 agtactttga tctccattta agcaggcacc tctgtcctgc gttccggagc tgcgttcccg
 181 atggtcctcc tttggctcac gctgctcctg atcgccctgc cctgtctcct gcaaacgaag
 241 gaagatccaa acccaccaat cacgaaccta aggatgaaag caaaggctca gcagttgacc
 301 tgggacctta acagaaatgt gaccgatatc gagtgtgtta aagacgccga ctattctatg
 361 ccggcagtga acaatagcta ttgccagttt ggagcaattt ccttatgtga agtgaccaac
 421 tacaccgtcc gagtggccaa cccaccattc tccacgtgga tcctcttccc tgagaacagt
 481 gggaagcctt gggcaggtgc ggagaatctg acctgctgga ttcatgacgt ggatttcttg
 541 agctgcagct gggcggtagg cccgggggcc cccgcggacg tccagtacga cctgtacttg
 601 aacgttgcca acaggcgtca acagtacgag tgtcttcact acaaaacgga tgctcaggga
```

[0212]

```
 661 acacgtatcg ggtgtcgttt cgatgacatc tctcgactct ccagcggttc tcaaagttcc
 721 cacatcctgg tgcggggcag gagcgcagcc ttcggtatcc cctgcacaga taagtttgtc
 781 gtcttttcac agattgagat attaactcca cccaacatga ctgcaaagtg taataagaca
 841 cattccttta tgcactggaa aatgagaagt catttcaatc gcaaatttcg ctatgagctt
 901 cagatacaaa agagaatgca gcctgtaatc acagaacagg tcagagacag aacctccttc
 961 cagctactca atcctggaac gtacacagta caaataagag cccgggaaag agtgtatgaa
1021 ttcttgagcg cctggagcac cccccagcgc ttcgagtgcg accaggagga gggcgcaaac
1081 acacgtgcct ggcggacgtc gctgctgatc gcgctgggga cgctgctggc cctggtctgt
1141 gtcttcgtga tctgcagaag gtatctggtg atgcagagac tctttccccg catccctcac
1201 atgaaagacc ccatcggtga cagcttccaa aacgacaagc tggtggtctg ggaggcgggc
1261 aaagccggcc tggaggagtg tctggtgact gaagtacagg tcgtgcagaa aacttgagac
1321 tggggttcag ggcttgtggg ggtctgcctc aatctccctg gccgggccag gcgcctgcac
1381 agactggctg ctggacctgc gcacgcagcc caggaatgga cattcctaac gggtggtggg
1441 catgggagat gcctgtgtaa tttcgtccga agctgccagg aagaagaaca gaactttgtg
1501 tgtttatttc atgataaagt gatttttttt tttttaaccc a
```

[0213]

[0214] "분화 클러스터 137(CD137) 폴리펩타이드"는 NCBI 참조 서열: NP_001552.2 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. CD137은 또한 4-1BB로도 알려져 있다. 예시적인 아미노산 서열이 아래에 제공된다.

[0215] >NP_001552.2 종양 괴사 인자 수용체 패밀리 구성원 9 전구체 [호모 사피엔스]

```
  1 MGNSCYNIVA TLLLVLNFER TRSLQDPCSN CPAGTFCDNN RNQICSPCPP NSFSSAGGQR
 61 TCDICRQCKG VFRTRKECSS TSNAECDCTP GFHCLGAGCS MCEQDCKQGQ ELTKKGCKDC
121 CFGTFNDQKR GICRPWTNCS LDGKSVLVNG TKERDVVCGP SPADLSPGAS SVTPPAPARE
181 PGHSPQIISF FLALTSTALL FLLFFLTLRF SVVKRGRKKL LYIFKQPFMR PVQTTQEEDG
241 CSCRFPEEEE GGCEL
```

[0216]

[0217] "분화 클러스터 137(CD137) 폴리뉴클레오타이드"는 CD137 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. 예시적인 CD137 핵산 서열이 아래에 제공된다.

[0218] > NM_001561.6 호모 사피엔스 TNF 수용체 패밀리 구성원 9(TNFRSF9), mRNA

```
   1 gcagaagcct gaagaccaag gagtggaaag ttctccggca gccctgagat ctcaagagtg
  61 acatttgtga gaccagctaa tttgattaaa attctcttgg aatcagcttt gctagtatca
 121 tacctgtgcc agatttcatc atgggaaaca gctgttacaa catagtagcc actctgttgc
 181 tggtcctcaa ctttgagagg acaagatcat tgcaggatcc ttgtagtaac tgcccagctg
 241 gtacattctg tgataataac aggaatcaga tttgcagtcc ctgtcctcca aatagtttct
 301 ccagcgcagg tggacaaagg acctgtgaca tatgcaggca gtgtaaaggt gttttcagga
 361 ccaggaagga gtgttcctcc accagcaatg cagagtgtga ctgcactcca gggtttcact
 421 gcctgggggc aggatgcagc atgtgtgaac aggattgtaa acaaggtcaa gaactgacaa
 481 aaaaaggttg taaagactgt tgctttggga catttaacga tcagaaacgt ggcatctgtc
 541 gaccctggac aaactgttct ttggatggaa agtctgtgct tgtgaatggg acgaaggaga
 601 gggacgtggt ctgtggacca tctccagccg acctctctcc gggagcatcc tctgtgaccc
 661 cgcctgcccc tgcgagagag ccaggacact ctccgcagat catctccttc tttcttgcgc
 721 tgacgtcgac tgcgttgctc ttcctgctgt tcttcctcac gctccgtttc tctgttgtta
 781 aacggggcag aaagaaactc ctgtatatat tcaaacaacc atttatgaga ccagtacaaa
 841 ctactcaaga ggaagatggc tgtagctgcc gatttccaga agaagaagaa ggaggatgtg
 901 aactgtgaaa tggaagtcaa tagggctgtt gggactttct tgaaaagaag caaggaaata
 961 tgagtcatcc gctatcacag ctttcaaaag caagaacacc atcctacata atacccagga
1021 ttcccccaac acacgttctt ttctaaatgc caatgagttg gcctttaaaa atgcaccact
1081 ttttttttt ttttgacagg gtctcactct gtcacccagg ctggagtgca gtggcaccac
1141 catggctctc tgcagccttg acctctggga gctcaagtga tcctcctgcc tcagtctcct
1201 gagtagctgg aactacaagg aagggccacc acacctgact aacttttttg ttttttgttt
1261 ggtaaagatg gcatttcacc atgttgtaca ggctggtctc aaactcctag gttcactttg
1321 gcctcccaaa gtgctgggat tacagacatg aactgccagg cccggccaaa ataatgcacc
1381 acttttaaca gaacagacag atgaggacag agctggtgat aaaaaaaaaa aaaaaaaagc
1441 attttctaga taccacttaa caggtttgag ctagtttttt tgaaatccaa agaaaattat
1501 agtttaaatt caattacata gtccagtggt ccaactataa ttataatcaa aatcaatgca
1561 ggtttgtttt ttggtgctaa tatgacatat gacaataagc cacgaggtgc agtaagtacc
1621 cgactaaagt ttccgtgggt tctgtcatgt aacacgacat gctccaccgt caggggggag
1681 tatgagcaga gtgcctgagt ttagggtcaa ggacaaaaaa cctcaggcct ggaggaagtt
1741 ttggaaagag ttcaagtgtc tgtatatcct atggtcttct ccatcctcac accttctgcc
1801 tttgtcctgc tccctttaa gccaggttac attctaaaaa ttcttaactt ttaacataat
1861 attttatacc aaagccaata aatgaactgc atatgatagg tatgaagtac agtgagaaaa
1921 ttaacacctg tgagctcatt gtcctaccac agcactagag tgggggccgc caaactccca
1981 tggccaaacc tggtgcacca tttgcctttg tttgtctgtt ggtttgcttg agacagtctt
2041 gctctgttgc ccaggctgga atggagtggc tattcacagg cacaatcata gcacacttta
2101 gccttaaact cctgggctca agtgatccac ccgcctcagt ctcccaagta gctgggatta
2161 caggtgcaaa cctggcatgc ctgccattgt ttggcttatg atctaaggat agctttttaa
2221 attttattca ttttattttt ttttgagaca gtgtctcact ctgtctccca ggctggagta
2281 cagtggtaca atcttggatc accgcctccc agtttcaagt gatctccctg cctcagcctc
2341 ctaagtagct gggactacag gtatgtgcca ccacgcctgg ctaatttta tatttttagt
```

[0219]

2401 agagacgggg tttcaccatg ttgtccaggc tggtctcaaa ctcctgacct caggtgatct
2461 gcccacctct gcctcccaaa gtgctgggat tacaggcatg agccaccatg cctggccatt
2521 tcttacactt ttgtatgaca tgcctattgc aagcttgcgt gcctctgtcc catgttattt
2581 tactctggga tttaggtgga gggagcagct tctatttgga acattggcca tcgcatggca
2641 aatgggtatc tgtcacttct gctcctattt agttggttct actataacct ttagagcaaa
2701 tcctgcagcc aagccaggca tcaatagggc agaaaagtat attctgtaaa taggggtgag
2761 gagaagatat ttctgaacaa tagtctactg cagtaccaaa ttgcttttca aagtggctgt
2821 tctaatgtac tcccgtcagt catataagtg tcatgtaagt atcccattga tccacatcct
2881 tgctaccctc tggtactatc aggtgccctt aattttgcca agccagtggg tatagaatga
2941 gatctcactg tggtcttagt ttgcatttgc ttggttactg atgagcacct tgtcaaatat
3001 ttatatacca tttgtgttta tttttttaaa taaaatgctt gctcatgctt ttttgcccat
3061 ttgcaaaaaa acttggggcc gggtgcagtg gctcatgcct gtagtcccag ctctttggga
3121 ggccaaggtg ggcagatcgc ttgagcccag gagttcgaga ccagccttgg caacatggcg
3181 aaaccctgtc tttacaaaaa atacaaaaat tagccgggtg tggtggtgtg cacctgaagt
3241 cccagctact cagtaggttc gctttgagcc tgggaggcag aggttgcagt gagctgggac
3301 cgcatcacta cacttcagcc tgggcaacag agaaaaacct tttctcagaa acaaacaaac
3361 ccaaatgtgg ttgtttgtcc tgattcctaa aaggtcttta tgtattctag ataataatct
3421 ttggtcagtt atatgtgtta aaaaatatct tctttgtggc caggcacggt agctcacacc
3481 tgtaatccca gcactttgcg gggctgaggt gggtggatca tctgaggtca agagttcaag
3541 atcagcctgg ccaacacagt gaaaccccat ctctactaaa catgtacaaa acttagctgg
3601 gtatggtggc gggtgcctgt aaccccagct gctccagagg ctgtggcaga agaatcgctt
3661 gaacccagga ggcagaggtt gcagcgagcc aagattgtgc cattgcactc cagactgggt
3721 gacaagagtg aaattctgcc tatctatcta tctatctatc tatatctata tatatatata
3781 tatatatcct ttgtaattta tttttccctt tttaaaattt tttataaaat tctttttat
3841 ttttatttttt agcagaggtg aggtttctga ggtttcatta tgttgcccag gctggtcttg
3901 aactcctgag ctcaagtgat cctcccacct cagccttcca aagtgctgga attgcagaca
3961 tgagccaccg cgcccctcct gtttttctct aattaatggt gtctttcttt gtctttctgg
4021 taataagcaa aaagttcttc atttgatttg gttaaattta taactgtttt ctcatatggt
4081 taacattttt tcttgcctgg ctaaagaaat ccttttctgc ccaatactat aaagaggttt
4141 gcccacattt tattccaaaa gttttaagtt ttgtctttca tcttgaagtc taatgtatca
4201 ggaactggct tttgtgcctg ttgggaggta gtgatccaat tccatgtctt gcatgtaggt
4261 aaccactggt ccctgcgcca tgtattcaat acgtcgtctt tctcctgcgg gtctgcaatc
4321 tcacctacca tccatcaagt ttccataggg ccatgggtct gcttctgggc tccctgttct
4381 gttccattgt caatttgtct atcctgtgcc agtatcacac tgtgtttatt acaatagctt
4441 tgtaacagct ctcgatatcc ggtaggacat ctccctccac cttctttttc tacttcagaa
4501 gtgtcttagc taggtcaggc acggtggctc acgcctgtaa tcccagcact ttgggaggcc
4561 gacgcggatg gatcacctga ggtcaggagt tttgagacag cctggccaac atggtgaaac
4621 cccatctcta ctaaaaaata caaaaattag tcaggcatgg tggcatgtgc ctgtaatccc
4681 agctatttgg gaggctgagg ccggagaatt gcttgaaccc gggggggcgga ggttgcagtg
4741 agccgagatc gtaccattgc actccagcct gggtgacaga gcgaaactct gtctcaggaa
4801 aaaaaagaaa agagatgtct tggttattct tggttcttta ttattcaata taaattttag
4861 aagctgaatt tgaaaagatt tggattggaa tttcattaaa tctacaggtc aatttaggga

[0220]

4921 gagttgataa ttttacagaa ttgagtcatc tggtgttcca ataagaataa gagaacaatt
4981 attggctgta caattcttgc caaatagtag gcaaagcaaa gcttaggaag tatactggtg
5041 ccatttcagg aacaaagcta ggtgcgaata tttttgtctt tctgaatcat gatgctgtaa
5101 gttctaaagt gatttctcct cttggctttg gacacatggt gtttaattac ctactgctga
5161 ctatccacaa acagaaagag actggtcatg ccccacaggg ttggggtatc caagataatg
5221 gagcgaggct ctcatgtgtc ctaggttaca caccgaaaat ccacagttta ttctgtgaag
5281 aaaggaggct atgtttatga tacagactgt gatattttta tcatagccta ttctggtatc
5341 atgtgcaaaa gctataaatg aaaaacacag gaacttggca tgtgagtcat tgctcccccc
5401 aaatgacaat taataaggaa ggaacattga gacagaataa aatgatcccc ttctgggttt
5461 aatttagaaa gttccataat taggtttaat agaaataaat gtaaatttct atgattaaaa
5521 ataaattagc acatttaggg atacacaaat tataaatcat tttctaaatg ctaaaaacaa
5581 gctcaggttt ttttcagaag aaagttttaa tttttttttct ttagtggaag atatcactct
5641 gacggaaagt tttgatgtga ggggcggatg actataaagt gggcatcttc ccccacagga
5701 agatgtttcc atctgtgggt gagaggtgcc caccgcagct agggcaggtt acatgtgccc
5761 tgtgtgtggt aggacttgga gagtgatctt tatcaacgtt tttatttaaa agactatcta
5821 ataaaacaca aaactatgat gttcacagga aaaaaagaat aagaaaaaaa ga

[0221]

[0222] "분화 클러스터 247(CD247) 폴리펩타이드"는 NCBI 참조 서열: NP_932170.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. CD137은 또한 CD3ζ으로도 알려져 있다. 예시적인 아미노산 서열이 아래에 제공된다.

[0223] > NP_932170.1 T-세포 표면 당단백질 CD3 제타 사슬 동형단백질 1 전구체 [호모 사피엔스]

```
  1 MKWKALFTAA ILQAQLPITE AQSFGLLDPK LCYLLDGILF IYGVILTALF LRVKFSRSAD
 61 APAYQQGQNQ LYNELNLGRR EEYDVLDKRR GRDPEMGGKP QRRKNPQEGL YNELQKDKMA
121 EAYSEIGMKG ERRRGKGHDG LYQGLSTATK DTYDALHMQA LPPR
```

[0224]

[0225] "분화 클러스터 247(CD247) 폴리뉴클레오타이드"는 CD247 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. 예시적인 CD247 핵산 서열이 아래에 제공된다.

[0226] > NM_ NM_198053.3 호모 사피엔스 CD247 분자(CD247), 전사체 변이체 1, mRNA

```
   1 aaccgtcccg gccaccgctg cctcagcctc tgcctcccag cctctttctg agggaaagga
  61 caagatgaag tggaaggcgc ttttcaccgc ggccatcctg caggcacagt tgccgattac
 121 agaggcacag agctttggcc tgctggatcc caaactctgc tacctgctgg atggaatcct
 181 cttcatctat ggtgtcattc tcactgcctt gttcctgaga gtgaagttca gcaggagcgc
 241 agacgccccc gcgtaccagc agggccagaa ccagctctat aacgagctca atctaggacg
 301 aagagaggag tacgatgttt tggacaagag acgtggccgg gaccctgaga tgggggggaaa
 361 gccgcagaga aggaagaacc ctcaggaagg cctgtacaat gaactgcaga aagataagat
 421 ggcggaggcc tacagtgaga ttgggatgaa aggcgagcgc cggaggggca aggggcacga
 481 tggcctttac cagggtctca gtacagccac caaggacacc tacgacgccc ttcacatgca
 541 ggccctgccc cctcgctaac agccagggga tttcaccact caaaggccag acctgcagac
 601 gcccagatta tgagacacag gatgaagcat ttacaacccg gttcactctt ctcagccact
 661 gaagtattcc cctttatgta caggatgctt tggttatatt tagctccaaa ccttcacaca
 721 cagactgttg tccctgcact ctttaaggga gtgtactccc agggcttacg gccctggcct
 781 tgggccctct ggtttgccgg tggtgcaggt agacctgtct cctggcggtt cctcgttctc
 841 cctgggaggc gggcgcactg cctctcacag ctgagttgtt gagtctgttt tgtaaagtcc
 901 ccagagaaag cgcagatgct agcacatgcc ctaatgtctg tatcactctg tgtctgagtg
 961 gcttcactcc tgctgtaaat ttggcttctg ttgtcacctt cacctccttt caaggtaact
1021 gtactgggcc atgttgtgcc tccctggtga gagggccggg cagaggggca gatggaaagg
1081 agcctaggcc aggtgcaacc agggagctgc aggggcatgg gaaggtgggc gggcagggga
1141 gggtcagcca gggcctgcga gggcagcggg agcctccctg cctcaggcct ctgtgccgca
1201 ccattgaact gtaccatgtg ctacaggggc cagaagatga acagactgac cttgatgagc
1261 tgtgcacaaa gtggcataaa aaacatgtgg ttacacagtg tgaataaagt gctgcggagc
1321 aagaggaggc cgttgattca cttcacgctt tcagcgaatg acaaaatcat ctttgtgaag
1381 gcctcgcagg aagacccaac acatgggacc tataactgcc cagcggacag tggcaggaca
1441 ggaaaaaccc gtcaatgtac taggatactg ctgcgtcatt acagggcaca ggccatggat
1501 ggaaaacgct ctctactctg ctttttttct actgttttaa tttatactgg catgctaaag
1561 ccttcctatt ttgcataata aatgcttcag tgaaaatgca
```

[0227]

[0228] "병용-투여" 또는 "병용 투여된"은 치료 과정 동안 2종 이상의 치료제 또는 약제학적 조성물을 투여하는 것을 지칭한다. 이러한 병용-투여는 동시 투여 또는 순차적 투여일 수 있다. 나중에 투여되는 치료제 또는 약제학적 조성물의 순차적 투여는 제1 약제학적 조성물 또는 치료제의 투여 후 치료 과정 동안 임의의 시간에 발생할 수 있다.

[0229] 용어 "보존적 아미노산 치환" 또는 "보존적 돌연변이"는 공통 특성을 갖는 또 다른 아미노산에 의한 한 아미노산의 대체를 지칭한다. 개별 아미노산 사이의 공통 특성을 정의하는 기능적 방식은 상동 유기체의 상응하는 단백질 사이의 정규화된 아미노산 변화의 빈도를 분석하는 것이다(Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979)]). 이러한 분석에 따르면, 아미노산의 그룹은 그룹 내의 아미노산이 서로 우선적으로 교환되므로 전체 단백질 구조에 미치는 영향이 가장 유사한 것으로 정의될 수 있다(상기 문헌[Schulz, G. E. 및 Schirmer, R. H.]). 보존적 돌연변이의 비제한적인 예는 아미노산의 아미노산 치환, 예를 들어, 양전하가 유지될 수 있도록 아르기닌의 라이신으로의 치환 및 그 반대; 음전하가 유지될 수 있도록 아스파르트산의 글루탐산으로의 치환 및 그 반대; 유리 -OH가 유지될 수 있도록 트레오닌

의 세린으로의 치환; 및 유리 $-NH_2$가 유지될 수 있도록 아스파라긴의 글루타민으로의 치환을 포함한다.

[0230] 본 명세서에서 상호교환적으로 사용되는 용어 "코딩 서열" 또는 "단백질 코딩 서열"은 단백질을 코딩하는 폴리뉴클레오타이드의 분절을 지칭한다. 영역 또는 서열은 개시 코돈에 의해 5' 말단에 더 가깝고 종결 코돈에 의해 3' 말단에 더 가깝게 결합된다. 코딩 서열은 또한 오픈 리딩 프레임으로도 지칭될 수 있다.

[0231] "코돈 최적화"는 천연 아미노산 서열을 유지하면서 관심 숙주 세포의 유전자에서 보다 빈번하게 또는 가장 빈번하게 사용되는 코돈을 갖는 천연 서열의 적어도 하나의 코돈(예를 들어, 약 1개, 2개, 3개, 4개, 5개, 10개, 15개, 20개, 25개, 50개 이상 또는 초과의 코돈)을 대체함으로써 해당 숙주 세포에서의 향상된 발현을 위해 핵산 서열을 변형시키는 과정을 의미한다. 다양한 종은 특정 아미노산의 소정의 코돈에 대해 특정한 편향을 나타낸다. 코돈 편향(유기체간 코돈 사용의 차이)은 종종 메신저 RNA(mRNA)의 번역의 효율성과 상관관계가 있으며, 이는 결국 무엇보다도 번역되는 코돈의 특성 및 특정 운반 RNA(tRNA) 분자의 이용 가능성에 달려 있는 것으로 여겨진다. 세포에서 선택된 tRNA의 우세는 일반적으로 펩타이드 합성에서 가장 자주 사용되는 코돈을 반영한다. 따라서, 유전자는 코돈 최적화를 기반으로 주어진 유기체에서 최적의 유전자 발현을 위해 맞춤화될 수 있다. 코돈 사용 표는, 예를 들어, www.kazusa.orjp/codon/(2002년 7월 9일 방문)에서 이용 가능한 "코돈 사용 데이터베이스"에서 쉽게 이용할 수 있으며, 이러한 표는 여러 방식으로 적합화될 수 있다. 문헌[Nakamura, Y., *et al.* "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000)] 참조. Gene Forge(압타겐(Aptagen); 펜실베니아주 제이코버스 소재)와 같은 특정 숙주 세포에서의 발현을 위한 특정 서열을 최적화하는 코돈을 위한 컴퓨터 알고리즘도 사용 가능하다. 일부 실시형태에서, 조작된 뉴클레이스를 암호화하는 서열에서 하나 이상의 코돈(예를 들어, 1개, 2개, 3개, 4개, 5개, 10개, 15개, 20개, 25개, 50개 이상 또는 모든 코돈)은 특정 아미노산에 대해 가장 자주 사용되는 코돈에 해당한다.

[0232] "사이티딘 데아미네이스"는 아미노기를 카보닐기로 전환시키는 탈아미노화 반응을 촉매할 수 있는 폴리펩타이드 또는 이의 단편을 의미한다. 일 실시형태에서, 사이티딘 데아미네이스는 사이토신을 우라실로 또는 5-메틸사이토신을 티민으로 전환한다. 페트로마이존 마리누스(*Petromyzon marinus*)로부터 유래되는 PmCDA1(페트로마이존 마리누스 사이토신 데아미네이스 1("PmCDA1(*Petromyzon marinus* cytosine deaminase 1)")) 또는 포유동물(예를 들어, 인간, 돼지, 소, 말, 원숭이 등)로부터 유래되는 AID(활성화-유도성 사이티딘 데아미네이스("AICD(Activation-induced cytidine deaminase)")) 및 APOBEC가 예시적인 사이티딘 데아미네이스이다.

[0233] PmCDA1의 염기 서열과 아미노산 서열 및 인간 AID의 염기 서열과 아미노산 서열이 아래에 나타나 있다.

[0234] >tr|A5H718|A5H718_PETMA 사이토신 데아미네이스 OS=페트로마이존 마리누스 OX=7757 PE=2 SV=1

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERG
IHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYY
EKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELS

[0235] IMIQVKILHTTKSPAV

[0236] >EF094822.1 페트로마이존 마리누스 단리물 PmCDA.21 사이토신 데아미네이스 mRNA, 완전 cds

TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAG
AGAGGACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACG
CTGAGTACGTGAGAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAA
CAAAAAATCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTGAACGTAGA
GCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGCGGGACAGAACGTGGAATTCACGCCG
AAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCACGATAAA
TTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAG
GAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATG
CGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAG
TGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATTGAATGAGAAT
AGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATGATTC
AGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGTTTTC

[0237]

[0238] >tr|Q6QJ80|Q6QJ80_인간 활성화-유도성 사이티딘 데아미네이스 OS=호모 사피엔스 OX=9606 GN=AICDA PE=2 SV=1

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRY
ISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRR
LHRAGVQIAIMTFKAPV

[0239]

[0240] >NG_011588.1:5001-15681 호모 사피엔스 활성화 유도성 사이티딘 데아미네이스(AICDA), 12번 염색체 상의 RefSeqGene(LRG_17)

AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGAC
ACTCTGGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTC
CTCTCAGAGCAAATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGAC
TGATAAGATCAGCTTGATCAATATGCATATATATTTTTTGATCTGTCTCCTTTTCTTCTATTCA
GATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCA
TGTGGCAAAAGAAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTACCATGGTTGAAA
CTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATCATGAAAAGGTAATA
GCAGTACTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTTGTAAATATTCAACAGTAA
AACAACTTGAAGACACACTTTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAA
TTTGTAATTTTAGAAATGCCAAGCATTCTAAATTAATTGCTTGAAAGTCACTATGATTGTGTCC
ATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAAGGCCCAATTGTTAGGCAGT
TAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTACCTCTTA
GGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCAT
AGAACTTATTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTA
TACTTTACCTCAATTTAACTCCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAA
TATATTTTTTTGTAGAGACAGGGTCTTAGCCCAGCCGAGGCTGGTCTCTAAGTCCTGGCCCAAG
CGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTATAGACATGAGCCATCACATCCAATAT
ACAGAATAAAGATTTTTAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCAGACAA
TGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTG
AAGGGTCCATTGGAAATACTTGTTCAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAA
ATTCAGAGAAAGACAGAAAAGGAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTAT
CATTAAGGAGGACAGTATCTGTAGAGCTCATTAGTGATGGCAAAATGACTTGGTCAGGATTATT
TTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGGTTCTGCCTCAGGGAGCA
CAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCTAC
TCAGGACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAAT
GGATCAACAAAGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATG
TGACAGAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAG
TCCTTTATCTATGCCACATCCTTCTTATCTATACTTCCAGGACACTTTTTCTTCCTTATGATAA
GGCTCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACACACACACAAAC
ACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCTCTGCCTTTCTCATCTACAC
AGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTA
ACACTGGGCCTCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGA
GCTTATATATACAAAGGGTTATTATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGG
TAATTACCATAAAAATTATTATCCTTTTAAAATAAAGCTAATTATTATTGGATCTTTTTTAGTA
[0241] TTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCTGTTACCCAGGCTG

GAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTG
CCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTA
GATTAAAATATGCATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTAT
ACTGGCAATAAATCTAGTTTGCTGCCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGT
GAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTGTGGTGGCTCACGCCTGTAATCCCAGCA
CTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCTGAGCAACAT
GGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCAGC
TACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGC
TTGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAAA
AATTAAAAATAAATGGAAACAACTACAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCT
GATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTACATTATTAAAATATCA
ATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACCTTTATTCACAAAA
CCCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATAT
CCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATG
ACGCTAAATGAAAGAGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAA
TGAAAGTAACTTATAGTTACAGAAAGCAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCC
CAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCA
ACACAGTGAAACTCCATTCTCCACAAAAATGGGAAAAAAAGAAAGCAAATCAGTGGTTGTCCTG
TGGGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATTCAGGTT
CTGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGCA
TCTTAAATGGGTGGAGTTTACTGTATGTAAATTATACCTCAATGTAAGAAAAAATAATGTGTAA
GAAAACTTTCAATTCTCTTGCCAGCAAACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAAT
TCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGGATAAATGCAT
TTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCATGCTGTAC
AGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTAC
CAGAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAG
CCAGAAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGAT
CATTTTGACTAGTTAAAAAAGCAGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAAAT
CATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACACAAGAATCTTAACAGTCATTGTC
ATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTTA
ATATTTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTATTCAAGTTTGACAATGATCAAGTAT
TACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTTGGTCTTGTTGCCCATGC
TGGAGTGGAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAGCTGT
CGCCTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCCACCACCACACTCGGCTAATGTTTGTA

[0242] TTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAGG

ATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGCCCGGCCAAG
TATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCCAGCCAGGTATTGCTCTTATA
CATTAAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGC
GGGCAGAACACCCGAGGTCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCT
ATTAAAAATACAAACATTACCTGGGCATGATGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGG
CTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTGCCTGAGCCTGGGAGGTTGAGGCTACAGT
AAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCGACAAAGTGAGACCGTAACAAAAAAAAAA
AAATTTAAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAAAGTGGCCTAAACACCACATTAA
AGAGTTTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCTTCAGCATGGGAATGGC
ATGGTGCACCTGGTTTTTGTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGA
CTGGAGGCAGACAGACCGGTTAAAAGGCCAGCACAACAGATAAGGAGGAAGAAGATGAGGGCTT
GGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAAT
CAACACATTTAGATGATTAATTAAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATGGTT
CCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGG
GCAGTTGAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAG
ACTGGAGGGGAAATCTGAATATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTGTT
TCGTTTTCTTGTTGAAGAACAAATTTAATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTG
GCAGGAGGTGACTGTCTTGTGGGTAAGGGTTTGGGGTCCTTGATGAGTATCTCTCAATTGGCCT
TAAATATAAGCAGGAAAAGGAGTTTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAGGGCTCA
GGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCCAAGTAATGACTTCCTTAAAA
AGCTGAAGGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTC
TCACCCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGC
TAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCTTT
TCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTTAA
TGGTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTTGGCAT
TTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCAT
ATTAGACATGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCA
CTCCTTCCTTCAGTGCCAAGAACAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGA
ATTGCCTTTGAGATTAATTAAGCTAAAAGCATTTTTATATGGGAGAATATTATCAGCTTGTCCA
AGCAAAAATTTTAAATGTGAAAAACAAATTGTGTCTTAAGCATTTTTGAAAATTAAGGAAGAAG
AATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCCAAATGATTTCTTTTCCCTCCTACT
CACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCAGAAAACTCAGAGAAGCCT
CGGCTGATGATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGACTTC
TTCACCAAAATCCAGATGGGTTTACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACAC
[0243] GCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAAGCATCCGTGGGGTGGAAGGTCATCGTCTG

GCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACATTTGTATTGAATACATC
CCAATCTCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCAAGC
ACTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTA
CCAAAATCTATTTCCCCTTTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCCACC
TAGTGGTCCAGGTATATTTCCACAATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCA
AAAGGTGCAAAAAGCAACTTCATAAACACAAATTAAATCTTCGGTGAGGTAGTGTGATGCTGCT
TCCTCCCAACTCAGCGCACTTCGTCTTCCTCATTCCACAAAAACCCATAGCCTTCCTTCACTCT
GCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGCTT
AGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATG
ACTAAGGCTACCAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTCTGTCTCTCCA
GAACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGC
CGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGG
CCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTG
TGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCC
ATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTCGG
GATTGCGATGCGGAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCT
GGTTCACCTCTGGAGCCGAAATTAAAGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACA
AGGCCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCTTCTTTCCCCTCGATTTGGAACCTGAA
CTGTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTTTTTTTTTTTTTGAAGATTATTTTTACT
GCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAA
TTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCTTCCTCGCTTTTTAA
ATTTTCTTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTT
CAATCACTCTCAGTTTTCATCTGATGAAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTG
CTGTTTCACCATTCAGAGCCCTCTGCTAAGGTTCCTTTTCCCTCCCTTTTCTTTCTTTTGTTGT
TTCACATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTCAGAATTCTTTTC
TCCTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCAAAAAAACTCTTT
CCCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCC
CACCGACCCCCAACCTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCC
CTGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTTTGATAGCAACTTCCAG
GAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAAGTCT
TCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTTATATACGAC
TCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTGC
AATTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACTGCAGGACCTGGG
AGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCC
[0244] TAAAAAGCATGGTGAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATTTGATTCATTTG

AGTTAACAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAA
CACAAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTG
ACCCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAA
GCATGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCA
TGGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACA
CCCTAATAATGGGTTGATGTCTGAAGTAGCAAATCTTCTGGAAACGCAAACTCTTTTAAGGAAG
TCCCTAATTTAGAAACACCCACAAACTTCACATATCATAATTAGCAAACAATTGGAAGGAAGTT
GCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCA
ATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAACTAT
ATAAGAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTA
GCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAA
CATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGG
TGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGA
GTTTGAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCA
GGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAA
CCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGA
GCAAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAAAGAGAACAATATTTGG
GAGAGAAGGATGGGGAAGCATTGCAAGGAAATTGTGCTTTATCCAACAAAATGTAAGGAGCCAA
TAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAACAACTGTCTTTGACAGTGAG
AAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCTTGCAATGAAGATG
AGCAGATCCACAGGAAAACTTGAATGCACAACTGTCTTATTTTAATCTTATTGTACATAAGTTT
GTAAAAGAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATTATTTTGCGTCTAAT
GATTTTTTATTAACATGATTTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAAT
TTATAACTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAATGGTGCTA
CGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTT
CAATCAGTTAAATAAATGATAAATAATTTTGGAAGCTGTGAAGATAAAATACCAAATAAAATAA
TATAAAAGTGATTTATATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

[0245]

[0246] 아포지단백질 B mRNA 편집 효소, 촉매 폴리펩타이드-유사(APOBEC) 단백질은 진화적으로 보존된 사이티딘 데아미네이스의 패밀리이다. 이 패밀리의 구성원은 C를 U로 편집하는 효소(C-to-U editing enzyme)이다. APOBEC 유사 단백질의 N-말단 도메인은 촉매 도메인인 반면, C-말단 도메인은 유사촉매 도메인(pseudocatalytic domain)이다. 보다 구체적으로, 촉매 도메인은 아연 의존성 사이티딘 데아미네이스 도메인이며, 사이티딘 탈아미노화에 중요하다. APOBEC 패밀리 구성원은 APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D("APOBEC3E"가 이제 이를 지칭함), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 및 활성화-유도성(사이티딘) 데아미네이스를 포함한다. 애드진으로부터 이용 가능한 SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, VRER-BE3, YE1-BE3, EE-BE3, YE2-BE3 및 YEE-BE3(플라스미드 85169, 85170, 85171, 85172, 85173, 85174, 85175, 85176, 85177)을 포함하지만 이들로 제한되지 않는 많은 변형된 사이티딘 데아미네이스가 상업적으로 이용 가능하다.

[0247] 본 개시내용의 양태에 따른 Cas9에 융합될 수 있는 다른 예시적인 데아미네이스가 본 명세서에 제공된다. 일부 실시형태에서, 각 서열의 활성 도메인은, 예를 들어, 국부화 신호가 없는 도메인(핵 방출 신호, 세포질 국부화 신호가 없는 핵 국부화 서열)이 사용될 수 있음을 이해하여야 한다.

[0248] 본 명세서에서 사용되는 용어 "데아미네이스" 또는 "데아미네이스 도메인"은 탈아미노화 반응을 촉매하는 단백질 또는 효소를 지칭한다. 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 사이티딘 또는 데옥시사이티딘의 우리딘 또는 데옥시우리딘으로의 가수분해성 탈아미노화를 각각 촉매하는 사이티딘 데아미네이스이다. 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 사이토신의 우라실로의 가수분해성 탈아미노화를 촉매하는 사이토신 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 아데노신 데아미네이스이며, 이는 아데닌의 하이포잔틴으로의 가수분해성 탈아미노화를 촉매한다. 일부 실시형태에서, 데아미네이스는 아데노신 데아미네이스이며, 이는 아데노신 또는 아데닌(A)의 이노신(I)으로의 가수분해성 탈아미노화를 촉매한다. 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 아데노신 또는 데옥시아데노신의 이노신 또는 데옥시이노신으로의 가수분해성 탈아미노화를 각각 촉매하는 아데노신 데아미네이스이다. 일부 실시형태에서, 아데

노신 데아미네이스는 데옥시리보핵산(DNA)에서 아데노신의 가수분해성 탈아미노화를 촉매한다. 본 명세서에 제공된 아데노신 데아미네이스(예를 들어, 조작된 아데노신 데아미네이스, 진화된 아데노신 데아미네이스)는 세균과 같은 임의의 유기체로부터 유래할 수 있다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균(*Escherichia coli*), 스타필로코커스 아우레우스, 살모넬라 티피무리움(*Salmonella typhimurium*), 쉬와넬라 푸트레파시엔스(*Shewanella putrefaciens*), 헤모필루스 인플루엔자에(*Haemophilus influenzae*) 또는 카울로박터 크레센투스(*Caulobacter crescentus*)와 같은 세균으로부터 유래한다.

[0249] 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스와 같은 유기체로부터의 자연 발생적 데아미네이스의 변이체이다. 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 자연에서 발생하지 않는다. 예를 들어, 일부 실시형태에서, 데아미네이스 또는 데아미네이스 도메인은 자연 발생적 데아미네이스와 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75% 적어도 80%, 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 적어도 99.1%, 적어도 99.2%, 적어도 99.3%, 적어도 99.4%, 적어도 99.5%, 적어도 99.6%, 적어도 99.7%, 적어도 99.8% 또는 적어도 99.9% 동일하다.

[0250] 예를 들어, 데아미네이스 도메인은 각각 그 전문이 참조에 의해 본 명세서에 원용되어 있는 PCT 특허 PCT/2017/045381(WO 2018/027078) 및 PCT/US2016/058344(WO 2017/070632)에 기술되어 있다. 또한, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)), 및 Rees, H.A., *et al.*, "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1]을 참조한다.

[0251] "검출하다"는 검출될 분석물질의 존재, 부재 또는 양을 식별하는 것을 지칭한다.

[0252] "검출 가능한 표지"는 관심 분자에 연결될 때 분광학적, 광화학적, 생화학적, 면역화학적 또는 화학적 수단을 통해 후자를 검출 가능하게 만드는 조성물을 의미한다. 예를 들어, 유용한 표지는 방사성 동위원소, 자성 비드, 금속 비드, 콜로이드 입자, 형광 염료, 고전자-밀도 시약, 효소(예를 들어, 효소-연결 면역흡착 검정(ELISA)에서 일반적으로 사용됨), 바이오틴, 디곡시게닌 또는 합텐을 포함한다.

[0253] "질환"은 세포, 조직 또는 기관의 정상적인 기능을 손상시키거나 또는 방해하는 임의의 병태 또는 장애를 의미한다. 일 실시형태에서, 질환은 신생물 또는 암이다. 일부 실시형태에서, 질환은 혈액암이다. "혈액암"은 면역계 세포의 악성 종양을 의미한다. 일부 실시형태에서, 혈액암은 백혈병, 골수종 및/또는 림프종이다. 림프종 및 백혈병은 "액체암" 또는 혈액에 존재하는 암의 예이며, 골수의 조혈 전구체 또는 혈액의 성숙 조혈 세포의 변형으로부터 유래된다. 백혈병은 림프성 또는 골수성 및 급성 또는 만성일 수 있다. 골수종의 경우, 변형된 세포는 골수에서 악성 세포의 분산된 집합체 또는 고형 종괴로서 존재할 수 있는 완전히 분화된 형질 세포이다. 림프종의 경우, 2차 림프구성 조직에서 변형된 림프구는 고형 종괴를 생성한다. 림프종은 호지킨 림프종(Hodgkin lymphoma: HL) 또는 비-호지킨 림프종(non-Hodgkin lymphoma: NHL)으로 분류된다.

[0254] 일부 실시형태에서, 혈액암은 B 세포암이다. 일부 실시형태에서, B 세포암은 림프종 또는 백혈병이다. 일부 경우에, 백혈병은 전백혈병을 포함한다. 일부 경우에, 백혈병은 급성 백혈병이다. 급성 백혈병은, 예를 들어, 급성 골수성 백혈병(AML)을 포함한다. 급성 백혈병은 또한, 예를 들어, 급성 림프성 백혈병 또는 급성 림프구성 백혈병(ALL)을 포함하고; ALL은 B-계통 ALL; T-계통 ALL; 및 T-세포 급성 림프구성 백혈병(T-ALL)을 포함한다.

[0255] 질환의 비제한적인 예는 T-세포 급성 림프아구성 백혈병(T-ALL), 용상 진균증(MF), 세자리 증후군(SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 $ALK^+$, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 $CD30^+$ 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종을 포함한다. 일부 실시형태에서, 질환은 액체 종양이다. 일부 실시형태에서, 질환은 T-세포 급성 림프아구성 백혈병(T-ALL)이다. 일부 실시형태에서, 질환은 급성 골수성 백혈병(AML)이다.

[0256] 본 명세서에서 사용되는 용어 "유효량"은 목적하는 생물학적 반응을 유도하기에 충분한 생물학적으로 활성인 작용제의 양을 지칭한다. 일부 실시형태에서, 유효량은 치료받지 않은 환자에 비해 질환의 증상을 개선하는데 필요한 양이다. 질환의 치료적 치료를 위해 본 발명을 실시하는데 사용되는 활성제(들)의 유효량은 투여 방식, 대상체의 연령, 체중 및 일반적인 건강에 따라 다양하다. 궁극적으로, 주치의 또는 수의사가 적절한 양 및 투여 양생법을 결정할 것이다. 이러한 양은 "유효한" 양으로 지칭된다. 일 실시형태에서, 유효량은 세포(예를 들어, 시험관내 또는 생체내 세포)에서 관심 유전자에 변경을 도입하기에 충분한 본 발명의 염기 편집기(예를 들어, 프로그램 가능한 DNA 결합 단백질, 핵염기 편집기 및 gRNA를 포함하는 융합 단백질)의 양이다. 일 실시형태에서, 유효량은 (예를 들어, 질환 또는 이의 증상 또는 병태를 감소시키거나 또는 제어하기 위한) 치료적 효과를 달성하는데 필요한 염기 편집기의 양이다. 이러한 치료적 효과는 대상체, 조직 또는 기관의 모든 세포에서 관심 유전자를 변경하기에 충분할 필요는 없으며, 대상체, 조직 또는 기관에 존재하는 세포의 약 1%, 5%, 10%, 25%, 50%, 75% 이상에서 관심 유전자를 변경하는 데에만 필요하다.

[0257] 일부 실시형태에서, 본 명세서에 제공된 융합 단백질, 예를 들어, nCas9 도메인 및 데아미네이스 도메인(예를 들어, 아데노신 데아미네이스 또는 사이티딘 데아미네이스)을 포함하는 핵염기 편집기의 유효량은 본 명세서에 기재된 핵염기 편집기에 특이적으로 결합되고 편집되는 표적 부위의 편집을 유도하기에 충분한 융합 단백질의 양을 지칭한다. 당업자에 의해 이해되는 바와 같이, 작용제, 예를 들어, 융합 단백질, 뉴클레이스, 하이브리드 단백질, 단백질 이량체, 단백질(또는 단백질 이량체)의 복합체 및 폴리뉴클레오타이드 또는 폴리뉴클레오타이드의 유효량은, 예를 들어, 목적하는 생물학적 반응, 예를 들어, 편집될 특정 대립유전자, 게놈 또는 표적 부위, 표적화되는 세포 또는 조직 및/또는 사용중인 작용제와 같은 다양한 인자에 따라 달라질 수 있다. CAR-T 세포의 맥락에서, "유효량"은 치료적 반응을 달성하기 위해 환자에게 투여하는데 필요한 세포의 양을 지칭한다.

[0258] 본 명세서에서 사용되는 "에피토프"는 항원 결정기를 의미한다. 에피토프는 이의 구조에 따라 항원을 인식하고 이에 결합할 특정 항체 분자를 결정하는 항원 분자의 일부이다.

[0259] "Fas 세포 표면 사멸 수용체(Fas Cell Surface Death Receptor; FAS) 폴리펩타이드"는 NCBI 등록 번호 NP_000034.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0260] > NP_000034.1 종양 괴사 인자 수용체 패밀리 구성원 6 동형단백질 1 전구체 [호모 사피엔스]

[0261]
```
  1 MLGIWTLLPL VLTSVARLSS KSVNAQVTDI NSKGLELRKT VTTVETQNLE GLHHDGQFCH
 61 KPCPPGERKA RDCTVNGDEP DCVPCQEGKE YTDKAHFSSK CRRCRLCDEG HGLEVEINCT
121 RTQNTKCRCK PNFFCNSTVC EHCDPCTKCE HGIIKECTLT SNTKCKEEGS RSNLGWLCLL
181 LLPIPLIVWV KRKEVQKTCR KHRKENQGSH ESPTLNPETV AINLSDVDLS KYITTIAGVM
241 TLSQVKGFVR KNGVNEAKID EIKNDNVQDT AEQKVQLLRN WHQLHGKKEA YDTLIKDLKK
301 ANLCTLAEKI QTIILKDITS DSENSNFRNE IQSLV
```

[0262] "Fas 세포 표면 사멸 수용체(FAS) 폴리뉴클레오타이드"는 FAS 폴리펩타이드를 암호화하는 핵산을 의미한다. 예시적인 FAS 핵산 서열이 아래에 제공된다.

[0263] > NM_000043.6 호모 사피엔스 Fas 세포 표면 사멸 수용체(FAS), 전사체 변이체 1, mRNA

[0264]
```
  1 ctcttctccc gcgggttggt ggacccgctc agtacggagt tggggaagct ctttcacttc
 61 ggaggattgc tcaacaacca tgctgggcat ctggaccctc ctacctctgg ttcttacgtc
121 tgttgctaga ttatcgtcca aaagtgttaa tgcccaagtg actgacatca actccaaggg
181 attggaattg aggaagactg ttactacagt tgagactcag aacttggaag gcctgcatca
241 tgatggccaa ttctgccata agccctgtcc tccaggtgaa aggaaagcta gggactgcac
301 agtcaatggg gatgaaccag actgcgtgcc ctgccaagaa gggaaggagt acacagacaa
361 agcccatttt tcttccaaat gcagaagatg tagattgtgt gatgaaggac atggcttaga
421 agtggaaata aactgcaccc ggacccagaa taccaagtgc agatgtaaac caaacttttt
481 ttgtaactct actgtatgtg aacactgtga cccttgcacc aaatgtgaac atggaatcat
541 caaggaatgc acactcacca gcaacaccaa gtgcaaagag gaaggatcca gatctaactt
601 ggggtggctt tgtcttcttc ttttgccaat tccactaatt gtttgggtga agagaaagga
661 agtacagaaa acatgcagaa agcacagaaa ggaaaaccaa ggttctcatg aatctccaac
721 tttaaatcct gaaacagtgg caataaattt atctgatgtt gacttgagta aatatatcac
781 cactattgct ggagtcatga cactaagtca agttaaaggc tttgttcgaa agaatggtgt
```

```
 841 caatgaagcc aaaatagatg agatcaagaa tgacaatgtc caagacacag cagaacagaa
 901 agttcaactg cttcgtaatt ggcatcaact tcatggaaag aaagaagcgt atgacacatt
 961 gattaaagat ctcaaaaaag ccaatctttg tactcttgca gagaaaattc agactatcat
1021 cctcaaggac attactagtg actcagaaaa ttcaaacttc agaaatgaaa tccaaagctt
1081 ggtctagagt gaaaaacaac aaattcagtt ctgagtatat gcaattagtg tttgaaaaga
1141 ttcttaatag ctggctgtaa atactgcttg gttttttact gggtacattt tatcatttat
1201 tagcgctgaa gagccaacat atttgtagat ttttaatatc tcatgattct gcctccaagg
1261 atgtttaaaa tctagttggg aaaacaaact tcatcaagag taaatgcagt ggcatgctaa
1321 gtacccaaat aggagtgtat gcagaggatg aaagattaag attatgctct ggcatctaac
1381 atatgattct gtagtatgaa tgtaatcagt gtatgttagt acaaatgtct atccacaggc
1441 taaccccact ctatgaatca atagaagaag ctatgacctt ttgctgaaat atcagttact
1501 gaacaggcag gccactttgc ctctaaatta cctctgataa ttctagagat tttaccatat
1561 ttctaaactt tgtttataac tctgagaaga tcatatttat gtaaagtata tgtatttgag
1621 tgcagaattt aaataaggct ctacctcaaa gacctttgca cagtttattg gtgtcatatt
1681 atacaatatt tcaattgtga attcacatag aaaacattaa attataatgt ttgactatta
1741 tatatgtgta tgcattttac tggctcaaaa ctacctactt ctttctcagg catcaaaagc
1801 attttgagca ggagagtatt actagagctt tgccacctct ccatttttgc cttggtgctc
1861 atcttaatgg cctaatgcac ccccaaacat ggaaatatca ccaaaaaata cttaatagtc
1921 caccaaaagg caagactgcc cttagaaatt ctagcctggt ttggagatac taactgctct
1981 cagagaaagt agctttgtga catgtcatga acccatgttt gcaatcaaag atgataaaat
2041 agattcttat ttttccccca cccccgaaaa tgttcaataa tgtcccatgt aaaacctgct
2101 acaaatggca gcttatacat agcaatggta aaatcatcat ctggatttag gaattgctct
2161 tgtcataccc ccaagtttct aagatttaag attctcctta ctactatcct acgtttaaat
2221 atctttgaaa gtttgtatta aatgtgaatt ttaagaaata atatttatat ttctgtaaat
2281 gtaaactgtg aagatagtta taaactgaag cagatacctg gaaccaccta aagaacttcc
2341 atttatggag gatttttttg ccccttgtgt ttggaattat aaaatatagg taaaagtacg
2401 taattaaata atgtttttgg tatttctggt tttctctttt ttggtagggg cttgcttttt
2461 ggttttgtct tccttttctc taactgatgc taaatataac ttgtctttaa tgcttcttgg
2521 atcccttaga aggtacttcc tttttaacct taacccttttt agtagttaaa taattatttc
2581 cataggttgc tattgccaag aagacctctt ccaaacagca catgattatt cgtcaaacag
2641 tttcgtattc cagatactgg aatgtggata agaaagtata catttcaagg ggtaggtttt
2701 attattaaga aagccaaatg aggattttga aatattcttt cctgcatatt atccattcta
2761 gctacatgct ggccagtggg ccacctttct tttctgcaat ttaatgctag taatatattc
2821 tatttaaccc atgagtccca aagtattagc atttcaacat gtaagcatgt cggtaagata
2881 gttgtgcttt gcttagggtt ccctcctgtg ttatggtctg gaaagtgtct ttaggcagaa
2941 agtctgagtg atcacagggt tcactcatta atttctcttt tctgagccat catagtctgt
3001 gctgtctgct ctccagtttt ctatttctag acagaagtag ggcaagttag gtactagtta
3061 ttcttcatgg ccagaagtgc aagttctact ttgcaagaca agattaagtt agagaacacc
3121 ctattccact ttggtgaact cagagcaaga actttgagtt cctttgggag gaagacagtg
3181 gagaagtctt tgtacttggt gatgtggttt ttttcctcat ggcttcacct agtggcccca
3241 agcatgactt ctcccatgtc aatgagcaca gccacattcc cgagttgagg tgaccccacg
3301 gtccagaatc atcctcattc tggtgaacct ggttctcttt gtggtgggca tactgggtag
```

[0265]

```
3361 gagaatcacc caaaggtcac ccatgagctg cagaaaaaaa ggctatttgc agaaggagct
3421 cacagatcac attgaaagca ttgcatattc aaacatcttg gtcttcttta ttggcatgcc
3481 cacagggtct tctgacctct gattagatca gacacttttt agatattgaa tcatcagttt
3541 ctgtacaact atctgaataa ggtatataat caatgaaatt tagaattttt ttctatgctt
3601 actcctgatt ggtaatttgt ttgggtttag aattctatac aaggccattt gtaattttcc
3661 tcagcacttt aaaaatatta aaccatgttt tcttaa
```

[0266]

[0267] "단편"은 폴리펩타이드 또는 핵산 분자의 부분을 의미한다. 이 부분은 참조 핵산 분자 또는 폴리펩타이드의 전체 길이의 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% 또는 90%를 포함한다. 단편은 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90 또는 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개, 900개 또는 1000개의 뉴클레오타이드 또는 아미노산을 포함할 수 있다.

[0268] "동족 살해"는 면역 세포의 자기-항원에 의해 유도된 살해를 비롯한 다른 면역 세포에 의한 면역 세포의 살해를 의미한다. 소정의 실시형태에서, 본 발명의 면역 세포는 키메라 항원 수용체(CAR)를 발현하는 면역 세포에 의해 인식되는 항원의 발현을 방지하거나 또는 감소시키도록 유전적으로 변형되어 동족 살해를 방지하거나 또는 감소시킨다. 다양한 실시형태에서, 동족 살해는 생체내(예를 들어, 대상체에서) 또는 생체외(예를 들어, 면역 세포

제제에서)에서 발생할 수 있다.

[0269] "이식편 대 숙주 질환"(GVHD)은 공여체의 이식된 세포가 숙주의 세포에 대해 면역 반응을 일으키는 병리학적 병태를 지칭한다.

[0270] "가이드 RNA" 또는 "gRNA"는 표적 서열에 특이적일 수 있으며 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 단백질(예를 들어, Cas9 또는 Cpf1)과 복합체를 형성할 수 있는 폴리뉴클레오타이드를 의미한다. 실시형태에서, 가이드 폴리뉴클레오타이드는 가이드 RNA(gRNA)이다. gRNA는 2개 이상의 RNA의 복합체로 또는 단일 RNA 분자로 존재할 수 있다. 단일 RNA 분자로 존재하는 gRNA는 단일-가이드 RNA(sgRNA)로 지칭될 수 있지만, "gRNA"는 단일 분자 또는 2개 이상의 분자의 복합체로 존재하는 가이드 RNA를 지칭하기 위해 상호교환적으로 사용된다. 전형적으로, 단일 RNA 종으로 존재하는 gRNA는 다음 2개의 도메인을 포함한다: (1) 표적 핵산과 상동성을 공유하는(그리고, 예를 들어, 표적에 대한 Cas9 복합체의 결합을 지시하는) 도메인; 및 (2) Cas9 단백질에 결합하는 도메인. 일부 실시형태에서, 도메인 (2)는 tracrRNA로 공지된 서열에 해당하며, 줄기-루프 구조를 포함한다. 예를 들어, 일부 실시형태에서, 도메인 (2)는 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Jinek *et al.*, Science 337:816-821(2012)]에 제공된 바와 같이 tracrRNA와 동일하거나 또는 상동이다. gRNA(예를 들어, 도메인 2를 포함하는 것)는 각각 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 "Switchable Cas9 Nucleases and Uses Thereof"라는 명칭의 US20160208288 및 "Delivery System For Functional Nucleases"라는 명칭의 US 제9,737,604호에서 발견할 수 있다. 일부 실시형태에서, gRNA는 도메인 (1) 및 (2) 중 2개 이상을 포함하며, "연장된 gRNA"로도 지칭될 수 있다. 연장된 gRNA는 본 명세서에 기재된 바와 같이 2개 이상의 Cas9 단백질에 결합하고, 2개 이상의 별개의 영역에서 표적 핵산에 결합할 것이다. gRNA는 상기 표적 부위에 대한 뉴클레이스/RNA 복합체의 결합을 매개하는 표적 부위를 보완하는 뉴클레오타이드 서열을 포함하여 뉴클레이스:RNA 복합체의 서열 특이성을 제공한다. 당업자에 의해 이해되는 바와 같이, RNA 폴리뉴클레오타이드 서열, 예를 들어, gRNA 서열은 DNA 폴리뉴클레오타이드 서열에 포함되는 핵염기 티민(T)보다는 핵염기 우라실(U), 피리미딘 유도체를 포함한다. RNA에서, 우라실은 아데닌과 염기쌍을 이루고, DNA 전사 동안 티민을 대체한다.

[0271] "이종이량체"는 야생형 TadA 도메인 및 TadA 도메인의 변이체(예를 들어, TadA*8) 또는 2개의 변이체 TadA 도메인(예를 들어, TadA*7.10 및 TadA*8 또는 2개의 TadA*8 도메인)과 같은 2개의 도메인을 포함하는 융합 단백질을 의미한다.

[0272] "숙주 대 이식편 질환"(HVGD)은 숙주의 면역계가 공여체의 이식된 세포에 대한 면역 반응을 생성하는 병리학적 병태를 지칭한다.

[0273] "혼성화"는 상보적 핵염기 사이의 왓슨-크릭, 후그스틴 또는 역 후그스틴 수소 결합일 수 있는 수소 결합을 의미한다. 예를 들어, 아데닌 및 티민은 수소 결합의 형성을 통해 쌍을 이루는 상보적인 핵염기이다.

[0274] "면역 세포"는 면역 반응을 일으킬 수 있는 면역계의 세포를 의미한다.

[0275] "면역 효과기 세포"는 일단 활성화되면 표적 세포에 면역 반응을 일으킬 수 있는 림프구를 의미한다. 일부 실시형태에서, 면역 효과기 세포는 효과기 T 세포이다. 일부 실시형태에서, 효과기 T 세포는 미경험 $CD8^+$ T 세포, 세포독성 T 세포, 자연 살해 T(NKT) 세포, 자연 살해(NK) 세포 또는 조절 T(Treg) 세포이다. 일부 실시형태에서, 면역 효과기 세포는 효과기 NK 세포이다. 일부 실시형태에서, 효과기 T 세포는 흉선세포, 미성숙 T 림프구, 성숙 T 림프구, 휴지기 T 림프구 또는 활성화된 T 림프구이다. 일부 실시형태에서, 면역 효과기 세포는 $CD4^+$ $CD8^+$ T 세포 또는 $CD4^-$ $CD8^-$ T 세포이다. 일부 실시형태에서, 면역 효과기 세포는 T 헬퍼 세포이다. 일부 실시형태에서, T 헬퍼 세포는 T 헬퍼 1(Th1), T 헬퍼 2(Th2) 세포 또는 CD4를 발현하는 헬퍼 T 세포(CD4+ T 세포)이다.

[0276] "면역 반응 조절 유전자" 또는 "면역 반응 조절인자"는 면역 반응의 조절에 관여하는 폴리펩타이드를 암호화하는 유전자를 의미한다. 면역 반응 조절 유전자는 여러 메커니즘에서 또는 상이한 수준에서 면역 반응을 조절할 수 있다. 예를 들어, 면역 반응 조절 유전자는 면역 세포, 예를 들어, T 세포의 활성화를 저해하거나 또는 촉진할 수 있다. 면역 반응 조절 유전자는 면역 세포의 활성화 역치를 증가 또는 감소시킬 수 있다. 일부 실시형태에서, 면역 반응 조절 유전자는 면역 세포 신호 형질도입 경로를 양성적으로 조절한다. 일부 실시형태에서, 면역 반응 조절 유전자는 면역 세포 신호 형질도입 경로를 음성적으로 조절한다. 일부 실시형태에서, 면역 반응 조절 유전자는 항원, 항체, 사이토카인 또는 신경내분비를 암호화한다.

[0277] "면역원성 유전자" 면역 반응을 유발할 수 있는 폴리펩타이드를 암호화하는 유전자를 의미한다. 예를 들어, 면역원성 유전자는 면역 반응을 유발하는 면역원을 암호화할 수 있다. 일부 실시형태에서, 면역원성 유전자는 세포 표면 단백질을 암호화한다. 일부 실시형태에서, 면역원성 유전자는 세포 표면 항원 또는 세포 표면 마커를 암호화한다. 일부 실시형태에서, 세포 표면 마커는 T 세포 마커 또는 B 세포 마커이다. 일부 실시형태에서, 면역원성 유전자는 CD2, CD3e, CD3 델타, CD3 감마, TRAC, TRBC1, TRBC2, CD4, CD5, CD7, CD8, CD19, CD23, CD27, CD28, CD30, CD33, CD52, CD70, CD127, CD122, CD130, CD132, CD38, CD69, CD11a, CD58, CD99, CD103, CCR4, CCR5, CCR6, CCR9, CCR10, CXCR3, CXCR4, CLA, CD161, B2M 또는 CIITA 폴리펩타이드를 암호화한다.

[0278] 용어 "염기 복구 저해제" 또는 "IBR(inhibitor of base repair)"은 핵산 복구 효소, 예를 들어, 염기 절단 복구(BER) 효소의 활성을 저해할 수 있는 단백질을 지칭한다. 일부 실시형태에서, IBR은 이노신 염기 절단 복구 저해제이다. 염기 복구의 예시적인 저해제는 APE1, Endo III, Endo IV, Endo V, Endo VIII, Fpg, hOGG1, hNEIL1, T7 Endo1, T4PDG, UDG, hSMUG1 및 hAAG의 저해제를 포함한다. 일부 실시형태에서, IBR은 Endo V 또는 hAAG의 저해제이다. 일부 실시형태에서, IBR은 촉매적으로 불활성인 EndoV 또는 촉매적으로 불활성인 hAAG이다. 일부 실시형태에서, 염기 복구 저해제는 Endo V 또는 hAAG의 저해제이다. 일부 실시형태에서, 염기 복구 저해제는 촉매적으로 불활성인 EndoV 또는 촉매적으로 불활성인 hAAG이다.

[0279] 일부 실시형태에서, 염기 복구 저해제는 우라실 글리코실레이스 저해제(UGI)이다. UGI는 우라실-DNA 글리코실레이스 염기-절단 복구 효소를 저해할 수 있는 단백질을 지칭한다. 일부 실시형태에서, UGI 도메인은 야생형 UGI 또는 야생형 UGI의 단편을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 UGI 단백질은 UGI의 단편 및 UGI 또는 UGI 단편에 상동성인 단백질을 포함한다. 일부 실시형태에서, 염기 복구 저해제는 이노신 염기 절단 복구 저해제이다. 일부 실시형태에서, 염기 복구 저해제는 "촉매적으로 불활성인 이노신 특이적 뉴클레이스" 또는 "활성이 없는 이노신 특이적 뉴클레이스"이다. 임의의 특정 이론에 얽매이지 않고, 촉매적으로 불활성인 이노신 글리코실레이스(예를 들어, 알킬 아데닌 글리코실레이스(AAG))는 이노신에 결합할 수 있지만, 무염기성 부위를 생성하거나 또는 이노신을 제거할 수 없으므로 DNA 손상/복구 메커니즘으로부터 새로 형성된 이노신 모이어티를 입체적으로 차단한다. 일부 실시형태에서, 촉매적으로 불활성인 이노신 특이적 뉴클레이스는 핵산에서 이노신에 결합할 수 있지만, 핵산을 절단하지 않는다. 비제한적인 예시적인 촉매적으로 불활성인 이노신 특이적 뉴클레이스는, 예를 들어, 인간으로부터 유래하는 촉매적으로 불활성인 알킬 아데노신 글리코실레이스(AAG 뉴클레이스) 및, 예를 들어, 대장균으로부터 유래하는 촉매적으로 불활성인 엔도뉴클레이스 V(EndoV 뉴클레이스)를 포함한다. 일부 실시형태에서, 촉매적으로 불활성인 AAG 뉴클레이스는 E125Q 돌연변이 또는 또 다른 AAG 뉴클레이스의 상응하는 돌연변이를 포함한다.

[0280] "증가하다"는 적어도 10%, 25%, 50%, 75% 또는 100%의 양의 변경을 의미한다.

[0281] "인테인"은 단백질 스플라이싱으로 알려진 과정에서 자체적으로 절단될 수 있으며 펩타이드 결합으로 나머지 단편(엑스테인)을 연결할 수 있는 단백질의 단편이다. 인테인은 또한 "단백질 인트론"으로도 지칭된다. 인테인이 자체적으로 절단되어 단백질의 나머지 부분을 연결하는 과정은 본 명세서에서 "단백질 스플라이싱" 또는 "인테인-매개성 단백질 스플라이싱"이라고 한다. 일부 실시형태에서, 전구체 단백질(인테인-매개성 단백질 스플라이싱 이전의 인테인 함유 단백질)의 인테인은 2개의 유전자로부터 유래한다. 이러한 인테인은 본 명세서에서 스플릿 인테인(예를 들어, 스플릿 인테인-N 및 스플릿 인테인-C)로 지칭된다. 예를 들어, 시아노박테리아에서, DNA 폴리머레이스 III의 촉매 서브유닛인 DnaE는 2개의 개별 유전자인 dnaE-n 및 dnaE-c에 의해 암호화된다. dnaE-n 유전자에 의해 암호화된 인테인은 본 명세서에서 "인테인-N"으로 지칭될 수 있다. dnaE-c 유전자에 의해 암호화된 인테인은 본 명세서에서 "인테인-C"로 지칭될 수 있다.

[0282] 다른 인테인 시스템이 또한 사용될 수 있다. 예를 들어, dnaE 인테인에 기반한 합성 인테인인 Cfa-N(예를 들어, 스플릿 인테인-N) 및 Cfa-C(예를 들어, 스플릿 인테인-C) 인테인 쌍이 기술되어 있다(예를 들어, 참조에 의해 본 명세서에 원용되어 있는 문헌[Stevens *et al.*, J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5]). 본 개시내용에 따라 사용될 수 있는 인테인 쌍의 비제한적인 예는 Cfa DnaE 인테인, Ssp GyrB 인테인, Ssp DnaX 인테인, Ter DnaE3 인테인, Ter ThyX 인테인, Rma DnaB 인테인 및 Cne Prp8 인테인을 포함한다(예를 들어, 참조에 의해 본 명세서에 원용되어 있는 미국 특허 제8,394,604호에 기술되어 있음).

[0283] 인테인의 예시적인 뉴클레오타이드 및 아미노산 서열은 아래에 제공된다.

[0284] DnaE 인테인-N DNA:

TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTG
TGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTCAGCC
AGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGT
CTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACG
[0285] AAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

[0286] DnaE 인테인-N 단백질:

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGS
[0287] LIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

[0288] DnaE 인테인-C DNA:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAA
[0289] GAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

[0290] 인테인-C:

[0291] MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

[0292] Cfa-N DNA:

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTG
TCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCC
CATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGC
ATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATG
[0293] AGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

[0294] Cfa-N 단백질:

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGS
[0295] IIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0296] Cfa-C DNA:

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAA
TATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTT
[0297] CCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

[0298] Cfa-C 단백질:

[0299] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

[0300] 인테인-N 및 인테인-C는 스플릿 Cas9의 N-말단 부분 및 스플릿 Cas9의 C-말단 부분의 연결을 위해 각각 스플릿 Cas9의 N-말단 부분 및 스플릿 Cas9의 C-말단 부분에 융합될 수 있다. 예를 들어, 일부 실시형태에서, 인테인-N은 스플릿 Cas9의 N-말단 부분의 C-말단에 융합되어, 즉, N--[스플릿 Cas9의 N-말단 부분]-[인테인-N]--C의 구조를 형성한다. 일부 실시형태에서, 인테인-C는 스플릿 Cas9의 C-말단 부분의 N-말단에 융합되어, 즉, N-[인테인-C]--[스플릿 Cas9의 C-말단 부분]-C의 구조를 형성한다. 인테인이 융합된 단백질(예를 들어, 스플릿 Cas9)을 연결하기 위한 인테인-매개성 단백질 스플라이싱의 메커니즘은, 예를 들어, 참조에 의해 본 명세서에 원용되어 있는 문헌[Shah *et al.*, Chem Sci. 2014; 5(1):446-461]에 기술되어 있는 바와 같이 당업계에 공지되어 있다. 인테인을 설계하고 사용하는 방법은 당업계에 공지되어 있으며, 예를 들어, 각각 그 전문이 참조에 의해 본 명세서에 원용되어 있는 WO2014004336, WO2017132580, US20150344549 및 US20180127780에 기술되어 있다.

[0301] 용어 "단리된", "정제된", 또는 "생물학적으로 순수한"은 천연 상태에서 발견되는 바와 같은 일반적으로 수반되

는 성분이 다양한 정도로 없는 물질을 지칭한다. "단리하다"는 원래의 공급원 또는 주변과의 분리 정도를 나타낸다. "정제하다"는 단리보다 더 높은 분리 정도를 나타낸다. "정제된" 또는 "생물학적으로 순수한" 단백질은 임의의 불순물이 단백질의 생물학적 특성에 실질적으로 영향을 미치거나 또는 다른 역효과를 일으키지 않도록 다른 물질이 충분히 없다. 즉, 본 발명의 핵산 또는 펩타이드는 재조합 DNA 기법에 의해 생성될 때 세포 물질, 바이러스 물질 또는 배양 배지가 실질적으로 없는 경우 또는 화학적으로 합성될 때 화학 전구체 또는 다른 화학 물질이 없는 경우 정제된다. 순도 및 균질성은 전형적으로 분석적 화학 기법, 예를 들어, 폴리아크릴아마이드 겔 전기영동 또는 고성능 액체 크로마토그래피를 사용하여 결정된다. 용어 "정제된"은 핵산 또는 단백질이 전기영동 겔에서 본질적으로 하나의 밴드를 생성함을 나타낼 수 있다. 변형, 예를 들어, 인산화 또는 글리코실화될 수 있는 단백질의 경우, 상이한 변형은 별도로 정제될 수 있는 상이한 단리된 단백질을 생성할 수 있다.

[0302] "단리된 폴리뉴클레오타이드"는 본 발명의 핵산 분자가 유래되는 유기체의 자연-발생적 게놈에서 유전자 측면에 있는 유전자가 없는 핵산(예를 들어, DNA)을 의미한다. 따라서, 상기 용어는, 예를 들어, 벡터에; 자가 복제 플라스미드 또는 바이러스에; 또는 원핵생물 또는 진핵생물의 게놈 DNA에 혼입되거나; 또는 다른 서열과 독립적인 별도의 분자(예를 들어, PCR 또는 제한 엔도뉴클레이스 분해에 의해 생성된 cDNA 또는 게놈 또는 cDNA 단편)로 존재하는 재조합 DNA를 포함한다. 또한, 용어는 DNA 분자로부터 전사된 RNA 분자뿐만 아니라 추가적인 폴리펩타이드 서열을 암호화하는 하이브리드 유전자의 일부인 재조합 DNA를 포함한다.

[0303] "단리된 폴리펩타이드"는 자연적으로 수반되는 성분으로부터 분리된 본 발명의 폴리펩타이드를 의미한다. 전형적으로, 폴리펩타이드는 단백질 및 자연적으로 회합된 자연-발생적 유기 분자가 중량 기준으로 적어도 60% 없을 때 단리된다. 바람직하게는, 제제는 중량 기준으로 적어도 75%, 보다 바람직하게는 적어도 90%, 가장 바람직하게는 적어도 99%가 본 발명의 폴리펩타이드이다. 본 발명의 단리된 폴리펩타이드는, 예를 들어, 천연 공급원으로부터의 추출함으로써, 이러한 폴리펩타이드를 암호화하는 재조합 핵산의 발현함으로써; 또는 단백질을 화학적으로 합성함으로써 얻어질 수 있다. 순도는 임의의 적절한 방법, 예를 들어, 칼럼 크로마토그래피, 폴리아크릴아마이드 겔 전기영동 또는 HPLC 분석에 의해 측정될 수 있다. "리더 펩타이드"는 새로 합성된 분비 또는 막 단백질을 막(예를 들어, 소포체 막)으로 그리고 이를 통해 지시하는 짧은 아미노산 서열(예를 들어, 대략 16개 내지 30개의 아미노산 길이)을 의미한다. 리더 펩타이드는 전형적으로 폴리펩타이드의 N-말단에 위치하며, 폴리펩타이드가 막을 통과한 후 신호 펩티데이스에 의해 제거될 수 있다. 리더 펩타이드 서열은 전형적으로 3개의 일반적인 구조적 특징을 포함한다: N-말단 극성 염기성 영역(n-영역), 소수성 코어 및 친수성 c-영역. 일부 실시형태에서, 본 발명의 CAR은 리더 펩타이드 서열(예를 들어, 항원 결합 도메인의 N-말단)을 포함한다. 예시적인 리더 펩타이드 아미노산 서열은 다음과 같다: METDTLLLWVLLLWVPGSTG.

[0304] 본 명세서에서 사용되는 용어 "링커"는 2개 분자 또는 모이어티를 연결하는 결합(예를 들어, 공유 결합), 화학기 또는 분자, 예를 들어, 단백질 복합체 또는 리보핵복합체의 2개의 성분 또는, 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인(예를 들어, dCas9) 및 데아미네이스 도메인(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스)과 같은 융합 단백질의 2개 도메인을 지칭한다. 링커는 염기 편집 시스템의 상이한 성분 또는 성분의 다른 부분을 연결할 수 있다. 예를 들어, 일부 실시형태에서, 링커는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 가이드 폴리뉴클레오타이드 결합 도메인과 데아미네이스의 촉매 도메인을 연결할 수 있다. 일부 실시형태에서, 링커는 CRISPR 폴리펩타이드와 데아미네이스를 연결할 수 있다. 일부 실시형태에서, 링커는 Cas9와 데아미네이스를 연결할 수 있다. 일부 실시형태에서, 링커는 dCas9와 데아미네이스를 연결할 수 있다. 일부 실시형태에서, 링커는 nCas9와 데아미네이스를 연결할 수 있다. 일부 실시형태에서, 링커는 가이드 폴리뉴클레오타이드와 데아미네이스를 연결할 수 있다. 일부 실시형태에서, 링커는 염기 편집 시스템의 탈아미노 성분과 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 성분을 연결할 수 있다. 일부 실시형태에서, 링커는 염기 편집 시스템의 탈아미노 성분과 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 성분의 RNA-결합 부분을 연결할 수 있다. 일부 실시형태에서, 링커는 염기 편집 시스템의 탈아미노 성분의 RNA-결합 부분과 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 성분의 RNA-결합 부분을 연결할 수 있다. 링커는 2개의 기, 분자 또는 다른 모이어티 사이에 위치하거나 또는 그 측면에 위치할 수 있으며, 공유 결합 또는 비공유 상호작용을 통해 각각 연결되어 둘을 연결한다. 일부 실시형태에서, 링커는 유기 분자, 기, 중합체 또는 화학 모이어티일 수 있다. 일부 실시형태에서, 링커는 폴리뉴클레오타이드일 수 있다. 일부 실시형태에서, 링커는 DNA 링커일 수 있다. 일부 실시형태에서, 링커는 RNA 링커일 수 있다. 일부 실시형태에서, 링커는 리간드에 결합할 수 있는 압타머를 포함할 수 있다. 일부 실시형태에서, 리간드는 탄수화물, 펩타이드, 단백질 또는 핵산일 수 있다. 일부 실시형태에서, 링커는 리보스위치로부터 유래될 수 있는 압타머를 포함할 수 있다. 압타머가 유래되는 리보스위치는 테오필린 리보스위치, 티아민 피로포스페이트(TPP) 리보스위치, 아데노신 코발라민(AdoCbl) 리보스위치, S-아데노실 메티오닌(SAM) 리보스위치, SAH 리보스위치, 플라

빈 모노뉴클레오타이드(FMN) 리보스위치, 테트라하이드로폴레이트 리보스위치, 라이신 리보스위치, 글리신 리보스위치, 퓨린 리보스위치, GlmS 리보스위치 또는 프리-퀘오신1(PreQ1) 리보스위치로부터 선택될 수 있다. 일부 실시형태에서, 링커는 폴리펩타이드 또는 폴리펩타이드 리간드와 같은 단백질 도메인에 결합된 압타머를 포함할 수 있다. 일부 실시형태에서, 폴리펩타이드 리간드는 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머레이스 Ku 결합 모티프 및 Ku 단백질, 텔로머레이스 Sm7 결합 모티프 및 Sm7 단백질 또는 RNA 인식 모티프일 수 있다. 일부 실시형태에서, 폴리펩타이드 리간드는 염기 편집 시스템 성분의 일부일 수 있다. 예를 들어, 핵염기 편집 성분은 데아미네이스 도메인 및 RNA 인식 모티프를 포함할 수 있다.

[0305] 일부 실시형태에서, 링커는 아미노산 또는 복수의 아미노산(예를 들어, 펩타이드 또는 단백질)일 수 있다. 일부 실시형태에서, 링커는 약 5개 내지 100개의 아미노산 길이, 예를 들어, 약 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 20개 내지 30개, 30개 내지 40개, 40개 내지 50개, 50개 내지 60개, 60개 내지 70개, 70개 내지 80개, 80개 내지 90 또는 90개 내지 100개의 아미노산 길이일 수 있다. 일부 실시형태에서, 링커는 약 100개 내지 150개, 150개 내지 200개, 200개 내지 250개, 250개 내지 300개, 300개 내지 350개, 350개 내지 400개, 400개 내지 450 또는 450개 내지 500개의 아미노산 길이일 수 있다. 더 길거나 또는 더 짧은 링커도 또한 상정될 수 있다.

[0306] 일부 실시형태에서, 링커는 Cas9 뉴클레이스 도메인을 포함하는 RNA-프로그램 가능한 뉴클레이스의 gRNA 결합 도메인과 핵산 편집 단백질(예를 들어, 사이티딘 또는 아데노신 데아미네이스)의 촉매 도메인을 연결한다. 일부 실시형태에서, 링커는 dCas9와 핵산 편집 단백질을 연결한다. 예를 들어, 링커는 2개의 기, 분자 또는 다른 모이어티 사이에 위치하거나 또는 그 측면에 위치하며, 공유 결합을 통해 각각에 연결되어 둘을 연결한다. 일부 실시형태에서, 링커는 아미노산 또는 복수의 아미노산(예를 들어, 펩타이드 또는 단백질)이다. 일부 실시형태에서, 링커는 유기 분자, 기, 중합체 또는 화학 모이어티이다. 일부 실시형태에서, 링커는 5개 내지 200개의 아미노산 길이, 예를 들어, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 25개, 35개, 45개, 50개, 55개, 60개, 60개, 65개, 70개, 70개, 75개, 80개, 85개, 90개, 90개, 95개, 100개, 101개, 102개, 103개, 104개, 105개, 110개, 120개, 130개, 140개, 150개, 160개, 175개, 180개, 190개 또는 200개의 아미노산 길이이다. 더 길거나 또는 더 짧은 링커가 또한 상정된다. 일부 실시형태에서, 링커는 XTEN 링커로도 지칭될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함한다. 일부 실시형태에서, 링커는 아미노산 서열 SGGS를 포함한다. 일부 실시형태에서, 링커는 $(SGGS)_n$, $(GGGS)_n$, $(GGGGS)_n$, $(G)_n$, $(EAAAK)_n$, $(GGS)_n$, SGSETPGTSESATPES, 또는 $(XP)_n$ 모티프 또는 임의의 이들의 조합을 포함하되, n은 독립적으로 1 내지 30의 정수이고, X는 임의의 아미노산이다. 일부 실시형태에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15이다. 일부 실시형태에서, 링커는 복수의 프롤린 잔기를 포함하고, 5개 내지 21개, 5개 내지 14개, 5개 내지 9개, 5개 내지 7개의 아미노산 길이이며, 예를 들어, PAPAP, PAPAPA, PAPAPAP, PAPAPAPA, $P(AP)_4$, $P(AP)_7$, $P(AP)_{10}$이다. 이러한 프롤린-풍부 링커는 또한 "강성(rigid)" 링커라고도 한다.

[0307] 일부 실시형태에서, 키메라 항원 수용체는 적어도 하나의 링커를 포함한다. 적어도 하나의 링커는 키메라 항원 수용체의 세포외 결합 도메인의 가변 중쇄(VH) 영역을 불변 중쇄(CH) 영역에 결합하거나 또는 연결한다. 링커는 또한 세포외 결합 도메인의 가변 경쇄(VL) 영역을 가변 불변(VC) 영역에 연결할 수 있다.

[0308] 일부 실시형태에서, 염기 편집기의 도메인은 SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS, 또는 GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS의 아미노산 서열을 포함하는 링커를 통해 융합된다.

[0309] 일부 실시형태에서, 염기 편집기의 도메인은 XTEN 링커로도 지칭될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다. 일부 실시형태에서, 링커는 아미노산 서열 SGGS를 포함한다. 일부 실시형태에서, 링커는 24개의 아미노산 길이이다. 일부 실시형태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPES를 포함한다. 일부 실시형태에서, 링커는 40개의 아미노산 길이이다. 일부 실시형태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS를 포함한다. 일부 실시형태에서, 링커는 64개의 아미노산 길이이다. 일부 실시형태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS를 포함한다. 일부 실시형태에서, 링커는 92개의 아미노산 길이이다. 일부 실시형태에서, 링커는 아미노산 서열 PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS를 포함한

다.

[0310] 본 명세서에서 사용되는 용어 "액체암"은, 예를 들어, 혈액, 림프 및 골수와 같은 체액에 존재하는 암세포를 지칭한다. 액체암은 백혈병, 골수종 및 액체 림프종을 포함하지만, 이들로 제한되지 않는다. 본 명세서에서 사용되는 바와 같은 액체암은 육종 및 암종과 같은 고형 종양 또는 낭종 또는 액체 영역을 포함하지 않는 고형 림프종을 포함하지 않는다. "액체암"은 재발성, 불응성 또는 전이성일 수 있다. 본 명세서에 기재된 방법으로 치료되는 액체암은, 예를 들어, 액체 림프종일 수 있으며; 액체 림프종은 낭종 또는 액체 영역을 포함하는 림프종을 포함한다.

[0311] "림프구-활성화 유전자 3(LAG-3) 폴리펩타이드"는 NCBI 등록 번호 NP_002277.4 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0312] > NP_002277.4 림프구 활성화 유전자 3 단백질 전구체 [호모 사피엔스]

```
  1 MWEAQFLGLL FLQPLWVAPV KPLQPGAEVP VVWAQEGAPA QLPCSPTIPL QDLSLLRRAG
 61 VTWQHQPDSG PPAAAPGHPL APGPHPAAPS SWGPRPRRYT VLSVGPGGLR SGRLPLQPRV
121 QLDERGRQRG DFSLWLRPAR RADAGEYRAA VHLRDRALSC RLRLRLGQAS MTASPPGSLR
181 ASDWVILNCS FSRPDRPASV HWFRNRGQGR VPVRESPHHH LAESFLFLPQ VSPMDSGPWG
241 CILTYRDGFN VSIMYNLTVL GLEPPTPLTV YAGAGSRVGL PCRLPAGVGT RSFLTAKWTP
301 PGGGPDLLVT GDNGDFTLRL EDVSQAQAGT YTCHIHLQEQ QLNATVTLAI ITVTPKSFGS
361 PGSLGKLLCE VTPVSGQERF VWSSLDTPSQ RSFSGPWLEA QEAQLLSQPW QCQLYQGERL
421 LGAAVYFTEL SSPGAQRSGR APGALPAGHL LLFLILGVLS LLLLVTGAFG FHLWRRQWRP
481 RRFSALEQGI HPPQAQSKIE ELEQEPEPEP EPEPEPEPEP EPEQL
```

[0313]

[0314] "림프구-활성화 유전자 3(LAG-3) 폴리뉴클레오타이드"는 LAG-3 폴리펩타이드를 암호화하는 핵산을 의미한다. 예시적인 LAG-3 핵산 서열이 아래에 제공된다.

[0315] > NM_002286.6 호모 사피엔스 림프구 활성화 3(LAG3), mRNA

```
   1 agagaccagc agaacggcat cccagccacg acggccactt tgctctgtct gctctccgcc
  61 acggccctgc tctgttccct gggacacccc cgcccccacc tcctcaggct gcctgatctg
 121 cccagctttc cagctttcct ctggattccg gcctctggtc atccctcccc accctctctc
 181 caaggccctc tcctggtctc cctttcttcta gaaccccttc ctccacctcc ctctctgcag
 241 aacttctcct ttaccccccа ccccccacca ctgccccctt tccttttctg acctcctttt
 301 ggagggctca gcgctgccca gaccatagga gagatgtggg aggctcagtt cctgggcttg
 361 ctgtttctgc agccgctttg ggtggctcca gtgaagcctc tccagccagg ggctgaggtc
 421 ccggtggtgt gggcccagga gggggctcct gcccagctcc cctgcagccc cacaatcccc
 481 ctccaggatc tcagccttct gcgaagagca ggggtcactt ggcagcatca gccagacagt
 541 ggcccgcccg ctgccgcccc cggccatccc ctggcccccg gccctcaccc ggcggcgccc
 601 tcctcctggg ggcccaggcc ccgccgctac acggtgctga gcgtgggtcc cggaggcctg
 661 cgcagcggga ggctgccсct gcagccccgc gtccagctgg atgagcgcgg ccggcagcgc
 721 ggggacttct cgctatggct gcgcccagcc cggcgcgcgg acgccggcga gtaccgcgcc
 781 gcggtgcacc tcagggaccg cgccctctcc tgccgcctcc gtctgcgcct gggccaggcc
 841 tcgatgactg ccagcccccc aggatctctc agagcctccg actgggtcat tttgaactgc
 901 tccttcagcc gccctgaccg cccagcctct gtgcattggt tccggaaccg gggccagggc
 961 cgagtccctg tccgggagtc cccccatcac cacttagcgg aaagcttcct cttcctgccc
1021 caagtcagcc ccatggactc tgggccctgg ggctgcatcc tcacctacag agatggcttc
1081 aacgtctcca tcatgtataa cctcactgtt ctgggtctgg agcccccaac tcccttgaca
1141 gtgtacgctg gagcaggttc cagggtgggg ctgccctgcc gcctgcctgc tggtgtgggg
1201 acccggtctt tcctcactgc caagtggact cctcctgggg gaggccctga cctcctggtg
1261 actggagaca atggcgactt taccccttcga ctagaggatg tgagccaggc ccaggctggg
1321 acctacacct gccatatcca tctgcaggaa cagcagctca atgccactgt cacattggca
1381 atcatcacag tgactcccaa atcctttggg tcacctggat ccctggggaa gctgctttgt
1441 gaggtgactc cagtatctgg acaagaacgc tttgtgtgga gctctctgga caccccatcc
1501 cagaggagtt tctcaggacc ttggctggag gcacaggagg cccagctcct ttcccagcct
1561 tggcaatgcc agctgtacca gggggagagg cttcttggag cagcagtgta cttcacagag
1621 ctgtctagcc caggtgccca acgctctggg agagccccag gtgccctccc agcaggccac
1681 ctcctgctgt ttctcatcct tggtgtcctt tctctgctcc ttttggtgac tggagccttt
1741 ggctttcacc tttggagaag acagtggcga ccaagacgat tttctgcctt agagcaaggg
1801 attcaccctc cgcaggctca gagcaagata gaggagctgg agcaagaacc ggagccggag
1861 ccggagccgg aaccggagcc cgagcccgag cccgagccgg agcagctctg acctggagct
1921 gaggcagcca gcagatctca gcagcccagt ccaaataaac tccctgtcag cagcaa
```

[0316]

[0317] "마커"는 질환 또는 장애와 관련된 발현 수준 또는 활성의 변경을 나타내는 임의의 단백질 또는 폴리뉴클레오타이드를 의미한다.

[0318] 본 명세서에서 사용되는 용어 "돌연변이"는 서열, 예를 들어, 핵산 또는 아미노산 서열 내의 잔기의 또 다른 잔기로의 치환 또는 서열 내의 하나 이상의 잔기의 결실 또는 삽입을 지칭한다. 전형적으로 본 명세서에 기재된 돌연변이는 원래의 잔기를 확인하고 이어서 서열 내 잔기의 위치를 확인하고, 새로 치환된 잔기의 정체를 확인함으로써 확인된다. 본 명세서에 제공된 아미노산 치환(돌연변이)을 만드는 다양한 방법이 당업계에 잘 알려져 있으며, 예를 들어, 문헌[Green and Sambrook, *Molecular Cloning: A Laboratory Manual* ($4^{th}$ ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 의해 제공된다. 일부 실시형태에서, 현재 개시된 염기 편집기는 의도하지 않은 점 돌연변이와 같은 상당수의 의도하지 않은 돌연변이를 생성하지 않고 핵산(예를 들어, 대상체의 게놈 내의 핵산)에서 점 돌연변이와 같은 "의도된 돌연변이"를 효율적으로 생성할 수 있다. 일부 실시형태에서, 의도된 돌연변이는 의도된 돌연변이를 생성하도록 특별히 설계된, 가이드 폴리뉴클레오타이드(예를 들어, gRNA)에 결합된 특정 염기 편집기(예를 들어, 사이티딘 염기 편집기 또는 아데노신 염기 편집기)에 의해 생성되는 돌연변이이다.

[0319] 일반적으로, 서열(예를 들어, 본 명세서에 기재된 바와 같은 아미노산 서열)에서 만들어지거나 또는 확인된 돌연변이는 참조(또는 야생형) 서열, 즉, 돌연변이를 포함하지 않는 서열과 관련하여 넘버링된다. 당업자는 참조 서열에 대하여 아미노산 및 핵산 서열에서 돌연변이의 위치를 결정하는 방법을 쉽게 이해할 것이다.

[0320] "신생물"은 비정상적인 성장 또는 증식을 나타내는 세포 또는 조직을 지칭한다. 용어 신생물은 암, 액체 및 고형 종양을 포함한다. 일부 실시형태에서, 신생물은 고형 종양이다. 다른 실시형태에서, 신생물은 액체 종양이다. 일부 실시형태에서, 신생물은 혈액암이다. 일부 실시형태에서, 혈액암은 백혈병, 골수종 및/또는 림프종이다. 일부 실시형태에서, 혈액암은 B 세포암이다. 일부 실시형태에서, B 세포암은 림프종 또는 백혈병이다. 일부 경우에, 백혈병은 전백혈병을 포함한다. 일부 경우에, 백혈병은 급성 백혈병이다. 급성 백혈병은, 예를 들어, 급성 골수성 백혈병(AML)을 포함한다. 급성 백혈병은 또한, 예를 들어, 급성 림프성 백혈병 또는 급성 림프구성 백혈병(ALL)을 포함하고; ALL은 B-계통 ALL; T-계통 ALL; 및 T-세포 급성 림프구성 백혈병(T-ALL)을 포함한다.

[0321] 신생물의 비제한적인 예는 T-세포 급성 림프아구성 백혈병(T-ALL), 용상 진균증(MF), 세자리 증후군(SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 $ALK^+$, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 $CD30^+$ 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종을 포함한다. 일부 실시형태에서, 신생물은 T-세포 급성 림프아구성 백혈병(T-ALL)이다. 일부 실시형태에서, 신생물은 급성 골수성 백혈병(AML)이다.

[0322] "활성화된 T 세포의 핵 인자 1(NFATc1) 폴리펩타이드"는 NCBI 등록 번호 NM_172390.2 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 활성화된 T 세포 DNA-결합 전사 복합체의 성분인 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0323] >NP_765978.1 활성화된 T-세포의 핵 인자, 세포질 1 동형단백질 A [호모 사피엔스]

MPSTSFPVPSKFPLGPAAAVFGRGETLGPAPRAGGTMKSAEEEHYGYASSNVSPALPLPTAHST
LPAPCHNLQTSTPGIIPPADHPSGYGAALDGGPAGYFLSSGHTRPDGAPALESPRIEITSCLGL
YHNNNQFFHDVEVEDVLPSSKRSPSTATLSLPSLEAYRDPSCLSPASSLSSRSCNSEASSYESN
YSYPYASPQTSPWQSPCVSPKTTDPEEGFPRGLGACTLLGSPRHSPSTSPRASVTEESWLGARS
SRPASPCNKRKYSLNGRQPPYSPHHSPTPSPHGSPRVSVTDDSWLGNTTQYTSSAIVAAINALT
TDSSLDLGDGVPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQHIRKGGFCD
QYLAVPQHPYQWAKPKPLSPTSYMSPTLPALDWQLPSHSGPYELRIEVQPKSHHRAHYETEGSR
GAVKASAGGHPIVQLHGYLENEPLMLQLFIGTADDRLLRPHAFYQVHRITGKTVSTTSHEAILS
NTKVLEIPLLPENSMRAVIDCAGILKLRNSDIELRKGETDIGRKNTRVRLVFRVHVPQPSGRTL
SLQVASNPIECSQRSAQELPLVEKQSTDSYPVVGGKKMVLSGHNFLQDSKVIFVEKAPDGHHVW
EMEAKTDRDLCKPNSLVVEIPPFRNQRITSPVHVSFYVCNGKRKRSQYQRFTYLPANGNAIFLT
VSREHERVGCFF

[0324]

[0325] "활성화된 T 세포의 핵 인자 1(NFATc1) 폴리뉴클레오타이드"는 NFATc1 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. NFATc1 유전자는 사이토카인 유전자, 특히 T-세포에서 IL-2 및 IL-4의 유도성 발현에 관여하는 단백질을 암호화한다. 예시적인 핵산 서열이 아래에 제공된다.

[0326] >NM_172390.2 호모 사피엔스 활성화된 T 세포 핵 인자 1(NFATC1), 전사체 변이체 1, mRNA

GGCGGGCGCTCGGCGACTCGTCCCCGGGGCCCCGCGCGGGCCCGGGCAGCAGGGGCGTGATGTCACGGCA
GGGAGGGGGCGCGGGAGCCGCCGGGCCGGCGGGGAGGCGGGGGAGGTGTTTTCCAGCTTTAAAAAGGCAG
GAGGCAGAGCGCGGCCCTGCGTCAGAGCGAGACTCAGAGGCTCCGAACTCGCCGGCGGAGTCGCCGCGCC
AGATCCCAGCAGCAGGGCGCGGGCACCGGGGCGCGGGCAGGGCTCGGAGCCACCGCGCAGGTCCTAGGGC
CGCGGCCGGGCCCCGCCACGCGCGCACACGCCCCTCGATGACTTTCCTCCGGGGCGCGCGGCGCTGAGCC
CGGGGCGAGGGCTGTCTTCCCGGAGACCCGACCCCGGCAGCGCGGGGCGGCCGCTTCTCCTGTGCCTCCG
CCCGCCGCTCCACTCCCCGCCGCCGCCGCGCGGATGCCAAGCACCAGCTTTCCAGTCCCTTCCAAGTTTC
CACTTGGCCCTGCGGCTGCGGTCTTCGGGAGAGGAGAAACTTTGGGGCCCGCGCCGCGCGCCGGCGGCAC
CATGAAGTCAGCGGAGGAAGAACACTATGGCTATGCATCCTCCAACGTCAGCCCCGCCCTGCCGCTCCCC
ACGGCGCACTCCACCCTGCCGGCCCCGTGCCACAACCTTCAGACCTCCACACCGGGCATCATCCCGCCGG
CGGATCACCCCTCGGGGTACGGAGCAGCTTTGGACGGTGGGCCCGCGGGCTACTTCCTCTCCTCCGGCCA
CACCAGGCCTGATGGGGCCCCTGCCCTGGAGAGTCCTCGCATCGAGATAACCTCGTGCTTGGGCCTGTAC
CACAACAATAACCAGTTTTTCCACGATGTGGAGGTGGAAGACGTCCTCCCTAGCTCCAAACGGTCCCCCT
CCACGGCCACGCTGAGTCTGCCCAGCCTGGAGGCCTACAGAGACCCCTCGTGCCTGAGCCCGGCCAGCAG
CCTGTCCTCCCGGAGCTGCAACTCAGAGGCCTCCTCCTACGAGTCCAACTACTCGTACCCGTACGCGTCC
CCCCAGACGTCGCCATGGCAGTCTCCCTGCGTGTCTCCCAAGACCACGGACCCCGAGGAGGGCTTTCCCC
GCGGGCTGGGGGCCTGCACACTGCTGGGTTCCCCGCGGCACTCCCCCTCCACCTCGCCCCGCGCCAGCGT
CACTGAGGAGAGCTGGCTGGGTGCCCGCTCCTCCAGACCCGCGTCCCCTTGCAACAAGAGGAAGTACAGC
CTCAACGGCCGGCAGCCGCCCTACTCACCCCACCACTCGCCCACGCCGTCCCCGCACGGCTCCCCGCGGG
TCAGCGTGACCGACGACTCGTGGTTGGGCAACACCACCCAGTACACCAGCTCGGCCATCGTGGCCGCCAT
CAACGCGCTGACCACCGACAGCAGCCTGGACCTGGGAGATGGCGTCCCTGTCAAGTCCCGCAAGACCACC
CTGGAGCAGCCGCCCTCAGTGGCGCTCAAGGTGGAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCC
CGGCCGACTTCGCGCCCGAAGACTACTCCTCTTTCCAGCACATCAGGAAGGGCGGCTTCTGCGACCAGTA
CCTGGCGGTGCCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCCCTGTCCCCTACGTCCTACATGAGC
CCGACCCTGCCCGCCCTGGACTGGCAGCTGCCGTCCCACTCAGGCCCGTATGAGCTTCGGATTGAGGTGC
AGCCCAAGTCCCACCACCGAGCCCACTACGAGACGGAGGGCAGCCGGGGGGCCGTGAAGGCGTCGGCCGG
AGGACACCCCATCGTGCAGCTGCATGGCTACTTGGAGAATGAGCCGCTGATGCTGCAGCTTTTCATTGGG
ACGGCGGACGACCGCCTGCTGCGCCCGCACGCCTTCTACCAGGTGCACCGCATCACAGGGAAGACCGTGT
CCACCACCAGCCACGAGGCCATCCTCTCCAACACCAAAGTCCTGGAGATCCCACTCCTGCCGGAGAACAG
CATGCGAGCCGTCATTGACTGTGCCGGAATCCTGAAACTCAGAAACTCCGACATTGAACTTCGGAAAGGA
GAGACGGACATCGGGAGGAAGAACACACGGGTACGGCTGGTGTTCCGCGTTCACGTCCCGCAACCCAGCG
GCCGCACGCTGTCCCTGCAGGTGGCCTCCAACCCCATCGAATGCTCCCAGCGCTCAGCTCAGGAGCTGCC
TCTGGTGGAGAAGCAGAGCACGGACAGCTATCCGGTCGTGGGCGGGAAGAAGATGGTCCTGTCTGGCCAC
AACTTCCTGCAGGACTCCAAGGTCATTTTCGTGGAGAAAGCCCCAGATGGCCACCATGTCTGGGAGATGG
AAGCGAAAACTGACCGGGACCTGTGCAAGCCGAATTCTCTGGTGGTTGAGATCCCGCCATTTCGGAATCA
GAGGATAACCAGCCCCGTTCACGTCAGTTTCTACGTCTGCAACGGGAAGAGAAAGCGAAGCCAGTACCAG

[0327] CGTTTCACCTACCTTCCCGCCAACGGTAACGCCATCTTTCTAACCGTAAGCCGTGAACATGAGCGCGTGG

GGTGCTTTTTCTAAAGACGCAGAAACGACGTCGCCGTAAAGCAGCGTGGCGTGTTGCACATTTAACTGTG
TGATGTCCCGTTAGTGAGACCGAGCCATCGATGCCCTGAAAAGGAAAGGAAAAGGGAAGCTTCGGATGCA
TTTTCCTTGATCCCTGTTGGGGGTGGGGGGCGGGGGTTGCATACTCAGATAGTCACGGTTATTTTGCTTC
TTGCGAATGTATAACAGCCAAGGGGAAAACATGGCTCTTCTGCTCCAAAAAACTGAGGGGGTCCTGGTGT
GCATTTGCACCCTAAAGCTGCTTACGGTGAAAAGGCAAATAGGTATAGCTATTTTGCAGGCACCTTTAGG

[0328] AATAAACTTTGCTTTTAAGCCTGTAAAAAAAAAAAAA

[0329] 용어 "비-보존적 돌연변이"는 상이한 그룹 사이의 아미노산 치환, 예를 들어, 트립토판을 라이신으로 또는 세린을 페닐알라닌 등으로의 치환을 포함한다. 이러한 경우에, 비-보존적 아미노산 치환은 기능적 변이체의 생물학적 활성을 방해하거나 또는 저해하지 않는 것이 바람직하다. 비-보존적 아미노산 치환은 기능적 변이체의 생물학적 활성이 야생형 단백질과 비교하여 증가되도록 기능적 변이체의 생물학적 활성을 향상시킬 수 있다.

[0330] 용어 "핵 국부화 서열", "핵 국부화 신호" 또는 "NLS"는 단백질의 세포 핵 내로의 도입을 촉진하는 아미노산 서열을 지칭한다. 핵 국부화 서열은 당업계에 공지되어 있으며, 예를 들어, 예시적인 핵 국부화 서열의 개시를 위해 그 내용이 참조에 의해 본 명세서에 원용되어 있는, 2000년 11월 23일자로 출원되고 2001년 5월 31일자로

WO/2001/038547로 공개된 Plank 등의 PCT 출원 PCT/EP2000/011690에 기술되어 있다. 다른 실시형태에서, NLS는, 예를 들어, 문헌[Koblan *et al.*, Nature Biotech. 2018 doi:10.1038/nbt.4172]에 의해 기술된 최적화된 NLS이다. 일부 실시형태에서, NLS는 아미노산 서열 PKKKRKVEGADKRTADGSEFESPKKKRKV, KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRK, PKKKRKV 또는 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC를 포함한다.

[0331] 본 명세서에서 사용되는 용어 "핵산" 및 "핵산 분자"는 핵염기 및 산성 모이어티, 예를 들어, 뉴클레오사이드, 뉴클레오타이드 또는 뉴클레오타이드의 중합체를 포함하는 화합물을 지칭한다. 전형적으로, 중합체성 핵산, 예를 들어, 3개 이상의 뉴클레오타이드를 포함하는 핵산 분자는 선형 분자이며, 여기서 인접 뉴클레오타이드는 포스포다이에스터 연결을 통해 서로 연결된다. 일부 실시형태에서, "핵산"은 개별 핵산 잔기(예를 들어, 뉴클레오타이드 및/또는 뉴클레오사이드)를 지칭한다. 일부 실시형태에서, "핵산"은 3개 이상의 개별 뉴클레오타이드 잔기를 포함하는 올리고뉴클레오타이드 사슬을 지칭한다. 본 명세서에서 사용되는 용어 "올리고뉴클레오타이드" 및 "폴리뉴클레오타이드"는 뉴클레오타이드의 중합체(예를 들어, 적어도 3개의 뉴클레오타이드의 스트링(string))를 지칭하기 위해 상호교환적으로 사용될 수 있다. 일부 실시형태에서, "핵산"은 RNA뿐만 아니라 단일 및/또는 이중-가닥 DNA를 포함한다. 핵산은, 예를 들어, 게놈, 전사체, mRNA, tRNA, rRNA, siRNA, snRNA, 플라스미드, 코스미드, 염색체, 염색분체 또는 다른 자연 발생적 핵산 분자의 맥락에서 자연 발생적일 수 있다. 다른 한편으로, 핵산 분자는 비-자연 발생적 분자, 예를 들어, 재조합 DNA 또는 RNA, 인공 염색체, 조작된 게놈 또는 이의 단편, 또는 합성 DNA, RNA, DNA/RNA 하이브리드, 또는 비-자연 발생적 뉴클레오타이드 또는 뉴클레오사이드일 수 있다. 또한, 용어 "핵산", "DNA", "RNA" 및/또는 유사한 용어는 핵산 유사체, 예를 들어, 포스포다이에스터 백본 이외의 것을 갖는 유사체를 포함한다. 핵산은 천연 공급원으로부터 정제되고, 재조합 발현 시스템을 사용하여 생성되고 선택적으로 정제될 수 있으며, 화학적으로 합성될 수 있다. 적절한 경우, 예를 들어, 화학적으로 합성된 분자의 경우에, 핵산은 화학적으로 변형된 염기 또는 당 및 백본 변형을 갖는 유사체와 같은 뉴클레오사이드 유사체를 포함할 수 있다. 핵산 서열은 달리 지시되지 않는 한 5'에서 3' 방향으로 존재한다. 일부 실시형태에서, 핵산은 천연 뉴클레오사이드(예를 들어, 아데노신, 티미딘, 구아노신, 사이티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시구아노신 및 데옥시사이티딘); 뉴클레오사이드 유사체(예를 들어, 2-아미노아데노신, 2-티오티미딘, 이노신, 피롤로-피리미딘, 3-메틸 아데노신, 5-메틸사이티딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-아이오도우리딘, C5-프로핀일-우리딘, C5-프로핀일-사이티딘, C5-메틸사이티딘, 2-아미노아데노신, 7-데아자아데노신, 7-데아자구아노신, 8-옥소아데노신, 8-옥소구아노신, O(6)-메틸구아닌 및 2-티오사이티딘); 화학적으로 변형된 염기; 생물학적으로 변형된 염기(예를 들어, 메틸화된 염기); 삽입된 염기; 변형된 당(예를 들어, 2'-리보스, 2'-데옥시리보스, 아라비노스 및 헥소스); 및/또는 변형된 포스페이트기(예를 들어, 포스포로티오에이트 및 5'-*N*-포스포라미다이트 연결)이거나 또는 이를 포함한다. 용어 "핵산 프로그램 가능 DNA 결합 단백질" 또는 "napDNAbp"는 napDNAbp를 특정 핵산 서열로 가이드하는 가이드 핵산 또는 가이드 폴리뉴클레오타이드(예를 들어, gRNA)와 같은 핵산(예를 들어, DNA 또는 RNA)과 회합하는 단백질을 지칭하기 위해 "폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인"과 상호교환적으로 사용될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인이다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그램 가능한 RNA 결합 도메인이다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 Cas9 단백질이다. Cas9 단백질은 Cas9 단백질을 가이드 RNA에 상보적인 특정 DNA 서열로 가이드하는 가이드 RNA와 회합할 수 있다. 일부 실시형태에서, napDNAbp는 Cas9 도메인, 예를 들어, 뉴클레이스 활성이 있는 Cas9, Cas9 닉케이스(nCas9) 또는 뉴클레이스 불활성 Cas9(dCas9)이다. 핵산 프로그램 가능 DNA 결합 단백질의 비제한적인 예는 Cas9(예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 및 Cas12j/CasΦ를 포함한다. Cas 효소의 비제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9(Csn1 또는 Csx12로도 알려져 있음), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, 유형 II Cas 효과기 단백질, 유형 V Cas 효과기 단백질, 유형 VI Cas 효과기 단백질, CARF, DinG, 이들의 상동체 또는 이들의 변형된 또는 조작된 버전을 포함한다. 다른 핵산 프로그램 가능 DNA 결합 단백질도 본 개시내용의 범위 내에 있지만, 이들은 본 개시내용에 구체적으로 열거되지 않을 수 있다. 예를 들어, 각각

그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Makarova *et al.* "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" *CRISPR J.* 2018 Oct;1:325-336. doi: 10.1089/crispr.2018.0033; Yan *et al.*, "Functionally diverse type V CRISPR-Cas systems" *Science*. 2019 Jan 4;363(6422):88-91. doi: 10.1126/science.aav7271]을 참조한다.

[0332] 본 명세서에서 상호교환적으로 사용되는 용어 "핵염기", "질소 염기" 또는 "염기"는 결과적으로 뉴클레오타이드의 성분인, 뉴클레오사이드를 형성하는 질소-함유 생물학적 화합물을 지칭한다. 염기쌍을 형성하고 서로 쌓이는 핵염기의 능력은 리보핵산(RNA) 및 데옥시리보핵산(DNA)과 같은 장쇄 나선 구조로 직접 이어진다. 5가지의 핵염기 아데닌(A), 사이토신(C), 구아닌(G), 티민(T) 및 우라실(U)을 1차 또는 표준(표준)이라고 한다. 아데닌과 구아닌은 퓨린으로부터 파생되고, 사이토신, 우라실 및 티민은 피리미딘으로부터 파생된다. DNA와 RNA는 또한 변형된 다른(1차가 아닌) 염기를 포함할 수 있다. 비제한적인 예시적인 변형된 핵염기는 하이포잔틴, 잔틴, 7-메틸구아닌, 5,6-다이하이드로우라실, 5-메틸사이토신(m5C) 및 5-하이드로메틸사이토신을 포함할 수 있다. 하이포잔틴 및 잔틴은 돌연변이 유발물질의 존재를 통해 생성될 수 있으며, 둘 다 탈아미노화(아민기를 카보닐기로 대체)를 통해 생성될 수 있다. 하이포잔틴은 아데닌으로부터 변형될 수 있다. 잔틴은 구아닌으로부터 변형될 수 있다. 우라실은 사이토신의 탈아미노화로부터 생성될 수 있다. "뉴클레오사이드"는 핵염기 및 5개의 탄소 당(리보스 또는 데옥시리보스)으로 구성된다. 뉴클레오사이드의 예는 아데노신, 구아노신, 우리딘, 사이티딘, 5-메틸우리딘(m5U), 데옥시아데노신, 데옥시구아노신, 티미딘, 데옥시우리딘 및 데옥시사이티딘을 포함한다. 변형된 핵염기를 갖는 뉴클레오사이드의 예는 이노신(I), 잔토신(X), 7-메틸구아노신(m7G), 다이하이드로우리딘(D), 5-메틸사이티딘(m5C) 및 슈도우리딘(Ψ)을 포함한다. "뉴클레오타이드"는 핵염기, 5개의 탄소 당(리보스 또는 데옥시리보스) 및 적어도 하나의 포스페이트기로 구성된다.

[0333] 본 명세서에서 사용되는 용어 "핵염기 편집 도메인" 또는 "핵염기 편집 단백질"은 RNA 또는 DNA에서 핵염기 변형, 예컨대, 사이토신(또는 사이티딘)에서 우라실(또는 우리딘) 또는 티민(또는 티미딘)으로의 변형, 아데닌(또는 아데노신)에서 하이포잔틴(또는 이노신) 탈아미노화뿐만 아니라 비-주형 뉴클레오타이드 추가 및 삽입을 촉매할 수 있는 단백질 또는 효소를 지칭한다. 일부 실시형태에서, 핵염기 편집 도메인은 데아미네이스 도메인(예를 들어, 아데닌 데아미네이스 또는 아데노신 데아미네이스; 또는 사이티딘 데아미네이스 또는 사이토신 데아미네이스)이다. 일부 실시형태에서, 핵염기 편집 도메인은 하나 이상의 데아미네이스 도메인(예를 들어, 아데닌 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 또는 사이토신 데아미네이스)이다. 일부 실시형태에서, 핵염기 편집 도메인은 자연 발생적 핵염기 편집 도메인일 수 있다. 일부 실시형태에서, 핵염기 편집 도메인은 자연 발생적 핵염기 편집 도메인으로부터 조작된 또는 진화된 핵염기 편집 도메인일 수 있다. 핵염기 편집 도메인은 세균, 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스와 같은 임의의 유기체로부터 유래할 수 있다.

[0334] 본 명세서에서 사용되는 "작용제를 얻는"에서와 같이 "얻는"은 작용제를 합성, 구매 또는 그렇지 않으면 획득하는 것을 포함한다. 본 명세서에서 사용되는 "환자" 또는 "대상체"는 질환 또는 장애로 진단되거나, 이를 가지거나 발병할 위험이 있거나 또는 이를 가지거나 발병할 것으로 의심되는 포유동물 대상체 또는 개체를 지칭한다. 일부 실시형태에서, 용어 "환자"는 질환 또는 장애가 발병할 가능성이 평균보다 높은 포유동물 대상체를 지칭한다. 예시적인 환자는 인간, 비-인간 영장류, 고양이, 개, 돼지, 소, 고양이, 말, 낙타, 라마, 염소, 양, 설치류(예를 들어, 마우스, 토끼, 래트 또는 기니피그) 및 본 명세서에 개시된 요법으로부터 혜택을 받을 수 있는 다른 포유동물일 수 있다. 예시적인 인간 환자는 남성 및/또는 여성일 수 있다.

[0335] "이를 필요로 하는 환자" 또는 "이를 필요로 하는 대상체"는 본 명세서에서 질환 또는 장애(예를 들어, T-세포 또는 NK-세포 악성종양)로 진단되거나, 이의 위험이 있거나 또는 이를 가지거나, 이를 갖는 것으로 미리 결정되거나, 또는 이를 갖는 것으로 의심되는 환자로 지칭된다.

[0336] 용어 "병원성 돌연변이", "병원성 변이체", "질환 유발 돌연변이", "질환 유발 변이체", "유해성 돌연변이" 또는 "소인성 돌연변이"는 소정의 질환 또는 장애에 대한 개체의 감수성 또는 소인을 증가시키는 유전적 변경 또는 돌연변이를 지칭한다. 일부 실시형태에서, 병원성 돌연변이는 유전자에 의해 암호화된 단백질에서 적어도 하나의 병원성 아미노산으로 치환된 적어도 하나의 야생형 아미노산을 포함한다.

[0337] 용어 "약제학적으로 허용 가능한 담체"는 신체의 한 부위(예를 들어, 전달 부위)에서 또 다른 부위(예를 들어, 신체의 기관, 조직 또는 일부)로 화합물을 운반 또는 수송하는데 관여하는 액체 또는 고형 충전제, 희석제, 부형제, 제조 보조제(예를 들어, 윤활제, 활석 마그네슘, 칼슘 또는 아연 스테아레이트 또는 스테아르산) 또는 용매 캡슐화 물질과 같은 약제학적으로 허용 가능한 물질, 조성물 또는 비히클을 의미한다. 약제학적으로 허용 가

능한 담체는 제형의 다른 성분과 상용성이며 대상체의 조직에 해롭지 않다는 의미(예를 들어, 생리학적으로 양립 가능함, 멸균성, 생리학적 pH 등)에서 "허용 가능"하다. "부형제", "담체", "약제학적으로 허용 가능한 담체", "비히클" 등과 같은 용어는 본 명세서에서 상호교환적으로 사용된다.

[0338] 용어 "약제학적 조성물"은 약제학적 용도를 위해 제형화된 조성물을 의미한다. 일부 실시형태에서, 약제학적 조성물은 약제학적으로 허용 가능한 담체를 더 포함한다. 일부 실시형태에서, 약제학적 조성물은 (예를 들어, 특이적 전달, 반감기의 증가 또는 다른 치료적 화합물을 위한) 추가적인 작용제를 포함한다.

[0339] "예정 세포사 1(PDCD1 또는 PD-1) 폴리펩타이드"는 NCBI 등록 번호 AJS10360.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는 단백질을 의미한다. PD-1 단백질은 면역 반응 동안 그리고 내성 조건에서 T 세포 기능 조절에 관여하는 것으로 생각된다. 예시적인 B2M 폴리펩타이드 서열이 아래에 제공된다.

[0340] >AJS10360.1 예정 세포사 1 단백질 [호모 사피엔스]

MQIPQAPWPVVWAVLQLGWRPGWFLDSPDRPWNPPTFSPALLVVTEGDNATFTCSFSNTSESFV
LNWYRMSPSNQTDKLAAFPEDRSQPGQDCRFRVTQLPNGRDFHMSVVRARRNDSGTYLCGAISL
APKAQIKESLRAELRVTERRAEVPTAHPSPSPRPAGQFQTLVVGVVGGLLGSLVLLVWVLAVIC
SRAARGTIGARRTGQPLKEDPSAVPVFSVDYGELDFQWREKTPEPPVPCVPEQTEYATIVFPSG

[0341] MGTSSPARRGSADGPRSAQPLRPEDGHCSWPL

[0342] "예정 세포사 1(PDCD1 또는 PD-1) 폴리뉴클레오타이드"는 PD-1 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. PDCD1 유전자는 항원-특이적 방식으로 T-세포 효과기 기능을 저해하는 저해성 세포 표면 수용체를 암호화한다. 예시적인 PDCD1 핵산 서열이 아래에 제공된다.

[0343] >AY238517.1 호모 사피엔스 예정 세포사 1(PDCD1) mRNA, 완전 cds

ATGCAGATCCCACAGGCGCCCTGGCCAGTCGTCTGGGCGGTGCTACAACTGGGCTGGCGGCCAG
GATGGTTCTTAGACTCCCCAGACAGGCCCTGGAACCCCCCCACCTTCTCCCCAGCCCTGCTCGT
GGTGACCGAAGGGGACAACGCCACCTTCACCTGCAGCTTCTCCAACACATCGGAGAGCTTCGTG
CTAAACTGGTACCGCATGAGCCCCAGCAACCAGACGGACAAGCTGGCCGCCTTCCCCGAGGACC
GCAGCCAGCCCGGCCAGGACTGCCGCTTCCGTGTCACACAACTGCCCAACGGGCGTGACTTCCA
CATGAGCGTGGTCAGGGCCCGGCGCAATGACAGCGGCACCTACCTCTGTGGGGCCATCTCCCTG
GCCCCCAAGGCGCAGATCAAAGAGAGCCTGCGGGCAGAGCTCAGGGTGACAGAGAGAAGGGCAG
AAGTGCCCACAGCCCACCCCAGCCCCTCACCCAGGCCAGCCGGCCAGTTCCAAACCCTGGTGGT
TGGTGTCGTGGGCGGCCTGCTGGGCAGCCTGGTGCTGCTAGTCTGGGTCCTGGCCGTCATCTGC
TCCCGGGCCGCACGAGGGACAATAGGAGCCAGGCGCACCGGCCAGCCCCTGAAGGAGGACCCCT
CAGCCGTGCCTGTGTTCTCTGTGGACTATGGGGAGCTGGATTTCCAGTGGCGAGAGAAGACCCC
GGAGCCCCCCGTGCCCTGTGTCCCTGAGCAGACGGAGTATGCCACCATTGTCTTTCCTAGCGGA
ATGGGCACCTCATCCCCCGCCCGCAGGGGCTCAGCTGACGGCCCTCGGAGTGCCCAGCCACTGA

[0344] GGCCTGAGGATGGACACTGCTCTTGGCCCCTCTGA

[0345] 용어 "단백질", "펩타이드", "폴리펩타이드" 및 이들의 문법적 등가물은 본 명세서에서 상호교환적으로 사용되며, 펩타이드(아마이드) 결합에 의해 함께 연결된 아미노산 잔기의 중합체를 지칭한다. 용어는 임의의 크기, 구조 또는 기능의 단백질, 펩타이드 또는 폴리펩타이드를 지칭한다. 전형적으로, 단백질, 펩타이드 또는 폴리펩타이드는 적어도 3개의 아미노산 길이일 수 있다. 단백질, 펩타이드 또는 폴리펩타이드는 개별 단백질 또는 단백질의 집합체를 지칭할 수 있다. 단백질, 펩타이드 또는 폴리펩타이드에 있는 하나 이상의 아미노산은, 예를 들어, 탄수화물기, 하이드록실기, 포스페이트기, 파르네실기, 아이소파르네실기, 지방산기, 접합을 위한 링커, 작용화 또는 다른 변형 등과 같은 화학 독립체의 추가에 의해 변형될 수 있다. 단백질, 펩타이드 또는 폴리펩타이드는 또한 단일 분자일 수 있거나 또는 다중-분자 복합체일 수 있다. 단백질, 펩타이드 또는 폴리펩타이드는 단지 자연 발생적 단백질 또는 펩타이드의 단편일 수 있다. 단백질, 펩타이드 또는 폴리펩타이드는 자연 발생, 재조합 또는 합성 또는 이들의 임의의 조합일 수 있다. 본 명세서에서 사용되는 용어 "융합 단백질"은 적어도 2개의 상이한 단백질로부터의 단백질 도메인을 포함하는 하이브리드 폴리펩타이드를 지칭한다. 하나의 단백질은 융합 단백질의 아미노-말단(N-말단) 부분 또는 카복시-말단(C-말단) 단백질에 위치하여 각각 아미노-말단 융합 단

백질 또는 카복시-말단 융합 단백질을 형성할 수 있다. 단백질은 상이한 도메인, 예를 들어, 핵산 결합 도메인(예를 들어, 단백질의 표적 부위에 대한 단백질의 결합을 지시하는 Cas9의 gRNA 결합 도메인) 및 핵산 절단 도메인 또는 핵산 편집 단백질의 촉매 도메인을 포함할 수 있다. 일부 실시형태에서, 단백질은 단백질 부분, 예를 들어, 핵산 결합 도메인을 구성하는 아미노산 서열 및 유기 화합물, 예를 들어, 핵산 절단제로 작용할 수 있는 화합물을 포함한다. 일부 실시형태에서, 단백질은 핵산, 예를 들어, RNA 또는 DNA와 복합체를 형성하거나 또는 이와 회합되어 있다. 본 명세서에 제공된 임의의 단백질은 당업계에 공지된 임의의 방법에 의해 생성될 수 있다. 예를 들어, 본 명세서에 제공된 단백질은 펩타이드 링커를 포함하는 융합 단백질에 특히 적합한 재조합 단백질 발현 및 정제를 통해 생성될 수 있다. 재조합 단백질 발현 및 정제를 위한 방법은 잘 알려져 있으며, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 의해 기술된 것들을 포함한다.

[0346] 본 명세서에 개시된 폴리펩타이드 및 단백질(기능적 부분 및 이의 기능적 변이체를 포함)은 하나 이상의 자연-발생적 아미노산 대신 합성 아미노산을 포함할 수 있다. 이러한 합성 아미노산은 당업계에 공지되어 있으며, 예를 들어, 아미노사이클로헥세인 카복실산, 노르류신, α-아미노 n-데칸산, 호모세린, S-아세틸아미노메틸-시스테인, 트랜스-3- 및 트랜스-4-하이드록시프롤린, 4-아미노페닐알라닌, 4-나이트로페닐알라닌, 4-클로로페닐알라닌, 4-카복시페닐알라닌, β-페닐세린 β-하이드록시페닐알라닌, 페닐글리신, α-나프틸알라닌, 사이클로헥실알라닌, 사이클로헥실글리신, 인돌린-2-카복실산, 1,2,3,4-테트라하이드로아이소퀴놀린-3-카복실산, 아미노말론산, 아미노말론산 모노아마이드, N'-벤질-N'-메틸-라이신, N',N'-다이벤질-라이신, 6-하이드록시라이신, 오르니틴, α-아미노사이클로펜테인 카복실산, α-아미노사이클로헥세인 카복실산, α-아미노사이클로헵테인 카복실산, α-(2-아미노-2-노보네인)-카복실산, α,γ-다이아미노뷰티르산, α,β-다이아미노프로피온산, 호모페닐알라닌 및 α-tert-뷰틸글리신을 포함한다. 폴리펩타이드 및 단백질은 폴리펩타이드 작제물의 하나 이상의 아미노산의 번역 후 변형과 관련될 수 있다. 번역 후 변형의 비제한적인 예는 인산화, 아세틸화 및 포밀화를 포함하는 아실화, 글리코실화(N-연결 및 O-연결 포함), 아마이드화, 하이드록실화, 메틸화 및 에틸화를 포함하는 알킬화, 유비퀴틴화, 피롤리돈 카복실산의 추가, 이황화 브리지의 형성, 황산화, 미리스토일화, 팔미토일화, 아이소프레닐화, 파르네실화, 제라닐화, 글리피화, 리포일화 및 아이오딘화를 포함한다.

[0347] "프로모터"는 핵산의 전사를 지시하는 핵산 제어 서열의 어레이를 의미한다. 프로모터는 전사의 시작 부위 근처에 필요한 핵산 서열을 포함한다. 프로모터는 또한 선택적으로 원위 인핸서 또는 억제 서열 요소를 포함한다. "구성적 프로모터"는 지속적으로 활성이며 외부 신호 또는 분자에 의한 조절을 받지 않는 프로모터이다. 대조적으로, "유도성 프로모터"의 활성은 외부 신호 또는 분자(예를 들어, 전사 인자)에 의해 조절된다. 예로서, 프로모터는 CMV 프로모터일 수 있다.

[0348] 단백질 또는 핵산의 맥락에서 본 명세서에서 사용되는 용어 "재조합"은 자연에서 발생하지 않지만 인간 조작의 생성물인 단백질 또는 핵산을 지칭한다. 예를 들어, 일부 실시형태에서, 재조합 단백질 또는 핵산 분자는 임의의 자연 발생적 서열과 비교하여 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개 또는 적어도 7개의 돌연변이를 포함하는 아미노산 또는 뉴클레오타이드 서열을 포함한다.

[0349] "감소시키다"는 적어도 10%, 25%, 50%, 75% 또는 100%의 음의 변경을 의미한다.

[0350] "참조"는 표준 또는 제어 조건을 의미한다. 일 실시형태에서, 참조는 야생형 또는 건강한 세포이다. 다른 실시형태에서, 제한 없이, 참조는 테스트 조건에 적용되지 않거나 또는 플라시보 또는 생리 식염수, 배지, 완충액에 적용된 처리되지 않은 세포 및/또는 관심 폴리뉴클레오타이드를 보유하지 않는 대조군 벡터이다.

[0351] "참조 서열"은 서열 비교를 위한 기초로 사용되는 정의된 서열이다. 참조 서열은 지정된 서열의 서브세트 또는 전체; 예를 들어, 전장 cDNA 또는 유전자 서열 또는 완전 cDNA 또는 유전자 서열의 분절일 수 있다. 폴리펩타이드의 경우, 참조 폴리펩타이드 서열의 길이는 일반적으로 적어도 약 16개의 아미노산, 적어도 약 20개의 아미노산, 적어도 약 25개의 아미노산, 약 35개의 아미노산, 약 50개의 아미노산 또는 약 100개의 아미노산일 것이다. 핵산의 경우, 참조 핵산 서열의 길이는 일반적으로 적어도 약 50개의 뉴클레오타이드, 적어도 약 60개의 뉴클레오타이드, 적어도 약 75개의 뉴클레오타이드 및 약 100개의 뉴클레오타이드 또는 약 300개의 뉴클레오타이드, 또는 그 주위 또는 그 사이의 임의의 정수일 것이다. 일부 실시형태에서, 참조 서열은 관심 단백질의 야생형 서열이다. 다른 실시형태에서, 참조 서열은 야생형 단백질을 암호화하는 폴리뉴클레오타이드 서열이다.

[0352] 용어 "RNA-프로그램 가능한 뉴클레이스" 및 "RNA-가이드된 뉴클레이스"는 절단을 위한 표적이 아닌 하나 이상의 RNA(들)와 함께 사용된다(예를 들어, 결합하거나 또는 회합함). 일부 실시형태에서, RNA-프로그램 가능한 뉴클

레이스는 RNA와 복합체를 형성할 때 뉴클레이스:RNA 복합체로 지칭될 수 있다. 전형적으로, 결합된 RNA(들)는 가이드 RNA(gRNA)로 지칭된다. gRNA는 2개 이상의 RNA의 복합체로 또는 단일 RNA 분자로 존재할 수 있다. 단일 RNA 분자로 존재하는 gRNA는 단일-가이드 RNA(sgRNA)로 지칭될 수 있지만, "gRNA"는 단일 분자 또는 2개 이상의 분자의 복합체로 존재하는 가이드 RNA를 지칭하는데 상호교환적으로 사용된다. 전형적으로, 단일 RNA 종으로 존재하는 gRNA는 다음 2개 도메인을 포함한다: (1) 표적 핵산과 상동성을 공유하는(그리고, 예를 들어, 표적에 대한 Cas9 복합체의 결합을 지시함) 도메인; 및 (2) Cas9 단백질에 결합하는 도메인. 일부 실시형태에서, 도메인 (2)는 tracrRNA로 알려진 서열에 해당하며, 줄기-루프 구조를 포함한다. 예를 들어, 일부 실시형태에서, 도메인 (2)는 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Jinek *et al.*, Science 337:816-821(2012)]에 제공되는 바와 같은 tracrRNA와 동일하거나 또는 상동이다. gRNA의 다른 예(예를 들어, 도메인 2를 포함하는 것)는 각각 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 2013년 9월 6일자로 출원된 "Switchable Cas9 Nucleases and Uses Thereof"라는 명칭의 미국 가출원 특허 제61/874,682호 및 2013년 9월 6일자로 출원된 "Delivery System For Functional Nucleases"라는 명칭의 미국 가출원 특허 제61/874,746호에서 발견할 수 있다. 일부 실시형태에서, gRNA는 도메인(1) 및 (2) 중 2개 이상을 포함하며, "연장된 gRNA"로 지칭될 수 있다. 예를 들어, 연장된 gRNA는, 예를 들어, 본 명세서에 기재된 바와 같이 2개 이상의 Cas9 단백질에 결합하고, 2개 이상의 별개의 영역에서 표적 핵산에 결합할 것이다. gRNA는 상기 표적 부위에 대한 뉴클레이스/RNA 복합체의 결합을 매개하여 뉴클레이스:RNA 복합체의 서열 특이성을 제공하는 표적 부위를 보완하는 뉴클레오타이드 서열을 포함한다.

[0353] 일부 실시형태에서, RNA-프로그램 가능한 뉴클레이스는 (CRISPR-관련 시스템) Cas9 엔도뉴클레이스, 예를 들어, 스트렙토코커스 피오게네스로부터의 Cas9(Csn1)(예를 들어, 문헌["Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti J.J., *et al.*, Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., *et al.*, Nature 471:602-607(2011)] 참조)이다. RNA-프로그램 가능한 뉴클레이스(예를 들어, Cas9)는 DNA 절단 부위를 표적화하기 위해 RNA:DNA 혼성화를 사용하기 때문에, 이들 단백질은 원칙적으로 가이드 RNA에 의해 지정된 임의의 서열에 대해 표적화될 수 있다. 부위-특이적 절단을 위해(예를 들어, 게놈을 변형시키기 위해) Cas9와 같은 RNA-프로그램 가능한 뉴클레이스를 사용하는 방법은 당업계에 공지되어 있다(예를 들어, 문헌[Cong, L. *et al.*, Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali, P. *et al.*, RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang, W.Y. *et al.*, Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek, M. *et al.*, RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo, J.E. *et al.*, Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic acids research (2013); Jiang, W. *et al.* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013)] 참조; 각각의 전체 내용은 참조에 의해 본 명세서에 원용되어 있음).

[0354] "신호전달 도메인"은 T 세포 효과기 기능 신호(예를 들어, 활성화 신호)를 형질도입하고, T 세포가 특수 기능을 수행하도록 지시하는 T 세포에서 발현되는 단백질의 세포내 부분을 의미한다. T 세포 활성화는 T 세포의 표면의 T 세포 수용체에 대한 동족 항원의 결합 및 T 세포의 표면의 공동자극 분자에 대한 동족 리간드의 결합을 비롯한 다수의 인자에 의해 유도될 수 있다. T 세포 공동-자극 분자는 공동-자극 리간드와 특이적으로 결합하여 증식과 같지만 이로 제한되지 않는 T 세포에 의한 공동-자극 반응을 매개하는 T 세포 상의 동족 결합 파트너이다. 공동-자극 분자는 MHC 클래스 I 분자를 포함하지만, 이로 제한되지 않는다. 일부 실시형태에서, 공동-자극 도메인은 CD2 세포질 도메인이다. T 세포의 활성화는 T 세포 증식 및 분화와 같은 면역 반응으로 이어진다(예를 들어, 문헌[Smith-Garvin *et al.*, Annu. Rev. Immunol., 27:591-619, 2009] 참조). 예시적인 T 세포 신호전달 도메인은 당업계에 공지되어 있다. 비제한적인 예는 CD2, CD3ζ, CD8, CD28, CD27, CD154, GITR(TNFRSF18), CD134(OX40) 및 CD137(4-1BB) 신호전달 도메인을 포함한다.

[0355] "단일-사슬 항체" 또는 "scFv"는 유전적으로 융합된 단일 사슬 분자로서 적합한 폴리펩타이드 링커에 의해 연결된 하나 이상의 항체의 VH 및 VL 도메인을 포함하는 유전적으로 조작된 분자를 의미한다(예를 들어, 문헌[Bird *et al.*, Science, 242:423-426, 1988; Huston *et al.*, Proc. Natl. Acad. Sci., 85:5879-5883, 1988: Ahmad *et al.*, Clin. Dev. Immunol., 2012, doi:10.1155/2012/980250: Marbry, IDrugs, 13:543-549, 2010] 참조). 일부 실시형태에서, scFv에서 VH-도메인 및 VL-도메인의 분자내 배향은 VH-도메인-링커 도메인-VL-도메인이다. 일부 실시형태에서, scFv에서 VH-도메인 및 VL-도메인의 분자내 배향은 VL-도메인-링커 도메인-VH-도메인이다.

[0356] 용어 "단일 뉴클레오타이드 다형성(single nucleotide polymorphism: SNP)"은 게놈의 특정 위치에서 발생하는 단일 뉴클레오타이드의 변이이며, 여기서 각 변이는 집단 내에서 어느 정도 감지할 수 있는 정도로 존재한다(예를 들어, 1% 초과). 예를 들어, 인간 게놈의 특정 염기 위치에서, C 뉴클레오타이드는 대부분의 개체에서 나타날 수 있지만, 소수의 개체에서는 A가 그 위치를 차지한다. 이는 이 특정 위치에 SNP가 있음을 의미하며, 2개의 가능한 뉴클레오타이드 변이 C 또는 A를 이 위치에 대한 대립유전자라고 한다. SNP는 질환에 대한 감수성의 차이를 뒷받침한다. 질병의 중증도 및 신체가 치료에 반응하는 방식 또한 유전적 변이의 징후이다. SNP는 유전자의 코딩 영역, 유전자의 비-코딩 영역 또는 유전자간 영역(유전자 사이의 영역)에 속할 수 있다. 일부 실시형태에서, 코딩 서열 내의 SNP는 유전 코드의 축퇴성으로 인해 생성되는 단백질의 아미노산 서열을 반드시 변경하지는 않는다. 코딩 영역의 SNP에는 다음 2가지 유형이 있다: 동의적(synonymous) 및 비동의적(nonsynonymous) SNP. 동의적 SNP는 단백질 서열에 영향을 미치지 않는 반면, 비동의적 SNP는 단백질의 아미노산 서열을 변경한다. 비동의적 SNP에는 다음 2가지 유형이 있다: 미스센스 및 넌센스. 단백질-코딩 영역에 없는 SNP는 여전히 유전자 스플라이싱, 전사 인자 결합, 메신저 RNA 분해 또는 논코딩 RNA의 서열에 영향을 미칠 수 있다. 이러한 유형의 SNP에 의해 영향을 받는 유전자 발현은 eSNP(발현 SNP)로 지칭되며, 유전자로부터 상류 또는 하류에 있을 수 있다. 단일 뉴클레오타이드 변이체(SNV)는 빈도의 임의의 제한이 없는 단일 뉴클레오타이드의 변이이며, 체세포에서 발생할 수 있다. 체세포 단일 뉴클레오타이드 변이(예를 들어, 암과 관련된)는 또한 단일-뉴클레오타이드 변경이라고도 할 수 있다.

[0357] "특이적으로 결합하는"은 본 발명의 폴리펩타이드 및/또는 핵산 분자를 인식하고 이에 결합하지만 샘플, 예를 들어, 생물학적 샘플에서 다른 분자를 실질적으로 인식하고 이에 결합하지 않는 핵산 분자, 폴리펩타이드 또는 이의 복합체(예를 들어, 핵산 프로그램 가능 DNA 결합 단백질, 가이드 핵산 및 키메라 항원 수용체), 화합물 또는 분자를 의미한다. 예를 들어, 키메라 항원 수용체는 세포 표면에 발현된 특정 마커에 특이적으로 결합하지만, 세포 표면의 다른 폴리펩타이드, 탄수화물, 지질 또는 임의의 다른 화합물에는 결합하지 않는다.

[0358] 본 발명의 방법에 유용한 핵산 분자는 본 발명의 폴리펩타이드 또는 이의 단편을 암호화하는 임의의 핵산 분자를 포함한다. 이러한 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적인 동일성을 나타낼 것이다. 내인성 서열에 대해 "실질적인 동일성"을 갖는 폴리뉴클레오타이드는 전형적으로 이중-가닥 핵산 분자의 적어도 하나의 가닥과 혼성화할 수 있다. 본 발명의 방법에 유용한 핵산 분자는 본 발명의 폴리펩타이드 또는 이의 단편을 암호화하는 임의의 핵산 분자를 포함한다. 이러한 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적인 동일성을 나타낼 것이다. 내인성 서열에 대해 "실질적인 동일성"을 갖는 폴리뉴클레오타이드는 전형적으로 이중-가닥 핵산 분자의 적어도 하나의 가닥과 혼성화할 수 있다. "혼성화하다"는 다양한 엄격한 조건하에 상보적 폴리뉴클레오타이드 서열(예를 들어, 유전자 본 명세서에 기재된) 또는 이의 부분 사이에서 이중-가닥 분자를 형성하는 쌍을 의미한다(예를 들어, 문헌[Wahl, G. M. and S. L. Berger (1987) Methods Enzymol. 152:399; Kimmel, A. R. (1987) Methods Enzymol. 152:507] 참조).

[0359] 예를 들어, 엄격한 염 농도는 일반적으로 약 750mM NaCl 및 75mM 트라이소듐 시트레이트 미만, 바람직하게는 약 500mM NaCl 및 50mM 트라이소듐 시트레이트 미만, 보다 바람직하게는 약 250mM NaCl 및 25mM 트라이소듐 시트레이트 미만일 것이다. 낮은 엄격성 혼성화는 유기 용매, 예를 들어, 폼아마이드의 부재하에 얻어질 수 있는 반면, 높은 엄격성 혼성화는 적어도 약 35% 폼아마이드, 보다 바람직하게는 적어도 약 50% 폼아마이드의 존재하에 얻어질 수 있다. 온도 조건은 일반적으로 적어도 약 30℃, 보다 바람직하게는 적어도 약 37℃, 가장 바람직하게는 적어도 약 42℃의 온도를 포함할 것이다. 혼성화 시간, 세제, 예를 들어, 소듐 도데실 설페이트(sodium dodecyl sulfate: SDS)의 농도 및 담체 DNA의 포함 또는 배제와 같은 다양한 추가적인 매개변수는 당업자에게 잘 알려져 있다. 필요에 따라 이러한 다양한 조건을 조합함으로써 다양한 수준의 엄격성이 달성된다. 일 실시형태에서, 혼성화는 30℃에서 750mM NaCl, 75mM 트라이소듐 시트레이트 및 1% SDS에서 일어날 것이다. 또 다른 실시형태에서, 혼성화는 37℃에서 500mM NaCl, 50mM 트라이소듐 시트레이트, 1% SDS, 35% 폼아마이드 및 100 ㎍/㎖ 변성 연어 정자 DNA(ssDNA)에서 일어날 것이다. 또 다른 실시형태에서, 혼성화는 42℃에서 250mM NaCl, 25mM 트라이소듐 시트레이트, 1% SDS, 50% 폼아마이드 및 200 ㎍/㎖ ssDNA에서 일어날 것이다. 이러한 조건에 대한 유용한 변이는 당업자에게 명백할 것이다.

[0360] 대부분의 적용의 경우, 혼성화에 뒤따르는 세척 단계는 또한 엄격성이 다양할 것이다. 세척 엄격성 조건은 염 농도 및 온도에 의해 정의될 수 있다. 위와 같이, 세척 엄격성은 염 농도를 감소시키거나 또는 온도를 증가시킴으로써 증가될 수 있다. 예를 들어, 세척 단계를 위한 엄격한 염 농도는 바람직하게는 약 30mM NaCl 및 3mM 트라이소듐 시트레이트 미만, 가장 바람직하게는 약 15mM NaCl 및 1.5mM 트라이소듐 시트레이트 미만일 것이다. 세척 단계를 위한 엄격한 온도 조건은 일반적으로 적어도 약 25℃, 보다 바람직하게는 적어도 약 42℃, 훨씬 더

바람직하게는 적어도 약 68℃의 온도를 포함할 것이다. 실시형태에서, 세척 단계는 25℃에서 30mM NaCl, 3mM 트라이소듐 시트레이트 및 0.1% SDS에서 일어날 것이다. 보다 바람직한 실시형태에서, 세척 단계는 42℃에서 15mM NaCl, 1.5mM 트라이소듐 시트레이트 및 0.1% SDS에서 일어날 것이다. 보다 바람직한 실시형태에서, 세척 단계는 68℃에서 15mM NaCl, 1.5mM 트라이소듐 시트레이트 및 0.1% SDS에서 일어날 것이다. 이러한 조건에 대한 추가적인 변이는 당업자에게 명백할 것이다. 혼성화 기법은 당업자에게 잘 알려져 있으며, 예를 들어, 문헌[Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel *et al.* (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); 및 Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York]에 기술되어 있다.

[0361] "스플릿"은 2개 이상의 단편으로 분할되는 것을 의미한다.

[0362] "스플릿 Cas9 단백질" 또는 "스플릿 Cas9"는 2개의 개별 뉴클레오타이드 서열에 의해 암호화되는 N-말단 단편 및 C-말단 단편으로 제공되는 Cas9 단백질을 지칭한다. Cas9 단백질의 N-말단 부분 및 C-말단 부분에 상응하는 폴리펩타이드는 "재구성된" Cas9 단백질을 형성하기 위해 스플라이싱될 수 있다. 특정 실시형태에서, Cas9 단백질은, 예를 들어, 각각 참조에 의해 본 명세서에 원용되어 있는 문헌[Nishimasu *et al.*, Cell, Volume 156, Issue 5, pp. 935-949, 2014]에 기술된 바와 같거나 또는 문헌[Jiang *et al.* (2016) Science 351: 867-871. PDB file: 5F9R]에 기술된 바와 같은 단백질의 무질서 영역 내에서 2개의 단편으로 분할된다. 일부 실시형태에서, 단백질은 약 아미노산 A292-G364, F445-K483 또는 E565-T637 사이의 SpCas9의 영역 내의 임의의 C, T, A 또는 S에서 또는 임의의 다른 Cas9, Cas9 변이체(예를 들어, nCas9, dCas9) 또는 다른 napDNAbp의 상응하는 위치에서 2개의 단편으로 분할된다. 일부 실시형태에서, 단백질은 SpCas9 T310, T313, A456, S469 또는 C574에서 2개의 단편으로 분할된다. 일부 실시형태에서, 단백질을 2개 단편으로 분할하는 과정은 단백질 "스플리팅"으로 지칭된다.

[0363] 다른 실시형태에서, Cas9 단백질의 N-말단 부분은 S. 피오게네스 Cas9 야생형(SpCas9)(NCBI 참조 서열: NC_002737.2, Uniprot 참조 서열: Q99ZW2)의 1번 내지 573번 또는 1번 내지 637번 아미노산을 포함하고, Cas9 단백질의 C-말단 부분은 SpCas9 야생형의 574번 내지 1368번 또는 638번 내지 1368번 아미노산의 일부 또는 이의 상응하는 위치를 포함한다.

[0364] 스플릿 Cas9의 C-말단 부분은 완전 Cas9 단백질을 형성하기 위해 스플릿 Cas9의 N-말단 부분과 연결될 수 있다. 일부 실시형태에서, Cas9 단백질의 C-말단 부분은 Cas9 단백질의 N-말단 부분이 끝나는 곳으로부터 시작된다. 이와 같이, 일부 실시형태에서, 스플릿 Cas9의 C-말단 부분은 spCas9의 (551번 내지 651번) 내지 1368번 아미노산의 일부를 포함한다. "(551번 내지 651번) 내지 1368번"은 아미노산 551번 내지 651번(포함) 사이의 아미노산에서 시작하여 1368번 아미노산에서 끝나는 것을 의미한다. 예를 들어, 스플릿 Cas9의 C-말단 부분은 spCas9의 551번 내지 1368번, 552번 내지 1368번, 553번 내지 1368번, 554번 내지 1368번, 555번 내지 1368번, 556번 내지 1368번, 557번 내지 1368번, 558번 내지 1368번, 559번 내지 1368번, 560번 내지 1368번, 561번 내지 1368번, 562번 내지 1368번, 563번 내지 1368번, 564번 내지 1368번, 565번 내지 1368번, 566번 내지 1368번, 567번 내지 1368번, 568번 내지 1368번, 569번 내지 1368번, 570번 내지 1368번, 571번 내지 1368번, 572번 내지 1368번, 573번 내지 1368번, 574번 내지 1368번, 575번 내지 1368번, 576번 내지 1368번, 577번 내지 1368번, 578번 내지 1368번, 579번 내지 1368번, 580번 내지 1368번, 581번 내지 1368번, 582번 내지 1368번, 583번 내지 1368번, 584번 내지 1368번, 585번 내지 1368번, 586번 내지 1368번, 587번 내지 1368번, 588번 내지 1368번, 589번 내지 1368번, 590번 내지 1368번, 591번 내지 1368번, 592번 내지 1368번, 593번 내지 1368번, 594번 내지 1368번, 595번 내지 1368번, 596번 내지 1368번, 597번 내지 1368번, 598번 내지 1368번, 599번 내지 1368번, 600번 내지 1368번, 601번 내지 1368번, 602번 내지 1368번, 603번 내지 1368번, 604번 내지 1368번, 605번 내지 1368번, 606번 내지 1368번, 607번 내지 1368번, 608번 내지 1368번, 609번 내지 1368번, 610번 내지 1368번, 611번 내지 1368번, 612번 내지 1368번, 613번 내지 1368번, 614번 내지 1368번, 615번 내지 1368번, 616번 내지 1368번, 617번 내지 1368번, 618번 내지 1368번, 619번 내지 1368번, 620번 내지 1368번, 621번 내지 1368번, 622번 내지 1368번, 623번 내지 1368번, 624번 내지 1368번, 625번 내지 1368번, 626번 내지 1368번, 627번 내지 1368번, 628번 내지 1368번, 629번 내지 1368번, 630번 내지 1368번, 631번 내지 1368번, 632번 내지 1368번, 633번 내지 1368번, 634번 내지 1368번, 635번 내지 1368번, 636번 내지 1368번, 637번 내지 1368번, 638번 내지 1368번, 639번 내지 1368번, 640번 내지 1368번, 641번 내지 1368번, 642번 내지 1368번, 643번 내지 1368번, 644번 내지 1368번, 645번 내지 1368번, 646번 내지 1368번, 647번 내지 1368번, 648

번 내지 1368번, 649번 내지 1368번, 650번 내지 1368번 또는 651번 내지 1368번 아미노산 중 어느 하나의 일부를 포함할 수 있다. 일부 실시형태에서, 스플릿 Cas9의 C-말단 부분 단백질은 아미노산 SpCas9의 574번 내지 1368번 또는 638번 내지 1368번 아미노산의 일부를 포함한다.

[0365] "대상체"는 인간 또는 비-인간 포유동물, 예컨대, 소, 말, 개, 양 또는 고양이를 포함하지만 이들로 제한되지 않는 포유동물을 의미한다. 대상체는 제한 없이 소, 염소, 닭, 말, 돼지, 토끼 및 양을 포함하여 노동력을 제공하며 식품과 같은 상품을 제공하기 위해 사육된 동물인 가축을 포함한다.

[0366] "실질적으로 동일한"은 참조 아미노산 서열(예를 들어, 본 명세서에 기재된 아미노산 서열 중 임의의 하나) 또는 핵산 서열(예를 들어, 본 명세서에 기재된 핵산 서열 중 임의의 하나)에 대해 적어도 50%의 동일성을 나타내는 폴리펩타이드 또는 핵산 분자를 의미한다. 일 실시형태에서, 이러한 서열은 비교를 위해 사용된 서열과 아미노산 수준 또는 핵산에서 적어도 60%, 80% 또는 85%, 90%, 95% 또는 심지어 99% 동일하다.

[0367] 서열 동일성은 전형적으로 서열 분석 소프트웨어(예를 들어, 위스콘신 53705, 매디슨, 1710 유니버시티 애비뉴, 위스콘신 대학 바이오테크놀로지 센터 유전학 컴퓨터 그룹의 서열 분석 소프트웨어 패키지, 아세포, BETFIT, COBALT, EMBOSS 바늘, GAP 또는 PILEUP/PRETTYBOX 프로그램)를 사용하여 측정된다. 이러한 소프트웨어는 다양한 치환, 결실 및/또는 기타 변형에 상동성의 정도를 할당함으로써 동일하거나 또는 유사한 서열을 매칭한다. 보존적 치환은 전형적으로 다음 그룹 내의 치환을 포함한다: 글리신, 알라닌; 발린, 아이소류신, 류신; 아스파르트산, 글루탐산, 아스파라긴, 글루타민; 세린, 트레오닌; 라이신, 아르기닌; 및 페닐알라닌, 타이로신. 동일성의 정도를 결정하기 위한 예시적인 접근법에서, 밀접하게 관련된 서열을 나타내는 $e^{-3}$과 $e^{-100}$ 사이의 확률 점수와 함께 아세포 프로그램이 사용될 수 있다.

[0368] COBALT는, 예를 들어, 다음 매개변수와 함께 사용된다:

[0369] a) 정렬 매개변수: 갭 페널티-11, -1 및 말단-갭 페널티-5, -1,

[0370] b) CDD 매개변수: RPS 아세포를 사용하고; 아세포 E-값은 0.003이며; 보존된 열을 찾아 다시 계산함, 그리고

[0371] c) 쿼리 클러스터링 매개변수: 쿼리 클러스터를 사용하고; 단어 크기 4; 최대 클러스터 거리 0.8; 알파벳 일반.

[0372] EMBOSS 바늘은, 예를 들어, 다음 매개변수와 함께 사용된다:

[0373] a) 행렬: BLOSUM62;

[0374] b) 갭 오픈: 10;

[0375] c) 갭 연장: 0.5;

[0376] d) 입출력 형식: 쌍;

[0377] e) 엔드 갭 페널티: 거짓(false);

[0378] f) 엔드 갭 오픈: 10; 및

[0379] g) 엔드 갭 연장: 0.5.

[0380] 용어 "표적 부위"는 핵염기 편집기에 의해 변형된 핵산 분자 내의 서열을 지칭한다. 일 실시형태에서, 표적 부위는 데아미네이스 또는 데아미네이스(예를 들어, 사이티딘 또는 아데닌 데아미네이스)를 포함하는 융합 단백질에 의해 탈아미노화된다.

[0381] "T 세포 수용체 알파 불변(TRAC) 폴리펩타이드"는 NCBI 등록 번호 P01848.2 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0382] >sp|P01848.2|TRAC_인간 RecName: Full=T 세포 수용체 알파 불변

[0383]
```
IQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNVSQSKDSDVYITDKTVLDMRSMDFKSNSAVAW
SNKSDFACANAFNNSIIPEDTFFPSPESSCDVKLVEKSFETDTNLNFQNLSVIGFRILLLKVAG
FNLLMTLRLWSS
```

[0384] "T 세포 수용체 알파 불변(TRAC) 폴리뉴클레오타이드"는 TRAC 폴리펩타이드를 암호화하는 핵산을 의미한다. 예

시적인 TRAC 핵산 서열은 아래에 제공된다.

[0385] UCSC 인간 게놈 데이터베이스, 유전자 ENSG00000277734.8 인간 T-세포 수용체 알파 사슬(TCR-알파)

catgctaatcctccggcaaacctctgtttcctcctcaaaaggcaggaggtcggaaagaataaac
aatgagagtcacattaaaaacacaaaatcctacggaaatactgaagaatgagtctcagcactaa
ggaaaagcctccagcagctcctgctttctgagggtgaaggatagacgctgtggctctgcatgac
tcactagcactctatcacggccatattctggcagggtcagtggctccaactaacatttgtttgg
tactttacagtttattaaatagatgtttatatggagaagctctcatttctttctcagaagagcc
tggctaggaaggtggatgaggcaccatattcattttgcaggtgaaattcctgagatgtaaggag
ctgctgtgacttgctcaaggccttatatcgagtaaacggtagtgctggggcttagacgcaggtg
ttctgatttatagttcaaaacctctatcaatgagagagcaatctcctggtaatgtgatagattt
cccaacttaatgccaacataccataaacctcccattctgctaatgcccagcctaagttggggag
accactccagattccaagatgtacagtttgctttgctgggccttttttcccatgcctgcctttac
tctgccagagttatattgctggggttttgaagaagatcctattaaataaaagaataagcagtat
tattaagtagccctgcatttcaggtttccttgagtggcaggccaggcctggccgtgaacgttca
ctgaaatcatggcctcttggccaagattgatagcttgtgcctgtccctgagtcccagtccatca
cgagcagctggtttctaagatgctatttcccgtataaagcatgagaccgtgacttgccagcccc
acagagccccgcccttgtccatcactggcatctggactccagcctgggttggggcaaagaggga
aatgagatcatgtcctaaccctgatcctcttgtcccacagATATCCAGAACCCTGACCCTGCCG
[0386] TGTACCAGCTGAGAGACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTC

TCAAACAAATGTGTCACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGAC
ATGAGGTCTATGGACTTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCAT
GTGCAAACGCCTTCAACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGgtaaggg
cagctttggtgccttcgcaggctgtttccttgcttcaggaatggccaggttctgcccagagctc
tggtcaatgatgtctaaaactcctctgattggtggtctcggccttatccattgccaccaaaacc
ctcttttactaagaaacagtgagccttgttctggcagtccagagaatgacacgggaaaaaagc
agatgaagagaaggtggcaggagagggcacgtggcccagcctcagtctctccaactgagttcct
gcctgcctgcctttgctcagactgtttgccccttactgctcttctaggcctcattctaagcccc
ttctccaagttgcctctccttatttctccctgtctgccaaaaaatctttcccagctcactaagt
cagtctcacgcagtcactcattaacccaccaatcactgattgtgccggcacatgaatgcaccag
gtgttgaagtggaggaattaaaaagtcagatgaggggtgtgcccagaggaagcaccattctagt
tgggggagcccatctgtcagctgggaaaagtccaaataacttcagattggaatgtgttttaact
cagggttgagaaaacagctaccttcaggacaaaagtcagggaagggctctctgaagaaatgcta
cttgaagataccagccctaccaagggcagggagaggaccctatagaggcctgggacaggagctc
aatgagaaaggagaagagcagcaggcatgagttgaatgaaggaggcagggccgggtcacagggc
cttctaggccatgagagggtagacagtattctaaggacgccagaaagctgttgatcggcttcaa
gcaggggagggacacctaatttgctttttctttttttttttttttttttttttttttttgagat
ggagttttgctcttgttgcccaggctggagtgcaatggtgcatcttggctcactgcaacctccg
cctcccaggttcaagtgattctcctgcctcagcctcccgagtagctgagattacaggcacccgc
caccatgcctggctaatttttttgtatttttagtagagacagggtttcactatgttggccaggct
ggtctcgaactcctgacctcaggtgatccacccgcttcagcctcccaaagtgctgggattacag
gcgtgagccaccacacccggcctgcttttcttaaagatcaatctgagtgctgtacggagagtgg
gttgtaagccaagagtagaagcagaaagggagcagttgcagcagagagatgatggaggcctggg
cagggtggtggcagggaggtaaccaacaccattcaggtttcaaaggtagaaccatgcagggatg
agaaagcaaagaggggatcaaggaaggcagctggattttggcctgagcagctgagtcaatgata
gtgccgtttactaagaagaaaccaaggaaaaaatttggggtgcagggatcaaaacttttggaa
catatgaaagtacgtgtttatactctttatggcccttgtcactatgtatgcctcgctgcctcca
ttggactctagaatgaagccaggcaagagcagggtctatgtgtgatggcacatgtggccagggt
catgcaacatgtactttgtacaaacagtgtatattgagtaaatagaaatggtgtccaggagccg
aggtatcggtcctgccagggccaggggctctccctagcaggtgctcatatgctgtaagttccct
ccagatctctccacaaggaggcatggaaaggctgtagttgttcacctgcccaagaactaggagg
tctggggtgggagagtcagcctgctctggatgctgaaagaatgtctgtttttccttttagAAAG
TTCCTGTGATGTCAAGCTGGTCGAGAAAAGCTTTGAAACAGgtaagacaggggtctagcctggg
tttgcacaggattgcggaagtgatgaacccgcaataaccctgcctggatgagggagtgggaaga
aattagtagatgtgggaatgaatgatgaggaatggaaacagcggttcaagacctgcccagagct

[0387]

gggtggggtctctcctgaatccctctcaccatctctgactttccattctaagcactttgaggat
gagtttctagcttcaatagaccaaggactctctcctaggcctctgtattcctttcaacagctcc
actgtcaagagagccagagagagcttctgggtggcccagctgtgaaatttctgagtcccttagg
gatagccctaaacgaaccagatcatcctgaggacagccaagaggttttgccttctttcaagaca
agcaacagtactcacataggctgtgggcaatggtcctgtctctcaagaatcccctgccactcct
cacacccaccctgggcccatattcatttccatttgagttgttcttattgagtcatccttcctgt
ggtagcggaactcactaaggggcccatctggacccgaggtattgtgatgataaattctgagcac
ctaccccatccccagaagggctcagaaataaaataagagccaagtctagtcggtgtttcctgtc
ttgaaacacaatactgttggccctggaagaatgcacagaatctgtttgtaaggggatatgcaca
gaagctgcaagggacaggaggtgcaggagctgcaggcctcccccacccagcctgctctgccttg
gggaaaaccgtgggtgtgtcctgcaggccatgcaggcctgggacatgcaagcccataaccgctg
tggcctcttggttttacagATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTCCGAA
TCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCTGAGg
tgaggggccttgaagctgggagtggggtttagggacgcgggtctctgggtgcatcctaagctct
gagagcaaacctccctgcagggtcttgcttttaagtccaaagcctgagcccaccaaactctcct
acttcttcctgttacaaattcctcttgtgcaataataatggcctgaaacgctgtaaaatatcct
catttcagccgcctcagttgcacttctcccctatgaggtaggaagaacagttgtttagaaacga
agaaactgaggccccacagctaatgagtggaggaagagagacacttgtgtacaccacatgcctt
gtgttgtacttctctcaccgtgtaacctcctcatgtcctctctccccagtacggctctcttagc
tcagtagaaagaagacattacactcatattacaccccaatcctggctagagtctccgcaccctc
ctcccccagggtccccagtcgtcttgctgacaactgcatcctgttccatcaccatcaaaaaaaa
actccaggctgggtgcgggggctcacacctgtaatcccagcactttgggaggcagaggcaggag
gagcacaggagctggagaccagcctgggcaacacagggagaccccgcctctacaaaaagtgaaa
aaattaaccaggtgtggtgctgcacacctgtagtcccagctacttaagaggctgagatgggagg
atcgcttgagccctggaatgttgaggctacaatgagctgtgattgcgtcactgcactccagcct
ggaagacaaagcaagatcctgtctcaaataataaaaaaaataagaactccagggtacatttgct
cctagaactctaccacatagccccaaacagagccatcaccatcacatccctaacagtcctgggt
cttcctcagtgtccagcctgacttctgttcttcctcattccagATCTGCAAGATTGTAAGACAG
CCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTCTCCCTCTCCAAACAGAGGGAACTC
TCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTGCCCCCCCCGGCAATGCCACCAACTG
GATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTGCCAAACACTGCTGCCACCCCCTCT
GTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGCAGAGGCAAGGCTGCTGCAGCCTCC
CCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCCTCCGCCATCCCACAGATGATGGAT
CTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGCAGAATGTTGTGAGGGGTTTATTTTTTTTTAA
TAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAGACGTGGGGGGAAATTATCTCATTA
TCGAGGCCCTGCTATGCTGTGTATCTGGGCGTGTTGTATGTCCTGCTGCCGATGCCTTCATTAA
AATGATTTGGAAGAGCAGA

[0388]

[0389] 위의 소문자로 된 뉴클레오타이드는 비번역 영역 또는 인트론이고, 대문자로 된 뉴클레오타이드는 엑손이다.

[0390] >X02592.1 T-세포 수용체 알파 사슬(TCR-알파)에 대한 인간 mRNA

TTTTGAAACCCTTCAAAGGCAGAGACTTGTCCAGCCTAACCTGCCTGCTGCTCCTAGCTCCTGA
GGCTCAGGGCCCTTGGCTTCTGTCCGCTCTGCTCAGGGCCCTCCAGCGTGGCCACTGCTCAGCC
ATGCTCCTGCTGCTCGTCCCAGTGCTCGAGGTGATTTTTACCCTGGGAGGAACCAGAGCCCAGT
CGGTGACCCAGCTTGGCAGCCACGTCTCTGTCTCTGAAGGAGCCCTGGTTCTGCTGAGGTGCAA
CTACTCATCGTCTGTTCCACCATATCTCTTCTGGTATGTGCAATACCCCAACCAAGGACTCCAG
CTTCTCCTGAAGTACACATCAGCGGCCACCCTGGTTAAAGGCATCAACGGTTTTGAGGCTGAAT
TTAAGAAGAGTGAAACCTCCTTCCACCTGACGAAACCCTCAGCCCATATGAGCGACGCGGCTGA
GTACTTCTGTGCTGTGAGTGATCTCGAACCGAACAGCAGTGCTTCCAAGATAATCTTTGGATCA
GGGACCAGACTCAGCATCCGGCCAAATATCCAGAACCCTGACCCTGCCGTGTACCAGCTGAGAG
ACTCTAAATCCAGTGACAAGTCTGTCTGCCTATTCACCGATTTTGATTCTCAAACAAATGTGTC
ACAAAGTAAGGATTCTGATGTGTATATCACAGACAAAACTGTGCTAGACATGAGGTCTATGGAC
TTCAAGAGCAACAGTGCTGTGGCCTGGAGCAACAAATCTGACTTTGCATGTGCAAACGCCTTCA
ACAACAGCATTATTCCAGAAGACACCTTCTTCCCCAGCCCAGAAAGTTCCTGTGATGTCAAGCT
GGTCGAGAAAAGCTTTGAAACAGATACGAACCTAAACTTTCAAAACCTGTCAGTGATTGGGTTC
CGAATCCTCCTCCTGAAAGTGGCCGGGTTTAATCTGCTCATGACGCTGCGGCTGTGGTCCAGCT
GAGATCTGCAAGATTGTAAGACAGCCTGTGCTCCCTCGCTCCTTCCTCTGCATTGCCCCTCTTC
TCCCTCTCCAAACAGAGGGAACTCTCCTACCCCCAAGGAGGTGAAAGCTGCTACCACCTCTGTG
CCCCCCCGGTAATGCCACCAACTGGATCCTACCCGAATTTATGATTAAGATTGCTGAAGAGCTG
CCAAACACTGCTGCCACCCCCTCTGTTCCCTTATTGCTGCTTGTCACTGCCTGACATTCACGGC
AGAGGCAAGGCTGCTGCAGCCTCCCCTGGCTGTGCACATTCCCTCCTGCTCCCCAGAGACTGCC
TCCGCCATCCCACAGATGATGGATCTTCAGTGGGTTCTCTTGGGCTCTAGGTCCTGGAGAATGT
TGTGAGGGGTTTATTTTTTTTTAATAGTGTTCATAAAGAAATACATAGTATTCTTCTTCTCAAG
ACGTGGGGGGAAATTATCTCATTATCGAGGCCCTGCTATGCTGTGTGTCTGGGCGTGTTGTATG
[0391] TCCTGCTGCCGATGCCTTCATTAAAATGATTTGGAA

[0392] "T 세포 수용체 베타 불변 1 폴리펩타이드(TRBC1)"는 NCBI 등록 번호 P01850 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0393] >sp|P01850|TRBC1_인간 T 세포 수용체 베타 불변 1 OS=호모 사피엔스 OX=9606 GN=TRBC1 PE=1 SV=4

DLNKVFPPEVAVFEPSEAEISHTQKATLVCLATGFFPDHVELSWWVNGKEVHSGVSTDPQPLKE
QPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAKPVTQIVSAEAWGRA
[0394] DCGFTSVSYQQGVLSATILYEILLGKATLYAVLVSALVLMAMVKRKDF

[0395] "T 세포 수용체 베타 불변 1 폴리뉴클레오타이드(TRBC1)"는 TRBC1 폴리펩타이드를 암호화하는 핵산을 의미한다. 예시적인 TRBC1 핵산 서열이 아래에 제공된다.

[0396] >X00437.1

CTGGTCTAGAATATTCCACATCTGCTCTCACTCTGCCATGGACTCCTGGACCTTCTGCTGTGTG
TCCCTTTGCATCCTGGTAGCGAAGCATACAGATGCTGGAGTTATCCAGTCACCCCGCCATGAGG
TGACAGAGATGGGACAAGAAGTGACTCTGAGATGTAAACCAATTTCAGGCCACAACTCCCTTTT
CTGGTACAGACAGACCATGATGCGGGGACTGGAGTTGCTCATTTACTTTAACAACAACGTTCCG
ATAGATGATTCAGGGATGCCCGAGGATCGATTCTCAGCTAAGATGCCTAATGCATCATTCTCCA
CTCTGAAGATCCAGCCCTCAGAACCCAGGGACTCAGCTGTGTACTTCTGTGCCAGCAGTTTCTC
GACCTGTTCGGCTAACTATGGCTACACCTTCGGTTCGGGGACCAGGTTAACCGTTGTAGAGGAC
CTGAACAAGGTGTTCCCACCCGAGGTCGCTGTGTTTGAGCCATCAGAAGCAGAGATCTCCCACA
CCCAAAAGGCCACACTGGTGTGCCTGGCCACAGGCTTCTTCCCCGACCACGTGGAGCTGAGCTG
GTGGGTGAATGGGAAGGAGGTGCACAGTGGGGTCAGCACAGACCCGCAGCCCCTCAAGGAGCAG
CCCGCCCTCAATGACTCCAGATACTGCCTGAGCAGCCGCCTGAGGGTCTCGGCCACCTTCTGGC
AGAACCCCCGCAACCACTTCCGCTGTCAAGTCCAGTTCTACGGGCTCTCGGAGAATGACGAGTG
GACCCAGGATAGGGCCAAACCCGTCACCCAGATCGTCAGCGCCGAGGCCTGGGGTAGAGCAGAC
TGTGGCTTTACCTCGGTGTCCTACCAGCAAGGGGTCCTGTCTGCCACCATCCTCTATGAGATCC
TGCTAGGGAAGGCCACCCTGTATGCTGTGCTGGTCAGCGCCCTTGTGTTGATGGCCATGGTCAA
GAGAAAGGATTTCTGAAGGCAGCCCTGGAAGTGGAGTTAGGAGCTTCTAACCCGTCATGGTTCA
ATACACATTCTTCTTTTGCCAGCGCTTCTGAAGAGCTGCTCTCACCTCTCTGCATCCCAATAGA
TATCCCCCTATGTGCATGCACACCTGCACACTCACGGCTGAAATCTCCCTAACCCAGGGGGAC

[0397]

[0398] "T 세포 수용체 베타 불변 2 폴리펩타이드(TRBC2)"는 NCBI 등록 번호 A0A5B9 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0399] >sp|A0A5B9|TRBC2_인간 T 세포 수용체 베타 불변 2 OS=호모 사피엔스 OX=9606 GN=TRBC2 PE=1 SV=2

DLKNVFPPKVAVFEPSEAEISHTQKATLVCLATGFYPDHVELSWWVNGKEVHSGVSTDPQPLKE
QPALNDSRYCLSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAKPVTQIVSAEAWGRA
DCGFTSESYQQGVLSATILYEILLGKATLYAVLVSALVLMAMVKRKDSRG

[0400]

[0401] "T 세포 수용체 베타 불변 2 폴리뉴클레오타이드(TRBC2)"는 TRAC 폴리펩타이드를 암호화하는 핵산을 의미한다. 예시적인 TRBC2 핵산 서열이 아래에 제공된다.

[0402] >NG_001333.2:655095-656583 호모 사피엔스 7번 염색체 상의 T 세포 수용체 베타 유전자좌(TRB)

AGGACCTGAAAAACGTGTTCCCACCCGAGGTCGCTGTGTTTGAGCCATCAGAAGCAGAGATCTC
CCACACCCAAAAGGCCACACTGGTATGCCTGGCCACAGGCTTCTACCCCGACCACGTGGAGCTG
AGCTGGTGGGTGAATGGGAAGGAGGTGCACAGTGGGGTCAGCACAGACCCGCAGCCCCTCAAGG
AGCAGCCCGCCCTCAATGACTCCAGATACTGCCTGAGCAGCCGCCTGAGGGTCTCGGCCACCTT
CTGGCAGAACCCCCGCAACCACTTCCGCTGTCAAGTCCAGTTCTACGGGCTCTCGGAGAATGAC
GAGTGGACCCAGGATAGGGCCAAACCCGTCACCCAGATCGTCAGCGCCGAGGCCTGGGGTAGAG
CAGGTGAGTGGGGCCTGGGGAGATGCCTGGAGGAGATTAGGTGAGACCAGCTACCAGGGAAAAT
GGAAAGATCCAGGTAGCGGACAAGACTAGATCCAGAAGAAAGCCAGAGTGGACAAGGTGGGATG
ATCAAGGTTCACAGGGTCAGCAAAGCACGGTGTGCACTTCCCCCACCAAGAAGCATAGAGGCTG
AATGGAGCACCTCAAGCTCATTCTTCCTTCAGATCCTGACACCTTAGAGCTAAGCTTTCAAGTC
TCCCTGAGGACCAGCCATACAGCTCAGCATCTGAGTGGTGTGCATCCCATTCTCTTCTGGGGTC
CTGGTTTCCTAAGATCATAGTGACCACTTCGCTGGCACTGGAGCAGCATGAGGGAGACAGAACC
AGGGCTATCAAAGGAGGCTGACTTTGTACTATCTGATATGCATGTGTTTGTGGCCTGTGAGTCT
GTGATGTAAGGCTCAATGTCCTTACAAAGCAGCATTCTCTCATCCATTTTTCTTCCCCTGTTTT
CTTTCAGACTGTGGCTTCACCTCCGGTAAGTGAGTCTCTCCTTTTTCTCTCTATCTTTCGCCGT
CTCTGCTCTCGAACCAGGGCATGGAGAATCCACGGACACAGGGGCGTGAGGGAGGCCAGAGCCA
CCTGTGCACAGGTGCCTACATGCTCTGTTCTTGTCAACAGAGTCTTACCAGCAAGGGGTCCTGT
CTGCCACCATCCTCTATGAGATCTTGCTAGGGAAGGCCACCTTGTATGCCGTGCTGGTCAGTGC
CCTCGTGCTGATGGCCATGGTAAGGAGGAGGGTGGGATAGGGCAGATGATGGGGGCAGGGGATG
GAACATCACACATGGGCATAAAGGAATCTCAGAGCCAGAGCACAGCCTAATATATCCTATCACC
TCAATGAAACCATAATGAAGCCAGACTGGGGAGAAAATGCAGGGAATATCACAGAATGCATCAT
GGGAGGATGGAGACAACCAGCGAGCCCTACTCAAATTAGGCCTCAGAGCCCGCCTCCCCTGCCC
TACTCCTGCTGTGCCATAGCCCCTGAAACCCTGAAAATGTTCTCTCTTCCACAGGTCAAGAGAA

[0403] AGGATTCCAGAGGCTAG

[0404] "tet 메틸사이토신 다이옥시게네이스 2(TET2) 폴리펩타이드"는 NCBI 등록 번호 FM992369.1 또는 이의 단편에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며, 메틸사이토신을 5-하이드록시메틸사이토신으로 전환하는 촉매 활성을 갖는 단백질을 의미한다. 유전자의 결함은 골수증식성 장애와 관련이 있으며, 사이토신을 메틸화하는 효소의 능력은 전사 조절에 기여한다. 예시적인 TET2 아미노산 서열이 아래에 제공된다.

[0405] >CAX30492.1 tet 종양 유전자 패밀리 구성원 2 [호모 사피엔스]

MEQDRTNHVEGNRLSPFLIPSPPICQTEPLATKLQNGSPLPERAHPEVNGDTKWHSFKSYYGIP
CMKGSQNSRVSPDFTQESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLKQDQKANGERRNFGVS
QERNPGESSQPNVSDLSDKKESVSSVAQENAVKDFTSFSTHNCSGPENPELQILNEQEGKSANY
HDKNIVLLKNKAVLMPNGATVSASSVEHTHGELLEKTLSQYYPDCVSIAVQKTTSHINAINSQA
TNELSCEITHPSHTSGQINSAQTSNSELPPKPAAVVSEACDADDADNASKLAAMLNTCSFQKPE
QLQQQKSVFEICPSPAENNIQGTTKLASGEEFCSGSSSNLQAPGGSSERYLKQNEMNGAYFKQS
SVFTKDSFSATTTPPPPSQLLLSPPPPLPQVPQLPSEGKSTLNGGVLEEHHHYPNQSNTTLLRE
VKIEGKPEAPPSQSPNPSTHVCSPSPMLSERPQNNCVNRNDIQTAGTMTVPLCSEKTRPMSEHL
KHNPPIFGSSGELQDNCQQLMRNKEQEILKGRDKEQTRDLVPPTQHYLKPGWIELKAPRFHQAE
SHLKRNEASLPSILQYQPNLSNQMTSKQYTGNSNMPGGLPRQAYTQKTTQLEHKSQMYQVEMNQ
GQSQGTVDQHLQFQKPSHQVHFSKTDHLPKAHVQSLCGTRFHFQQRADSQTEKLMSPVLKQHLN
QQASETEPFSNSHLLQHKPHKQAAQTQPSQSSHLPQNQQQQQKLQIKNKEEILQTFPHPQSNND
QQREGSFFGQTKVEECFHGENQYSKSSEFETHNVQMGLEEVQNINRRNSPYSQTMKSSACKIQV
SCSNNTHLVSENKEQTTHPELFAGNKTQNLHHMQYFPNNVIPKQDLLHRCFQEQEQKSQQASVL
QGYKNRNQDMSGQQAAQLAQQRYLIHNHANVFPVPDQGGSHTQTPPQKDTQKHAALRWHLLQKQ
EQQQTQQPQTESCHSQMHRPIKVEPGCKPHACMHTAPPENKTWKKVTKQENPPASCDNVQQKSI
IETMEQHLKQFHAKSLFDHKALTLKSQKQVKVEMSGPVTVLTRQTTAAELDSHTPALEQQTTSS
EKTPTKRTAASVLNNFIESPSKLLDTPIKNLLDTPVKTQYDFPSCRCVEQIIEKDEGPFYTHLG
AGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRER
AGHTCEAAVIVILILVWEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPET
CGASFSFGCSWSMYYNGCKFARSKIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDA
YNNQIEYEHRAPECRLGLKEGRPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGG
KPEDEQLHVLPLYKVSDVDEFGSVEAQEEKKRSGAIQVLSSFRRKVRMLAEPVKTCRQRKLEAK
KAAAEKLSSLENSSNKNEKEKSAPSRTKQTENASQAKQLAELLRLSGPVMQQSQQPQPLQKQPP
QPQQQQRPQQQQPHHPQTESVNSYSASGSTNPYMRRPNPVSPYPNSSHTSDIYGSTSPMNFYST
[0406]

SSQAAGSYLNSSNPMNPYPGLLNQNTQYPSYQCNGNLSVDNCSPYLGSYSPQSQPMDLYRYPSQ
DPLSKLSLPPIHTLYQPRFGNSQSFTSKYLGYGNQNMQGDGFSSCTIRPNVHHVGKLPPYPTHE
MDGHFMGATSRLPPNLSNPNMDYKNGEHHSPSHIIHNYSAAPGMFNSSLHALHLQNKENDMLSH
TANGLSKMLPALNHDRTACVQGGLHKLSDANGQEKQPLALVQGVASGAEDNDEVWSDSEQSFLD
PDIGGVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGLALWEAK
MAEKAREKEEECEKYGPDYVPQKSHGKKVKREPAEPHETSEPTYLRFIKSLAERTMSVTTDSTV
[0407] TTSPYAFTRVTGPYNRYI

[0408] "tet 메틸사이토신 다이옥시게네이스 2(TET2) 폴리뉴클레오타이드"는 TET2 폴리펩타이드를 암호화하는 핵산 분자를 의미한다. TET 폴리펩타이드는 메틸사이토신 다이옥시게네이스를 암호화하며, 전사 조절 활성을 갖는다. 예시적인 TET2 핵산이 아래에 제공된다.

[0409] >FM992369.1 호모 사피엔스 tet 종양 유전자 패밀리 구성원 2(TET2 유전자)에 대한 mRNA

CCGTGCCATCCCAACCTCCCACCTCGCCCCCAACCTTCGCGCTTGCTCTGCTTCTTCTCCCAGG
GGTGGAGACCCGCCGAGGTCCCCGGGGTTCCCGAGGGCTGCACCCTTCCCCGCGCTCGCCAGCC
CTGGCCCCTACTCCGCGCTGGTCCGGGCGCACCACTCCCCCCGCGCCACTGCACGGCGTGAGGG
CAGCCCAGGTCTCCACTGCGCGCCCCGCTGTACGGCCCCAGGTGCCGCCGGCCTTTGTGCTGGA
CGCCCGGTGCGGGGGGCTAATTCCCTGGGAGCCGGGGCTGAGGGCCCCAGGGCGGCGGCGCAGG
CCGGGGCGGAGCGGGAGGAGGCCGGGGCGGAGCAGGAGGAGGCCCGGGCGGAGGAGGAGAGCCG
GCGGTAGCGGCAGTGGCAGCGGCGAGAGCTTGGGCGGCCGCCGCCGCCTCCTCGCGAGCGCCGC
GCGCCCGGGTCCCGCTCGCATGCAAGTCACGTCCGCCCCCTCGGCGCGGCCGCCCCGAGACGCC
GGCCCCGCTGAGTGATGAGAACAGACGTCAAACTGCCTTATGAATATTGATGCGGAGGCTAGGC
TGCTTTCGTAGAGAAGCAGAAGGAAGCAAGATGGCTGCCCTTTAGGATTTGTTAGAAAGGAGAC
CCGACTGCAACTGCTGGATTGCTGCAAGGCTGAGGGACGAGAACGAGGCTGGCAAACATTCAGC
AGCACACCCTCTCAAGATTGTTTACTTGCCTTTGCTCCTGTTGAGTTACAACGCTTGGAAGCAG
GAGATGGGCTCAGCAGCAGCCAATAGGACATGATCCAGGAAGAGCAAATTCAACTAGAGGGCAG
CCTTGTGGATGGCCCCGAAGCAAGCCTGATGGAACAGGATAGAACCAACCATGTTGAGGGCAAC

[0410] AGACTAAGTCCATTCCTGATACCATCACCTCCCATTTGCCAGACAGAACCTCTGGCTACAAAGC

TCCAGAATGGAAGCCCACTGCCTGAGAGAGCTCATCCAGAAGTAAATGGAGACACCAAGTGGCA
CTCTTTCAAAAGTTATTATGGAATACCCTGTATGAAGGGAAGCCAGAATAGTCGTGTGAGTCCT
GACTTTACACAAGAAAGTAGAGGGTATTCCAAGTGTTTGCAAAATGGAGGAATAAAACGCACAG
TTAGTGAACCTTCTCTCTCTGGGCTCCTTCAGATCAAGAAATTGAAACAAGACCAAAAGGCTAA
TGGAGAAAGACGTAACTTCGGGGTAAGCCAAGAAAGAAATCCAGGTGAAAGCAGTCAACCAAAT
GTCTCCGATTTGAGTGATAAGAAAGAATCTGTGAGTTCTGTAGCCCAAGAAAATGCAGTTAAAG
ATTTCACCAGTTTTTCAACACATAACTGCAGTGGGCCTGAAAATCCAGAGCTTCAGATTCTGAA
TGAGCAGGAGGGGAAAAGTGCTAATTACCATGACAAGAACATTGTATTACTTAAAAACAAGGCA
GTGCTAATGCCTAATGGTGCTACAGTTTCTGCCTCTTCCGTGGAACACACACATGGTGAACTCC
TGGAAAAAACACTGTCTCAATATTATCCAGATTGTGTTTCCATTGCGGTGCAGAAAACCACATC
TCACATAAATGCCATTAACAGTCAGGCTACTAATGAGTTGTCCTGTGAGATCACTCACCCATCG
CATACCTCAGGGCAGATCAATTCCGCACAGACCTCTAACTCTGAGCTGCCTCCAAAGCCAGCTG
CAGTGGTGAGTGAGGCCTGTGATGCTGATGATGCTGATAATGCCAGTAAACTAGCTGCAATGCT
AAATACCTGTTCCTTTCAGAAACCAGAACAACTACAACAACAAAAATCAGTTTTTGAGATATGC
CCATCTCCTGCAGAAAATAACATCCAGGGAACCACAAAGCTAGCGTCTGGTGAAGAATTCTGTT
CAGGTTCCAGCAGCAATTTGCAAGCTCCTGGTGGCAGCTCTGAACGGTATTTAAAACAAAATGA
AATGAATGGTGCTTACTTCAAGCAAAGCTCAGTGTTCACTAAGGATTCCTTTTCTGCCACTACC
ACACCACCACCACCATCACAATTGCTTCTTTCTCCCCCTCCTCCTCTTCCACAGGTTCCTCAGC
TTCCTTCAGAAGGAAAAAGCACTCTGAATGGTGGAGTTTTAGAAGAACACCACCACTACCCCAA
CCAAAGTAACACAACACTTTTAAGGGAAGTGAAAATAGAGGGTAAACCTGAGGCACCACCTTCC
CAGAGTCCTAATCCATCTACACATGTATGCAGCCCTTCTCCGATGCTTTCTGAAAGGCCTCAGA
ATAATTGTGTGAACAGGAATGACATACAGACTGCAGGGACAATGACTGTTCCATTGTGTTCTGA
GAAAACAAGACCAATGTCAGAACACCTCAAGCATAACCCACCAATTTTTGGTAGCAGTGGAGAG
CTACAGGACAACTGCCAGCAGTTGATGAGAAACAAAGAGCAAGAGATTCTGAAGGGTCGAGACA
AGGAGCAAACACGAGATCTTGTGCCCCCAACACAGCACTATCTGAAACCAGGATGGATTGAATT
GAAGGCCCCTCGTTTTCACCAAGCGGAATCCCATCTAAAACGTAATGAGGCATCACTGCCATCA
ATTCTTCAGTATCAACCCAATCTCTCCAATCAAATGACCTCCAAACAATACACTGGAAATTCCA
ACATGCCTGGGGGGCTCCCAAGGCAAGCTTACACCCAGAAAACAACACAGCTGGAGCACAAGTC
ACAAATGTACCAAGTTGAAATGAATCAAGGGCAGTCCCAAGGTACAGTGGACCAACATCTCCAG
TTCCAAAAACCCTCACACCAGGTGCACTTCTCCAAAACAGACCATTTACCAAAAGCTCATGTGC
AGTCACTGTGTGGCACTAGATTTCATTTTCAACAAAGAGCAGATTCCCAAACTGAAAAACTTAT
GTCCCCAGTGTTGAAACAGCACTTGAATCAACAGGCTTCAGAGACTGAGCCATTTTCAAACTCA
CACCTTTTGCAACATAAGCCTCATAAACAGGCAGCACAAACACAACCATCCCAGAGTTCACATC
TCCCTCAAAACCAGCAACAGCAGCAAAAATTACAAATAAAGAATAAAGAGGAAATACTCCAGAC
TTTTCCTCACCCCCAAAGCAACAATGATCAGCAAAGAGAAGGATCATTCTTTGGCCAGACTAAA

[0411]

GTGGAAGAATGTTTTCATGGTGAAAATCAGTATTCAAAATCAAGCGAGTTCGAGACTCATAATG
TCCAAATGGGACTGGAGGAAGTACAGAATATAAATCGTAGAAATTCCCCTTATAGTCAGACCAT
GAAATCAAGTGCATGCAAAATACAGGTTTCTTGTTCAAACAATACACACCTAGTTTCAGAGAAT
AAAGAACAGACTACACATCCTGAACTTTTTGCAGGAAACAAGACCCAAAACTTGCATCACATGC
AATATTTTCCAAATAATGTGATCCCAAAGCAAGATCTTCTTCACAGGTGCTTTCAAGAACAGGA
GCAGAAGTCACAACAAGCTTCAGTTCTACAGGGATATAAAAATAGAAACCAAGATATGTCTGGT
CAACAAGCTGCGCAACTTGCTCAGCAAAGGTACTTGATACATAACCATGCAAATGTTTTTCCTG
TGCCTGACCAGGGAGGAAGTCACACTCAGACCCCTCCCCAGAAGGACACTCAAAAGCATGCTGC
TCTAAGGTGGCATCTCTTACAGAAGCAAGAACAGCAGCAAACACAGCAACCCCAAACTGAGTCT
TGCCATAGTCAGATGCACAGGCCAATTAAGGTGGAACCTGGATGCAAGCCACATGCCTGTATGC
ACACAGCACCACCAGAAAACAAAACATGGAAAAAGGTAACTAAGCAAGAGAATCCACCTGCAAG
CTGTGATAATGTGCAGCAAAAGAGCATCATTGAGACCATGGAGCAGCATCTGAAGCAGTTTCAC
GCCAAGTCGTTATTTGACCATAAGGCTCTTACTCTCAAATCACAGAAGCAAGTAAAAGTTGAAA
TGTCAGGGCCAGTCACAGTTTTGACTAGACAAACCACTGCTGCAGAACTTGATAGCCACACCCC
AGCTTTAGAGCAGCAAACAACTTCTTCAGAAAAGACACCAACCAAAAGAACAGCTGCTTCTGTT
CTCAATAATTTTATAGAGTCACCTTCCAAATTACTAGATACTCCTATAAAAAATTTATTGGATA
CACCTGTCAAGACTCAATATGATTTCCCATCTTGCAGATGTGTAGAGCAAATTATTGAAAAAGA
TGAAGGTCCTTTTTATACCCATCTAGGAGCAGGTCCTAATGTGGCAGCTATTAGAGAAATCATG
GAAGAAAGGTTTGGACAGAAGGGTAAAGCTATTAGGATTGAAAGAGTCATCTATACTGGTAAAG
AAGGCAAAAGTTCTCAGGGATGTCCTATTGCTAAGTGGGTGGTTCGCAGAAGCAGCAGTGAAGA
GAAGCTACTGTGTTTGGTGCGGGAGCGAGCTGGCCACACCTGTGAGGCTGCAGTGATTGTGATT
CTCATCCTGGTGTGGGAAGGAATCCCGCTGTCTCTGGCTGACAAACTCTACTCGGAGCTTACCG
AGACGCTGAGGAAATACGGCACGCTCACCAATCGCCGGTGTGCCTTGAATGAAGAGAGAACTTG
CGCCTGTCAGGGGCTGGATCCAGAAACCTGTGGTGCCTCCTTCTCTTTTGGTTGTTCATGGAGC
ATGTACTACAATGGATGTAAGTTTGCCAGAAGCAAGATCCCAAGGAAGTTTAAGCTGCTTGGGG
ATGACCCAAAAGAGGAAGAGAAACTGGAGTCTCATTTGCAAAACCTGTCCACTCTTATGGCACC
AACATATAAGAAACTTGCACCTGATGCATATAATAATCAGATTGAATATGAACACAGAGCACCA
GAGTGCCGTCTGGGTCTGAAGGAAGGCCGTCCATTCTCAGGGGTCACTGCATGTTTGGACTTCT
GTGCTCATGCCCACAGAGACTTGCACAACATGCAGAATGGCAGCACATTGGTATGCACTCTCAC
TAGAGAAGACAATCGAGAATTTGGAGGAAAACCTGAGGATGAGCAGCTTCACGTTCTGCCTTTA
TACAAAGTCTCTGACGTGGATGAGTTTGGGAGTGTGGAAGCTCAGGAGGAGAAAAAACGGAGTG
GTGCCATTCAGGTACTGAGTTCTTTTCGGCGAAAAGTCAGGATGTTAGCAGAGCCAGTCAAGAC
TTGCCGACAAAGGAAACTAGAAGCCAAGAAAGCTGCAGCTGAAAAGCTTTCCTCCCTGGAGAAC
AGCTCAAATAAAAATGAAAAGGAAAAGTCAGCCCCATCACGTACAAAACAAACTGAAAACGCAA
GCCAGGCTAAACAGTTGGCAGAACTTTTGCGACTTTCAGGACCAGTCATGCAGCAGTCCCAGCA

[0412]

GCCCCAGCCTCTACAGAAGCAGCCACCACAGCCCCAGCAGCAGCAGAGACCCCAGCAGCAGCAG

CCACATCACCCTCAGACAGAGTCTGTCAACTCTTATTCTGCTTCTGGATCCACCAATCCATACA

TGAGACGGCCCAATCCAGTTAGTCCTTATCCAAACTCTTCACACACTTCAGATATCTATGGAAG

CACCAGCCCTATGAACTTCTATTCCACCTCATCTCAAGCTGCAGGTTCATATTTGAATTCTTCT

AATCCCATGAACCCTTACCCTGGGCTTTTGAATCAGAATACCCAATATCCATCATATCAATGCA

ATGGAAACCTATCAGTGGACAACTGCTCCCCATATCTGGGTTCCTATTCTCCCCAGTCTCAGCC

GATGGATCTGTATAGGTATCCAAGCCAAGACCCTCTGTCTAAGCTCAGTCTACCACCCATCCAT

ACACTTTACCAGCCAAGGTTTGGAAATAGCCAGAGTTTTACATCTAAATACTTAGGTTATGGAA

ACCAAAATATGCAGGGAGATGGTTTCAGCAGTTGTACCATTAGACCAAATGTACATCATGTAGG

GAAATTGCCTCCTTATCCCACTCATGAGATGGATGGCCACTTCATGGGAGCCACCTCTAGATTA

CCACCCAATCTGAGCAATCCAAACATGGACTATAAAAATGGTGAACATCATTCACCTTCTCACA

TAATCCATAACTACAGTGCAGCTCCGGGCATGTTCAACAGCTCTCTTCATGCCCTGCATCTCCA

AAACAAGGAGAATGACATGCTTTCCCACACAGCTAATGGGTTATCAAAGATGCTTCCAGCTCTT

AACCATGATAGAACTGCTTGTGTCCAAGGAGGCTTACACAAATTAAGTGATGCTAATGGTCAGG

AAAAGCAGCCATTGGCACTAGTCCAGGGTGTGGCTTCTGGTGCAGAGGACAACGATGAGGTCTG

GTCAGACAGCGAGCAGAGCTTTCTGGATCCTGACATTGGGGGAGTGGCCGTGGCTCCAACTCAT

GGGTCAATTCTCATTGAGTGTGCAAAGCGTGAGCTGCATGCCACAACCCCTTTAAAGAATCCCA

ATAGGAATCACCCCACCAGGATCTCCCTCGTCTTTTACCAGCATAAGAGCATGAATGAGCCAAA

ACATGGCTTGGCTCTTTGGGAAGCCAAAATGGCTGAAAAAGCCCGTGAGAAAGAGGAAGAGTGT

GAAAAGTATGGCCCAGACTATGTGCCTCAGAAATCCCATGGCAAAAAAGTGAAACGGGAGCCTG

CTGAGCCACATGAAACTTCAGAGCCCACTTACCTGCGTTTCATCAAGTCTCTTGCCGAAAGGAC

CATGTCCGTGACCACAGACTCCACAGTAACTACATCTCCATATGCCTTCACTCGGGTCACAGGG

CCTTACAACAGATATATATGAAGATATATATGATATCACCCCCTTTTGTTGGTTACCTCACTTG

AAAAGACCACAACCAACCTGTCAGTAGTATAGTTCTCATGACGTGGGCAGTGGGGAAAGGTCAC

AGTATTCATGACAAATGTGGTGGGAAAAACCTCAGCTCACCAGCAACAAAAGAGGTTATCTTAC

CATAGCACTTAATTTTCACTGGCTCCCAAGTGGTCACAGATGGCATCTAGGAAAAGACCAAAGC

ATTCTATGCAAAAAGAAGGTGGGGAAGAAAGTGTTCCGCAATTTACATTTTTAAACACTGGTTC

TATTATTGGACGAGATGATATGTAAATGTGATCCCCCCCCCCCGCTTACAACTCTACACATCTG

TGACCACTTTTAATAATATCAAGTTTGCATAGTCATGGAACACAAATCAAACAAGTACTGTAGT

ATTACAGTGACAGGAATCTTAAAATACCATCTGGTGCTGAATATATGATGTACTGAAATACTGG

AATTATGGCTTTTTGAAATGCAGTTTTTACTGTAATCTTAACTTTTATTTATCAAAATAGCTAC

AGGAAACATGAATAGCAGGAAAACACTGAATTTGTTTGGATGTTCTAAGAAATGGTGCTAAGAA

AATGGTGTCTTTAATAGCTAAAAATTTAATGCCTTTATATCATCAAGATGCTATCAGTGTACTC

CAGTGCCCTTGAATAATAGGGGTACCTTTTCATTCAAGTTTTTATCATAATTACCTATTCTTAC

[0413] ACAAGCTTAGTTTTTAAAATGTGGACATTTTAAAGGCCTCTGGATTTTGCTCATCCAGTGAAGT

CCTTGTAGGACAATAAACGTATATATGTACATATATACACAAACATGTATATGTGCACACACAT
GTATATGTATAAATATTTTAAATGGTGTTTTAGAAGCACTTTGTCTACCTAAGCTTTGACAACT
TGAACAATGCTAAGGTACTGAGATGTTTAAAAAACAAGTTTACTTTCATTTTAGAATGCAAAGT
TGATTTTTTTAAGGAAACAAAGAAAGCTTTTAAAATATTTTTGCTTTTAGCCATGCATCTGCTG
ATGAGCAATTGTGTCCATTTTTAACACAGCCAGTTAAATCCACCATGGGGCTTACTGGATTCAA
GGGAATACGTTAGTCCACAAAACATGTTTTCTGGTGCTCATCTCACATGCTATACTGTAAAACA
GTTTTATACAAAATTGTATGACAAGTTCATTGCTCAAAAATGTACAGTTTTAAGAATTTTCTAT
TAACTGCAGGTAATAATTAGCTGCATGCTGCAGACTCAACAAAGCTAGTTCACTGAAGCCTATG
CTATTTTATGGATCATAGGCTCTTCAGAGAACTGAATGGCAGTCTGCCTTTGTGTTGATAATTA
TGTACATTGTGACGTTGTCATTTCTTAGCTTAAGTGTCCTCTTTAACAAGAGGATTGAGCAGAC
TGATGCCTGCATAAGATGAATAAACAGGGTTAGTTCCATGTGAATCTGTCAGTTAAAAAGAAAC
AAAAACAGGCAGCTGGTTTGCTGTGGTGGTTTTAAATCATTAATTTGTATAAAGAAGTGAAAGA
GTTGTATAGTAAATTAAATTGTAAACAAAACTTTTTTAATGCAATGCTTTAGTATTTTAGTACT
GTAAAAAAATTAAATATATACATATATATATATATATATATATATATATATATGAGTTTGAAGC
AGAATTCACATCATGATGGTGCTACTCAGCCTGCTACAAATATATCATAATGTGAGCTAAGAAT
TCATTAAATGTTTGAGTGATGTTCCTACTTGTCATATACCTCAACACTAGTTTGGCAATAGGAT
ATTGAACTGAGAGTGAAAGCATTGTGTACCATCATTTTTTTCCAAGTCCTTTTTTTTATTGTTA
AAAAAAAAAGCATACCTTTTTTCAATACTTGATTTCTTAGCAAGTATAACTTGAACTTCAACCT
TTTTGTTCTAAAAATTCAGGGATATTTCAGCTCATGCTCTCCCTATGCCAACATGTCACCTGTG
TTTATGTAAAATTGTTGTAGGTTAATAAATATATTCTTTGTCAGGGATTTAACCCTTTTATTTT
GAATCCCTTCTATTTTACTTGTACATGTGCTGATGTAACTAAAACTAATTTTGTAAATCTGTTG
GCTCTTTTTATTGTAAAGAAAAGCATTTTAAAAGTTTGAGGAATCTTTTGACTGTTTCAAGCAG
GAAAAAAAAATTACATGAAAATAGAATGCACTGAGTTGATAAAGGGAAAAATTGTAAGGCAGGA
GTTTGGCAAGTGGCTGTTGGCCAGAGACTTACTTGTAACTCTCTAAATGAAGTTTTTTTGATCC
TGTAATCACTGAAGGTACATACTCCATGTGGACTTCCCTTAAACAGGCAAACACCTACAGGTAT
GGTGTGCAACAGATTGTACAATTACATTTTGGCCTAAATACATTTTTGCTTACTAGTATTTAAA
ATAAATTCTTAATCAGAGGAGGCCTTTGGGTTTTATTGGTCAAATCTTTGTAAGCTGGCTTTTG
TCTTTTTAAAAAATTTCTTGAATTTGTGGTTGTGTCCAATTTGCAAACATTTCCAAAAATGTTT
GCTTTGCTTACAAACCACATGATTTTAATGTTTTTTGTATACCATAATATCTAGCCCCAAACAT
TTGATTACTACATGTGCATTGGTGATTTTGATCATCCATTCTTAATATTTGATTTCTGTGTCAC
CTACTGTCATTTGTTAAACTGCTGGCCAACAAGAACAGGAAGTATAGTTTGGGGGGTTGGGGAG
AGTTTACATAAGGAAGAGAAGAAATTGAGTGGCATATTGTAAATATCAGATCTATAATTGTAAA
TATAAAACCTGCCTCAGTTAGAATGAATGGAAAGCAGATCTACAATTTGCTAATATAGGAATAT
CAGGTTGACTATATAGCCATACTTGAAAATGCTTCTGAGTGGTGTCAACTTTACTTGAATGAAT
[0414] TTTTCATCTTGATTGACGCACAGTGATGTACAGTTCACTTCTGAAGCTAGTGGTTAACTTGTGT

AGGAAACTTTTGCAGTTTGACACTAAGATAACTTCTGTGTGCATTTTTCTATGCTTTTTTAAAA
ACTAGTTTCATTTCATTTTCATGAGATGTTTGGTTTATAAGATCTGAGGATGGTTATAAATACT
GTAAGTATTGTAATGTTATGAATGCAGGTTATTTGAAAGCTGTTTATTATTATATCATTCCTGA
TAATGCTATGTGAGTGTTTTTAATAAAATTTATATTTATTTAATGCACTCTAAGTGTTGTCTTC
[0415] CT

[0416] "형질전환 성장 인자 수용체 2(TGFBRII) 폴리펩타이드"는 NCBI 등록 번호 ABG65632.1 또는 이의 단편에 대해 적어도 약 85%의 서열 동일성을 가지며 면역억제 활성을 갖는 단백질을 의미한다. 예시적인 아미노산 서열이 아래에 제공된다.

[0417] >ABG65632.1 형질전환 성장 인자 베타 수용체 II [호모 사피엔스]

MGRGLLRGLWPLHIVLWTRIASTIPPHVQKSVNNDMIVTDNNGAVKFPQLCKFCDVRFSTCDNQ
KSCMSNCSITSICEKPQEVCVAVWRKNDENITLETVCHDPKLPYHDFILEDAASPKCIMKEKKK
PGETFFMCSCSSDECNDNIIFSEEYNTSNPDLLLVIFQVTGISLLPPLGVAISVIIIFYCYRVN
RQQKLSSTWETGKTRKLMEFSEHCAIILEDDRSDISSTCANNINHNTELLPIELDTLVGKGRFA
EVYKAKLKQNTSEQFETVAVKIFPYEEYASWKTEKDIFSDINLKHENILQFLTAEERKTELGKQ
YWLITAFHAKGNLQEYLTRHVISWEDLRKLGSSLARGIAHLHSDHTPCGRPKMPIVHRDLKSSN
ILVKNDLTCCLCDFGLSLRLDPTLSVDDLANSGQVGTARYMAPEVLESRMNLENVESFKQTDVY
SMALVLWEMTSRCNAVGEVKDYEPPFGSKVREHPCVESMKDNVLRDRGRPEIPSFWLNHQGIQM
VCETLTECWDHDPEARLTAQCVAERFSELEHLDRLSGRSCSEEKIPEDGSLNTTK

[0418]

[0419] "형질전환 성장 인자 수용체 2(TGFBRII) 폴리뉴클레오타이드"는 TGFBRII 폴리펩타이드를 암호화하는 핵산을 의미한다. TGFBRII 유전자는 세린/트레오닌 카이네이스 활성을 갖는 막관통 단백질을 암호화한다. 예시적인 TGFBRII 핵산이 아래에 제공된다.

[0420] >M85079.1 인간 TGF-베타 유형 II 수용체 mRNA, 완전 cds

GTTGGCGAGGAGTTTCCTGTTTCCCCCGCAGCGCTGAGTTGAAGTTGAGTGAGTCACTCGCGCG
CACGGAGCGACGACACCCCCGCGCGTGCACCCGCTCGGGACAGGAGCCGGACTCCTGTGCAGCT
TCCCTCGGCCGCCGGGGGCCTCCCCGCGCCTCGCCGGCCTCCAGGCCCCTCCTGGCTGGCGAGC
GGGCGCCACATCTGGCCCGCACATCTGCGCTGCCGGCCCGGCGCGGGGTCCGGAGAGGGCGCGG
CGCGGAGCGCAGCCAGGGGTCCGGGAAGGCGCCGTCCGTGCGCTGGGGGCTCGGTCTATGACGA
GCAGCGGGGTCTGCCATGGGTCGGGGGCTGCTCAGGGGCCTGTGGCCGCTGCACATCGTCCTGT
GGACGCGTATCGCCAGCACGATCCCACCGCACGTTCAGAAGTCGGTTAATAACGACATGATAGT
CACTGACAACAACGGTGCAGTCAAGTTTCCACAACTGTGTAAATTTTGTGATGTGAGATTTTCC
ACCTGTGACAACCAGAAATCCTGCATGAGCAACTGCAGCATCACCTCCATCTGTGAGAAGCCAC
AGGAAGTCTGTGTGGCTGTATGGAGAAAGAATGACGAGAACATAACACTAGAGACAGTTTGCCA
TGACCCCAAGCTCCCCTACCATGACTTTATTCTGGAAGATGCTGCTTCTCCAAAGTGCATTATG
AAGGAAAAAAAAAAGCCTGGTGAGACTTTCTTCATGTGTTCCTGTAGCTCTGATGAGTGCAATG
ACAACATCATCTTCTCAGAAGAATATAACACCAGCAATCCTGACTTGTTGCTAGTCATATTTCA
AGTGACAGGCATCAGCCTCCTGCCACCACTGGGAGTTGCCATATCTGTCATCATCATCTTCTAC
TGCTACCGCGTTAACCGGCAGCAGAAGCTGAGTTCAACCTGGGAAACCGGCAAGACGCGGAAGC
TCATGGAGTTCAGCGAGCACTGTGCCATCATCCTGGAAGATGACCGCTCTGACATCAGCTCCAC
GTGTGCCAACAACATCAACCACAACACAGAGCTGCTGCCCATTGAGCTGGACACCCTGGTGGGG
AAAGGTCGCTTTGCTGAGGTCTATAAGGCCAAGCTGAAGCAGAACACTTCAGAGCAGTTTGAGA
CAGTGGCAGTCAAGATCTTTCCCTATGAGGAGTATGCCTCTTGGAAGACAGAGAAGGACATCTT
CTCAGACATCAATCTGAAGCATGAGAACATACTCCAGTTCCTGACGGCTGAGGAGCGGAAGACG
GAGTTGGGGAAACAATACTGGCTGATCACCGCCTTCCACGCCAAGGGCAACCTACAGGAGTACC
TGACGCGGCATGTCATCAGCTGGGAGGACCTGCGCAAGCTGGGCAGCTCCCTCGCCCGGGGGAT
TGCTCACCTCCACAGTGATCACACTCCATGTGGGAGGCCCAAGATGCCCATCGTGCACAGGGAC
CTCAAGAGCTCCAATATCCTCGTGAAGAACGACCTAACCTGCTGCCTGTGTGACTTTGGGCTTT
CCCTGCGTCTGGACCCTACTCTGTCTGTGGATGACCTGGCTAACAGTGGGCAGGTGGGAACTGC
AAGATACATGGCTCCAGAAGTCCTAGAATCCAGGATGAATTTGGAGAATGCTGAGTCCTTCAAG
CAGACCGATGTCTACTCCATGGCTCTGGTGCTCTGGGAAATGACATCTCGCTGTAATGCAGTGG
GAGAAGTAAAAGATTATGAGCCTCCATTTGGTTCCAAGGTGCGGGAGCACCCCTGTGTCGAAAG
CATGAAGGACAACGTGTTGAGAGATCGAGGGCGACCAGAAATTCCCAGCTTCTGGCTCAACCAC
CAGGGCATCCAGATGGTGTGTGAGACGTTGACTGAGTGCTGGGACCACGACCCAGAGGCCCGTC
TCACAGCCCAGTGTGTGGCAGAACGCTTCAGTGAGCTGGAGCATCTGGACAGGCTCTCGGGGAG
GAGCTGCTCGGAGGAGAAGATTCCTGAAGACGGCTCCCTAAACACTACCAAATAGCTCTTATGG
GGCAGGCTGGGCATGTCCAAAGAGGCTGCCCCTCTCACCAAA

[0421]

[0422] "Ig 및 ITIM 도메인을 갖는 T 세포 면역수용체(TIGIT) 폴리펩타이드"는 NCBI 등록 번호 ACD74757.1 또는 이의 단편에 대해 적어도 약 85%의 서열 동일성을 가지며 면역조절 활성을 갖는 단백질을 의미한다. 예시적인 TIGIT 아미노산 서열이 아래에 제공된다.

[0423] >ACD74757.1 Ig 및 ITIM 도메인을 갖는 T 세포 면역수용체 [호모 사피엔스]

[0424]
```
MRWCLLLIWAQGLRQAPLASGMMTGTIETTGNISAEKGGSIILQCHLSSTTAQVTQVNWEQQDQ
LLAICNADLGWHISPSFKDRVAPGPGLGLTLQSLTVNDTGEYFCIYHTYPDGTYTGRIFLEVLE
SSVAEHGARFQIPLLGAMAATLVVICTAVIVVVALTRKKKALRIHSVEGDLRRKSAGQEEWSPS
APSPPGSCVQAEAAPAGLCGEQRGEDCAELHDYFNVLSYRSLGNCSFFTETG
```

[0425] "Ig 및 ITIM 도메인을 갖는 T 세포 면역수용체(TIGIT) 폴리뉴클레오타이드"는 TIGIT 폴리펩타이드를 암호화하는 핵산을 의미한다. TIGIT 유전자는 신생물 및 T 세포 고갈과 관련되는 저해성 면역 수용체를 암호화한다. 예시적인 핵산 서열이 아래에 제공된다.

[0426] >EU675310.1 호모 사피엔스 Ig 및 ITIM 도메인을 갖는 T 세포 면역수용체(TIGIT) mRNA, 완전 cds

[0427]
```
CGTCCTATCTGCAGTCGGCTACTTTCAGTGGCAGAAGAGGCCACATCTGCTTCCTGTAGGCCCT
CTGGGCAGAAGCATGCGCTGGTGTCTCCTCCTGATCTGGGCCCAGGGGCTGAGGCAGGCTCCCC
TCGCCTCAGGAATGATGACAGGCACAATAGAAACAACGGGGAACATTTCTGCAGAGAAAGGTGG
CTCTATCATCTTACAATGTCACCTCTCCTCCACCACGGCACAAGTGACCCAGGTCAACTGGGAG
CAGCAGGACCAGCTTCTGGCCATTTGTAATGCTGACTTGGGGTGGCACATCTCCCCATCCTTCA
AGGATCGAGTGGCCCCAGGTCCCGGCCTGGGCCTCACCCTCCAGTCGCTGACCGTGAACGATAC
AGGGGAGTACTTCTGCATCTATCACACCTACCCTGATGGGACGTACACTGGGAGAATCTTCCTG
GAGGTCCTAGAAAGCTCAGTGGCTGAGCACGGTGCCAGGTTCCAGATTCCATTGCTTGGAGCCA
TGGCCGCGACGCTGGTGGTCATCTGCACAGCAGTCATCGTGGTGGTCGCGTTGACTAGAAAGAA
GAAAGCCCTCAGAATCCATTCTGTGGAAGGTGACCTCAGGAGAAAATCAGCTGGACAGGAGGAA
TGGAGCCCCAGTGCTCCCTCACCCCCAGGAAGCTGTGTCCAGGCAGAAGCTGCACCTGCTGGGC
TCTGTGGAGAGCAGCGGGGAGAGGACTGTGCCGAGCTGCATGACTACTTCAATGTCCTGAGTTA
CAGAAGCCTGGGTAACTGCAGCTTCTTCACAGAGACTGGTTAGCAACCAGAGGCATCTTCTGG
```

[0428] 본 명세서에서 사용되는 "형질도입"은 바이러스 벡터를 통해 유전자 또는 유전 물질을 세포로 전달하는 것을 의미한다.

[0429] 본 명세서에서 사용되는 "형질전환"은 외인성 핵산의 도입에 의해 생성된 세포에 유전적 변화를 도입하는 과정을 지칭한다.

[0430] "형질감염"은 화학적 또는 물리적 수단을 통해 유전자 또는 유전 물질을 세포로 전달하는 것을 지칭한다.

[0431] "전위"는 비-상동 염색체 사이의 핵산 분절의 재배열을 의미한다.

[0432] "막관통 도메인"은 세포 또는 바이러스 또는 바이러스-유사 입자의 지질 이중층과 같은 지질 이중층에 삽입되는 아미노산 서열을 의미한다. 막관통 도메인은 관심 단백질(예를 들어, CAR)을 막에 고정하는데 사용될 수 있다. 막관통 도메인은 천연 또는 합성 공급원으로부터 유래될 수 있다. 공급원이 천연인 경우, 도메인은 임의의 막-결합 또는 막관통 단백질로부터 유래될 수 있다. 개시된 CAR에서 사용하기 위한 막관통 도메인은 적어도 T-세포 수용체의 알파, 베타 또는 제타 사슬, CD28, CD3 엡실론, CD45, CD4, CD5, CD8, CD9, CD16, CD22, CD33, CD37, CD64, CD80, CD86, CD134, CD137, CD154의 막관통 영역(들)을 포함할 수 있다. 일부 실시형태에서, 막관통 도메인은 CD4, CD8α, CD28 및 CD3ζ로부터 유래된다. 일부 실시형태에서, 막관통 도메인은 CD8α 힌지 및 막관통 도메인이다.

[0433] 본 명세서에서 사용되는 용어 "치료하다", "치료하는", "치료" 등은 장애 및/또는 이와 관련된 증상을 감소 또는 개선시키거나 목적하는 약리학적 및/또는 생리학적 효과를 얻는 것을 지칭한다. 배제되지는 않지만, 장애 또는 병태를 치료하는 것이 이와 관련된 장애, 병태 또는 증상을 완전히 제거하는 것을 요구하지 않는다는 것이 이해될 것이다. 일부 실시형태에서, 효과는 치료적이며, 즉, 제한 없이, 효과는 질환 및/또는 질환에서 기인하는 유해 증상을 치유하거나 또는 이의 강도를 부분적으로 또는 완전히 감소, 약화, 폐지, 경감, 완화, 저감시킨다. 일부 실시형태에서, 효과는 예방적이며, 즉, 효과는 질환 또는 병태의 발생 또는 재발을 보호하거나 또는

예방한다. 이를 위해, 현재 개시된 방법은 치료적 유효량의 본 명세서에 기재된 조성물을 투여하는 단계를 포함한다.

[0434] 용어 "우라실 글리코실레이스 저해제" 또는 "UGI"는 우라실-절단 복구 시스템을 저해하는 작용제를 의미한다. 일 실시형태에서, 작용제는 숙주 우라실-DNA 글리코실레이스에 결합하고 DNA로부터 우라실 잔기의 제거를 방지하는 단백질 또는 이의 단편이다. 실시형태에서, UGI는 우라실-DNA 글리코실레이스 염기-절단 복구 효소를 저해할 수 있는 단백질, 이의 단편 또는 도메인이다. 일부 실시형태에서, UGI 도메인은 야생형 UGI 또는 이의 변형된 버전을 포함한다. 일부 실시형태에서, UGI 도메인은 아래에 기재된 예시적인 아미노산 서열의 단편을 포함한다. 일부 실시형태에서, UGI 단편은 아래에 제공되는 예시적인 UGI 서열의 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100%를 포함하는 아미노산 서열을 포함한다. 일부 실시형태에서, UGI는 아래에 기재된 바와 같은 예시적인 UGI 아미노산 서열 또는 이의 단편과 상동인 아미노산 서열을 포함한다. 일부 실시형태에서, UGI 또는 이의 일부는 야생형 UGI 또는 아래 기재된 바와 같은 UGI 서열 또는 이의 일부와 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 적어도 99.5%, 적어도 99.9% 또는 100% 동일하다. 예시적인 UGI는 다음과 같은 아미노산 서열을 포함한다:

[0435] >sp|P14739|UNGI_BPPB2 우라실-DNA 글리코실레이스 저해제

[0436]
```
MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT
SD APEYKPWALVIQDSNGENKIKML.
```

[0437] 용어 "벡터"는 핵산 서열을 세포 내로 도입하여 형질전환된 세포를 생성하는 수단을 지칭한다. 벡터는 플라스미드, 트랜스포존, 파지, 바이러스, 리포솜 및 에피솜을 포함한다. "발현 벡터"는 수용체 세포에서 발현되는 뉴클레오타이드 서열을 포함하는 핵산 서열이다. 발현 벡터는 시작, 종결, 인핸서, 프로모터 및 분비 서열과 같은 도입된 서열의 발현을 촉진하고/하거나 용이하게 하는 추가적인 핵산 서열을 포함할 수 있다.

[0438] "T 세포 수용체 관련 단백질 카이네이스 70의 제타 사슬(ZAP70) 폴리펩타이드"는 NCBI 등록 번호 AAH53878.1에 대해 적어도 약 85%의 아미노산 서열 동일성을 가지며 카이네이스 활성을 갖는 단백질을 의미한다. 예시적인 아미노산이 아래에 제공된다.

[0439] >AAH53878.1 제타-사슬(TCR) 관련 단백질 카이네이스 70kDa [호모 사피엔스]

[0440]
```
MPDPAAHLPFFYGSISRAEAEEHLKLAGMADGLFLLRQCLRSLGGYVLSLVHDVRFHHFPIERQ
LNGTYAIAGGKAHCGPAELCEFYSRDPDGLPCNLRKPCNRPSGLEPQPGVFDCLRDAMVRDYVR
QTWKLEGEALEQAIISQAPQVEKLIATTAHERMPWYHSSLTREEAERKLYSGAQTDGKFLLRPR
KEQGTYALSLIYGKTVYHYLISQDKAGKYCIPEGTKFDTLWQLVEYLKLKADGLIYCLKEACPN
SSASNASGAAAPTLPAHPSTLTHPQRRIDTLNSDGYTPEPARITSPDKPRPMPMDTSVYESPYS
DPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQIDVAIKVLKQGTEKADTEEMM
REAQIMHQLDNPYIVRLIGVCQAEALMLVMEMAGGGPLHKFLVGKREEIPVSNVAELLHQVSMG
MKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDSYYTARSAGKWPLKWYAPECIN
FRKFSSRSDVWSYGVTMWEALSYGQKPYKKMKGPEVMAFIEQGKRMECPPECPPELYALMSDCW
IYKWEDRPDFLTVEQRMRACYYSLASKVEGPPGSTQKAEAACA
```

[0441] "T 세포 수용체 관련 단백질 카이네이스 70의 제타 사슬(ZAP70) 폴리뉴클레오타이드"는 ZAP70 폴리펩타이드를 암호화하는 핵산을 의미한다. ZAP70 유전자는 T 세포 발달 및 림프구 활성화에 관여하는 타이로신 카이네이스를 암호화한다. 기능적 ZAP10의 부재는 CD8+ T 세포의 결핍을 특징으로 하는 심각한 복합 면역결핍으로 이어질 수 있다. 예시적인 ZAP70 핵산 서열이 아래에 제공된다.

[0442] >BC053878.1 호모 사피엔스 제타-사슬(TCR) 관련 단백질 카이네이스 70kDa, mRNA(cDNA 클론 MGC:61743 IMAGE:5757161), 완전 cds

GCTTGCCGGAGCTCAGCAGACACCAGGCCTTCCGGGCAGGCCTGGCCCACCGTGGGCCTCAGAG
CTGCTGCTGGGGCATTCAGAACCGGCTCTCCATTGGCATTGGGACCAGAGACCCCGCAAGTGGC
CTGTTTGCCTGGACATCCACCTGTACGTCCCCAGGTTTCGGGAGGCCCAGGGGCGATGCCAGAC
CCCGCGGCGCACCTGCCCTTCTTCTACGGCAGCATCTCGCGTGCCGAGGCCGAGGAGCACCTGA
AGCTGGCGGGCATGGCGGACGGGCTCTTCCTGCTGCGCCAGTGCCTGCGCTCGCTGGGCGGCTA
TGTGCTGTCGCTCGTGCACGATGTGCGCTTCCACCACTTTCCCATCGAGCGCCAGCTCAACGGC
ACCTACGCCATTGCCGGCGGCAAAGCGCACTGTGGACCGGCAGAGCTCTGCGAGTTCTACTCGC
GCGACCCCGACGGGCTGCCCTGCAACCTGCGCAAGCCGTGCAACCGGCCGTCGGGCCTCGAGCC
GCAGCCGGGGGTCTTCGACTGCCTGCGAGACGCCATGGTGCGTGACTACGTGCGCCAGACGTGG
AAGCTGGAGGGCGAGGCCCTGGAGCAGGCCATCATCAGCCAGGCCCCGCAGGTGGAGAAGCTCA
TTGCTACGACGGCCCACGAGCGGATGCCCTGGTACCACAGCAGCCTGACGCGTGAGGAGGCCGA

[0443]

GCGCAAACTTTACTCTGGGGCGCAGACCGACGGCAAGTTCCTGCTGAGGCCGCGGAAGGAGCAG
GGCACATACGCCCTGTCCCTCATCTATGGGAAGACGGTGTACCACTACCTCATCAGCCAAGACA
AGGCGGGCAAGTACTGCATTCCCGAGGGCACCAAGTTTGACACGCTCTGGCAGCTGGTGGAGTA
TCTGAAGCTGAAGGCGGACGGGCTCATCTACTGCCTGAAGGAGGCCTGCCCCAACAGCAGTGCC
AGCAACGCCTCAGGGGCTGCTGCTCCCACACTCCCAGCCCACCCATCCACGTTGACTCATCCTC
AGAGACGAATCGACACCCTCAACTCAGATGGATACACCCCTGAGCCAGCACGCATAACGTCCCC
AGACAAACCGCGGCCGATGCCCATGGACACGAGCGTGTATGAGAGCCCCTACAGCGACCCAGAG
GAGCTCAAGGACAAGAAGCTCTTCCTGAAGCGCGATAACCTCCTCATAGCTGACATTGAACTTG
GCTGCGGCAACTTTGGCTCAGTGCGCCAGGGCGTGTACCGCATGCGCAAGAAGCAGATCGACGT
GGCCATCAAGGTGCTGAAGCAGGGCACGGAGAAGGCAGACACGGAAGAGATGATGCGCGAGGCG
CAGATCATGCACCAGCTGGACAACCCCTACATCGTGCGGCTCATTGGCGTCTGCCAGGCCGAGG
CCCTCATGCTGGTCATGGAGATGGCTGGGGGCGGGCCGCTGCACAAGTTCCTGGTCGGCAAGAG
GGAGGAGATCCCTGTGAGCAATGTGGCCGAGCTGCTGCACCAGGTGTCCATGGGGATGAAGTAC
CTGGAGGAGAAGAACTTTGTGCACCGTGACCTGGCGGCCCGCAACGTCCTGCTGGTTAACCGGC
ACTACGCCAAGATCAGCGACTTTGGCCTCTCCAAAGCACTGGGTGCCGACGACAGCTACTACAC
TGCCCGCTCAGCAGGGAAGTGGCCGCTCAAGTGGTACGCACCCGAATGCATCAACTTCCGCAAG
TTCTCCAGCCGCAGCGATGTCTGGAGCTATGGGGTCACCATGTGGGAGGCCTTGTCCTACGGCC
AGAAGCCCTACAAGAAGATGAAAGGGCCGGAGGTCATGGCCTTCATCGAGCAGGGCAAGCGGAT
GGAATGCCCACCAGAGTGTCCACCCGAACTGTACGCACTCATGAGTGACTGCTGGATCTACAAG
TGGGAGGATCGCCCCGACTTCCTGACCGTGGAGCAGCGCATGCGAGCCTGTTACTACAGCCTGG
CCAGCAAGGTGGAAGGGCCCCCAGGCAGCACACAGAAGGCTGAGGCTGCCTGTGCCTGAGCTCC
CGCTGCCCAGGGGAGCCCTCCACACCGGCTCTTCCCCACCCTCAGCCCCACCCCAGGTCCTGCA
GTCTGGCTGAGCCCTGCTTGGTTGTCTCCACACACAGCTGGGCTGTGGTAGGGGGTGTCTCAGG
CCACACCGGCCTTGCATTGCCTGCCTGGCCCCCTGTCCTCTCTGGCTGGGGAGCAGGGAGGTCC
GGGAGGGTGCGGCTGTGCAGCCTGTCCTGGGCTGGTGGCTCCCGGAGGGCCCTGAGCTGAGGGC
ATTGCTTACACGGATGCCTTCCCCTGGGCCCTGACATTGGAGCCTGGGCATCCTCAGGTGGTCA
GGCGTAGATCACCAGAATAAACCCAGCTTCCCTCTTGAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

[0444]

[0445] 본 명세서에서 변수의 임의의 정의에서 화학 그룹의 목록의 인용은 임의의 단일 그룹 또는 열거된 그룹의 조합으로서 해당 변수의 정의를 포함한다. 본 명세서에서 변수 또는 양태에 대한 실시형태의 언급은 임의의 단일 실시형태로 또는 임의의 다른 실시형태 또는 이들의 일부와의 조합으로 해당 실시형태를 포함한다.

[0446] 본 명세서에 제공된 임의의 조성물 또는 방법은 본 명세서에 제공된 임의의 기타 조성물 및 방법 중 하나 이상과 조합될 수 있다.

[0447] DNA 편집은 유전자 수준에서 병원성 돌연변이를 수정함으로써 질환 상태를 변경하는 실행 가능한 수단으로 등장하였다. 최근까지, 모든 DNA 편집 플랫폼은 지정된 게놈 부위에서 DNA 이중 가닥 파손(double strand break: DSB)을 유도하고 내인성 DNA 복구 경로에 의존하여 반확률적 방식으로 제품 결과를 결정함으로써, 복잡한 유전자 산물의 집단을 생성하는 기능을 하였다. 정확한 사용자-정의된 복구 결과가 상동성 유도 복구(homology

directed repair: HDR) 경로를 통해 달성될 수 있지만, 치료적으로 관련된 세포 유형에서 HDR을 사용한 고효율 복구를 방해하는 여러 가지 문제가 있다. 실제로, 이 경로는 경쟁하는 오류가 발생하기 쉬운 비-상동 말단 연결 경로에 비해 비효율적이다. 또한, HDR은 세포 주기의 G1 및 S 단계로 엄격하게 제한되어, 유사분열 후 세포에서 DSB의 정확한 복구를 방지한다. 그 결과, 이들 집단에서 고효율로 사용자-정의된 프로그램 가능한 방식으로 게놈 서열을 변경하는 것이 어렵거나 또는 불가능한 것으로 입증되었다.

[0448] **참조에 의한 원용**

[0449] 본 명세서에서 언급된 모든 간행물, 특허 및 특허 출원은 각각의 개별 간행물, 특허 또는 특허 출원이 참조에 의해 원용되는 것으로 구체적이고 개별적으로 표시된 것과 동일한 정도로 참조에 의해 본 명세서에 원용된다.

## *도면의 간단한 설명*

[0450] 본 발명의 신규한 특징은 첨부된 청구범위에 구체적으로 제시되어 있다. 본 발명의 특징 및 이점에 대한 더 나은 이해는 본 발명의 원리가 이용되는 예시적인 실시형태를 제시하는 다음의 상세한 설명 및 첨부 도면을 참조하여 얻어질 것이다:

도 1a 내지 도 1b는 T 세포 기능에 영향을 미치는 세 가지 단백질의 예시이다. 도 1a는 이식편 대 숙주 질환에서 핵심 성분인 TRAC 단백질의 예시이다. 도 1b는 숙주의 CD8+ T 세포에 의해 인식될 수 있는 유핵 세포에 존재하는 MHC 클래스 1 항원 제시 복합체의 성분인 B2M 단백질의 예시이다. 도 1c는 PDCD1 유전자의 발현을 유도하는 T 세포 신호전달의 예시이며, 생성된 PD-1 단백질은 T 세포 신호전달을 저해하는 작용을 한다.

도 2는 염기 편집 후 표적 유전자의 발현이 넉다운된 세포의 백분율의 그래프이다. "EP"는 전기천공법을 나타낸다.

도 3은 형질도입되지 않은 세포 또는 BE4 염기 편집 시스템 또는 Cas9 뉴클레이스로 형질도입된 세포에서 관찰된 유전자 변형의 유형의 백분율의 그래프이다.

도 4는 BE4를 스플라이싱 부위 수용체(SA) 또는 공여체(SD)로 지시하는 BE4 및 sgRNA로 형질도입된 세포에서 유세포 분석(FC)에 의해 결정된 바와 같이 표적 단백질 발현에 대해 음성이거나 또는 종결 코돈을 생성하는 세포의 백분율에 의해 측정된 바와 같은 표적 뉴클레오타이드 변형 백분율을 나타내는 그래프이다. 대조군 세포는 모의 전기천공(EP)되었다.

도 5는 스플라이싱 부위 수용체(SA), 스플라이스 공여체(SD)를 파괴하거나 또는 종결 코돈을 생성하는 BE4 시스템의 다이어그램이다.

도 6은 표적 유전자를 파괴하기 위해 사용된 sgRNA의 표적외 결합 부위를 요약하는 차트이다.

도 7은 감소된 단백질 발현을 나타내는 BE4 또는 Cas9로 편집된 세포의 백분율의 유세포 분석(FC) 데이터를 요약하는 그래프이다. 세포는 B2M 또는 CD3에 게이팅되었으며, 후자는 TRAC 발현에 대한 프록시이다.

도 8a는 편집되지 않은 대조군 세포의 FACS 데이터의 산점도이다. 도 8b는 B2M, TRAC 및 PD1 유전자좌에서 편집된 세포의 FACS 데이터의 산점도이다.

도 9는 CAR-T 면역요법에 부정적인 영향을 미칠 수 있는 특정 유전자를 변형시키기 위한 본 명세서에 기재된 염기 편집 기법의 효과를 나타내는 그래프이다.

도 10은 유전자 변형 및 전위를 검출하고 정량화하기 위한 액적 디지털 PCR(droplet digital PCR: ddPCR) 프로토콜을 도시하는 다이어그램이다.

도 11은 BE4 시스템 또는 Cas9 시스템을 사용하여 편집된 세포의 차세대 시퀀싱(NGS) 분석 또는 ddPCR로부터 생성된 데이터를 보여주는 2개의 그래프를 나타낸다.

도 12는 Cbl-b가 T 세포 활성화를 억제하는 역할을 예시하는 개략도이다.

도 13은 스플라이싱 부위의 파괴에 의한 Cbl-b 넉다운의 효율성을 도시하는 그래프이다. SA = 스플라이스 수용체; SD = 스플라이스 공여체; STOP - 생성된 종결 코돈; 2° 단독 = 2차 항체 단독; C373은 기능 변이체(C373R)의 상실을 지칭함; RL1-A::APC-A = 레이저; ICS = 세포내 염색.

도 14는 T 세포의 GRIN2B 및 DNMT1 유전자에서 Cas12b-매개성 indel의 비율을 나타내는 그래프이다. EP는 전기천공법을 나타낸다.

도 15는 bvCas12b 및 TRAC, GRIN2B 및 DNMT1에 특이적이고 CD3에 대해 게이팅된 가이드 RNA로 전기천공법(EP)에 의해 형질도입된 세포의 형광 보조 세포 분류(FACS) 데이터를 요약하는 그래프이다.

도 16은 CAR 발현을 나타내는 CAR-P2A-mCherry 렌티바이러스가 형질도입된 세포의 형광 보조 세포 분류 데이터의 산점도이다.

도 17은 폴리(1,8-옥테인다이올 시트레이트)(POC) 렌티바이러스 벡터로 형질도입된 세포에서 CAR 발현을 보여주는 형광 보조 세포 분류 데이터의 산점도이다.

도 18은 BE4가 높은 생성물 순도로 효율적이고, 지속 가능한 유전자 넉아웃을 생성하였음을 보여주는 그래프이다.

도 19a는 BE4 또는 spCas9에 의한 유전자 넉아웃으로 인해 단백질의 표면 발현의 상실을 보여주는 대표적인 FACS 분석이다. 도 19b는 BE4 또는 spCas9에 의한 유전자 넉아웃이 B2M 표면 발현의 상실을 일으킴을 보여주는 그래프이다.

도 20은 B2M, TRAC 및 PD-1 표적 부위의 위치를 나타내는 개략도이다. 전위는 B2M, TRAC 및 PD-1 서열이 재조합될 때 검출될 수 있다.

도 21은 다중 염기 편집이 세포 확장을 크게 손상시키지 않음을 보여주는 그래프이다.

도 22는 BE4가 spCas9와 유사한 표적에 대한 편집 효율 및 세포 표현형을 갖는 삼중-편집된 T 세포를 생성하였음을 나타내는 그래프이다.

도 23은 삼중-편집된 $CD3^-$, $B2M^-$, $PD1^-$ T 세포의 생성을 보여주는 유세포 분석을 도시한다.

도 24는 BE4 및 Cas9 편집된 세포에서 CAR 발현을 보여주는 유세포 분석을 도시한다.

도 25는 CAR-T 세포 살해 또는 항원 양성 세포를 보여주는 그래프이다.

도 26은 Cas12b 및 BE4가 T 세포에서 효율적인 다중 편집을 위해 쌍을 이룰 수 있음을 보여주는 그래프이다.

도 27은 Cas12 뉴클레이스 및 유전자좌를 표적화하는 sgRNA의 존재하에 CAR을 암호화하는 이중-가닥 DNA 주형을 세포 내로 도입함으로써 Cas12b가 키메라 항원 수용체(CAR)의 유전자좌 내로의 삽입을 지시할 수 있음을 보여주는 그래프이다.

도 28a 내지 도 28b는 CAR-T 세포를 조작하기 위한 유전자를 침묵시키는 염기 편집 접근법을 보여주는 묘사이다. 도 28a는 CAR-T 편집을 위한 표적을 예시한다. 도 28b는 염기 편집기를 사용한 침묵을 위한 두 가지 전략: CBE로 종결 코돈의 생성 및 CBE를 이용한 스플라이스 파괴를 보여준다.

도 29는 AML에서 질환의 클론성을 처리하는데 다수의 CAR-T가 필요한 예시적인 만화 표현이다.

도 30은 T-ALL CAR-T 조합에 대한 동족 살해를 제거하기 위해 다중 편집이 필요할 수 있는 예시적인 만화 표현이다.

도 31a 내지 도 31b는 T-ALL에 대한 4개의 동시 염기 편집을 테스트한 실험 결과를 보여주며; CAR-T는 뉴클레이스와 비교하여 수율에 영향을 미치지 않는다. 도 31a는 전기천공법 사용하지 않음(EP 사용하지 않음; 어두운 원), EP 단독(어두운 사각형), CBE 변이체 1(밝은 똑바른 삼각형), CBE 변이체 2(밝은 역삼각형) 및 Cas9(밝은 다이아몬드)에 대한 쿼드 T-ALL 편집의 경합에서 이론적인 수율을 보여준다. 도 31b는 전기천공법 후 세포의 생존능력을 보여준다. 24시간(hr), 48시간, 72시간, 96시간 및 168시간에서 각 결과 세트에 대해, 결과는 오른쪽에서 왼쪽으로 EP 사용하지 않음, EP 단독, CBE 변이체 1, CBE 변이체 2 및 Cas9로 표시된다.

도 32는 BE4에 의한 90% 초과의 쿼드 넉아웃을 보여준다. 전기천공법은 5M 세포 규모에서 규모에서 수행되었고, 쿼드 편집은 단일 전기천공법(EP) 단계에서 수행되었다. 90% 초과의 편집 효율이 rBE4를 사용하여 4개의 표적 모두에 대해 달성되었다. 각 결과 그룹에 대해, 데이터는 왼쪽에서 오른쪽으로: PD1, CD7, TRAC 및 CD52이다.

도 33은 염기 편집이 세포 수율에 차이를 일으키지 않음을 보여준다. 데이터는 EP를 사용하지 않음(원), EP 단독(사각형), rBE4(삼각형), ppBE4(역삼각형) 및 Cas9(다이아몬드)이다.

도 34는 편집된 CAR-T가 모델 종양 세포 상의 CD3 및 CD7을 표적화함을 보여준다. 데이터는 UTD 1:1(원), 7CAR8 1:1(사각형) 및 3CAR2 1:1(삼각형)이다.

도 35는 T-ALL 종양 세포를 표적화하기 위한 예시적인 CD7 CAR-T 세포의 개략도이다.

도 36은 CD7 CAR-T 세포를 갖는 환자를 치료하기 위한 임상 프로토콜을 도시하는 흐름도이다.

도 37a 내지 도 37c는 CD7 CAR-T 세포 생산을 도시한다. 도 37a는 TALL017 이종 해동된 CD7 CAR-T 세포를 생산하기 위한 프로토콜을 도시하는 흐름도이다. 도 37b는 TALL083 CD7 CAR-T 세포를 생산하기 위한 프로토콜을 도시하는 흐름도이다. 도 37c는 TALL017 CAR-T 세포가 해동 후 고도로 활성화됨을 보여주는 형광 보조 세포 분류 데이터의 산점도이다.

도 38은 차세대 시퀀싱(NGS)에 의한 TALL017 및 TALL038 CD7 CAR-T 세포에서의 총 편집을 도시하는 그래프이다.

도 39a 내지 도 39b는 TALL017 및 TALL038 CD7 CAR-T 세포에서 TCRα/β, CD7 및 CD52의 해동 후 24시간의 단백질 발현을 도시한다. 도 39a는 TALL017 및 TALL038 CD7 CAR-T 세포에 대한 형광 보조 세포 분류 데이터의 산점도이다. 도 39b는 FACS를 통한 잔류 단백질 발현을 도시하는 그래프이다.

도 40은 정체 패널에 대한 게이팅 방식을 도시한다.

도 41은 CD2+/- 및 CD56+/- 세포의 혼합물로서 최종 세포 생성물의 정체를 도시하는 그래프이다.

도 42는 해동 후 24시간 및 48시간의 CD7 CAR-T 세포 발현을 도시하는 그래프이다.

도 43은 TALL017 CD7 CAR-T 세포와 비교하여 TALL038 CD7 CAR-T 세포에서 해동 후 CD25 발현이 더 낮음을 도시하는 그래프를 포함한다.

도 44는 시험관내 특성화를 위한 CD7 CAR-T 비드 기반 효능 프로토콜을 도시하는 흐름도이다.

도 45a 내지 도 45b는 TALL038 CD7 CAR-T 세포가 CD7 항원에 대한 반응과 비교하여 IFNɣ, TNFα 및 IL-2를 방출함을 도시한다. 도 45a는 TALL038 CD7 CAR-T 세포에 의한 IFNɣ의 방출을 도시하는 그래프이다. 도 45b는 TALL038 CD7 CAR-T 세포에 의한 TNFα, IL-10 및 IL-2의 방출을 도시하는 그래프이다.

도 46은 종양 세포의 CAR 지시된 T-세포 살해의 측정을 위한 공동-배양 프로토콜을 도시하는 개략도이다.

도 47a 내지 도 47b는 TALL038 CD7 CAR-T 세포가 TALL017 CAR T 세포와 비교하여 재처리될 때 증가된 CCRF를 나타냄을 도시한다. 도 47a는 1차 처리를 도시하는 그래프이다. 도 47b는 2차 처리를 도시하는 그래프이다.

도 48은 CCRF 이식 후 제10일/-1일부터 CD7 CAR-T 처리까지의 생물발광 광 데이터(평균, SEM)를 도시하는 그래프이다. 총 플럭스는 선형 척도로 측정된다.

도 49a 내지 도 49b는 CCRF 이식 후 제19일/CD7 CAR-T 처리 후 제8일에서의 생물발광 광 데이터(평균, SEM)를 도시한다. 도 49a는 평균 종양 부담을 도시하는 그래프이다. 총 플럭스는 선형 척도로 측정된다. 도 49b는 평균 마우스 체중을 도시하는 그래프이다.

도 50a 내지 도 50b는 CCRF 이식 후 제38일/CD7 CAR-T 처리 후 제27일에서의 생물발광 광 데이터(평균, SEM)를 도시한다. 도 50a는 평균 종양 부담을 도시하는 그래프이다. 총 플럭스는 선형 척도로 측정된다. 도 50b는 평균 종양 부담을 도시하는 그래프이다. 총 플럭스는 로그 척도로 측정된다.

도 51a 내지 도 51b는 CCRF 이식 후 제38일/CD7 CAR-T 처리 후 제27일에서의 생물발광 광 데이터(개별 마우스)를 도시한다. 도 51a는 평균 종양 부담을 도시하는 그래프이다. 총 플럭스는 선형 척도로 측정된다. 도 51b는 평균 종양 부담을 도시하는 그래프이다. 총 플럭스는 로그 척도로 측정된다.

도 52a 내지 도 52b는 CCRF 이식 후 제38일/CD7 CAR-T 처리 후 제27일에서의 생물발광 광 데이터(개별 마우스)를 도시한다. 도 52a는 평균 종양 부담을 도시하는 그래프를 포함한다. 총 플럭스는 선형 척도(위) 및 로그 척도(아래)로 측정된다. 도 52b는 평균 종양 부담을 도시하는 그래프를 포함한다. 총 플럭스는 선형 척도(위) 및 로그 척도(아래)로 측정된다.

도 53은 CD5 gRNA 후보물질의 편집 효율을 도시하는 그래프이다.

도 54는 BE4와 조합된 CD5 gRNA 후보물질 g103 및 g104의 NGS를 통한 편집 효율을 도시하는 그래프이다.

도 55는 예시적인 CD5 CAR 작제물의 생산을 도시하는 개략도이다.

도 56은 CD5 gRNA 후보물질 g103 및 g104를 사용한 렌티바이러스 벡터(LVV) 스크리닝으로부터의 CAR 발현을 도시하는 그래프이다.

도 57은 힌지 서열에 기초한 CAR LVV 발현을 도시하는 그래프이다.

도 58a 내지 도 58d는 라이브 이미징 세포 살해 검정에서 편집 및 편집되지 않은 작제물이 CD5+ CCRF에 대해 완전히 세포독성임을 도시하는 그래프이다.

도 59a 내지 도 59b는 CD5+ CCRF-CEM 백혈병 세포주의 존재하에 sgRNA 103 또는 sgRNA 104를 사용한 CD5 CAR LVV 작제물에 의한 사이토카인(IFNɣ)의 생산을 도시한다. 도 59(왼쪽)는 CD5 CAR LVV(LV63-69) 형질도입된 T 세포 단독 또는 sgRNA 103을 사용하여 편집된 CCRF 세포를 사용한 IFNɣ 생산을 도시하는 그래프이다. 도 59(오른쪽)는 CD5 CAR LVV(LV63-69) 형질도입된 T 세포 단독 또는 sgRNA 104를 사용하여 편집된 CCRF 세포를 사용한 IFNɣ 생산을 도시하는 그래프이다. 도 59(아래)는 편집되지 않은 CD5 CAR LVV(LV63-69) 형질도입된 T 세포 또는 CCRF 세포를 사용한 IFNɣ 생산을 도시하는 그래프이다. 도 59b(위)는 sgRNA 103 또는 104를 사용하여 편집되지 않거나 또는 편집된 CD5 CAR LVV(LV63-69) 형질도입된 T 세포 단독을 사용한 IFNɣ 생산을 도시하는 그래프이다. 도 59b(아래)는 sgRNA 103 또는 104를 사용하여 편집되지 않거나 또는 편집된 CD5 CAR LVV(LV63-69) 형질도입된 CCRF 세포를 사용한 IFNɣ 생산을 도시하는 그래프이다.

## *발명을 실시하기 위한 구체적인 내용*

[0451] 본 발명은 향상된 항-신생물 활성, 면역 억제에 대한 저항성을 갖고 이식편 대 숙주 반응 또는 숙주 대 이식편 반응 또는 이들의 조합을 유발할 위험이 감소된 유전적으로 변형된 면역 세포를 특징으로 한다. 본 발명은 또한 이러한 변형된 면역 세포(예를 들어, 면역 효과기 세포, 예컨대, T 세포)를 생산하고 사용하는 방법을 특징으로 한다.

[0452] 일 실시형태에서, 이식편 대 숙주 질환(GVHD)을 갖거나 또는 이를 발병하는 경향을 갖는 대상체에는 기능적 TRAC가 결핍되거나 또는 이의 수준이 감소된 CAR-T 세포가 투여된다. 일 실시형태에서, 숙주 대 이식편 질환(HVGD)을 갖거나 또는 이를 발병하는 경향을 갖는 대상체에는 기능적 베타2 마이크로글로불린(B2M)이 결핍되거나 또는 이의 수준이 감소된 CAR-T 세포가 투여된다.

[0453] 키메라 항원 수용체를 발현하고 이들의 발현이 면역 세포 기능에 미칠 수 있는 부정적인 영향을 감소시키기 위해 특정 유전자를 넉아웃 또는 넉다운시키기 위한 면역 효과기 세포의 변형은 본 명세서에 기재된 바와 같은 사이티딘 데아미네이스 또는 아데노신 데아미네이스를 포함하는 염기 편집 시스템을 사용하여 달성된다.

[0454] 자가, 환자-유래 키메라 항원 수용체-T 세포(CAR-T) 요법은 일부 혈액암을 치료하는데 있어서 놀라운 효능을 보여주었다. 이러한 제품은 환자에게 상당한 임상적 이점을 제공하였지만, 개별화된 요법을 만들어내야 하는 필요성은 상당한 제조 문제와 재정적 부담을 야기한다. 동종이계 CAR-T 요법은 자가 제품과 유사한 임상 효능 프로파일을 갖는 동시에 건강한 단일 공여체로부터 유래되는 세포로 많은 환자를 치료함으로써 이러한 문제에 대한 잠재적인 해결책으로서 개발되어 상품 비용 및 로트간 변동성을 크게 줄였다.

[0455] 대부분의 1세대 동종이계 CAR-T는 표적 T 세포 집단에 2개 이상의 표적화된 게놈 DNA 이중 가닥 파손(DSB)을 도입하기 위해 뉴클레이스를 사용하고, 오류가 발생하기 쉬운 DNA 복구에 의존하여 반-확률론적 방식으로 표적 유전자를 넉아웃시키는 돌연변이를 생성한다. 이러한 뉴클레이스-기반 유전자 넉아웃 전략은 이식편-대-숙주-질환 및 CAR-T의 숙주 거부의 위험을 줄이는 것을 목표로 한다. 그러나, 다중 DSB의 동시 유도는 균형 및 불균형 전위와 같은 대규모 게놈 재배열을 포함하는 최종 세포 생성물과 역위 및 대규모 결실을 포함하는 국부 재배열이 비교적 풍부하게 생성된다. 또한, 유도된 DSB에 의해 동시 유전적 변형의 수가 증가함에 따라, 처리된 세포 집단에서 상당한 유전독성이 관찰되었다. 이는 각 제조 실행에서 세포 확장 가능성을 상당히 감소시켜 건강한 공여체당 치료될 수 있는 환자의 수를 감소시킬 가능성이 있다.

[0456] 염기 편집기(BE)는 DSB를 생성하지 않고도 표적 게놈 DNA의 매우 효율적인 사용자-정의된 변형을 가능하게 할 수 있는 새로운 유전자 편집 시약의 한 종류이다. 본 명세서에서, 염기 편집 기술을 사용하여 검출 가능한 게놈 재배열을 줄이거나 또는 제거하는 동시에 또한 세포 확장을 개선함으로써 동종이계 CAR-T 세포를 생산하는 대체 수단이 제안된다. 본 명세서에 나타낸 바와 같이, 뉴클레이스-단독 편집 전략과 대조적으로, 염기 편집에 의한 하나 이상, 예를 들어, 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 이상의 유전자 좌위의 동시 변형은 검출 가능한 전위 이벤트 없이 고효율 유전자 넉아웃을 생성한다.

[0457] 일부 실시형태에서, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 본 명세서에 제공된 염기 편집 조성물 및 방법으로 면역 세포에서 변형된다. 일부 실시형태에서, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 ACAT1, ACLY, ADORA2A, AXL, B2M, BATF, BCL2L11, BTLA, CAMK2D, cAMP, CASP8, CBLB, CCR5, CD2, CD3D, CD3E,

CD3G, CD4, CD5, CD7, CD8A, CD33, CD38, CD52, CD70, CD82, CD86, CD96, CD123, CD160, CD244, CD276, CDK8, CDKN1B, Chi3l1, CIITA, CISH, CSF2CSK, CTLA-4, CUL3, Cyp11a1, DCK, DGKA, DGKZ, DHX37, ELOB(TCEB2), ENTPD1(CD39), FADD, FAS, GATA3, IL6, IL6R, IL10, IL10RA, IRF4, IRF8, JUNB, Lag3, LAIR-1(CD305), LDHA, LIF, LYN, MAP4K4, MAPK14, MCJ, MEF2D, MGAT5, NR4A1, NR4A2, NR4A3, NT5E(CD73), ODC1, OTULINL(FAM105A), PAG1, PDCD1, PDIA3, PHD1(EGLN2), PHD2(EGLN1), PHD3(EGLN3), PIK3CD, PIKFYVE, PPARa, PPARd, PRDMI1, PRKACA, PTEN, PTPN2, PTPN6, PTPN11, PVRIG(CD112R), RASA2, RFXANK, SELPG/PSGL1, SIGLEC15, SLA, SLAMF7, SOCS1, Spry1, Spry2, STK4, SUV39, H1TET2, TGFbRII, TIGIT, Tim-3, TMEM222, TNFAIP3, TNFRSF8(CD30), TNFRSF10B, TOX, TOX2, TRAC, TRBC1, TRBC2, UBASH3A, VHL, VISTA, XBP1, YAP1 및 ZC3H12A로부터 선택된다. 일부 실시형태에서, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 CD3, CD5, CD7, CD33, CD123, TRAC, LAG-3, FAS, CD52, TRBC1, TRBC2, B2M 및 CIITA 및 PD-1로부터 선택된다. 일부 실시형태에서, 변형된 면역 세포는 CD5에 변형을 포함하고, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 TRAC, LAG-3, FAS, CD52, TRBC1, TRBC2, B2M 및 CIITA 및 PD-1로부터 선택된다. 일부 실시형태에서, 변형된 면역 세포는 CD7에 변형을 포함하고, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 TRAC, LAG-3, FAS, CD52, TRBC1, TRBC2, B2M 및 CIITA 및 PD-1로부터 선택된다. 일부 실시형태에서, 변형된 면역 세포는 CD33에 변형을 포함하고, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 TRAC, LAG-3, FAS, CD52, TRBC1, TRBC2, B2M 및 CIITA 및 PD-1로부터 선택된다. 일부 실시형태에서, 변형된 면역 세포는 CD3에 변형을 포함하고, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 TRAC, LAG-3, FAS, CD52, TRBC1, TRBC2, B2M 및 CIITA 및 PD-1로부터 선택된다. 일부 실시형태에서, 변형된 면역 세포는 CD123에 변형을 포함하고, 적어도 하나 이상의 유전자 또는 이의 조절 요소는 TRAC, LAG-3, FAS, CD52, TRBC1, TRBC2, B2M 및 CIITA 및 PD-1로부터 선택된다. 유전자의 다중 편집은 개선된 치료적 특성을 갖는 CAR-T 세포 요법의 생성에 유용할 수 있다. 이 방법은 다중-편집된 T 세포 생성물의 알려진 한계를 해결하며, 차세대 정밀 세포-기반 요법을 향한 유망한 발전이다.

[0458] 일 양태에서, 범용 CAR-T 세포가 본 명세서에 제공된다. 일부 실시형태에서, 본 명세서에 기재된 CAR-T 세포는 동종이계 세포이다. 일부 실시형태에서, 범용 CAR-T 세포는 목적하는 CAR을 발현하는데 사용될 수 있으며, 공여체 및 수용체의 면역원성 적합성과 무관하게 보편적으로 적용될 수 있는 동종이계 T 세포이다. 동종이계 면역 세포는 하나 이상의 공여체로부터 유래될 수 있다. 소정의 실시형태에서, 동종이계 면역 세포는 단일 인간 공여체로부터 유래된다. 예를 들어, 동종이계 T 세포는 단일 건강한 인간 공여체의 PBMC로부터 유래될 수 있다. 소정의 실시형태에서, 동종이계 면역 세포는 여러 인간 공여체로부터 유래된다. 일부 실시형태에서, 범용 CAR-T 세포는 다중 유전자 좌위, 예를 들어, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 이상의 유전자 좌위에서 동시 편집을 도입하기 위해 유전자 변형을 사용함으로써 본 명세서에 기재된 바와 같이 생성될 수 있다. 본 명세서에 기재된 바와 같은 변형 또는 동시 변형은 염기 편집기에 의해 생성된 염기 편집과 같은 유전자 편집일 수 있다. 염기 편집기는 C 염기 편집기 또는 염기 편집기일 수 있다. 본 명세서에서 논의된 바와 같이, 염기 편집은 유전자가 발현되지 않도록 유전자 파괴를 달성하는데 사용될 수 있다. 염기 편집에 의한 변형은 유전자 발현의 감소를 달성하기 위해 사용될 수 있다. 일부 실시형태에서, 염기 편집기는 편집된 유전자가 구조적으로 또는 기능적으로 생존 가능한 단백질 생성물을 생성하지 않도록 하는 유전자 변형을 도입하는데 사용될 수 있다. 일부 실시형태에서, 본 명세서에 기재된 동시 변형과 같은 변형은 유전자 산물의 발현 또는 기능이 임의의 방식으로 변경되도록 하는 염기 편집과 같은 유전자 편집을 포함할 수 있다. 예를 들어, 유전자 산물의 발현은 기준치 발현 수준과 비교하여 향상되거나 또는 상향조절될 수 있다. 일부 실시형태에서, 유전자 산물의 활성 또는 기능은 염기 편집 또는 협력하여 작용하는 다중 염기 편집 이벤트의 결과로서 상향조절될 수 있다.

[0459] 일부 실시형태에서, 범용 CAR-T 세포의 생성은 긴급한 사용을 위해 생성하기 어려울 수 있는 자가 T 세포(CAR-T)에 비해 유리할 수 있다. 자가 T 세포를 조작하여 CAR을 발현시키고 최종적으로 의료 응급시 이식을 위해 목적하는 세포 생성물을 얻는 불확실성과 관련된 여러 상황에서 자가 세포 준비보다 동종이계 접근법이 선호된다. 그러나, 동종이계 T 세포 또는 "상용(off-the-shelf)" T 세포의 경우, CAR-T 세포에 대한 숙주의 반응성(HVGD)뿐만 아니라 숙주 세포에 대한 동종이계 T 세포의 잠재적 적대성(GVHD)을 신중하게 대처하는 것이 중요하다. 시나리오가 주어지면, (a) 동족 살해 저항성 면역 세포를 생성하기 위해 항원의 발현을 감소시키거나 또는 하향조절하는 것이 가능하도록; (b) 내인성 T 세포 수용체, 예를 들어, TCR 알파 사슬(예컨대, TRAC의 염기 편집을 통해서) 또는 TCR 베타 사슬(예컨대, TRBC1/TRBC2의 염기 편집을 통해서)가 없거나 또는 이를 적은 양으로 발현하는 플랫폼 세포 유형을 생성하는 것이 가능하도록; 그리고/또는 (c) 숙주 조직 시스템과 양립할 수 없는 항원의 발현을 감소시키거나 또는 하향 조절하는 것 및 그 반대의 경우도 가능하도록, 염기 편집을 성공적으로 사용하여 다중 동시 유전자 편집 이벤트를 생성할 수 있다.

[0460] 일부 실시형태에서, 본 명세서에 기재된 방법은 CAR-T를 발현하는 자가 T 세포를 생성하는데 사용될 수 있다.

일부 실시형태에서, 다중 염기 편집 이벤트는 단일 전기천공법 이벤트에서 달성될 수 있으며, 이에 의해 전기천공법 이벤트 관련 독성이 감소될 수 있다. 외인성 유전 물질을 세포 내로 혼입시키기 위한 임의의 공지된 방법을 사용하여 전기천공법을 대체할 수 있고, 이로써 당업계에 공지된 이러한 방법은 임의의 본 명세서에 기재된 방법에서 사용하기 위해 상정된다.

[0461] 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에게 CD2의 수준이 부족하거나 또는 감소되고, CD2 공동-자극 도메인을 포함하는 CD2 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 대상체에게 투여된 CD2 변형된 면역 세포는 본 명세서에 제공된 염기 편집 조성물 및 방법으로 하나 이상의 유전자 또는 조절 요소(예를 들어, *CD52*, *TRAC*, *PD-1*)에서 추가로 변형된다.

[0462] 본 명세서에 나타낸 바와 같이, CAR 삽입과 조합된 염기 편집은 최소한의 게놈 재배열로 동족 살해 저항성 동종이계 T 세포를 생성하기 위한 유용한 전략이다. 유전자의 다중 편집은 또한 개선된 치료적 특성을 갖는 CAR-T 세포 요법의 생성에 유용할 수 있다. 이 방법은 CAR-T 요법의 알려진 한계를 해결하고, 차세대 정밀 세포 기반 요법을 향한 유망한 발전이다.

[0463] 한 실험에서, 염기 편집기 BE4는 T 세포(TRAC, B2M 및 PD-1)에서 3개의 세포 표면 표적에 대한 고효율 다중 염기 편집을 보였으며, 단일 전기천공법에서 유전자 발현을 각각 95%, 95% 및 88%로 넉아웃시켜 B2M 및 CD3의 단백질 발현이 감소된 세포의 비율이 높은 세포 집단을 생성하였다. 이들 유전자 각각을 편집하는 것은 개선된 치료적 특성을 갖는 CAR-T 세포 요법의 생성에 유용할 수 있다. 각각의 유전자는 이중 가닥 파손의 생성 없이 단일 표적화된 염기 변화(C에서 T로)에 의해 침묵화되었다. 그 결과, BE4-처리된 세포는 또한 임의의 측정 가능한 전위(대규모 게놈 재배열)를 나타내지 않은 반면, 뉴클레이스를 이용한 동일한 세 가지 편집을 받은 세포는 검출 가능한 게놈 재배열을 보여주었다.

[0464] 따라서, T*RAC* 유전자의 뉴클레이스-기반 넉아웃과 2개 추가적인 유전자 동시 BE-매개성 넉아웃을 결합하면 최소한의 게놈 재배열을 갖는 균질한 동종이계 T 세포 집단이 생성되어, T*RAC* 유전자좌에 CAR 이식유전자의 표적화된 삽입이 가능하다. 종합하면, 이는 단독으로 또는 단일 뉴클레이스 넉아웃과 CAR 삽입을 조합한 염기 편집이 뉴클레이스-단독 접근법과 비교하여 최소한의 게놈 재배열을 갖는 동종이계 T 세포를 생성하기 위한 유용한 전략임을 입증한다. 이러한 방법은 다중-편집된 T 세포 생성물의 알려진 한계를 해결하며, 차세대 정밀 세포 기반 요법을 향한 유망한 발전이다.

[0465] **키메라 항원 수용체 및 CAR-T 세포**

[0466] 본 발명은 키메라 항원 수용체(CAR)를 발현하는, 본 명세서에 기재된 핵염기 편집기를 사용하여 변형된 면역 세포를 제공한다. 키메라 항원 수용체를 발현하도록 하는 면역 세포의 변형은 면역 세포의 면역반응 활성을 향상시킬 수 있되, 키메라 항원 수용체는 항원의 에피토프에 대한 친화성을 갖고, 항원은 유기체의 변경된 적합성(fitness)과 관련된다. 예를 들어, 키메라 항원 수용체는 신생물 세포에서 발현되는 단백질 상의 에피토프에 대해 친화성을 가질 수 있다. CAR-T 세포는 주요 조직적합성 복합체(MHC)와 독립적으로 작용할 수 있기 때문에, 활성화된 CAR-T 세포는 항원을 발현하는 신생물 세포를 죽일 수 있다. CAR-T 세포의 직접적인 작용은 MHC가 면역 세포에 항원을 제시하는 것에 반응하여 진화된 신생물 세포 방어 메커니즘을 회피한다.

[0467] 일부 실시형태에서, 본 발명은 자가면역 반응에 관여하는 B 세포(예를 들어, 대상체 자신의 조직에 대해 생성된 항체를 발현하는 대상체의 B 세포)를 표적으로 하는 키메라 항원 수용체를 발현하는 면역 효과기 세포를 제공한다.

[0468] 일부 실시형태는 자가 면역 세포 면역요법을 포함하되, 면역 세포는 표면 마커를 발현하는 암성 또는 그렇지 않으면 변경된 세포를 특징으로 하는 질환 또는 변경된 적합성을 갖는 대상체로부터 얻어진다. 얻어진 면역 세포는 키메라 항원 수용체를 발현하도록 유전적으로 변형되고, 특정 항원에 대해 효과적으로 재지시된다. 따라서, 일부 실시형태에서, 면역 세포는 CAR-T 면역요법을 필요로 하는 대상체로부터 얻어진다. 일부 실시형태에서, 이러한 자가 면역 세포는 대상체로부터 얻어진 직후에 배양되고 변형된다. 다른 실시형태에서, 자가 세포가 얻어진 다음, 향후 사용을 위해 보관된다. 이러한 관행은 향후 면역 세포 수를 감소시킬 병행 치료를 받을 수 있는 개체에게 권장될 수 있다. 동종이계 면역 세포 면역요법에서, 면역 세포는 치료를 받을 대상체가 아닌 다른 공여체로부터 얻어질 수 있다. 일부 실시형태에서, 면역 세포는 건강한 대상체 또는 공여체로부터 얻어지고, 키메라 항원 수용체를 발현하도록 유전적으로 변형되며, 특정 항원에 대해 효과적으로 재지시된다. 면역 세포는, 키메라 항원 수용체를 발현하도록 변형된 후, 신생물(예를 들어, 백혈병)을 치료하기 위해 대상체에게 투여된다.

일부 실시형태에서, 키메라 항원 수용체를 발현하도록 변형되는 면역 세포는 면역 세포의 기존의 스톡 배양물로부터 얻어질 수 있다.

[0469] 면역 세포 및/또는 면역 효과기 세포는 당업계에 공지된 표준 기법을 사용하여 대상체 또는 공여체로부터 수집된 샘플로부터 단리 또는 정제될 수 있다. 예를 들어, 면역 효과기 세포는 적혈구를 용해시키고 원심분리에 의해 말초 단핵 혈액 세포를 제거함으로써 전혈 샘플로부터 단리 또는 정제될 수 있다. 면역 효과기 세포는 CD25, CD3, CD4, CD8, CD28, CD45RA 또는 CD45RO과 같은 세포-특이적 마커를 기반으로 면역 효과기 세포를 단리하는 선택적 정제 방법을 사용하여 추가로 단리 또는 정제될 수 있다. 일 실시형태에서, CD25+는 조절 T 세포를 선별하기 위한 마커로서 사용된다. 일 실시형태에서, $CD4^+$는 T 세포를 선별하기 위한 마커로서 사용된다. 일 실시형태에서, $CD8^+$는 T 세포를 선별하기 위한 마커로서 사용된다. 일 실시형태에서, $CD4^+$ 및 $CD8^+$는 T 세포를 선별하기 위한 마커로서 사용된다. 일 실시형태에서, $CD4^+$ 및 $CD25^+$ T 세포를 선별하기 위한 마커로서 사용된다.

[0470] 또 다른 실시형태에서, 본 발명은 TCRαβ 표면 발현을 담당하는 TCR 불변 영역(TRAC)에 표적화된 유전자 넉아웃을 갖는 T 세포를 제공한다. TCRαβ-결핍 CAR T 세포는 동종이계 면역요법과 양립할 수 있다(문헌[Qasim *et al.*, Sci. Transl. Med. 9, eaaj2013 (2017); Valton *et al.*, Mol Ther. 2015 Sep; 23(9): 1507-1518]). 원하는 경우, 잔여 TCRαβ T 세포는 GVHD의 위험을 최소화하기 위해 CliniMACS 자성 비드 고갈을 사용하여 제거된다. 또 다른 실시형태에서, 본 발명은 생체외에서 선별된 공여체 T 세포를 제공하여 수용체 조혈 세포에서 발현되는 소수의 조직적합성 항원을 인식하고, 이에 의해 이식 후 이환율 및 사망률의 주요 원인인 이식편-대-숙주 질환(GVHD)의 위험을 최소화한다(문헌[Warren *et al.*, Blood 2010;115(19):3869-3878]). 면역 효과기 세포를 단리하거나 또는 정제하기 위한 또 다른 기법은 유세포 분석이다. 형광 활성화된 세포 분류에서, 면역 효과기 세포 마커에 대해 친화성을 갖는 형광 표지된 항체는 샘플의 면역 효과기 세포를 표지하는데 사용된다. 마커를 발현하는 세포에 적절한 게이팅 전략이 세포를 분리하는데 사용된다. 예를 들어, T 림프구는, 예를 들어, 면역 효과기 세포 마커(예를 들어, CD4, CD8, CD28, CD45)에 특이적인 형광 표지된 항체 및 상응하는 게이팅 전략을 사용함으로써 샘플의 다른 세포로부터 분리될 수 있다. 일 실시형태에서, CD45 게이팅 전략이 사용된다. 일부 실시형태에서, 면역 효과기 세포에 특이적인 다른 마커에 대한 게이팅 전략이 CD45 게이팅 전략 대신에 또는 이와 조합하여 사용된다. 일 실시형태에서, CD4 게이팅 전략이 사용된다. 일 실시형태에서, CD8 게이팅 전략이 사용된다. 일 실시형태에서, CD25 게이팅 전략이 사용된다. 일 실시형태에서, CD4 및 CD8 게이팅 전략이 사용된다. 일 실시형태에서, CD4 및 CD25 게이팅 전략이 사용된다. 일부 실시형태에서, 면역 효과기 세포에 특이적인 다른 마커에 대한 게이팅 전략이 CD4, CD25 및/또는 CD8 게이팅 전략 대신에 또는 이와 조합하여 사용된다. 일부 실시형태에서, 도 40에 제공된 바와 같은 게이팅 전략이 사용된다.

[0471] 본 발명에서 상정되는 면역 효과기 세포는 효과기 T 세포이다. 일부 실시형태에서, 효과기 T 세포는 미경험 $CD8^+$ T 세포, 세포독성 T 세포, 자연 살해 T(NKT) 세포, 자연 살해(NK) 세포 또는 조절 T(Treg) 세포이다. 일부 실시형태에서, 효과기 T 세포는 흉선세포, 미성숙 T 림프구, 성숙 T 림프구, 휴지기 T 림프구 또는 활성화된 T 림프구이다. 일부 실시형태에서, 면역 효과기 세포는 $CD4^+$ $CD8^+$ T 세포 또는 $CD4^-$ $CD8^-$ T 세포이다. 일부 실시형태에서, 면역 효과기 세포는 T 헬퍼 세포이다. 일부 실시형태에서, T 헬퍼 세포는 T 헬퍼 1(Th1), T 헬퍼 2(Th2) 세포 또는 CD4를 발현하는 헬퍼 T 세포(CD4+ T 세포)이다. 일부 실시형태에서, 면역 효과기 세포는 효과기 NK 세포이다. 일부 실시형태에서, 면역 효과기 세포는 T 세포의 임의의 다른 서브세트이다. 변형된 면역 효과기 세포는 키메라 항원 수용체에 더하여, 외인성 사이토카인, 상이한 키메라 수용체 또는 면역 효과기 세포 신호전달 또는 기능을 향상시킬 수 있는 임의의 다른 작용제를 발현할 수 있다. 예를 들어, 키메라 항원 수용체와 사이토카인의 동시-발현은 표적 세포를 용해하는 CAR-T 세포의 능력을 향상시킬 수 있다.

[0472] 본 발명에서 상정된 바와 같은 키메라 항원 수용체는 세포외 결합 도메인, 막관통 도메인 및 세포내 도메인을 포함한다. 세포외 결합 도메인에 대한 항원의 결합은 CAR-T 세포를 활성화하고, CAR-T 세포 증식, 사이토카인 생산 및 항원 발현 세포의 사멸을 초래하는 기타 과정을 포함하는 효과기 반응을 생성할 수 있다. 본 발명의 일부 실시형태에서, 키메라 항원 수용체는 링커를 더 포함한다. 일부 실시형태에서, 링커는 $(GGGGS)_n$ 링커이다. 일부 실시형태에서, 링커는 $(GGGGS)_3$ 링커이다. 일부 실시형태에서, 본 발명의 CAR은 리더 펩타이드 서열(예를 들어, 항원 결합 도메인에 대한 N-말단)을 포함한다. 예시적인 리더 펩타이드 아미노산 서열은 METDTLLLWVLLLWVPGSTG이다.

[0473] 본 명세서에서 상정되는 키메라 항원 수용체의 세포외 결합 도메인은 특정 항원에 대해 친화성을 갖는 항체 또

는 이의 항원 결합 단편 아미노산 서열을 포함한다. 다양한 실시형태에서, CAR은 5T4에 특이적으로 결합한다. 예시적인 항-5T4 CAR은 제한 없이 CART-5T4(옥스포드 바이오메디카 상장법인(Oxford BioMedica plc)) 및 UCART-5T4(셀렉티스 에스에이(Cellectis SA))를 포함한다.

[0474] 다양한 실시형태에서, CAR은 알파-태아단백질에 특이적으로 결합한다. 예시적인 항-알파-태아단백질 CAR은 제한 없이 ET-1402(유레카 테라퓨틱스 주식회사(Eureka Therapeutics Inc))를 포함한다.

[0475] 다양한 실시형태에서, CAR은 Axl에 특이적으로 결합한다. 예시적인 항-Axl CAR은 제한 없이 CCT-301-38(F1 온콜로지 주식회사(F1 Oncology Inc))를 포함한다.

[0476] 다양한 실시형태에서, CAR은 B7H6에 특이적으로 결합한다. 예시적인 항-B7H6 CAR은 제한 없이 CYAD-04(셀야드 에스에이(Celyad SA))를 포함한다.

[0477] 다양한 실시형태에서, CAR은 BCMA에 특이적으로 결합한다. 예시적인 항-BCMA CAR은 제한 없이 ACTR-087 + SEA-BCM(시애틀 제네틱스 주식회사(Seattle Genetics Inc)), ALLO-715(셀렉티스 에스에이), ARI-0002(아우구스트 피 수니어 생의학 연구소(Institut d'Investigacions Biomediques August Pi I Sunyer)), bb-2121(블루버드 바이오 주식회사(블루버드 바이오 Inc)), bb-21217(블루버드 바이오 주식회사), CART-BCM(펜실베니아 대학), CT-053(카스겐 테라퓨틱스 유한 회사(Carsgen Therapeutics Ltd)), Descartes-08(카테이션 테라퓨틱스(Cartesian Therapeutics)), FCARH-143(주노 테라퓨틱스 주식회사(Juno Therapeutics Inc)), ICTCAR-032(이노베이티브 셀 테라퓨틱스 유한 책임회사(Innovative Cellular Therapeutics Co Ltd)), IM21 CART(베이징 이뮤노차이나 메디칼 사이언스 앤 테크놀로지 유한 책임회사(Beijing Immunochina Medical Science & Technology Co Ltd)), JCARH-125(메모리얼 슬론-케터링 암 센터(Memorial Sloan-Kettering 암 Center)), KITE-585(카이트 파르마 주식회사(Kite Pharma Inc)), LCAR-B38M(난징 레전드 바이오테크 유한 책임회사(Nanjing Legend Biotech Co Ltd)), LCAR-B4822M(난징 레전드 바이오테크 유한 책임회사), MCARH-171(메모리얼 슬론-케터링 암 센터), P-BCMA-101(포세이다 테라퓨틱스 주식회사(Poseida Therapeutics Inc)), P-BCMA-ALLO1(포세이다 테라퓨틱스 주식회사), spCART-269(상하이 유니카-테라피 바이오-메디신 테크놀로지 유한 책임회사(Shanghai Unicar-Therapy Bio-medicine Technology Co Ltd)) 및 BCMA02/bb2121(블루버드 바이오 주식회사)을 포함한다. BCMA02/bb2121 CAR의 폴리펩타이드 서열이 아래에 제공된다:

MALPVTALLLPLALLLHAARPDIVLTQSPPSLAMSLGKRATISCRASESVTILGSHLIHW
YQQKPGQPPTLLIQLASNVQTGVPARFSGSGSRTDFTLTIDPVEEDDVAVYYCLQSRTIP
RTFGGGTKLEIKGSTSGSGKPGSGEGSTKGQIQLVQSGPELKKPGETVKISCKASGYTFT
DYSINWVKRAPGKGLKWMGWINTETREPAYAYDFRGRFAFSLETSASTAYLQINNLKYED
TATYFCALDYSYAMDYWGQGTSVTVSSAAATTTPAPRPPTPAPTIASQPLSLRPEACRPA
AGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQT
TQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRR
GRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKD
TYDALHMQALPPR

[0478]

[0479] 다양한 실시형태에서, CAR은 CCK2R에 특이적으로 결합한다. 예시적인 항-CCK2R CAR은 제한 없이 엔도사이트(Endocyte)/퍼듀(Purdue)(퍼듀 대학(Purdue University))의 항-CCK2R CAR-T 어댑터 분자(CAM) + 항-FITC CAR T-세포 요법(암)을 포함한다.

[0480] 다양한 실시형태에서, CAR은 CD 항원에 특이적으로 결합한다. 예시적인 항-CD 항원 CAR은 제한 없이 VM-802(바이로메드 유한 책임회사(ViroMed Co Ltd))를 포함한다.

[0481] 다양한 실시형태에서, CAR은 CD123에 특이적으로 결합한다. 예시적인 항-CD123 CAR은 제한 없이 MB-102(포트레스 바이오테크 주식회사(Fortress Biotech Inc)), RNA CART123(펜실베니아 대학), SFG-iMC-CD123.zet(벨리쿰 파마슈티칼 주식회사(Bellicum Pharmaceuticals Inc)) 및 UCART-123(셀렉티스 에스에이)을 포함한다.

[0482] 다양한 실시형태에서, CAR은 CD133에 특이적으로 결합한다. 예시적인 항-CD133 CAR은 제한 없이 KD-030(난징 카이디 바이오테크 주식회사)을 포함한다.

[0483] 다양한 실시형태에서, CAR은 CD138에 특이적으로 결합한다. 예시적인 항-CD138 CAR은 제한 없이 ATLCAR.CD138(유엔씨 린버거 통합 암 센터(UNC Lineberger Comprehensive Cancer Center)) 및 CART-138(중국 PLA 종합병원(Chinese PL일반적인 Hospital))을 포함한다.

[0484] 다양한 실시형태에서, CAR은 CD171에 특이적으로 결합한다. 예시적인 항-CD171 CAR은 제한 없이 JCAR-023(주노 테라퓨틱스 주식회사)을 포함한다.

[0485] 다양한 실시형태에서, CAR은 CD19에 특이적으로 결합한다. 예시적인 항-CD19 CAR은 제한 없이 1928z-41BBL(메모리얼 슬론-케터링 암 센터), 1928z-E27(메모리얼 슬론-케터링 암 센터), 19-28z-T2(광저우 생명의학 및 건강 연구소(Guangzhou Institutes of Biomedicine and Health)), 4G7-CARD(칼리지 런던 대학(University College London)), 4SCAR19(센젠 제노-이뮨 의학 연구소(Shenzhen Geno-Immune Medical Institute)), ALLO-501(화이자 주식회사(Pfizer Inc)), ATA-190(QIMR 베르그호퍼 의료 연구소(QIMR Berghofer Medical Research Institute)), AUTO-1(칼리지 런던 대학), AVA-008(아박타 유한회사(Avacta Ltd)), 엑시캅타진 실로류셀(axicabtagene ciloleucel)(카이트 파르마 주식회사), BG-T19(광저우 바이오-진 테크놀로지 유한 책임회사(Guangzhou Bio-gene Technology Co Ltd)), BinD-19(센젠 빈데바이오 유한회사(Shenzhen BinDeBio Ltd.)), BPX-401(벨리쿰 파마슈티칼 주식회사), CAR19h28TM41BBz(웨스트미드 의학 연구소(Westmead Institute for Medical Research)), C-CAR-011(중국 PLA 종합병원), CD19CART(이노베이티브 셀 테라퓨틱스 유한 책임회사), CIK-CAR.CD19(포뮬라 파마슈티칼 주식회사(Formula Pharmaceuticals Inc)), CLIC-1901(오타와 병원 연구소(Ottawa Hospital Research Institute)), CSG-CD19(카스겐 테라퓨틱스 유한 회사), CTL-119(펜실베니아 대학), CTX-101(크리스퍼 테라퓨틱스 아게(CRISPR Therapeutics AG)), DSCAR-01(상하이 레인 바이오테크놀로지(Shanghai Hrain Biotechnology)), ET-190(유레카 테라퓨틱스 주식회사), FT-819(메모리얼 슬론-케터링 암 센터), ICAR-19(이뮨 셀 테라피 주식회사(Immune Cell Therapy Inc)), IM19 CAR-T(베이징 이뮤노차이나 메디칼 사이언스 앤 테크놀로지 유한 책임회사), JCAR-014(주노 테라퓨틱스 주식회사), JWCAR-029(밍주 테라퓨틱스(상하이) 유한 책임회사(MingJu Therapeutics (Shanghai) Co., Ltd)), KD-C-19(난징 카이디 바이오테크 주식회사), LinCART19(아이셀진 테라퓨틱스(iCell Gene Therapeutics)), 리소캅타진 마라류셀(lisocabtagene maraleucel)(주노 테라퓨틱스 주식회사), 일치CART(상하이 레인 바이오테크놀로지), MB-CART19.1(상하이 아동 의료 센터(Shanghai Children's Medical Center)), PBCAR-0191(프리시전 바이오사이언시스 주식회사(Precision BioSciences Inc)), PCAR-019(퍼슨젠 바이오메디신(쑤저우) 유한 책임회사(PersonGen Biomedicine (Suzhou) Co Ltd)), pCAR-19B(총칭 프리시전 바이오테크 유한 책임회사(Chongqing Precision Biotech Co Ltd)), PZ-01(핀즈 라이프테크놀로지 유한 책임회사(Pinze Lifetechnology Co Ltd)), RB-1916(리퓨즈 바이오테크놀로지스 주식회사(Refuge Biotechnologies Inc)), SKLB-083019(청두 인허 바이오메디칼 유한 책임회사(Chengdu Yinhe Biomedical Co Ltd)), spCART-19(상하이 유니카-테라피 바이오-메디신 테크놀로지 유한 책임회사), TBI-1501(타카라 바이오 주식회사(Takara Bio Inc)), TC-110(TCR2 테라퓨틱스 주식회사(TCR2 Therapeutics Inc)), TI-1007(티이뮨 바이오테크 주식회사(Timmune Biotech Inc)), 티사젠렉류셀(tisagenlecleucel)(펜실베니아 대학 아브람슨 암 센터(Abramson Cancer Center of the University of Pennsylvania)), U-CART(상하이 바이오레이 래보래토리 주식회사(Shanghai Bioray Laboratory Inc)), UCART-19(우겐 주식회사(Wugen Inc)), UCART-19(셀렉티스 에스에이), 바다캅타진 레라류셀(vadacabtagene leraleucel)(메모리얼 슬론-케터링 암 센터), XLCART-001(난징 의과 대학(Nanjing Medical University)) 및 인누오카티-19(yinnuokati-19)(센젠 이노베이션 이뮤노테크놀로지 유한 책임회사(Shenzhen Innovation Immunotechnology Co Ltd))를 포함한다.

[0486] 다양한 실시형태에서, CAR은 CD2에 특이적으로 결합한다. 예시적인 항-CD2 CAR은 제한 없이 UCART-2(우겐 주식회사)를 포함한다.

[0487] 다양한 실시형태에서, CAR은 CD20에 특이적으로 결합한다. 예시적인 항-CD20 CAR은 제한 없이 ACTR-087(싱가포르 국제 대학(National University of Singapore)), ACTR-707(우넘 테라퓨틱스 주식회사(Unum Therapeutics Inc)), CBM-C20.1(중국 PLA 종합병원), MB-106(프레드 허친슨 암 연구 센터(Fred Hutchinson Cancer Research Center)) 및 MB-CART20.1(밀텐이 바이오테크 게엠베하(Miltenyi Biotec GmbH))을 포함한다.

[0488] 다양한 실시형태에서, CAR은 CD22에 특이적으로 결합한다. 예시적인 항-CD22 CAR은 제한 없이 펜실베니아 대학(University of Pennsylvania)의 항-CD22 CAR T-세포 요법(B-세포 급성 림프아구성 백혈병), CD22-CART(상하이 유니카-테라피 바이오-메디신 테크놀로지 유한 책임회사), JCAR-018(오퍼스 바이오 주식회사(Opus Bio Inc)), MendCART(상하이 레인 바이오테크놀로지) 및 UCART-22(셀렉티스 에스에이)를 포함한다.

[0489] 다양한 실시형태에서, CAR은 CD30에 특이적으로 결합한다. 예시적인 항-CD30 CAR은 제한 없이 ATLCAR.CD30(유엔씨 린버거 통합 암 센터), CBM-C30.1(중국 PLA 종합병원) 및 Hu30-CD28zet(국립 암 연구소(National Cancer Institute))를 포함한다.

[0490] 다양한 실시형태에서, CAR은 CD33에 특이적으로 결합한다. 예시적인 항-CD33 CAR은 제한 없이 TC 바이오팜/칼리

지 런던 대학(University College London)의 항-CD33 CAR 감마 델타 T-세포 요법(급성 골수성 백혈병), CAR33VH(오퍼스 바이오 주식회사), CART-33(중국 PLA 종합병원), CIK-CAR.CD33(포뮬라 파마슈티칼 주식회사), UCART-33(셀렉티스 에스에이) 및 VOR-33(콜롬비아 대학(Columbia University))을 포함한다.

[0491] 다양한 실시형태에서, CAR은 CD38에 특이적으로 결합한다. 예시적인 항-CD38 CAR은 제한 없이 UCART-38(셀렉티스 에스에이)을 포함한다.

[0492] 다양한 실시형태에서, CAR은 CD38 A2에 특이적으로 결합한다. 예시적인 항-CD38 A2 CAR은 제한 없이 T-007(티엔케이 테라퓨틱스 주식회사(TNK Therapeutics Inc))을 포함한다.

[0493] 다양한 실시형태에서, CAR은 CD4에 특이적으로 결합한다. 예시적인 항-CD4 CAR은 제한 없이 CD4CAR(아이셀 진 테라퓨틱스)을 포함한다.

[0494] 다양한 실시형태에서, CAR은 CD44에 특이적으로 결합한다. 예시적인 항-CD44 CAR은 제한 없이 CAR-CD44v6(H 산 라파엘레 과학 연구소(Istituto Scientifico H San Raffaele))을 포함한다.

[0495] 다양한 실시형태에서, CAR은 CD5에 특이적으로 결합한다. 예시적인 항-CD5 CAR은 제한 없이 CD5CAR(아이셀 진 테라퓨틱스)을 포함한다. 예시적인 CD5 CAR 아미노산 서열이 아래에 제공된다:

[0496] >5CAR-CH3-CD28TM-CD28-CD3Z

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL 300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ 360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKFWVLVVVGGVLACYSLLVTVAFIIFWV 420
421 RSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQ 480
481 LYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGE 540
541 RRRGKGHDGLYQGLSTATKDTYDALHMQALPPR 573
```

[0497]

[0498] >5CAR-CD8aH-CD28TM-CD28-CD3Z

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA 300
301 AGGAVHTRGLDFACDFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPR 360
361 RPGPTRKHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDK 420
421 RRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTAT 480
481 KDTYDALHMQALPPR 495
```

[0499]

[0500] >5CAR-CD28H-CD28TM-CD28-CD3Z

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAIEVMYPPPYLDNEKSNGTIIHVKGKHLCPS 300
301 PLFPGPSKPFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTR 360
361 KHYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDP 420
421 EMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDA 480
481 LHMQALPPR 489
```

[0501]

[0502] >5CAR-CH3-CD8aTM-41BB-CD3Z

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPAEPKSPDKTHTCPGQPREPQVYTLPPSRDEL 300
301 TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ 360
361 QGNVFSCSVMHEALHNHYTQKSLSLSPGKKDPKPTTTPAPRPPTPAPTIASQPLSLRPEA 420
421 CRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMR 480
481 PVQTTQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVL 540
541 DKRRGRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLST 600
601 ATKDTYDALHMQALPPR 617
```

[0503]

[0504] >5CAR-CD8aH-CD8aTM-41BB-CD3Z

```
  1 MEFGLSWLFLVAILKGVQCIDAMGNIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNW 60
 61 VKQAPGKGLRWMGWINTHTGEPTYADDFKGRFAFSLETSASTAYLQINNLKNEDTATYFC 120
121 TRRGYDWYFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIKMTQSPSSMYASLGERVTITC 180
181 KASQDINSYLSWFHHKPGKSPKTLIYRANRLVDGVPSRFSGSGSGQDYSLTISSLDYEDM 240
241 GIYYCQQYDESPWTFGGGTKLEMKGSGDPATTTPAPRPPTPAPTIASQPLSLRPEACRPA 300
301 AGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCKRGRKKLLYIFKQPFMRPVQT 360
361 TQEEDGCSCRFPEEEEGGCELRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRR 420
421 GRDPEMGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKD 480
481 TYDALHMQALPPRX 494
```

[0505]

[0506] 다양한 실시형태에서, CAR은 CD7에 특이적으로 결합한다. 예시적인 항-CD7 CAR은 제한 없이 CAR-pNK(퍼슨젠 바이오메디신(쑤저우) 유한 책임회사) 및 CD7.CAR/28제타 CAR T 세포(베일러 의과 대학(Baylor College of Medicine)), UCART7(워싱턴 대학교 세인트루이스(Washington University in St Louis))을 포함한다. 예시적인 CD7 CAR 아미노산 서열은 다음과 같다:

[0507] >7CAR8

```
  1 MALPVTALLLPLALLLHAARPGSDIELTQSPAIMSASLGEEITLTCSASSSVSYMHWYQQ 60
 61 KSGTSPKLLIYSTSNLASGVPSRFSGSGSGTFYSLTISSVEAEDAADYYCHQWSSYTFGG 120
121 GTKLEIKRGGGGSGGGGSGGGGSQVKLQESGGGLVKPGGSLKLSCAASGFTFSSYAMSWV 180
181 RQTPEKRLEWVATISSGGSYTYYPDSVKGRFTISRDNAKNTLYLQMSSLRSEDTAMYYCA 240
241 RQDGYYPGWFANWGQGTTVTVSSAAATTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGA 300
301 VHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCRSKRSRLLHSDYMNMTPRRPGPTRK 360
361 HYQPYAPPRDFAAYRSRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPE 420
421 MGGKPRRKNPQEGLYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDAL 480
481 HMQALPPR 488
```

[0508]

[0509] 다양한 실시형태에서, CAR은 CDH17에 특이적으로 결합한다. 예시적인 항-CDH17 CAR은 제한 없이 ARB-001.T(아벨르 유한회사(Arbele Ltd))를 포함한다.

[0510] 다양한 실시형태에서, CAR은 CEA에 특이적으로 결합한다. 예시적인 항-CEA CAR은 제한 없이 HORC-020(휴오리진 주식회사(HumOrigin Inc))을 포함한다.

[0511] 다양한 실시형태에서, CAR은 키메라 TGF-베타 수용체(CTBR)에 특이적으로 결합한다. 예시적인 항-키메라 TGF-베타 수용체(CTBR) CAR은 제한 없이 CAR-CTBR T 세포(블루버드 바이오 주식회사)를 포함한다.

[0512] 다양한 실시형태에서, CAR은 Claudin18.2에 특이적으로 결합한다. 예시적인 항-Claudin18.2 CAR은 제한 없이 CAR-CLD18 T-세포(카스겐 테라퓨틱스 유한 회사) 및 KD-022(난징 카이디 바이오테크 주식회사)를 포함한다.

[0513] 다양한 실시형태에서, CAR은 CLL1에 특이적으로 결합한다. 예시적인 항-CLL1 CAR은 제한 없이 키트E-796(카이트 파르마 주식회사)을 포함한다.

[0514] 다양한 실시형태에서, CAR은 DLL3에 특이적으로 결합한다. 예시적인 항-DLL3 CAR은 제한 없이 AMG-119(암젠 주식회사(Amgen Inc))를 포함한다.

[0515] 다양한 실시형태에서, CAR은 이중 BCMA/TACI(APRIL)에 특이적으로 결합한다. 예시적인 항-이중 BCMA/TACI(APRIL) CAR은 제한 없이 AUTO-2(오토러스 테라퓨틱스 리미티드(Autolus Therapeutics Limited))를 포함한다.

[0516] 다양한 실시형태에서, CAR은 이중 CD19/CD22에 특이적으로 결합한다. 예시적인 항-이중 CD19/CD22 CAR은 제한 없이 AUTO-3(오토러스 테라퓨틱스 리미티드) 및 LCAR-L10D(난징 레전드 바이오테크 유한 책임회사)를 포함한다.

[0517] 다양한 실시형태에서, CAR은 CD19에 특이적으로 결합한다.

[0518] 다양한 실시형태에서, CAR은 이중 CLL1/CD33에 특이적으로 결합한다. 예시적인 항-이중 CLL1/CD33 CAR은 제한 없이 ICG-136(아이셀 진 테라퓨틱스)을 포함한다.

[0519] 다양한 실시형태에서, CAR은 이중 EpCAM/CD3에 특이적으로 결합한다. 예시적인 항-이중 EpCAM/CD3 CAR은 제한 없이 IKT-701(아이셀 케알렉스 테라퓨틱스(Icell Kealex Therapeutics))을 포함한다.

[0520] 다양한 실시형태에서, CAR은 이중 ErbB/4ab에 특이적으로 결합한다. 예시적인 항-이중 ErbB/4ab CAR은 제한 없이 LEU-001(킹스 칼리지 런던(King's College London))을 포함한다.

[0521] 다양한 실시형태에서, CAR은 이중 FAP/CD3에 특이적으로 결합한다. 예시적인 항-이중 FAP/CD3 CAR은 제한 없이 IKT-702(아이셀 케알렉스 테라퓨틱스)를 포함한다.

[0522] 다양한 실시형태에서, CAR은 EBV에 특이적으로 결합한다. 예시적인 항-EBV CAR은 제한 없이 TT-18(테사 테라퓨틱스 유한회사(Tessa Therapeutics Pte Ltd))을 포함한다.

[0523] 다양한 실시형태에서, CAR은 EGFR에 특이적으로 결합한다. 예시적인 항-EGFR CAR은 제한 없이 블루버드 바이오(bluebird bio Inc)의 항-EGFR CAR T-세포 요법(CBLB MegaTAL, 암), 상하이 셀 테라피 연구소(Shanghai Cell Therapy Research Institute)의 CTLA-4 관문 저해제 + PD-1 관문 저해제 mAb를 발현하는 항-EGFR CAR T-세포 요법(EGFR-양성 진행성 고형 종양), CSG-EGFR(카스겐 테라퓨틱스 유한 회사) 및 EGFR-IL12-CART(프리진(센젠) 바이오테크놀로지 유한 책임회사(Pregene (Shenzhen) Biotechnology Co Ltd))를 포함한다.

[0524] 다양한 실시형태에서, CAR은 EGFRvIII에 특이적으로 결합한다. 예시적인 항-EGFRvIII CAR은 제한 없이 KD-

035(난징 카이디 바이오테크 주식회사) 및 UCART-EgfrVIII(셀렉티스 에스에이)을 포함한다.

[0525] 다양한 실시형태에서, CAR은 Flt3에 특이적으로 결합한다. 예시적인 항-Flt3 CAR은 제한 없이 ALLO-819(화이자 주식회사) 및 AMG-553(암젠 주식회사)을 포함한다.

[0526] 다양한 실시형태에서, CAR은 폴레이트 수용체에 특이적으로 결합한다. 예시적인 항-폴레이트 수용체 CAR은 제한 없이 EC17/CAR T(엔도사이트 주식회사)를 포함한다.

[0527] 다양한 실시형태에서, CAR은 G250에 특이적으로 결합한다. 예시적인 항-G250 CAR은 제한 없이 에라스무스 의료 센터(Erasmus Medical Center)(다니엘 덴 호드 암 센터(Daniel den Hoed Cancer Center))의 자가 T-림프구 세포 요법(G250-scFV-형질도입됨, 신세포 암종)을 포함한다.

[0528] 다양한 실시형태에서, CAR은 GD2에 특이적으로 결합한다. 예시적인 항-GD2 CAR은 제한 없이 1RG-CART(칼리지 런던 대학), 4SCAR-GD2(센젠 제노-이뮨 의학 연구소), C7R-GD2.CART 세포(베일러 의과 대학), CMD-501(베일러 의과 대학), CSG-GD2(카스겐 테라퓨틱스 유한 회사), GD2-CART01(밤비노 게수 병원 및 연구소(Bambino Gesu Hospital and Research Institute)), GINAKIT 세포(베일러 의과 대학), iC9-GD2-CAR-IL-15 T-세포(유엔씨 린버거 통합 암 센터) 및 IKT-703(아이셀 케알렉스 테라퓨틱스)을 포함한다.

[0529] 다양한 실시형태에서, CAR은 GD2 및 MUC1에 특이적으로 결합한다. 예시적인 항-GD2/MUC1 CAR은 제한 없이 PSMA CAR-T(펜실베니아 대학)를 포함한다.

[0530] 다양한 실시형태에서, CAR은 GPC3에 특이적으로 결합한다. 예시적인 항-GPC3 CAR은 제한 없이 ARB-002.T(아벨르 유한회사), CSG-GPC3(카스겐 테라퓨틱스 유한 회사), GLYCAR(베일러 의과 대학) 및 TT-14(테사 테라퓨틱스 유한 회사)를 포함한다.

[0531] 다양한 실시형태에서, CAR은 Her2에 특이적으로 결합한다. 예시적인 항-Her2 CAR은 제한 없이 ACTR-087 + 트라스트주맙(trastuzumab)(우넘 테라퓨틱스 주식회사), ACTR-707 + 트라스트주맙(우넘 테라퓨틱스 주식회사), CIDeCAR(벨리쿰 파마슈티칼 주식회사), MB-103(머스탕 바이오 주식회사(Mustang Bio Inc)), RB-H21(리퓨즈 바이오테크놀로지스 주식회사) 및 TT-16(베일러 의과 대학)을 포함한다.

[0532] 다양한 실시형태에서, CAR은 IL13R에 특이적으로 결합한다. 예시적인 항-IL13R CAR은 제한 없이 MB-101(시티 오브 호프(City of Hope)) 및 YYB-103(유영 파마슈티칼스 유한 책임회사(YooYoung Pharmaceuticals Co Ltd))을 포함한다.

[0533] 다양한 실시형태에서, CAR은 인테그린 베타-7에 특이적으로 결합한다. 예시적인 항-인테그린 베타-7 CAR은 제한 없이 MMG49 CAR T-세포 요법(오사카 대학)을 포함한다.

[0534] 다양한 실시형태에서, CAR은 LC 항원에 특이적으로 결합한다. 예시적인 항-LC 항원 CAR은 제한 없이 VM-803(바이로메드 유한 책임회사) 및 VM-804(바이로메드 유한 책임회사)를 포함한다.

[0535] 다양한 실시형태에서, CAR은 메소텔린에 특이적으로 결합한다. 예시적인 항-메소텔린 CAR은 제한 없이 CARMA-hMeso(존스 홉킨스 대학(Johns Hopkins University)), CSG-MESO(카스겐 테라퓨틱스 유한 회사), iCasp9M28z(메모리얼 슬론-케터링 암 센터), KD-021(난징 카이디 바이오테크 주식회사), m-28z-T2(광저우 생명의학 및 건강 연구소), MesoCART(펜실베니아 대학), meso-CAR-T + PD-78(미르이뮨 엘엘씨(MirImmune LLC)), RB-M1(리퓨즈 바이오테크놀로지스 주식회사) 및 TC-210(TCR2 테라퓨틱스 주식회사)을 포함한다.

[0536] 다양한 실시형태에서, CAR은 MUC1에 특이적으로 결합한다. 예시적인 항-MUC1 CAR은 제한 없이 광저우 안지 바이오메디칼 테크놀로지/테크놀로지 시드니 대학(Guangzhou Anjie Biomedical Technology Co LTD)의 항-MUC1 CAR T-세포 요법 + PD-1 넉아웃 T 세포 요법(식도암/NSCLC), ICTCAR-043(이노베이티브 셀 테라퓨틱스 유한 책임회사), ICTCAR-046(이노베이티브 셀 테라퓨틱스 유한 책임회사), P-MUC1C-101(포세이다 테라퓨틱스 주식회사) 및 TAB-28z(온코탭 주식회사(OncoTab Inc))를 포함한다.

[0537] 다양한 실시형태에서, CAR은 MUC16에 특이적으로 결합한다. 예시적인 항-MUC16 CAR은 제한 없이 4H1128Z-E27(유레카 테라퓨틱스 주식회사) 및 JCAR-020(메모리얼 슬론-케터링 암 센터)을 포함한다.

[0538] 다양한 실시형태에서, CAR은 nfP2X7에 특이적으로 결합한다. 예시적인 항-nfP2X7 CAR은 제한 없이 BIL-022c(바이오셉터 인터내셔날 유한회사(Biosceptre International Ltd))를 포함한다.

[0539] 다양한 실시형태에서, CAR은 PSCA에 특이적으로 결합한다. 예시적인 항-PSCA CAR은 제한 없이 BPX-601(벨리쿰

파마슈티칼 주식회사)을 포함한다.

[0540] 다양한 실시형태에서, CAR은 PSMA에 특이적으로 결합한다. CIK-CAR.PSM(포뮬라 파마슈티칼 주식회사) 및 P-PSMA-101(포세이다 테라퓨틱스 주식회사)을 포함한다.

[0541] 다양한 실시형태에서, CAR은 ROR1에 특이적으로 결합한다. 예시적인 항-ROR1 CAR은 제한 없이 JCAR-024(프레드 허친슨 암 연구 센터)를 포함한다.

[0542] 다양한 실시형태에서, CAR은 ROR2에 특이적으로 결합한다. 예시적인 항-ROR2 CAR은 제한 없이 CCT-301-59(F1 온콜로지 주식회사)를 포함한다.

[0543] 다양한 실시형태에서, CAR은 SLAMF7에 특이적으로 결합한다. 예시적인 항-SLAMF7 CAR은 제한 없이 UCART-CS1(셀렉티스 에스에이)을 포함한다.

[0544] 다양한 실시형태에서, CAR은 TRBC1에 특이적으로 결합한다. 예시적인 항-TRBC1 CAR은 제한 없이 AUTO-4(오토러스 테라퓨틱스 리미티드)를 포함한다.

[0545] 다양한 실시형태에서, CAR은 TRBC2에 특이적으로 결합한다. 예시적인 항-TRBC2 CAR은 제한 없이 AUTO-5(오토러스 테라퓨틱스 리미티드)를 포함한다.

[0546] 다양한 실시형태에서, CAR은 TSHR에 특이적으로 결합한다. 예시적인 항-TSHR CAR은 제한 없이 ICTCAT-023(이노베이티브 셀 테라퓨틱스 유한 책임회사)을 포함한다.

[0547] 다양한 실시형태에서, CAR은 VEGFR-1에 특이적으로 결합한다. 예시적인 항-VEGFR-1 CAR은 제한 없이 SKLB-083017(쓰촨 대학교(Sichuan University))을 포함한다.

[0548] 다양한 실시형태에서, CAR은 AT-101(앱클론 주식회사(AbClon Inc)); AU-101, AU-105 및 AU-180(오로라 바이오팜 주식회사(Aurora Biopharma Inc)); CARMA-0508(카리스마 테라퓨틱스(Carisma Therapeutics)); CAR-T(페이트 테라퓨틱스 주식회사(Fate Therapeutics Inc)); CAR-T(셀 디자인 랩스 주식회사(Cell 설계 Labs Inc)); CM-CX1(셀다라 메디칼 엘엘씨(Celdara Medical LLC)); CMD-502, CMD-503 및 CMD-504(베일러 의과 대학); CSG-002 및 CSG-005(카스겐 테라퓨틱스 유한 회사); ET-1501, ET-1502 및 ET-1504(유레카 테라퓨틱스 주식회사); FT-61314(페이트 테라퓨틱스 주식회사); GB-7001(상하이 진캠 유한 책임회사(Shanghai GeneChem Co Ltd)); IMA-201(이매틱스 바이오테크놀로지스 게엠베하(Immatics Biotechnologies GmbH)); IMM-005 및 IMM-039(이뮤놈 주식회사(Immunome Inc)); ImmuniCAR(TC 바이오팜 유한회사); NT-0004 및 NT-0009(바이오엔테크 셀 앤 진 테라피 게엠베하(BioNTech Cell and Gene Therapies GmbH)), OGD-203(OGD2 파르마 에스에이에스(OGD2 Pharma SAS)), PMC-005B(파멥신(PharmAbcine)) 및 TI-7007(티이뮨 바이오테크 주식회사)이다.

[0549] 본 명세서에 기재된 키메라 항원 수용체를 암호화하는 핵산이 또한 본 명세서에 제공된다. 일부 실시형태에서, 핵산은 단리 또는 정제된다. 핵산의 생체외 전달은 당업계에 공지된 방법을 사용하여 달성될 수 있다. 예를 들어, 대상체로부터 얻어진 면역 세포는 키메라 항원 수용체를 암호화하는 핵산 벡터로 형질전환될 수 있다. 그런 다음, 벡터를 사용하여 수용체 면역 세포를 형질전환시켜 이들 세포가 키메라 항원 수용체를 발현하도록 할 수 있다. 면역 세포를 형질전환시키는 효율적인 수단은 형질감염 및 형질도입을 포함한다. 이러한 방법은 당업계에 잘 알려져 있다. 예를 들어, 키메라 항원 수용체를 암호화하는 핵산 분자(및 염기 편집기를 암호화하는 핵산(들))를 전달하기 위해 적용 가능한 방법은 각각 그 전문이 참조에 의해 본 명세서에 원용되어 있는 국제 출원 PCT/US2009/040040 및 미국 특허 제8,450,112호; 제9,132,153호; 및 제9,669,058호에서 찾을 수 있다. 추가적으로, 염기 편집기를 암호화하는 핵산을 전달하기 위한 본 명세서에 기재된 이들 방법 및 벡터는 키메라 항원 수용체를 암호화하는 핵산을 전달하는데 적용될 수 있다.

[0550] ***세포외 결합 도메인***

[0551] 본 발명의 키메라 항원 수용체는 세포외 결합 도메인을 포함한다. 본 명세서에서 상정된 키메라 항원 수용체의 세포외 결합 도메인은 특정 항원에 대해 친화성을 갖는 항체 또는 이의 항원 결합 단편의 아미노산 서열을 포함한다. 일부 실시형태에서, 항원은 CD3이다. 일부 실시형태에서, 항원은 CD5이다. 일부 실시형태에서, 항원은 CD7이다. 일부 실시형태에서, 항원은 CD33이다. 일부 실시형태에서, 항원은 CD123이다.

[0552] 일부 실시형태에서, 키메라 항원 수용체는 항체의 아미노산 서열을 포함한다. 일부 실시형태에서, 키메라 항원 수용체는 항체의 항원 결합 단편의 아미노산 서열을 포함한다. 세포외 결합 도메인의 항체(또는 이의 단편) 부분은 항원의 에피토프를 인식하고 이에 결합한다. 일부 실시형태에서, 키메라 항원 수용체의 항체 단편 부분은

단일 사슬 가변 단편(scFv)이다. scFv는 단일클론 항체의 경쇄 및 가변 단편을 포함한다. 다른 실시형태에서, 키메라 항원 수용체의 항체 단편 부분은 하나 이상의 세포외 결합 도메인을 포함할 수 있어 하나 이상의 항원에 동시에 결합할 수 있는 다중사슬 가변 단편이다. 다중 사슬 가변 단편 실시형태에서, 힌지 영역은 다양한 가변 단편을 분리하여 필요한 공간 배열과 유연성을 제공할 수 있다.

[0553] 다른 실시형태에서, 키메라 항원 수용체의 항체 부분은 적어도 하나의 중쇄 사슬 및 적어도 하나의 경쇄 사슬을 포함한다. 일부 실시형태에서, 키메라 항원 수용체의 항체 부분은 이황화 브리지에 의해 연결된 2개 중쇄 사슬 및 2개 경쇄 사슬을 포함하되, 경쇄 사슬은 이황화 브리지에 의해 중쇄 사슬 중 하나에 각각 연결된다. 일부 실시형태에서, 경쇄 사슬은 불변 영역 및 가변 영역을 포함한다. 항체의 가변 영역에 있는 상보성 결정 영역은 특정 항원에 대한 항체의 친화성을 담당한다. 따라서, 상이한 항원을 인식하는 항체는 상이한 상보성 결정 영역을 포함한다. 상보성 결정 영역은 세포외 결합 도메인의 가변 도메인에 존재하고, 가변 도메인(즉, 가변 중쇄 및 가변 경쇄)은 링커 또는 일부 실시형태에서 이황화 브리지와 연결될 수 있다. 일부 실시형태에서, 가변 중쇄 사슬 및 가변 경쇄 사슬은 $(GGGGS)_n$ 링커에 의해 연결되되, N은 1 내지 10의 정수이다. 일부 실시형태에서, 링커는 $(GGGGS)_3$ 링커이다.

[0554] 일부 실시형태에서, 세포외 도메인에 의해 인식되고 이에 결합된 항원은 단백질 또는 펩타이드, 핵산, 지질 또는 다당류이다. 항원은 병원성 세균 또는 바이러스에서 발현되는 것들과 같이 이종성일 수 있다. 항원은 또한 합성일 수도 있고; 예를 들어, 일부 개체는 합성 라텍스에 극도의 알레르기가 있으며, 이 항원에 노출되면 극도의 면역 반응을 일으킬 수 있다. 일부 실시형태에서, 항원은 자가조직이며, 병든 또는 그렇지 않으면 변경된 세포에서 발현된다. 예를 들어, 일부 실시형태에서, 항원은 신생물 세포에서 발현된다. 일부 실시형태에서, 신생물 세포는 고형 종양 세포이다. 다른 실시형태에서, 신생물 세포는 액체 종양 세포이다. 다른 실시형태에서, 신생물 세포는 B 세포암과 같은 혈액암이다. 일부 실시형태에서, B 세포암은 림프종 또는 백혈병이다.

[0555] 본 명세서에 기재된 방법으로 치료되는 액체암은, 예를 들어, 백혈병일 수 있다. 일부 경우에, 백혈병은 전백혈병을 포함한다. 일부 경우에, 백혈병은 급성 백혈병이다. 급성 백혈병은, 예를 들어, 급성 골수성 백혈병(AML)을 포함한다. 급성 백혈병은 또한, 예를 들어, 급성 림프성 백혈병 또는 급성 림프구성 백혈병(ALL)을 포함하고; ALL은 B-계통 ALL; T-계통 ALL; 및 T-세포 급성 림프구성 백혈병(T-ALL)을 포함한다.

[0556] 신생물의 비제한적인 예는 T-세포 급성 림프아구성 백혈병(T-ALL), 용상 진균증(MF), 세자리 증후군(SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 $ALK^+$, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 $CD30^+$ 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종을 포함한다. 일부 실시형태에서, 신생물은 T-세포 급성 림프아구성 백혈병(T-ALL)이다. 일부 실시형태에서, 신생물은 급성 골수성 백혈병(AML)이다.

[0557] 항체-항원 상호작용은 수소 결합, 정전기적 또는 소수성 상호작용 또는 반 데르 발스 힘으로 인한 비공유 상호작용이다. 항원에 대한 키메라 항원 수용체의 세포외 결합 도메인의 친화성은 다음 공식으로 계산될 수 있다:

[0558] $K_A$ = [항체-항원]/[항체][항원], 여기서

[0559] [Ab] = 항체 상의 비점유 결합 부위의 몰농도;

[0560] [Ag] = 항원 상의 비점유 결합 부위의 몰농도; 및

[0561] [Ab-Ag] = 항체-항원 복합체의 몰농도.

[0562] 항체-항원 상호작용은 또한 항체로부터 항원의 해리에 기초하여 특성화될 수 있다. 해리 상수($K_D$)는 해리 속도에 대한 결합 속도의 비이며, 친화성 상수에 반비례한다. 따라서, $K_D = 1/K_A$이다. 당업자는 이러한 개념에 익숙할 것이며, ELISA 검정과 같은 전통적인 방법을 사용하여 이러한 상수를 계산할 수 있음을 알 것이다.

[0563] ***막관통 도메인***

[0564] 본 발명의 키메라 항원 수용체는 막관통 도메인을 포함한다. 본 명세서에 기재된 키메라 항원 수용체의 막관통 도메인은 CAR-T 세포 지질 이중층 세포막에 걸쳐 있으며, 세포외 결합 도메인 및 세포내 신호전달 도메인을 분리한다. 일부 실시형태에서, 이러한 도메인은 막관통 도메인을 갖는 다른 수용체로부터 유래되는 반면, 다른 실

시형태에서, 이러한 도메인은 합성이다. 일부 실시형태에서, 막관통 도메인은 비-인간 막관통 도메인으로부터 유래될 수 있고, 일부 실시형태에서는 인간화된다. "인간화된"은 인간 대상체에서 보다 안정적으로 또는 효율적으로 발현되도록 최적화된 막관통 도메인을 암호화하는 핵산의 서열을 갖는 것을 의미한다. 일부 실시형태에서, 막관통 도메인은 인간 면역 효과기 세포에서 발현되는 또 다른 막관통 단백질로부터 유래된다. 이러한 단백질의 예는 T 세포 수용체(TCR) 복합체의 서브유닛, PD1 또는 임의의 분화 클러스터 단백질 또는 면역 효과기 세포에서 발현되며 막관통 도메인을 갖는 다른 단백질을 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 막관통 도메인은 합성일 것이며, 이러한 서열은 많은 소수성 잔기를 포함할 것이다.

[0565] 개시된 CAR에서 사용하기 위한 막관통 도메인은 적어도 T-세포 수용체의 알파, 베타 또는 제타 사슬, CD28, CD3 엡실론, CD45, CD4, CD5, CD8, CD9, CD16, CD22, CD33, CD37, CD64, CD80, CD86, CD134, CD137, CD154의 막관통 영역(들)을 포함할 수 있다. 일부 실시형태에서, 막관통 도메인은 CD4, CD8α, CD28 또는 CD3ζ로부터 유래된다. 일부 실시형태에서, 막관통 도메인은 CD28 막관통 도메인이다. 일부 실시형태에서, 막관통 도메인은 CD8α 막관통 도메인이다.

[0566] 일부 실시형태에서, 막관통 도메인은 CD8α 힌지 및 막관통 도메인이다. 일부 실시형태에서, CD8α 힌지 및 막관통 도메인은 아래 제공되는 바와 같은 예시적인 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 동일하다:

[0567] SDPTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDIYIWAPLAGTCGVLLL
SLVITLYC

[0568] 키메라 항원 수용체는 일부 실시형태에서 막관통 도메인과 세포외 도메인, 세포내 도메인 또는 둘 모두 사이에 스페이서를 포함하도록 설계된다. 이러한 스페이서는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 아미노산 길이일 수 있다. 일부 실시형태에서, 스페이서는 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90개 또는 100개의 아미노산 길이일 수 있다. 또 다른 실시형태에서, 스페이서는 100개 내지 500개의 아미노산 길이일 수 있다. 스페이서는 한 도메인을 또 다른 도메인에 연결하고, 키메라 항원 수용체 기능을 향상시키거나 또는 최적화하기 위해 이러한 연결된 도메인을 배치하는데 사용되는 임의의 폴리펩타이드일 수 있다. 일부 실시형태에서, 힌지/스페이서는 CH3, CD8α 또는 CD28로부터 선택된다.

[0569] ***세포내 신호전달 도메인***

[0570] 본 발명의 키메라 항원 수용체는 세포내 신호전달 도메인을 포함한다. 세포내 신호전달 도메인은 T 세포 효과기 기능 신호(예를 들어, 활성화 신호)를 변환하고 T 세포가 특화된 기능을 수행하도록 지시하는 T 세포에서 발현되는 단백질의 세포내 부분이다. T 세포 활성화는 T 세포의 표면의 T 세포 수용체에 대한 동족 항원의 결합 및 T 세포의 표면의 공동자극 분자에 대한 동족 리간드의 결합을 포함하는 다수의 인자에 의해 유도될 수 있다. T 세포 공동-자극 분자는 공동-자극 리간드와 특이적으로 결합하여 증식과 같지만 이로 제한되지 않는 T 세포에 의한 공동-자극 반응을 매개하는 T 세포 상의 동족 결합 파트너이다. 공동-자극 분자는 MHC 클래스 I 분자를 포함하지만, 이로 제한되지 않는다. T 세포의 활성화는 T 세포 증식 및 분화와 같은 면역 반응으로 이어진다(예를 들어, 문헌[Smith-Garvin *et al.*, Annu. Rev. Immunol., 27:591-619, 2009] 참조). 예시적인 T 세포 신호전달 도메인은 당업계에 공지되어 있다. 비제한적인 예는 CD3ζ, CD8, CD28, CD27, CD154, GITR(TNFRSF18), CD134(OX40) 및 CD137(4-1BB) 신호전달 도메인을 포함한다.

[0571] 본 명세서에서 상정된 키메라 항원 수용체의 세포내 신호전달 도메인은 1차 신호전달 도메인을 포함한다. 일부 실시형태에서, 키메라 항원 수용체는 1차 신호전달 도메인 및 2차 또는 공동-자극 신호전달 도메인을 포함한다.

[0572] 일부 실시형태에서, 1차 신호전달 도메인은 하나 이상의 면역수용체 타이로신-기반 활성화 모티프 또는 ITAM을 포함한다. 일부 실시형태에서, 1차 신호전달 도메인은 하나 이상의 ITAM을 포함한다. 키메라 항원 수용체에 혼입된 ITAM은 다른 세포 수용체로부터의 ITAM으로부터 유래될 수 있다. 일부 실시형태에서, ITAM을 포함하는 1차 신호전달 도메인은 CD3γ, CD3ε, CD3ζ 또는 CD3δ와 같은 TCR 복합체의 서브유닛으로부터 유래될 수 있다(도 1a 참조). 일부 실시형태에서, ITAM을 포함하는 1차 신호전달 도메인은 FcRγ, FcRβ, CD5, CD22, CD79a, CD79b 또는 CD66d로부터 유래될 수 있다.

[0573] 일부 실시형태에서, 1차 신호전달 도메인은 CD8, CD28, CD134(OX40), CD137(4-1BB) 및 CD3ζ로 이루어진 군으로부터 선택된다. 일부 실시형태에서, 1차 신호전달 도메인은 CD3ζ 신호전달 도메인이다. 일부 실시형태에서, CD3ζ 신호전달 도메인은 아래 제공되는 바와 같은 예시적인 아미노산 서열과 적어도 85%, 적어도 90%, 적어도

95%, 적어도 98%, 적어도 99% 또는 100% 동일하다:

[0574] RVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEGLYNELQK
DKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR

[0575] 일부 실시형태에서, 1차 신호전달 도메인은 CD134(OX40) 신호전달 도메인이다. 일부 실시형태에서, CD134(OX40) 신호전달 도메인은 아래 제공되는 바와 같은 예시적인 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 동일하다:

[0576] RRDQRLPPDAHKPPGGGSFRTPIQEEQADAHSTLAKI

[0577] 일부 실시형태에서, 2차 또는 공동-자극, 신호전달 도메인은 4-1BB, CD2, CD4, CD28, CDS, CD8α, CD83, CD134, CD137, ICOS 또는 CD154로부터 유래된다. 일부 실시형태에서, 2차 신호전달 도메인은 CD28 신호전달 도메인이다. 일부 실시형태에서, CD28 신호전달 도메인은 아래 제공되는 바와 같은 예시적인 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 동일하다:

[0578] SKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS

[0579] 일부 실시형태에서, 2차 신호전달 도메인은 CD137(4-1BB) 신호전달 도메인이다. 일부 실시형태에서, CD137(4-1BB) 신호전달 도메인은 아래 제공되는 바와 같은 예시적인 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 동일하다:

[0580] KRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL

[0581] 일부 실시형태에서, CD137(4-1BB) 신호전달 도메인은 아래 제공되는 바와 같은 예시적인 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 동일하다:

[0582] RFSVVKRGRKKLLYIFKQPFMRPVQTTQEEDGCSCRFPEEEEGGCEL.

[0583] 일부 실시형태에서, CAR은 하나 이상의 신호전달 도메인을 포함한다. 일부 실시형태에서, CAR은 4-1BB 신호전달 도메인 및 CD3ζ 신호전달 도메인을 포함한다. 일부 실시형태에서, CAR은 CD28 신호전달 도메인 및 CD3ζ 신호전달 도메인을 포함한다.

[0584] ***면역 세포에서 표적 유전자의 편집***

[0585] 본 발명은 키메라 항원(CAR) 및 하나 이상의 편집된 유전자, 이들의 하나 이상의 조절 요소 또는 이들의 조합을 포함하는 면역 세포를 제공하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 CAR 및 동족 살해에 대한 저항성을 제공하고, 면역 세포 기능, 면역억제 또는 저해 또는 이들의 조합에 대한 저항성을 향상시키는 변경된 내인성 유전자를 포함한다. 일부 실시형태에서, 면역 세포는 T 세포이다. 일부 실시형태에서, 면역 세포는 CAR-T 세포이다. 일부 실시형태에서, 면역 세포는 NK 세포이다. 일부 실시형태에서, 각 편집된 유전자는 단일 염기 편집을 포함할 수 있다. 일부 실시형태에서, 각 편집된 유전자는 유전자의 상이한 영역에 다중 염기 편집을 포함할 수 있다.

[0586] 일부 실시형태에서, 단일 변형 이벤트(예컨대, 전기천공법)는 하나 이상의 유전자 편집을 도입할 수 있다. 일부 실시형태에서, 적어도 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개 이상의 편집이 하나 이상의 유전자에 동시에 도입될 수 있다. 일부 실시형태에서, 임의의 전술한 유전자 편집으로부터 선택되는 편집된 유전자를 포함하는 임의의 면역 세포를 포함하지만 이들로 제한되지 않는 면역 세포가 CAR-T의 기능을 향상시키거나 또는 세포의 면역억제 또는 저해를 감소시키는 다른 유전자에서 돌연변이를 생성하도록 편집될 수 있다.

[0587] 일부 실시형태에서, CAR-T 세포는 유사한 CAR-T 세포와 비교하여 증가된 동족 살해 저항성을 갖지만, 본 명세서에 기재된 바와 같은 하나 이상의 편집된 유전자를 추가로 갖지 않는다. 일부 실시형태에서, CAR-T 세포는 유사한 CAR-T 세포와 비교하여 감소된 면역원성을 갖지만, 본 명세서에 기재된 바와 같은 하나 이상의 편집된 유전자를 추가로 갖지 않는다. 일부 실시형태에서, CAR-T 세포는 유사한 CAR-T 세포와 비교하여 낮은 활성화 역치를 갖지만, 본 명세서에 기재된 바와 같은 하나 이상의 편집된 유전자를 추가로 갖지 않는다. 일부 실시형태에서, CAR-T 세포는 유사한 CAR-T 세포와 비교하여 증가된 항-신생물 활성을 갖지만, 본 명세서에 기재된 바와 같은 하나 이상의 편집된 유전자를 추가로 갖지 않는다.

[0588] 일부 실시형태에서, 내인성 유전자 또는 이의 조절 요소에 적어도 하나의 변형을 갖는 면역 세포가 본 명세서에 제공된다. 일부 실시형태에서, 면역 세포는 내인성 유전자 또는 이의 조절 요소에 적어도 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개 이상의 추가의 변형을 포함할 수 있다.

[0589] 일부 실시형태에서, 하나 이상의 유전자 또는 이들의 하나 이상의 조절 요소 또는 이들의 조합은 CD3 항원(CD3); CD5 항원(CD5); CD7 항원(CD7); CD33 항원(CD33); CD52 항원(CD52); CD123 항원(CD123); T 세포 수용체 알파 불변(TRAC); 예정 세포사 1(PDCD1 또는 PD-1); Fas 세포 표면 사멸 수용체(FAS); 림프구-활성화 유전자 3(LAG-3); 클래스 II 주요 조직적합성 복합체 전사촉진인자(CIITA); T 세포 수용체 베타 불변 1(TRBC1); T 세포 수용체 베타 불변 2(TRBC2); 및 베타-2 마이크로글로불린(B2M)으로 이루어진 군으로부터 선택될 수 있다. 일부 실시형태에서, CD3, CD5, CD7, CD33 또는 CD123은 편집된다. 일부 실시형태에서, 면역 세포는 편집된 *CD3* 유전자 및 추가적으로, 적어도 하나의 편집된 유전자를 포함한다. 일부 실시형태에서, 면역 세포는 편집된 *CD5* 유전자 및 추가적으로, 적어도 하나의 편집된 유전자를 포함한다. 일부 실시형태에서, 면역 세포는 편집된 *CD7* 유전자 및 추가적으로, 적어도 하나의 편집된 유전자를 포함한다. 일부 실시형태에서, 면역 세포는 편집된 *CD33* 유전자 및 추가적으로, 적어도 하나의 편집된 유전자를 포함한다. 일부 실시형태에서, 면역 세포는 편집된 *CD123* 유전자 및 추가적으로, 적어도 하나의 편집된 유전자를 포함한다. 적어도 하나의 편집된 유전자는 선행하는 단락에서 언급된 유전자의 목록으로부터 선택될 수 있다. 일부 실시형태에서, CD3, CD5, CD7, CD33 또는 CD123은 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 중 하나 이상과 조합하여 편집된다.

[0590] 다양한 실시형태에서, 변형된 면역 세포는 CD5, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD5, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 중 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD5에서의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 일 실시형태에서, 변형된 면역 세포는 CD5, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1에서의 돌연변이를 포함한다.

[0591] 다양한 실시형태에서, 변형된 면역 세포는 CD7, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD7, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 중 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD7에서의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 일 실시형태에서, 변형된 면역 세포는 CD7, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1에서의 돌연변이를 포함한다.

[0592] 다양한 실시형태에서, 변형된 면역 세포는 CD3, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD3, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 중 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD3에서의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 일 실시형태에서, 변형된 면역 세포는 CD3, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1에서의 돌연변이를 포함한다.

[0593] 다양한 실시형태에서, 변형된 면역 세포는 CD33, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD33, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 중 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD33에서의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 일 실시형태에서, 변형된 면역 세포는 CD33, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1에서의 돌연변이를 포함한다.

[0594] 다양한 실시형태에서, 변형된 면역 세포는 CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 중 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는

10개에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD123에서의 돌연변이 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 일 실시형태에서, 변형된 면역 세포는 CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PD1에서의 돌연변이를 포함한다.

[0595] 일부 실시형태에서, 적어도 하나의 변형은 단일 핵염기 변형이다. 일부 실시형태에서, 변경된 내인성 유전자는 염기 편집에 의해 생성될 수 있다. 일부 실시형태에서, 염기 편집은 유전자 발현을 감소시키거나 약화시킬 수 있다. 일부 실시형태에서, 염기 편집은 유전자 활성화를 감소시키거나 약화시킬 수 있다. 일부 실시형태에서, 염기 편집은 유전자 산물의 기능을 감소시키거나 약화시킬 수 있다. 일부 다른 실시형태에서, 염기 편집은 유전자 발현을 활성화하거나 또는 향상시킬 수 있다. 일부 실시형태에서, 염기 편집은 유전자 산물의 기능을 증가시킬 수 있다.

[0596] 내인성 면역 세포 수용체뿐만 아니라 키메라 항원 수용체를 발현하는 동종이계 면역 세포는 숙주 세포를 인식하여 공격할 수 있으며, 이 상황은 이식편 대 숙주 질환(GVHD)으로 지칭된다. 면역 세포 수용체 복합체의 알파 성분은 TRAC 유전자에 의해 암호화되며, 일부 실시형태에서, 이 유전자는 TCR 복합체의 알파 서브유닛이 기능하지 않거나 또는 존재하지 않도록 편집된다. 이 서브유닛은 내인성 면역 세포 신호전달에 필요하기 때문에, 이 유전자를 편집하면 동종이계 면역 세포에 의해 야기되는 이식편 대 숙주 질환의 위험을 줄일 수 있다.

[0597] 일부 실시형태에서, 동족 살해 저항성을 제공하거나, 면역 세포의 기능을 향상시키거나 또는 면역억제 또는 저해를 감소시키기 위한 유전자의 편집은 세포가 키메라 항원 수용체를 발현하도록 형질전환되기 전에 면역 세포에서 일어날 수 있다. 다른 양태에서, 면역 세포의 기능을 향상시키거나 또는 면역억제 또는 저해를 감소시키기 위한 유전자의 편집은, 즉, 면역 세포가 키메라 항원 수용체를 발현하도록 형질전환된 후에 CAR-T 세포에서 일어날 수 있다.

[0598] 본 발명의 일부 실시형태에서, *CD5* 유전자가 넉아웃 또는 넉다운 발현을 위해 CAR-T 세포에서 편집된다. 그런 다음, CAR-T는 CD5 scFv를 갖는 키메라 항원 수용체를 발현하도록 형질전환된다. *CD5* 유전자의 발현을 넉아웃 또는 넉다운시킴으로써, 변형된 CAR-T 세포가 동족 살해를 일으킬 가능성이 줄어든다.

[0599] 본 발명의 일부 실시형태에서, *CD7* 유전자가 넉아웃 또는 넉다운 발현을 위해 CAR-T 세포에서 편집된다. 그런 다음, CAR-T는 CD7 scFv를 갖는 키메라 항원 수용체를 발현하도록 형질전환된다. *CD7* 유전자의 발현을 넉아웃 또는 넉다운시킴으로써, 변형된 CAR-T 세포가 동족 살해를 일으킬 가능성이 줄어든다.

[0600] 본 발명의 일부 실시형태에서, *CD33* 유전자가 넉아웃 또는 넉다운 발현을 위해 CAR-T 세포에서 편집된다. 그런 다음, CAR-T는 CD33 scFv를 갖는 키메라 항원 수용체를 발현하도록 형질전환된다. *CD33* 유전자의 발현을 넉아웃 또는 넉다운시킴으로써, 변형된 CAR-T 세포가 동족 살해를 일으킬 가능성이 줄어든다.

[0601] 본 발명의 일부 실시형태에서, *CD3* 유전자가 넉아웃 또는 넉다운 발현을 위해 CAR-T 세포에서 편집된다. 그런 다음, CAR-T는 CD3 scFv을 갖는 키메라 항원 수용체를 발현하도록 형질전환된다. *CD3* 유전자의 발현을 넉아웃 또는 넉다운시킴으로써, 변형된 CAR-T 세포가 동족 살해를 일으킬 가능성이 줄어든다.

[0602] 본 발명의 일부 실시형태에서, *CD123* 유전자가 넉아웃 또는 넉다운 발현을 위해 CAR-T 세포에서 편집된다. 그런 다음, CAR-T는 CD123 scFv를 갖는 키메라 항원 수용체를 발현하도록 형질전환된다. *CD123* 유전자의 발현을 넉아웃 또는 넉다운시킴으로써, 변형된 CAR-T 세포가 동족 살해를 일으킬 가능성이 줄어든다.

[0603] 숙주 면역 세포는 잠재적으로 동종이계 CAR-T 세포를 비-자기로 인식하고, 비-자기 세포를 제거하기 위해 면역 반응을 유도할 수 있다. B2M은 거의 모든 유핵 세포에서 발현되며, MHC 클래스 I 복합체와 관련이 있다(도 1b). 순환 숙주 $CD8^+$ T 세포는 이러한 B2M 단백질을 비-자기로 인식하여 동종이계 세포를 죽일 수 있다. 이러한 이식 거부를 극복하기 위해, 일부 실시형태에서, *B2M* 유전자는 발현을 넉아웃 또는 넉다운시키도록 편집된다. 일부 실시형태에서, 면역 세포가 내인성 기능적 B2M을 발현하지 않도록 편집된 *B2M* 유전자를 갖는 면역 세포가 본 명세서에 제공된다. 일부 실시형태에서, CAR-T 세포가 내인성 B2M의 감소되거나 또는 무시해도 될 정도의 발현을 나타내거나 또는 이를 전혀 발현하지 않도록 편집된 *B2M* 유전자를 갖는 CAR-T 세포가 본 명세서에 제공된다.

[0604] 일부 실시형태에서, 면역 세포는 키메라 항원 수용체 및 하나 이상의 편집된 유전자, 이들의 조절 요소 또는 이들의 조합을 포함한다. 편집된 유전자는 면역 반응 조절 유전자, 면역원성 유전자, 관문 저해제 유전자, 면역 반응에 관련된 유전자, 세포 표면 마커, 예를 들어, T 세포 표면 마커 또는 이들의 임의의 조합일 수 있다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체, 및 활성화된 T 세포 증식, 알파-베타 T 세포 활성화, 감마-

델타 T 세포 활성화, T 세포 증식의 양성 조절, T-헬퍼 세포 증식 또는 분화의 음성 조절, 또는 이들의 조절 요소 또는 이들의 조합과 관련된 편집된 유전자를 포함한다. 일부 실시형태에서, 편집된 유전자는 관문 저해제 유전자, 예를 들어, 예컨대, *PD1* 유전자, *PDC1* 유전자, 또는 이들의 형성 또는 활성화 경로와 관련되거나 또는 이를 조절하는 구성원일 수 있다. 일부 실시형태에서, 편집된 유전자는 *TRAC* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD5* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD7* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD33* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD3* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD123* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *B2M* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CIITA* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *TRBC1/2* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD5* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD7* 유전자이다. 일부 실시형태에서, 편집된 유전자는 *CD52* 유전자이다. 일부 실시형태에서, *PD-1*, *CD2*, *CD3*, *CD5*, *CD7*, *CD52*, *B2M*, *TRBC1/2*, *CIITA* 및 *TRAC* 또는 이들의 조합으로부터 선택되는 적어도 하나의 유전자가 편집된다. 일부 실시형태에서, *PD-1*, *CD2*, *CD52* 및 *TRAC* 유전자가 편집된다. 일부 실시형태에서, *PD-1*, *CD2*, *CD52*, *B2M*, *TRBC1/2*, *CIITA* 및 *TRAC* 유전자가 편집된다. 일부 실시형태에서, *PD-1*, *CD5*, *CD52* 및 *TRAC* 유전자가 편집된다. 일부 실시형태에서, *PD-1*, *CD3*, *CD7* 및 *CD52* 유전자가 편집된다.

[0605] 일부 실시형태에서, 내인성 유전자의 편집은 유전자의 발현을 감소시킨다. 일부 실시형태에서, 내인성 유전자의 편집은 변형이 없는 대조군 세포와 비교하여 유전자의 발현을 적어도 50%만큼 감소시킨다. 일부 실시형태에서, 내인성 유전자의 편집은 변형이 없는 대조군 세포와 비교하여 유전자의 발현을 적어도 60%만큼 감소시킨다. 일부 실시형태에서, 내인성 유전자의 편집은 변형이 없는 대조군 세포와 비교하여 유전자의 발현을 적어도 70%만큼 감소시킨다. 일부 실시형태에서, 내인성 유전자의 편집은 변형이 없는 대조군 세포와 비교하여 유전자의 발현을 적어도 80%만큼 감소시킨다. 일부 실시형태에서, 내인성 유전자의 편집은 변형이 없는 대조군 세포와 비교하여 유전자의 발현을 적어도 90%만큼 감소시킨다. 일부 실시형태에서, 내인성 유전자의 편집은 변형이 없는 대조군 세포와 비교하여 유전자의 발현을 적어도 100%만큼 감소시킨다. 일부 실시형태에서, 내인성 유전자의 편집은 유전자 발현을 제거한다.

[0606] 본 발명의 일부 실시형태에서, *PDCD1* 유전자는 넉아웃 또는 넉다운 발현을 위해 CAR-T 세포에서 편집된다. *PDCD1* 유전자는 면역 세포에서 발현되는 면역계 관문인 세포 표면 수용체 PD-1을 암호화하며, 이는 항원 특이적 면역 세포의 세포자멸사를 촉진시킴으로써 자가면역을 감소시키는데 관여한다. *PDCD1* 유전자의 발현을 넉아웃 또는 넉다운시킴으로써, 변형된 CAR-T 세포가 세포자멸사 일으킬 가능성이 적고, 이의 증식 가능성이 높으며, 예정 세포사 면역 관문을 벗어날 수 있다.

[0607] CBLB 유전자는 면역 효과기 세포 활성화를 저해하는데 중요한 역할을 하는 E3 유비퀴틴 라이게이스를 암호화한다. **도 1c**을 참조하면, *CBLB* 단백질은 면역 효과기 세포 내성을 초래하는 신호전달 경로를 선호하며, 면역 효과기 세포 활성화를 유도하는 신호전달을 능동적으로 저해한다. 면역 효과기 세포 활성화는 CAR-T 세포가 이식 후 생체내에서 증식하기 위해 필요하기 때문에, 본 발명의 일부 실시형태에서, *CBLB* 유전자는 발현을 넉아웃 또는 넉다운시키도록 편집된다.

[0608] 일부 실시형태에서, 면역 세포가 내인성 기능적 T 세포 수용체 알파 사슬을 발현하지 않도록 편집된 *TRAC* 유전자(여기서, T*RAC* 유전자는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 이상의 염기 편집을 포함할 수 있음)를 갖는 면역 세포가 본 명세서에 제공된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체(CAR-T 세포)를 발현하는 T 세포이다. 일부 실시형태에서, CAR-T 세포가 내인성 T 세포 수용체 알파 단백질의 감소되거나 또는 무시해도 될 정도의 발현을 나타내거나 또는 이를 전혀 발현하지 않도록 *TRAC* 유전자에 염기 편집이 있는 CAR-T 세포가 본 명세서에 제공된다.

[0609] 일부 실시형태에서, 면역 세포가 내인성 기능적 클래스 II, 주요 조직적합성 복합체, 전사촉진인자(trans활성인자)를 발현하지 않도록 편집된 *CIITA* 유전자를 갖는 면역 세포가 본 명세서에 제공된다. 일부 실시형태에서, CAR-T 세포가 내인성 클래스 II, 주요 조직적합성 복합체, 전사촉진인자의 감소되거나 또는 무시해도 될 정도의 발현을 나타내거나 또는 이를 전혀 발현하지 않도록 편집된 *CIITA* 유전자를 갖는 CAR-T 세포가 본 명세서에 제공된다.

[0610] 일부 실시형태에서, 면역 세포가 내인성 기능적 T 세포 수용체 베타 사슬을 발현하지 않도록 편집된 *TRBC1* 또는 *TRBC2* 유전자를 갖는 면역 세포가 본 명세서에 제공된다. 일부 실시형태에서, CAR-T 세포가 내인성 T 세포 수용체 베타 사슬의 감소되거나 또는 무시해도 될 정도의 발현을 나타내거나 또는 이를 전혀 발현하지 않도록 편집된 *TRBC1/TRBC2* 유전자를 갖는 CAR-T 세포가 본 명세서에 제공된다.

[0611] 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRAC, B2M, PDCD1, CBLB 유전자 또는 이들의 조합을 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRAC 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRAC 및 B2M 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRAC 및 PDCD1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRAC 및 CBLB 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRAC, B2M 및 PDCD1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRAC, B2M 및 CBLB 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포 또는 면역 효과기 세포는 키메라 항원 수용체와 편집된 TRAC, PDCD1 및 CBLB 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원과 편집된 TRAC, B2M, PDCD1 및 CBLB 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 B2M 유전자이되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 B2M 및 PDCD1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 B2M 및 CBLB 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 B2M, PDCD1 및 CBLB 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 PDCD 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 PDCD 및 CBLB 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CBLB를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다.

[0612] 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CD5 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CD7 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CD33 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CD3 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CD123 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 FAS 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 LAG-3 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CIITA 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRBC1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TRBC2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 CD52 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 하나 이상의 편집된 CD3, CD5, CD7, CD33, CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및/또는 PD1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다.

[0613] 염기 편집은 유전자의 임의의 적합한 위치 또는 유전자의 조절 요소에 위치할 수 있다. 따라서, 예를 들어, 개시 코돈에서 단일 염기 편집은 유전자의 발현을 완전히 제거할 수 있음을 이해할 수 있다. 일부 실시형태에서, 변경된 내인성 유전자는 엑손, 인트론, 엑손-인트론 접합부 또는 이들의 조절 요소에서 변형 또는 편집될 수 있다. 변형은 유전자 또는 이의 조절 요소에서 단일 핵염기에 대해 편집될 수 있다. 변형은 엑손, 하나 이상의 엑손, 인트론 또는 하나 이상의 인트론, 또는 이들의 조합에 있을 수 있다. 변형은 유전자의 오픈 리딩 프레임에 있을 수 있다. 변형은 유전자의 비번역 영역, 예를 들어, 3'-UTR 또는 5'-UTR에 있을 수 있다. 일부 실시형태에서, 변형은 내인성 유전자의 조절 요소에 있다. 일부 실시형태에서, 변형은 프로모터, 인핸서, 작동자, 침묵자, 절연인자, 종결자, 전사 개시 서열, 번역 개시 서열(예를 들어, 코작 서열) 또는 이들의 임의의 조합에 있다. 일부 실시형태에서, 염기 편집은 조기 종결 코돈을 엑손에 도입하여 번역된 생성물의 결핍을 초래하거나, 또는 잘못 접혀서 분해에 의해 제거될 수 있는 절단을 초래하거나 또는 쉽게 분해되는 불안정한 mRNA를 생성할 수 있

다.

[0614] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *PDC1/PD-1* 유전자의 엑손 1, 엑손 2 또는 엑손 3 또는 엑손 4 또는 엑손 5에서 수행될 수 있다. 일부 실시형태에서, 인간 *PDC1/PD-1* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1* 유전자의 염기 편집은 엑손 5 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5 또는 이들의 임의의 조합에서 인간 *PDC1/PD-1* 유전자에 대해 수행될 수 있다.

[0615] 일부 실시형태에서, 인간 *PDC1/PD-1A* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번, 7번, 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1A* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번, 7번, 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1A* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 7번, 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1A* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 5번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *PDC1/PD-1A* 유전자의 염기 편집은 엑손 5를 표적화하는 가이드 RNA 스페이서 서열의 5번 또는 8번 위치를 편집함으로써 수행된다.

[0616] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *CD7* 유전자의 엑손 1, 엑손 2 또는 엑손 3에서 수행될 수 있다. 일부 실시형태에서, 인간 *CD7* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD7* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD7* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3 또는 이들의 임의의 조합에서 인간 *CD7* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *CD7* 유전자의 염기 편집은 엑손 1 내의 4번, 8번, 9번 위치에서 수행된다. 일부 실시형태에서, 인간 *CD7* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번, 7번, 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD7* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 9번 위치를 편집함으로써 수행된다.

[0617] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *LAG-3* 유전자의 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 6, 엑손 7 또는 엑손 8에서 수행될 수 있다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 5 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 6 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 7 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 8 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 6, 엑손 7, 엑손 8 또는 이들의 임의의 조합에서 인간 *LAG-3* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번, 6번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 4를 표적화하는 가이드 RNA 스페이서 서열의 4번, 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 5를 표적화하는 가이드 RNA 스페이서 서열의 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 6을 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 7을 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *LAG-3* 유전자의 염기 편집은 엑손 8을 표적화하는 가이드 RNA 스페이서 서열의 8번 위치를 편집함으로써 수행된다.

[0618] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *CD33* 유전자의 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5 또는 엑손 6에서 수행될 수 있다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 1 내의 부위에서 수

행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 5 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 6 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 6 또는 이들의 임의의 조합에서 인간 *CD33* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 7번, 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번, 6번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번, 6번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 4를 표적화하는 가이드 RNA 스페이서 서열의 6번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 5를 표적화하는 가이드 RNA 스페이서 서열의 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD33* 유전자의 염기 편집은 엑손 6을 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번 또는 6번 위치를 편집함으로써 수행된다.

[0619] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *CD123* 유전자의 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 7, 엑손 8, 엑손 10 또는 엑손 11에서 수행될 수 있다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 5 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 7 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 8 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 10 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 11 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 7, 엑손 8, 엑손 10, 엑손 11 또는 이들의 임의의 조합에서 인간 *CD123* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 4를 표적화하는 가이드 RNA 스페이서 서열의 5번 또는 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 5를 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 5번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 7을 표적화하는 가이드 RNA 스페이서 서열의 5번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 8을 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 10을 표적화하는 가이드 RNA 스페이서 서열의 4번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD123* 유전자의 염기 편집은 엑손 11을 표적화하는 가이드 RNA 스페이서 서열의 5번 또는 8번 위치를 편집함으로써 수행된다.

[0620] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *FAS* 유전자의 엑손 1 또는 엑손 3 또는 엑손 4 또는 엑손 5 또는 엑손 6 또는 엑손 7 또는 엑손 8 또는 엑손 9에서 수행될 수 있다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 5 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *FAS*의 염기 편집은 엑손 6 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *FAS*의 염기 편집은 엑손 7 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *FAS*의 염기 편집은 엑손 8 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *FAS*의 염기 편집은 엑손 9 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 3, 엑손 4, 엑손 5, 엑손 6, 엑손 7, 엑손 8, 엑손 9 또는 이들의 임의의 조합에서 인간 *FAS* 유전자에 대해 수행될 수 있다.

[0621] 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은

엑손 4를 표적화하는 가이드 RNA 스페이서 서열의 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 5를 표적화하는 가이드 RNA 스페이서 서열의 5번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 6을 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 7을 표적화하는 가이드 RNA 스페이서 서열의 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 8을 표적화하는 가이드 RNA 스페이서 서열의 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *FAS* 유전자의 염기 편집은 엑손 9를 표적화하는 가이드 RNA 스페이서 서열의 5번 또는 6번 위치를 편집함으로써 수행된다.

[0622] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *CD52* 유전자의 엑손 1, 엑손 2 또는 엑손 3에서 수행될 수 있다. 일부 실시형태에서, 인간 *CD52* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD52* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD7* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3 또는 이들의 임의의 조합에서 인간 *CD52* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *CD52* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD52* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번 또는 7번 위치를 편집함으로써 수행된다.

[0623] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *CD5* 유전자의 엑손 1, 엑손 2 또는 엑손 3 또는 엑손 4 또는 엑손 5 또는 엑손 6 또는 엑손 7 또는 엑손 8 또는 엑손 9 또는 엑손 10에서 수행될 수 있다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 5 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 6 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 7 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 8 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 9 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 10 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 6, 엑손 7, 엑손 8, 엑손 9, 엑손 10 또는 이들의 임의의 조합에서 인간 *CD5* 유전자에 대해 수행될 수 있다.

[0624] 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 5번 및/또는 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 5번 및/또는 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 5, 6, 8 및/또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 4를 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 5번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 5를 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번, 7번, 8번 또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 6을 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번, 6번, 7번, 8번 및/또는 9번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 7을 표적화하는 가이드 RNA 스페이서 서열의 4번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 8을 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 9를 표적화하는 가이드 RNA 스페이서 서열의 6번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CD5* 유전자의 염기 편집은 엑손 10을 표적화하는 가이드 RNA 스페이서 서열의 9번 위치를 편집함으로써 수행된다.

[0625] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *TRAC* 유전자(UCSC 게놈 데이터베이스 ENSG00000277734.8)의 엑손 1 또는 엑손 2 또는 엑손 3 또는 엑손 4에서 수행될 수 있다. 일부 실시형태에서, 인간 *TRAC* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *TRAC* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *TRAC* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부

실시형태에서, 인간 *TRAC* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4 또는 이들의 임의의 조합에서 인간 *TRAC* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *TRAC* 유전자의 염기 편집은 엑손 1 내에서 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번 또는 9번 위치를 편집함으로써 수행된다.

[0626] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *B2M* 유전자(염색체 15, NC_000015.10, 44711492 내지 44718877; 예시적인 mRNA 서열 NM_004048)의 엑손 1, 엑손 2, 엑손 3 또는 엑손 4에서 수행될 수 있다. 일부 실시형태에서, 인간 *B2M* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *B2M* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *B2M* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *B2M* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4 또는 이들의 임의의 조합에서 인간 B2M 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *B2M* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 5번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *B2M* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번 또는 9번 위치를 편집함으로써 수행된다.

[0627] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *CIITA* 유전자의 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 6, 엑손 7, 엑손 8, 엑손 9, 엑손 10, 엑손 11, 엑손 12, 엑손 13, 엑손 14, 엑손 15, 엑손 16, 엑손 17, 엑손 18 또는 엑손 19에서 수행될 수 있다. 일부 실시형태에서, 인간 *CD52* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 4 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 5 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 6 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 7 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 8 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 9 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 10 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 11 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 12 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 13 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 14 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 15 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 16 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 17 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 18 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 19 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3, 엑손 4, 엑손 5, 엑손 6, 엑손 7, 엑손 8, 엑손 9, 엑손 10, 엑손 11, 엑손 12, 엑손 13, 엑손 14, 엑손 15, 엑손 16, 엑손 17, 엑손 18, 엑손 19 또는 이들의 임의의 조합에서 인간 *CIITA* 유전자에 대해 수행될 수 있다.

[0628] 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 6번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 4를 표적화하는 가이드 RNA 스페이서 서열의 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 7을 표적화하는 가이드 RNA 스페이서 서열의 4번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 8을 표적화하는 가이드 RNA 스페이서 서열의 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 9를 표적화하는 가이드 RNA 스페이서 서열의 4번, 6번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 10을 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번 또는 7번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 11을 표적화하는 가이드 RNA 스페이서 서열의 4번, 5번, 6번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 12를 표적화하는 가이드 RNA 스페이서 서열의 6번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 14를 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 5번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 15를 표적화하는 가이드 RNA 스페이서 서열의 4번, 7번 또는 8번 위치를 편집함으로써 수행된

다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 16을 표적화하는 가이드 RNA 스페이서 서열의 5번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 17을 표적화하는 가이드 RNA 스페이서 서열의 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 18을 표적화하는 가이드 RNA 스페이서 서열의 5번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *CIITA* 유전자의 염기 편집은 엑손 19를 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번, 7번 또는 8번 위치를 편집함으로써 수행된다.

[0629] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *TRBC1* 유전자의 엑손 1, 엑손 2 또는 엑손 3에서 수행될 수 있다. 일부 실시형태에서, 인간 *TRBC1* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *TRBC1* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *TRBC1* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3 또는 이들의 임의의 조합에서 인간 *TRBC1* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *TRBC1* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *TRBC1* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *TRBC1* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 4번 또는 5번 위치를 편집함으로써 수행된다.

[0630] 일부 실시형태에서, 염기 편집은, 예를 들어, 인간 *TRBC2* 유전자의 엑손 1, 엑손 2 또는 엑손 3에서 수행될 수 있다. 일부 실시형태에서, 인간 *TRBC2* 유전자의 염기 편집은 엑손 1 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *TRBC2* 유전자의 염기 편집은 엑손 2 내의 부위에서 수행된다. 일부 실시형태에서, 인간 *TRBC2* 유전자의 염기 편집은 엑손 3 내의 부위에서 수행된다. 일부 실시형태에서, 하나 이상의 염기 편집 작용이 엑손 1, 엑손 2, 엑손 3 또는 이들의 임의의 조합에서 인간 *TRBC2* 유전자에 대해 수행될 수 있다. 일부 실시형태에서, 인간 *TRBC2* 유전자의 염기 편집은 엑손 1을 표적화하는 가이드 RNA 스페이서 서열의 5번, 6번, 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *TRBC2* 유전자의 염기 편집은 엑손 2를 표적화하는 가이드 RNA 스페이서 서열의 7번 또는 8번 위치를 편집함으로써 수행된다. 일부 실시형태에서, 인간 *TRBC2* 유전자의 염기 편집은 엑손 3을 표적화하는 가이드 RNA 스페이서 서열의 4번 위치를 편집함으로써 수행된다.

[0631] 일부 실시형태에서, 염기 편집은 인트론에서 수행될 수 있다. 예를 들어, 염기 편집은 인트론에서 수행될 수 있다. 일부 실시형태에서, 염기 편집은 인트론 내의 부위에서 수행될 수 있다. 일부 실시형태에서, 염기 편집은 부위 하나 이상의 인트론 부위에서 수행될 수 있다. 일부 실시형태에서, 염기 편집은 유전자의 다중 인트론 중 임의의 엑손에서 수행될 수 있다. 일부 실시형태에서, 하나 이상의 염기 편집은 엑손, 인트론 또는 엑손과 인트론의 임의의 조합에서 수행될 수 있다.

[0632] 일부 실시형태에서, 변형 또는 염기 편집은 프로모터 부위 내에 있을 수 있다. 일부 실시형태에서, 염기 편집은 대체 프로모터 부위 내에 도입될 수 있다. 일부 실시형태에서, 염기 편집은 인핸서와 같은 5' 조절 요소에 있을 수 있다. 일부 실시형태에서, 염기 편집은 핵산 결합 단백질의 결합 부위를 파괴하기 위해 도입될 수 있다. 예시적인 핵산 결합 단백질은 다른 많은 것들 중에서 폴리머레이스, 뉴클레이스, 자이레이스, 토포아이소머레이스, 메틸레이스 또는 메틸 트랜스퍼레이스, 전사 인자, 인핸서, PABP, 아연 핑거 단백질일 수 있다.

[0633] 일부 실시형태에서, 염기 편집은 표적 단백질 발현을 침묵시키기 위한 스플라이스 파괴에 사용될 수 있다. 일부 실시형태에서, 염기 편집은 스플라이스 수용체-스플라이스 공여체(splice acceptor-splice donor: SA-SD) 부위를 생성할 수 있다. SA-SD를 생성하거나 또는 SA-SD 부위에서 표적화된 염기 편집은 유전자의 발현을 감소시킬 수 있다. 일부 실시형태에서, 염기 편집기(예를 들어, ABE, CBE)는 각 인트론의 처음 및 마지막 두 뉴클레오타이드인 스플라이스 수용체 및 스플라이스 공여체 부위를 구성하는 다이뉴클레오타이드 모티프를 표적화하는데 사용된다. 예를 들어, *CD2*의 엑손 3 스플라이스 공여체(SD) 부위는 염기 편집을 위해 표적화될 수 있다. 일부 실시형태에서, 스플라이스 파괴는 아데노신 염기 편집기(ABE)로 달성된다. 일부 실시형태에서, 스플라이스 파괴는 사이티딘 염기 편집기(CBE)로 달성된다. 일부 실시형태에서, 염기 편집기(예를 들어, ABE, CBE)는 종결 코돈을 생성하여 엑손을 편집하는데 사용된다. 조기 종결 코돈은 엑손, 인트론 또는 비번역 영역에서 생성될 수 있다. 일부 실시형태에서, 염기 편집은 하나 이상의 대체 리딩 프레임에 하나 이상의 종결 코돈을 도입하는데 사용될 수 있다. 예를 들어, 조기 종결 코돈은 엑손 2 내의 8번 위치, 엑손 3 내의 4번 위치, 엑손 3 내의 6번 위치, 엑손 3 내의 9번 위치, 엑손 4 내의 4번 위치, 엑손 4 내의 5번 위치 또는 엑손 5 내의 4번 위치에 도입될 수 있다. 일부 실시형태에서, 종결 코돈은 아데노신 염기 편집기(ABE)에 의해 생성된다. 일부 실시형태에서, 종결 코돈은 사이티딘 염기 편집기(CBE)에 의해 생성된다. 일부 실시형태에서, CBE는 종결 코돈을 생성하기 위해

다음 편집(밑줄 친 글꼴로 표시) 중 어느 하나를 생성한다: CAG→TAG; CAA→TAA; CGA→TGA; TGG→TGA; TGG→TAG; 또는 TGG→TAA.

[0634] 일부 실시형태에서, 변형/염기 편집은, 예를 들어, 폴리아데닐화(폴리-A) 부위에서 3'-UTR에 도입될 수 있다. 일부 실시형태에서, 염기 편집은 5'-UTR 영역에서 수행될 수 있다.

[0635] 일부 실시형태에서, 임의의 전술한 유전자 편집을 포함하는 임의의 면역 세포를 포함하지만 이들로 제한되지 않는 면역 세포는 CAR-T의 기능을 향상시키거나 또는 세포의 면역억제 또는 저해를 감소시키는 다른 유전자에서 돌연변이를 생성하도록 편집될 수 있다. 예를 들어, 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2, ZAP70, NFATc1, TET2 유전자 또는 이들의 조합을 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2 및 ZAP70 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2 및 ZAP70 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2 및 NFATC1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2 및 TET2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2, ZAP70 및 NFATC1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2, ZAP70 및 TET2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TGFBR2, NFATC1 및 TET2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 및 편집된 TGFBR2, ZAP70, NFATC1 및 TET2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 ZAP70 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 ZAP70 및 NFATC1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 ZAP70 및 TET2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 ZAP70, PDCD1 및 TET2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 PDCD1 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 PDCD1 및 TET2 유전자를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 면역 세포는 키메라 항원 수용체와 편집된 TET2를 포함하되, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다.

[0636] 일부 실시형태에서, 키메라 항원 수용체는 TRAC 유전자에 삽입된다. 이는 이점이 있다. 첫째, TRAC는 면역 세포에서 고도로 발현되기 때문에, 키메라 항원 수용체는 수용체의 발현이 TRAC 프로모터에 의해 구동되도록 이의 작제물이 키메라 항원 수용체를 TRAC 유전자에 삽입하도록 설계될 때 유사하게 발현될 것이다. 둘째, 키메라 항원 수용체를 TRAC 유전자로 삽입하는 것은 TRAC 발현을 넉아웃시킬 것이다. 일부 실시형태에서, 본 명세서에 기재된 유전자 편집 시스템이 키메라 항원 수용체를 TRAC 유전자좌로 삽입하는데 사용될 수 있다. TRAC 유전자좌에 특이적인 gRNA는 가이드 유전자 편집 시스템을 유전자좌로 가이드하고, 이중-가닥 DNA 절단을 개시할 수 있다. 특정 실시형태에서, gRNA는 Cas12b와 함께 사용된다. 다양한 실시형태에서, 유전자 편집 시스템은 CAR 수용체를 암호화하는 서열을 갖는 핵산과 함께 사용된다. 예시적인 가이드 RNA가 다음 표 1A에 제공된다.

표 1

[표 1A]

가이드 RNA

| | | | |
|---|---|---|---|
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACAGAGUCUCUCAGCUGGUACAC | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 1 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGGAAACUCCUAUUGCUGGACGAUGUCUCUUACG AGGCAUUAGCACACCGAUUUUGAUUCUCAAACA | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 2 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACUCAAACAAAUGUGCACAAAG | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 3 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACUCAAACAAAUGUGUCACAAAG | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 4 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACUUUGAGAAUCAAAAUCGGUA | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 5 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACUGAUGUGUAUAUCACAGACAA | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 6 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCAGUUGCUCCAGGCCACAGCAU | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 7 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACUUCCAGAAGACACCUUCUUCC | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 8 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACCAGAAGACACCUUCUUCCCCA | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 9 (엑손 2) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAGAAACUCCUAUUGCUGGACGAUGUCUCUUAC GAGGCAUUAGCACGGUUCCGAAUCCUCCUGA | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 10 (엑손 4) |
| GUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUG CAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGA GGCAUUAGCACGGAACCCAAUCACUGACAGGU | ATTN | BhCas12b 뉴클레이스 | TRAC KO gRNA 11 (엑손 4) |

[0637]

[0638] 키메라 항원 수용체 및 gRNA 표적화 서열 측면에 있는 TRAC DNA의 연장된 신장부를 포함하는 핵산을 암호화하는 DNA 작제물. 이론에 얽매이지 않고, 작제물은 상보적 TRAC 서열에 결합하며, 작제물의 TRAC 서열에 근접하여 존재하는 키메라 항원 수용체 DNA는 그 뒤에 병변 부위에 삽입되어, TRAC 유전자를 효과적으로 넉아웃시키고 키메라 항원 수용체 핵산을 넉인시킨다. 표 1B는 키메라 항원 수용체 핵산의 삽입을 가능하게 하는 TRAC 유전자좌로 염기 편집 기구를 가이드할 수 있는 TRAC 유전자에 대한 가이드 RNA를 제공한다. 처음 11개의 gRNA는 BhCas12b 뉴클레이스에 대한 것이다. 11개의 두 번째 세트는 BvCas12b 뉴클레이스에 대한 것이다. 이들은 모두 염기 편집이 아니라 이중 가닥 파손을 생성하여 TRAC에 CAR을 삽입하기 위한 것이다.

[표 1B]

TRAC 가이드 RNA

| 가이드 RNA | 유전자 엑손 |
|---|---|
| **GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACA**GAGTCTCTCAGCTGGT ACA | TRAC KO gRNA 1 |
| **GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACA**CCGATTTTGATTCTCA AAC | TRAC KO gRNA 2 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACTGATTCTCAAACAAATG TGT | TRAC KO gRNA 3 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAG**CAC**TCAAACAAATGTGTCAC AAA | TRAC KO gRNA 4 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGTTTGAGAATCAAAATC GGT | TRAC KO gRNA 5 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACTGATGTGTATATCACAG ACA | TRAC KO gRNA 6 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGTTGCTCCAGGCCACAG CAC | TRAC KO gRNA 7 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACTTCCAGAAGACACCTTC TTC | TRAC KO gRNA 8 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACCAGAAGACACCTTCTTC CCC | TRAC KO gRNA 9 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGGTTCCGAATCCTCCTC CTG | TRAC KO gRNA 10 |
| GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAA CTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACGGAACCCAATCACTGAC AGG | TRAC KO gRNA 11 |
| **GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACA**GAGTCTCTCAGCTGGTACA | TRAC KO gRNA 1 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACACCGATTTTGATTCTCAAAC | TRAC KO gRNA 2 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACTGATTCTCAAACAAATGTGT | TRAC KO gRNA 3 |

[0639]

| 가이드 RNA | 유전자 엑손 |
|---|---|
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACTCAAACAAATGTGTCACAAA | TRAC KO gRNA 4 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACGTTTGAGAATCAAAATCGGT | TRAC KO gRNA 5 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACTGATGTGTATATCACAGACA | TRAC KO gRNA 6 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACGTTGCTCCAGGCCACAGCAC | TRAC KO gRNA 7 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACTTCCAGAAGACACCTTCTTC | TRAC KO gRNA 8 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACCAGAAGACACCTTCTTCCCC | TRAC KO gRNA 9 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACGGTTCCGAATCCTCCTCCTG | TRAC KO gRNA 10 |
| GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACCCACC ACAGGAGCACCTGAAAACAGGTGCTTGGCACGGAACCCAATCACTGACAGG | TRAC KO gRNA 11 |

[0640]

[0641] 처음 11개의 gRNA는 BhCas12b 뉴클레이스에 대한 것이다. 11개의 gRNA의 두 번째 세트는 BvCas12b 뉴클레이스에 대한 것이다. 첫 번째 예에서 스캐폴드 서열은 굵게 표시되어 있다.

[0642] 일부 실시형태에서, 본 발명의 키메라 항원 수용체를 암호화하는 핵산은 BE4 염기 편집기를 사용하여 TRAC 유전자좌에 표적화될 수 있다. 일부 실시형태에서, 키메라 항원 수용체는 CRISPR/Cas9 염기 편집 시스템을 사용하여 TRAC 유전자좌에 표적화된다.

[0643] 위에 기재된 유전자 편집을 생성하기 위해, 면역 세포는 대상체로부터 수집되고, 2개 이상의 가이드 RNA 및 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)과 사이티딘 데아미네이스 또는 아데노신 데아미네이스를 포함하는 핵염기 편집기 폴리펩타이드와 접촉된다. 일부 실시형태에서, 수집된 면역 세포는 적어도 하나의 핵산과 접촉되되, 적어도 하나의 핵산은 2개 이상의 가이드 RNA, 및 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)과 사이티딘 데아미네이스를 포함하는 핵염기 편집기 폴리펩타이드를 암호화한다. 일부 실시형태에서, gRNA는 뉴클레오타이드 유사체를 포함한다. 이러한 뉴클레오타이드 유사체는 세포 처리로부터 gRNA의 분해를 저해할 수 있다. 표 2는 gRNA에 사용되는 표적 서열을 제공한다.

표 2

예시적인 표적 서열

| 표적 단백질 | 표적 잔기 | gRNA 표적 | 염기 편집 | 코돈 변화 | 잔기 기능 |
|---|---|---|---|---|---|
| NFATC1 | R118 | CTCGATGCGAGGACTCTCCA | BE | CGC>CAC | 칼시뉴린 결합 |
| | I119 | TCTCGATGCGAGGACTCTCC | ABE | ATC>ACC | 칼시뉴린 결합 |
| | E120 | CATCGAGATAACCTCGTGCT | ABE | GAG>GGG | 칼시뉴린 결합 |
| | S172 | TGGCCGGGCTCAGGCACGAG | BE | AGC>AAC | 인산화 |
| | W396 | GCCCACTGGTAGGGGTGCTG | ABE | TGG>CGG | 칼시뉴린 결합 |
| | R439 | TGGGCTCGGTGGTGGGACTT | BE | CGA>CAA | DNA 결합 |
| | H441 | CGAGCCCACTACGAGACGGA | ABE | CAC>CGC | DNA 결합 |
| | Y442 | CTCGTAGTGGGCTCGGTGGT | ABE | TAC>CAC | DNA 결합 |
| | K452 | GCCGTGAAGGCGTCGGCCGG | ABE | AAG>GGG | DNA 결합 |
| | R540 | GTTTCTGAGTTTCAGGATTC | BE | AGA>AAA | DNA 결합 |
| | R555 | CATCGGGAGGAAGAACACAC | ABE | AGG>GGG | DNA 결합 |
| | K556 | GGAGGAAGAACACACGGGTA | ABE | AAG>GGG | DNA 결합 |
| | Q589 | GAGCGCTGGGCTGCATCAGA | BE | CAG>CAT | DNA 결합 |
| NFATC2 | E114 | TGATCTCGATCCGAGGGCTC | BE | GAG>AAA | 칼시뉴린 결합 |
| | I115 | ACGGAGTGATCTCGATCCGA | ABE | ATC>ACC | 칼시뉴린 결합 |
| | R253 | GCGGAGGCATTCGTGCGCCG | ABE | AGG>GGG | NLS |
| | S99 | GCCGCGCTCAGAAACTTCTG | BE | AGC>AAC | 인산화 |
| | S107 | GGGCCTCGGGCCTGAGCCCT | BE | TCG>TTG | 인산화 |

[0644]

| 표적 단백질 | 표적 잔기 | gRNA 표적 | 염기 편집 | 코돈 변화 | 잔기 기능 |
|---|---|---|---|---|---|
| | S148 | CCTCGGGCTGGCGGCCACCC | BE | AGC>AAC | 인산화 |
| | S236 | CCACTCGCCCGTGCCCCGTC | BE | TCG>TTG | 인산화 |
| | S255 | GCATTCGTGCGCCGAGGCCT | BE | TCG>TTG | 인산화 |
| | S268 | GAGCCTCACCCCAGCGCTCC | BE | TCA>TTA | 인산화 |
| | S274 | GAGGGGCTCCGGGAGCGCTG | BE | AGC>AAC | 인산화 |
| | S326 | AGGGCTGGTCTTCCACATCT | BE | AGC>AAC | 인산화 |
| NFATC4 | S213 | GCGGGGAGCCCAGGCCAAAG | ABE | TCC>CCC | 인산화 |
| AKT1 | T305 | GCCACCATGAAGACCTTTTG | BE | ACC>ATT | 인산화 |
| | T312 | TTGCGGCACACCTGAGTACC | BE | ACA>ATA | 인산화 |
| | S473 | GTAGGAGAACTGGGGGAAGT | ABE | TCC>CCC | 인산화 |
| | Y474 | CTCCTACTCGGCCAGCGGCA | ABE | TAC>TGC | 인산화 |
| AKT2 | T309 | GAAAACCTTCTGTGGGACCC | BE | ACC>ATT | 인산화 |
| | S474 | AGTAGGAGAACTGGGGGAAG | ABE | TCC>CCC | 인산화 |
| BLIMP1 | C608 (ZF2) | GTTGCAAGTCTGACATTTGA | ABE | TGC>CGC | DNA 결합 |
| | C608 (ZF2) | GTTGCAAGTCTGACATTTGA | BE | TGC>TAC | DNA 결합 |
| | H621 (ZF2) | GAAACACTACCTGGTACACA | BE | CAC>TAT | DNA 결합 |
| | C636 (ZF3) | TGTGGCAGACCTACAGTGTA | BE | TGC>TAC | DNA 결합 |
| | C664 (ZF4) | GGGCACACCTTGCATTGGTA | ABE | TGC>CGC | DNA 결합 |
| | 스플라이스 부위 1 | CTGCGCACCTGGCATTCATG | BE | | |

[0645]

| 표적 단백질 | 표적 잔기 | gRNA 표적 | 염기 편집 | 코돈 변화 | 잔기 기능 |
|---|---|---|---|---|---|
| GCN2 카이네이스 (IDO 경로) | 엑손 1 SD | CCTACCGGTCCGCAAGCGTC | BE | | 넉아웃 |
| | 엑손 2 SD | ACTCACACATCTGGATAGGT | BE | | 넉아웃 |
| | 엑손 5 SD | GACTTACCTAGACCTTCCTG | BE | | 넉아웃 |
| CBL-B | C373 | AATCTTACAGAGCTGAAAAG | BE | TGT>TAT | E3 유비퀴틴 라이게이스 |
| | Y665.1 | CATCATATTCTTCACTTCCA | ABE | TAT>CAC | |
| | Y665.2 | AAGAATATGATGTTCCTCCC | ABE | TAT>TGT | |
| | K907 | CCCCTAAACCACGACCGCGC | ABE | AAA>GGG | |
| | R911 | TCCTGCGCGGTCGTGGTTTA | BE | CGC>CAC | |
| SHP1 | Y377 | CCCTACTCTGTGACCAACTG | ABE | TAC>TGC | |
| IRF4 | R96 | CGCAGGCGCGTCTTCCAGGT | BE | CGC>CAC | DNA 결합 |
| | R98 | GCACCGCAGGCGCGTCTTCC | BE | CGG>CAG | DNA 결합 |
| | K103 | GAACAAGAGCAATGACTTTG | ABE | AAG>GGG | DNA 결합 |
| PD1 | 엑손 1 STOP | CACCTACCTAAGAACCATCC | BE | | 넉아웃 |
| | 엑손 2 STOP | GGGGTTCCAGGGCCTGTCTG | BE | | 넉아웃 |
| TET2 | H1386 | GACTTGCACAACATGCAGAA | BE | CAC>TAC | DNA 결합 |
| | R1302 | TTGCCAGAAGCAAGATCCCA | ABE | AGA>GGG | DNA 결합 |
| | S1290 | CCATGAACAACCAAAAGAGA | ABE | TCA>CCA | DNA 결합 |
| SMARCA4 | T353 | TCACCCCCATCCAGAAGCCG | BE | ACC>ATT | 인산화 |
| | S610 | ATCTGGCTGGTCTCGTCCAG | BE | AGC>ATC | 인산화 |

[0646]

| 표적 단백질 | 표적 잔기 | gRNA 표적 | 염기 편집 | 코돈 변화 | 잔기 기능 |
|---|---|---|---|---|---|
| | S613 | GATGAGCGACCTCCCGGTGA | ABE | AGC>GGC | 인산화 |
| | S695 | AGACAGCGATGACGTCTCTG | ABE | AGC>GGC | 인산화 |
| | S699 | ACGTCTCTGAGGTGGACGCG | BE | TCT>TTT | 인산화 |
| | S1452 | TTAGGGGAGAGTTTCTCGGC | ABE | TCC>CCC | 인산화 |
| | S1575 | GGAGAGTGAGGAGGAGGAAG | ABE | AGT>GGT | 인산화 |
| | S1586 | AAGGCTCCGAATCCGAATCT | BE | TCC>TTT | 인산화 |
| | S1627 | ATCGTCACTCACGACCGGCT | BE | AGT>AAT | 인산화 |
| | S1631 | TGACAGTGAGGAGGAACAAG | ABE | AGT>GGT | 인산화 |
| CDK4 | P173 | CACCCGTGGTTGTTACACTC | BE | CCC>CTT | |
| ZAP70 | S144 | CATCAGCCAGGCCCCGCAGG | ABE | AGC>TGC | 인산화 |
| | Y292 | GGTGTATCCATCTGAGTTGA | ABE | TAC>CAC | 인산화 |
| | Y292 | GGGTGTATCCATCTGAGTTG | ABE | TAC>CAC | 인산화 |
| | R360 | GCGCAAGAAGCAGATCGACG | BE | CGC>TGC | 고차형 활성 |
| | Y598 | TTACTACAGCCTGGCCAGCA | ABE | TAC>TGC | 인산화 |

[0647]

[표 2B]

| 유전자 | gRNA 명칭 | gRNA 표적 | 배향 | 표적 염기(들) | 예측 결과 |
|---|---|---|---|---|---|
| *PDCD1* | Ex. 1 SD | CACCTACCTAAGAACCATCC | 안티센스 | C7 | 스플라이스 공여체 파괴: GT → AT |
| *PDCD1* | Ex. 2 SA | GGAGTCTGAGAGATGGAGAG | 안티센스 | C6 | 스플라이스 수여체 파괴: AG → AA |
| *PDCD1* | Ex. 3 SA | TTCTCTCTGGAAGGGCACAA | 안티센스 | C7 | 스플라이스 수여체 파괴: AG → AA |
| *PDCD1* | Ex. 3 SD | GACGTTACCTCGTGCGGCCC | 안티센스 | C8 | 스플라이스 공여체 파괴: GT → AT |
| *PDCD1* | Ex. 4 SA | CCTGCAGAGAAACACACTTG | 안티센스 | C2 | 스플라이스 수여체 파괴 AG → AA |
| *PDCD1* | Ex. 2 pmSTOP | GGGGTTCCAGGGCCTGTCTG | 안티센스 | C7, C8 | pmSTOP 유도 : TGG (Trp) → TAG, TGA, TAA |
| *PDCD1* | Ex. 3 pmSTOP_1 | CAGTTCCAAACCCTGGTGGT | 센스 | C7 | pmSTOP 유도 : CAA (Gln) → TAA |
| *PDCD1* | Ex. 3 pmSTOP_2 | GGACCCAGACTAGCAGCACC | 안티센스 | C5, C6 | pmSTOP 유도 : TGG (Trp) → TAG, TGA, TAA |
| *TRAC* | Ex. 1 SD | CTTACCTGGGCTGGGGAAGA | 안티센스 | C5 | 스플라이스 공여체 파괴: GT → AT |
| *TRAC* | Ex. 3 SA | TTCGTATCTGTAAAACCAAG | 안티센스 | C8 | 스플라이스 수여체 파괴 AG → AA |
| *TRAC* | Ex. 3 pmSTOP_1 | TTTCAAAACCTGTCAGTGAT | 센스 | C4 | pmSTOP 유도 : CAA (Gln) → TAA |
| *TRAC* | Ex. 3 pmSTOP_2 | TTCAAAACCTGTCAGTGATT | 센스 | C3 | pmSTOP 유도 : CAA (Gln) → TAA |

[0648]

[0649] 본 발명에서 사용되는 사이티딘 및 아데노신 데아미네이스 핵염기 편집기는 단일 가닥 DNA를 비롯한 DNA에 작용할 수 있다. 면역 세포의 표적 핵염기 서열에 변형을 생성하기 위해 이들을 사용하는 방법이 제시된다.

[0650] 소정의 실시형태에서, 본 명세서에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성을 개선하는 하나 이상

의 특징을 포함한다. 예를 들어, 본 명세서에 제공된 임의의 융합 단백질은 뉴클레이스 활성이 감소된 Cas9 도메인을 포함할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 뉴클레이스 활성(dCas9)이 없는 Cas9 도메인 또는 Cas9 닉케이스(nCas9)로도 지칭되는 이중나선 DNA 분자의 한 가닥을 절단하는 Cas9 도메인을 가질 수 있다. 임의의 특정 이론에 얽매이지 않고, 촉매 잔기(예를 들어, H840)의 존재는 표적화된 핵염기 반대편의 편집되지 않은(예를 들어, 비-메틸화된) 가닥을 절단하는 Cas9의 활성을 유지한다. 촉매 잔기의 돌연변이(예를 들어, D10에서 A10으로)는 표적화된 잔기를 포함하는 편집된 가닥의 절단을 방지한다. 이러한 Cas9 변이체는 gRNA-정의된 표적 서열을 기반으로 하는 특정 위치에서 단일-가닥 DNA 파손(닉)을 생성할 수 있으며, 이는 편집되지 않은 가닥의 복구를 유도하여 궁극적으로 편집되지 않은 가닥의 핵염기 변화를 초래한다.

[0651] **핵염기 편집기**

[0652] 폴리뉴클레오타이드의 표적 뉴클레오타이드 서열을 편집, 변형 또는 변경하기 위한 염기 편집기 또는 핵염기 편집기가 본 명세서에 개시된다. 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 및 핵염기 편집 도메인(예를 들어, 아데노신 데아미네이스 또는 사이티딘 데아미네이스)을 포함하는 핵염기 편집기 또는 염기 편집기가 본 명세서에 기재된다. 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 결합된 가이드 폴리뉴클레오타이드(예를 들어, gRNA)와 함께 사용되는 경우, 표적 폴리뉴클레오타이드 서열에 특이적으로 결합할 수 있고(즉, 결합된 가이드 핵산의 염기와 표적 폴리뉴클레오타이드 서열의 염기 사이의 상보적 염기 페어링을 통해), 이에 의해 염기 편집기를 편집하고자 하는 표적 핵산 서열로 국부화할 수 있다. 일부 실시형태에서, 표적 폴리뉴클레오타이드 서열은 단일-가닥 DNA 또는 이중-가닥 DNA를 포함한다. 일부 실시형태에서, 표적 폴리뉴클레오타이드 서열은 RNA를 포함한다. 일부 실시형태에서, 표적 폴리뉴클레오타이드 서열은 DNA-RNA 하이브리드를 포함한다.

[0653] ***폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인***

[0654] 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 또한 RNA에 결합하는 핵산 프로그램 가능한 단백질을 포함할 수 있음을 이해하여야 한다. 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 RNA로 가이드하는 핵산과 회합될 수 있다. 다른 핵산 프로그램 가능 DNA 결합 단백질도 또한 본 개시내용의 범위 내에 있지만, 본 개시내용에 구체적으로 열거되지는 않는다.

[0655] 염기 편집기의 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 그 자체로 하나 이상의 도메인을 포함할 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 하나 이상의 뉴클레이스 도메인을 포함할 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 뉴클레이스 도메인은 엔도뉴클레이스 또는 엑소뉴클레이스를 포함할 수 있다. 본 명세서에서 용어 "엑소뉴클레이스"는 자유 말단(free end)으로부터 핵산(예를 들어, RNA 또는 DNA)을 소화시킬 수 있는 단백질 또는 폴리펩타이드를 지칭하고, 용어 "엔도뉴클레이스"는 핵산(예를 들어, DNA 또는 RNA)의 내부 영역을 촉매화(예를 들어, 절단)할 수 있는 단백질 또는 폴리펩타이드를 지칭한다. 일부 실시형태에서, 엔도뉴클레이스는 이중-가닥 핵산의 단일 가닥을 절단할 수 있다. 일부 실시형태에서, 엔도뉴클레이스는 이중-가닥 핵산 분자 중 두 가닥을 절단할 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데옥시리보뉴클레이스일 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 리보뉴클레이스일 수 있다.

[0656] 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 뉴클레이스 도메인은 표적 폴리뉴클레오타이드의 0개, 1개 또는 2개의 가닥을 절단할 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 닉케이스 도메인을 포함할 수 있다. 본 명세서에서 용어 "닉케이스"는 이중나선 핵산 분자(예를 들어, DNA)에서 2개 가닥 중 하나의 가닥만을 절단할 수 있는 뉴클레이스 도메인을 포함하는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 지칭한다. 일부 실시형태에서, 닉케이스는 활성 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 하나 이상의 돌연변이를 도입함으로써 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 완전히 촉매적으로 활성인(예를 들어, 천연) 형태로부터 유래될 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인이 Cas9로부터 유래되는 닉케이스 도메인을 포함하는 경우, Cas9-유래 닉케이스 도메인은 840번 위치에 D10A 돌연변이 및 히스티딘을 포함할 수 있다. 이러한 실시형태에서, 잔기 H840은 촉매 활성을 보유하며, 이에 의해 핵산 이중체의 단일 가닥을 절단할 수 있다. 또 다른 예에서, Cas9-유래 닉케이스 도메인은 H840A 돌연변이를 포함할 수 있는 반면, 10번 위치의 아미노산 잔기는 D를 유지한다. 일부 실시형태에서, 닉케

이스는 닉케이스 활성에 필요하지 않은 뉴클레이스 도메인의 전부 또는 일부를 제거함으로써 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 완전히 촉매적으로 활성(예를 들어, 천연)인 형태로부터 유래될 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인이 Cas9로부터 유래되는 닉케이스 도메인을 포함하는 경우, Cas9-유래 닉케이스 도메인은 RuvC 도메인 또는 HNH 도메인의 전부 또는 일부를 포함할 수 있다.

[0657] 예시적인 촉매적으로 활성인 Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD

[0658] ATLIHQSITGLYETRIDLSQLGGD.

[0659] 따라서, 닉케이스 도메인을 포함하는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기는 특정 폴리뉴클레오타이드 표적 서열(예를 들어, 결합된 가이드 핵산의 상보적 서열에 의해 결정됨)에 단일-가닥 DNA 파손(닉)을 생성할 수 있다. 일부 실시형태에서, 닉케이스 도메인(예를 들어, Cas9-유래 닉케이스 도메인)을 포함하는 염기 편집기에 의해 절단된 핵산 이중체 표적 폴리뉴클레오타이드 서열의 가닥은 염기 편집기에 의해 편집되지 않은 가닥이다(즉, 염기 편집기에 의해 절단된 가닥은 편집될 염기를 포함하는 가닥과 반대임). 다른 실시형태에서, 닉케이스 도메인(예를 들어, Cas9-유래 닉케이스 도메인)을 포함하는 염기 편집기는 편집을 위해 표적화되는 DNA 분자의 가닥을 절단할 수 있다. 이러한 실시형태에서, 비-표적화된 가닥은 절단되지 않는다.

[0660] 또한, 촉매적으로 활성이 없는(즉, 표적 폴리뉴클레오타이드 서열을 절단할 수 없는) 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기가 본 명세서에 제공된다. 본 명세서에서 용어 "촉매적으로 활성이 없는" 및 "뉴클레이스 활성이 없는"은 하나 이상의 돌연변이 및/또는 결실을 가져 핵산의 가닥을 절단할 수 없는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 지칭하기 위해 상호교환적으로 사용된다. 일부 실시형태에서, 촉매적으로 활성이 없는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 염기 편집기는 하나 이상의 뉴클레이스 도메인에서 특정 점 돌연변이의 결과로 뉴클레이스 활성이 부족할 수 있다. 예를 들어, Cas9 도메인을 포함하는 염기 편집기의 경우에, Cas9는 D10A 돌연변이와 H840A 돌연변이 둘 다를 포함할 수 있다. 이러한 돌연변이는 두 뉴클레이스 도메인을 모두 불활성화하여 뉴클레이스 활성의 손실을 초래한다. 다른 실시형태에서, 촉매적으로 활성이 없는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 촉매 도메인(예를 들어, RuvC1 및/또는 HNH 도메인)의 전부 또는 일부의 하나 이상의 결실을 포함할 수 있다. 추가의 실시형태에서, 촉매적으로 활성이 없는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 점 돌연변이(예를 들어, D10A 또는 H840A)뿐만 아니라 뉴클레이스 도메인의

전부 또는 일부의 결실을 포함한다.

[0661] 또한, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 이전의 기능적 버전으로부터 촉매적으로 활성이 없는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 생성할 수 있는 돌연변이가 본 명세서에서 상정된다. 예를 들어, 촉매적으로 활성이 없는 Cas9("dCas9")의 경우에, D10A 및 H840A 이외의 돌연변이를 갖는 변이체가 제공되어 뉴클레이스 불활성화 Cas9가 생성된다. 이러한 돌연변이는 예로서 D10 및 H840에서의 다른 아미노산 치환 또는 Cas9의 뉴클레이스 도메인 내의 다른 치환(예를 들어, HNH 뉴클레이스 하위도메인 및/또는 RuvC1 하위도메인에서의 치환)을 포함한다. 추가적인 적합한 뉴클레이스-불활성 dCas9 도메인은 본 개시내용 및 당업계의 지식에 기초하여 당업자에게 명백할 수 있으며, 본 개시내용의 범위 내에 있다. 이러한 추가의 예시적인 적합한 뉴클레이스-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하지만, 이들로 제한되지 않는다(예를 들어, 문헌[Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838] 참조, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있음).

[0662] 염기 편집기에 혼입될 수 있는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 비제한적인 예는 CRISPR 단백질-유래 도메인, 제한 뉴클레이스, 메가뉴클레이스, TAL 뉴클레이스(TALEN) 및 아연 핑거 뉴클레이스(ZFN)를 포함한다. 일부 실시형태에서, 염기 편집기는 결합된 가이드 핵산을 통해 핵산의 CRISPR(즉, 클러스터링된 규칙적 간격의 짧은 회문 반복부)-매개성 변형 동안 핵산 서열에 결합할 수 있는 천연 또는 변형된 단백질 또는 이의 일부를 포함하는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 포함한다. 이러한 단백질은 본 명세서에서 "CRISPR 단백질"로 지칭된다. 따라서, CRISPR 단백질의 전부 또는 일부를 포함하는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기(즉, 염기 편집기의 "CRISPR 단백질-유래 도메인"으로도 지칭되는 CRISPR 단백질의 전부 또는 일부를 도메인으로 포함하는 염기 편집기)가 본 명세서에 개시된다. 염기 편집기에 혼입되는 CRISPR 단백질-유래 도메인은 CRISPR 단백질의 야생형 또는 천연 버전과 비교하여 변형될 수 있다. 예를 들어, 아래 기재된 바와 같이, CRISPR 단백질-유래 도메인은 CRISPR 단백질의 야생형 또는 천연 버전에 비해 하나 이상의 돌연변이, 삽입, 결실, 재배열 및/또는 재조합을 포함할 수 있다.

[0663] CRISPR은 이동성 유전 요소(바이러스, 전이성 요소 및 접합성 플라스미드)에 대한 보호를 제공하는 적응 면역 시스템이다. CRISPR 클러스터는 스페이서, 선행 이동성 요소에 상보적인 서열 및 표적 침입 핵산을 포함한다. CRISPR 클러스터는 전사되어 CRISPR RNA(crRNA)로 처리된다. 유형 II CRISPR 시스템에서, pre-crRNA의 올바른 처리에는 트랜스-암호화된 작은 RNA(tracrRNA), 내인성 리보뉴클레이스 3(rnc) 및 Cas9 단백질이 필요하다. tracrRNA는 pre-crRNA의 리보뉴클레이스 3-보조 처리를 위한 가이드 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 원형 dsDNA 표적을 핵내 분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥이 먼저 핵내 분해적으로 절단된 다음, 3'-5' 핵외 분해적으로 트리밍된다. 사실상, DNA-결합 및 절단에는 전형적으로 단백질 및 두 RNA가 모두 필요하다. 그러나, 단일 가이드 RNA("sgRNA", 또는 간단히 "gRNA")는 두 crRNA 및 tracrRNA 모두의 양태를 단일 RNA 종에 혼입하도록 조작될 수 있다. 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Jinek M., *et al.* Science 337:816-821(2012)] 참조. Cas9는 CRISPR 반복 서열의 짧은 모티프(PAM 또는 프로토스페이서 인접 모티프)를 인식하여 자기 대 비-자기를 구별하는데 도움이 된다.

[0664] 일부 실시형태에서, 본 명세서에 기재된 방법은 조작된 Cas 단백질을 이용할 수 있다. 가이드 RNA(gRNA)는 Cas-결합에 필요한 스캐폴드 서열 및 변형될 게놈 표적을 정의하는 사용자-정의된 대략 20개의 뉴클레오타이드 스페이서로 구성된 짧은 합성 RNA이다. 따라서, 당업자는 Cas 단백질의 게놈 표적을 변경할 수 있고, 특이성은 나머지 게놈과 비교하여 gRNA 표적화 서열이 게놈 표적에 대해 얼마나 특이적인지에 의해 부분적으로 결정된다.

[0665] 일부 실시형태에서, gRNA 스캐폴드 서열은 다음과 같다:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA

[0666] CCGAGUCGGUGCUUUU.

[0667] 실시형태에서, RNA 스캐폴드는 줄기 루프를 포함한다. 실시형태에서, RNA 스캐폴드는 다음 핵산 서열을 포함한다:

[0668] GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAACGAAACUUACACAGUUACUUAAAUCUUGC
AGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUG.

[0669] 실시형태에서, RNA 스캐폴드는 표준(canonical) 줄기 루프를 포함한다. 실시형태에서, RNA 스캐폴드는 다음 핵산 서열을 포함한다:

[0670] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA
CCGAGUCGGUGCU*mU*mU*mU

[0671] 여기서, m=2'-O-메틸 변형이고, *=3' 포스포로티오에이트 뉴클레오타이드간 연결(즉, 여기에 나타낸 바와 같이 처음 3' 말단 RNA 잔기에서).

[0672] 실시형태에서, RNA 스캐폴드는 다음 핵산 서열을 포함한다:

[0673] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA
CCGAGUCGGUGCUUUU

[0674] 여기서, m=2'-O-메틸 변형이고, *=3' 포스포로티오에이트 뉴클레오타이드간 연결(즉, 여기에 나타낸 바와 같이 처음 3' 말단 RNA 잔기에서).

[0675] 실시형태에서, S. 피오게네스 sgRNA 스캐폴드 폴리뉴클레오타이드 서열은 다음과 같다:

[0676] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA
CCGAGUCGGUGC.

[0677] 실시형태에서, S. 아우레우스 sgRNA 스캐폴드 폴리뉴클레오타이드 서열은 다음과 같다:

[0678] GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUC
UCGUCAACUUGUUGGCGAGA.

[0679] 실시형태에서, BhCas12b sgRNA 스캐폴드는 다음 폴리뉴클레오타이드 서열을 갖는다:

[0680] GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCUAU
UGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC.

[0681] 실시형태에서, BvCas12b sgRNA 스캐폴드는 다음 폴리뉴클레오타이드 서열을 갖는다:

[0682] GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGAG
CACCUGAAAACAGGUGCUUGGCAC.

[0683] 실시형태에서, RNA 스캐폴드는 비-표준 서열을 포함한다. 실시형태에서, RNA 스캐폴드는 다음 핵산 서열을 포함한다:

[0684] GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGGA
CCGAGUCGGUGCU*mU*mU*mU

[0685] 여기서, m=2'-O-메틸 변형이고, *=3' 포스포로티오에이트 뉴클레오타이드간 연결(즉, 여기에 나타낸 바와 같이 처음 3' 말단 RNA 잔기에서).

[0686] 일부 실시형태에서, 염기 편집기에 혼입되는 CRISPR 단백질-유래 도메인은 결합된 가이드 핵산과 함께일 때 표적 폴리뉴클레오타이드에 결합할 수 있는 엔도뉴클레이스(예를 들어, 데옥시리보뉴클레이스 또는 리보뉴클레이스)이다. 일부 실시형태에서, 염기 편집기에 혼입되는 CRISPR 단백질-유래 도메인은 결합된 가이드 핵산과 함께일 때 표적 폴리뉴클레오타이드에 결합할 수 있는 닉케이스이다. 일부 실시형태에서, 염기 편집기에 혼입되는 CRISPR 단백질-유래 도메인은 결합된 가이드 핵산과 함께일 때 표적 폴리뉴클레오타이드에 결합할 수 있는 촉매적으로 활성이 없는 도메인이다. 일부 실시형태에서, 염기 편집기의 CRISPR 단백질 유래 도메인에 의해 결합된 표적 폴리뉴클레오타이드는 DNA이다. 일부 실시형태에서, 염기 편집기의 CRISPR 단백질-유래 도메인에 의해 결합된 표적 폴리뉴클레오타이드는 RNA이다.

[0687] 본 명세서에서 사용될 수 있는 Cas 단백질은 클래스 1 및 클래스 2를 포함한다. Cas 단백질의 비제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas9(Csn1 또는 Csx12로도 알려져 있음), Cas10, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 및 Cas12j/CasΦ, CARF, DinG, 이들의 상동체 또는 이들의 변형된 버전을 포함한다. 변형되지 않은 CRISPR 효소는 RuvC 및 HNH의 2개의 기능적 엔도뉴클레이스 도메인을 갖는 Cas9와 같은 DNA 절단 활성을 갖는다. CRISPR 효소는 표적 서열 내 및/또는 표적 서열의 상보체 내와 같은 표적 서열에서 하나 또는 두 가닥 모두의 절단을 지시할 수 있다. 예를 들어, CRISPR 효소는 표적 서열의 첫 번째 또는 마지막 뉴클레오타이드로부터의 약 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 15개, 20개, 25개, 50개, 100개, 200개, 500개 이상의 염기쌍 내에서 하나 또는 두 가닥 모두의 절단을 지시할 수 있다.

[0688] 돌연변이된 CRISPR 효소가 표적 서열을 포함하는 표적 폴리뉴클레오타이드의 하나 또는 두 가닥 모두를 절단하는 능력이 결여되도록 상응하는 야생형 효소에 대해 돌연변이된 CRISPR 효소를 암호화하는 벡터가 사용될 수 있다. Cas9는 야생형의 예시적인 Cas9 폴리펩타이드(예를 들어, S. 피오게네스로부터의 Cas9)에 대해 적어도 또는 적어도 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%의 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 지칭할 수 있다. Cas9는 야생형의 예시적인 Cas9 폴리펩타이드(예를 들어, S. 피오게네스 유래)에 대해 최대 또는 최대 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%의 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 지칭할 수 있다. Cas9는 결실, 삽입, 치환, 변이체, 돌연변이, 융합, 키메라 또는 이들의 임의의 조합과 같은 아미노산 변화를 포함할 수 있는 Cas9 단백질의 야생형 또는 변형된 형태를 지칭할 수 있다.

[0689] 일부 실시형태에서, 염기 편집기의 CRISPR 단백질-유래 도메인은 코리네박테리움 울서란스(NCBI 참조: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아(NCBI 참조: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라(NCBI 참조: NC_021284.1); 프레보텔라 인테르메디아(NCBI 참조: NC_017861.1); 스피로플라스마 타이와넨세(NCBI 참조: NC_021846.1); 스트렙토코커스 이니아에(NCBI 참조: NC_021314.1); 벨리엘라 발티카(NCBI 참조: NC_018010.1); 싸이크로플렉서스 토르퀴스(*Psychroflexus torquis*)(NCBI 참조: NC_018721.1); 스트렙토코커스 써모필루스(NCBI 참조: YP_820832.1); 리스테리아 이노쿠아(NCBI 참조: NP_472073.1); 캄필로박터 제주니(NCBI 참조: YP_002344900.1); 네이세리아 메닌지티디스(NCBI 참조: YP_002342100.1), 스트렙토코커스 피오게네스 또는 스타필로코커스 아우레우스로부터의 Cas9의 전부 또는 일부를 포함할 수 있다.

[0690] ***핵염기 편집기의 Cas9 도메인***

[0691] Cas9 뉴클레이스 서열 및 구조는 당업자에게 잘 알려져 있다(예를 들어, 문헌["Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., *et al.*, Nature 471:602-607(2011); 및 "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., *et al.*, Science 337:816-821(2012)] 참조, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있음). Cas9 오쏘로그는 S. 피오게네스 및 S. 써모필루스를 포함하지만 이로 제한되지 않는 다양한 종에서 설명되었다. 추가적인 적합한 Cas9 뉴클레이스 및 서열은 본 개시내용에 기초하여 당업자에게 명백할 것이며, 이러한 Cas9 뉴클레이스 및 서열은 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737]에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다.

[0692] 일부 실시형태에서, 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)은 Cas9 도메인이다. 비제한적인, 예시적인 Cas9 도메인이 본 명세서에 제공된다. Cas9 도메인은 뉴클레이스 활성이 있는 Cas9 도메인, 뉴클레이스 불활성 Cas9 도메인(dCas9) 또는 Cas9 닉케이스(nCas9)일 수 있다. 일부 실시형태에서, Cas9 도메인은 뉴클레이스 활성 도메인이다. 예를 들어, Cas9 도메인은 이중나선 핵산의 두 가닥(예를 들어, 이중나선 DNA 분자의 두 가닥)을 절단하는 Cas9 도메인일 수 있다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 기재된 바와 같은 아미노산 서열 중 어느 하나를 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 1개, 2개, 3개, 4개,

5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 적어도 900개, 적어도 1000개, 적어도 1100개 또는 적어도 1200개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0693] 일부 실시형태에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시형태에서, 단백질은 다음 2개의 Cas9 도메인 중 하나를 포함한다: (1) Cas9의 gRNA 결합 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 실시형태에서, Cas9 또는 이들의 단편을 포함하는 단백질은 "Cas9 변이체"로 지칭된다. Cas9 변이체는 Cas9 또는 이의 단편과 상동성을 공유한다. 예를 들어, Cas9 변이체는 야생형 Cas9와 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일하다. 일부 실시형태에서, Cas9 변이체는 야생형 Cas9와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 아미노산 변화를 가질 수 있다. 일부 실시형태에서, Cas9 변이체는 단편이 야생형 Cas9의 상응하는 단편과 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일하도록 Cas9의 단편(예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함한다. 일부 실시형태에서, 단편은 상응하는 야생형 Cas9의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 동일, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5%이다. 일부 실시형태에서, 단편은 적어도 100개의 아미노산 길이이다. 일부 실시형태에서, 단편은 적어도 100개, 150개, 200개, 250개, 300개, 350개, 400개, 450개, 500개, 550개, 600개, 650개, 700개, 750개, 800개, 850개, 900개, 950개, 1000개, 1050개, 1100개, 1150개, 1200개, 1250개 또는 적어도 1300개의 아미노산 길이이다.

[0694] 일부 실시형태에서, Cas9 본 명세서에 제공된 바와 같은 융합 단백질은 Cas9 단백질의 전장 아미노산 서열, 예를 들어, 본 명세서에 제공된 Cas9 서열 중 하나를 포함한다. 다른 실시형태에서, 그러나, 본 명세서에 제공된 바와 같은 융합 단백질은 전장 Cas9 서열을 포함하지 않지만 이의 하나 이상의 단편만을 포함한다. 적합한 Cas9 도메인 및 Cas9 단편의 예시적인 아미노산 서열이 본 명세서에 제공되며, Cas9 도메인 및 단편의 추가적인 적합한 서열은 당업자에게 명백할 것이다.

[0695] Cas9 단백질은 Cas9 단백질을 가이드 RNA에 상보적인 특정 DNA 서열로 가이드하는 가이드 RNA와 회합할 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 Cas9 도메인, 예를 들어, 뉴클레이스 활성이 있는 Cas9, Cas9 닉케이스(nCas9) 또는 뉴클레이스 불활성 Cas9(dCas9)이다. 핵산 프로그램 가능 DNA 결합 단백질의 예는 제한 없이 Cas9(예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 및 Cas12j/CasΦ를 포함한다. Cas 효소의 비제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9(Csn1 또는 Csx12로도 알려져 있음), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, Cs0, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, 유형 II Cas 효과기 단백질, 유형 V Cas 효과기 단백질, 유형 VI Cas 효과기 단백질, CARF, DinG, 이들의 상동체 또는 이들의 변형된 또는 조작된 버전을 포함한다.

[0696] 일부 실시형태에서, 야생형 Cas9는 스트렙토코커스 피오게네스(NCBI 참조 서열: NC_017053.1)로부터의 Cas9에 해당한다. 예시적인 스트렙토코커스 피오게네스 Cas9(spCas9) 핵산 서열이 아래에 제공된다:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCA
CTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTAT
CAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTC
AAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTT
TTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGT
GGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTAT
CATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGA
GGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTAC
AATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTG
CACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAA
TGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATT
TATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGC
TATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCA
ATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAAC
AACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATAT
TGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGAT
GGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTG
ACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGA
AGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATT
CCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTG
AAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATT
TATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGT
TTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAA
TGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGAT
AGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGC
TAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATAT
TGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCT
CACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTT
[0697] GAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTT

AAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTA
ACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACT
GGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACA
ACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAAT
TAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTA
TCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTA
AGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATA
AGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGT
CAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTT
GATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAAC
GCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAA
TACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAA
TTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATC
ATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACT
TGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCT
GAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCA
AAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGA
AACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATG
CCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTT
TACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGG
TGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAA
TCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTG
AAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCAT
TAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCC
GGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAG
CTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGA
GCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATT
TTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATAC
GTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCC
ACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAG
[0698] GTGACTGA

[0699] 예시적인 스트렙토코커스 피오게네스 Cas9(spCas9) 아미노산 서열이 아래에 제공된다:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIY
NQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD

[0700]

[0701] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0702] 일부 실시형태에서, 야생형 Cas9는 다음 뉴클레오타이드 서열에 해당하거나 또는 이를 포함한다:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAA
CCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGAT
TAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTG
AAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTT
TTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGT
CGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATAT
CATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGG
ACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGA
GGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTAT

[0703]

AATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCG
CCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAA
TGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTC
GACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATC
TACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGC
AATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCA
ATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGC
AACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATAT
TGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGAT
GGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCG
ACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGA
GGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATA
CCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCG
AAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTT
CATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGT
TTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCA
TGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGAC
CAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGAT
TCTGTCGAGATCTCCGGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCC
TAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATAT
AGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCT
CACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT
TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCT
AAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTC
AAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGA
ATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCT
AGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAA
ACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAG
AACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACT
TTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGT
TTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACA
ATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGT
CGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAG
TTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTA

[0704]

AACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAAT
GAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCA
AAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACC
ACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAA
GCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAA
AGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCT
TTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGG
GGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCC
ATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGA
TTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTA
CGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGA
AAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTT
TTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCAT
AATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGC
GCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATT
TAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGT
TGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTC
ATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCA
TACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGC
ATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGAC
GCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTG
GGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGA
TTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

[0705]

[0706] 일부 실시형태에서, 야생형 Cas9는 다음 아미노산 서열에 해당하거나 또는 이를 포함한다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD

[0707]

SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[0708]

[0709] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인).

[0710] 일부 실시형태에서, 야생형 Cas9는 스트렙토코커스 피오게네스(NCBI 참조 서열: NC_002737.2(다음과 같은 뉴클레오타이드 서열))로부터의 Cas9에 해당한다:

[0711] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCA
CTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTAT
CAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTC
AAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTT
TTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGT
GGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTAT
CATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGA
GGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTAC
AATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTG
CACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAA
TGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATT
TATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGC
TATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCA
ATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAAC
AACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATAT
TGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGAT

GGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTG
ACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGA
AGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATT
CCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTG
AAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATT
TATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGT
TTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAA
TGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGAT
AGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGC
TAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATAT
TGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCT
CACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTT
GAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTT
AAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAA
ATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATT
GGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAG
ACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAG
AATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCT
CTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGT
TTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACA
ATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGT
AGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAG
TTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCA
AACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCAT
GAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCT
AAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACC
ATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAA
ACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAG
TCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCT
TCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGG
GGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCC
ATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAA

[0712]

TTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATA
TGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGG
AAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCT
TTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAAT
CATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGT
GCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATT
TAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGT
GGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTT
ATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAA
TACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGAT
GCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAG

[0713]

GAGGTGACTGA

[0714] 일부 실시형태에서, 야생형 Cas9는 스트렙토코커스 피오게네스(Uniprot 참조 서열: Q99ZW2(다음과 같은 아미노

산 서열))로부터의 Cas9에 해당한다:

[0715]
```
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
```

[0716]
```
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD
```

[0717] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0718] 일부 실시형태에서, Cas9는 코리네박테리움 울서란스(NCBI 참조: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아(NCBI 참조: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라(NCBI 참조: NC_021284.1); 프레보텔라 인테르메디아(NCBI 참조: NC_017861.1); 스피로플라스마 타이와넨세(NCBI 참조: NC_021846.1); 스트렙토코커스 이니아에(NCBI 참조: NC_021314.1); 벨리엘라 발티카(NCBI 참조: NC_018010.1); 싸이크로플렉서스 토르퀴스(NCBI 참조: NC_018721.1); 스트렙토코커스 써모필루스(NCBI 참조: YP_820832.1), 리스테리아 이노쿠아(NCBI 참조: NP_472073.1), 캄필로박터 제주니(NCBI 참조: YP_002344900.1) 또는 네이세리아 메닌지티디스(NCBI 참조: YP_002342100.1)로부터의 Cas9 또는 임의의 다른 유기체로부터의 Cas9를 지칭한다.

[0719] 이들의 변이체 및 상동체를 비롯한 추가적인 Cas9 단백질(예를 들어, 뉴클레이스 활성이 없는 Cas9(dCas9), Cas9 닉케이스(nCas9) 또는 뉴클레이스 활성이 있는 Cas9는 본 개시내용의 범위 내에 있음을 이해하여야 한다. 예시적인 Cas9 단백질은 제한 없이 아래에 제공되는 것들을 포함한다. 일부 실시형태에서, Cas9 단백질은 뉴클레이스 활성이 없는 Cas9(dCas9)이다. 일부 실시형태에서, Cas9 단백질은 Cas9 닉케이스(nCas9)이다. 일부 실시형태에서, Cas9 단백질은 뉴클레이스 활성이 있는 Cas9이다.

[0720] 일부 실시형태에서, Cas9 도메인은 뉴클레이스-불활성 Cas9 도메인(dCas9)이다. 예를 들어, dCas9 도메인은 이중나선 핵산 분자 가닥을 절단하지 않고 이중나선 핵산 분자에 결합(예를 들어, gRNA 분자를 통해)할 수 있다. 일부 실시형태에서, 뉴클레이스-불활성 dCas9 도메인은 본 명세서에 기재된 아미노산 서열의 D10X 돌연변이 및 H840X 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열의 상응하는 돌연변이를 포함하되, X는 임의의 아미노산 변화이다. 일부 실시형태에서, 뉴클레이스-불활성 dCas9 도메인은 본 명세서에 기재된 아미노산 서열의 D10A 돌연변이 및 H840A 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열의 상응하는 돌연변이를 포함한다. 일례로서, 뉴클레이스-불활성 Cas9 도메인은 클로닝 벡터 pPlatTET-gRNA2(등록 번호 BAV54124)에 기재된 아미노산 서열을 포함한다.

[0721] 예시적인 촉매적으로 불활성인 Cas9(dCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[0722]

[0723] (예를 들어, 문헌[Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell*. 2013; 152(5):1173-83] 참조, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있음).

[0724] 추가적인 적합한 뉴클레이스-불활성 dCas9 도메인은 본 개시내용 및 당업계의 지식에 기초하여 당업자에게 명백할 것이며 본 개시내용의 범위 내에 있다. 이러한 추가적인 예시적인 적합한 뉴클레이스-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하지만, 이들로 제한되지 않는다(예를 들어, 문헌[Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838] 참조, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있음).

[0725] 일부 실시형태에서, Cas9 뉴클레이스는 불활성(예를 들어, 불활성화된) DNA 절단 도메인을 가지며, 즉, Cas9는 "nCas9" 단백질("닉케이스" Cas9의 경우)로도 지칭되는 닉케이스이다. 뉴클레이스-불활성화 Cas9 단백질은 "dCas9" 단백질(뉴클레이스-"활성이 없는" Cas9의 경우) 또는 촉매적으로 불활성인 Cas9로 상호교환적으로 지칭될 수 있다. 불활성 DNA 절단 도메인을 갖는 Cas9 단백질(또는 이의 단편)을 생성하는 방법은 공지되어 있다(예를 들어, 문헌[Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell*. 28;152(5):1173-83] 참조, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있음). 예를 들어, Cas9의 DNA 절단 도메인은 2개 하위도메인, HNH 뉴클레이스 하위도메인 및 RuvC1 하위도메인을 포함하는 것으로 알려져 있다. HNH 하위도메인은 gRNA에 상보적인 가닥을 절단하는 반면, RuvC1 하위도메인은 비-상보적 가닥을 절단한다. 이들 하위도메인 내의 돌연변이는 Cas9의 뉴클레이스 활성을 침묵시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 S. 피오게네스 Cas9의 뉴클레이스 활성을 완전히 불활성화한다(문헌[Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell*. 28;152(5):1173-83 (2013)]).

[0726] 일부 실시형태에서, dCas9 도메인은 본 명세서에 제공된 dCas9 도메인 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개,

11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상 또는 그 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 적어도 900개, 적어도 1000개, 적어도 1100개 또는 적어도 1200개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0727] 일부 실시형태에서, dCas9는 Cas9 뉴클레이스 활성을 불활성화하는 하나 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 상응하거나 또는 부분적으로 또는 전체적으로 이를 포함한다. 일부 실시형태에서, 뉴클레이스-불활성 dCas9 도메인은 본 명세서에 기재된 아미노산 서열의 D10X 돌연변이 및 H840X 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열의 상응하는 돌연변이를 포함하되, X는 임의의 아미노산 변화이다. 일부 실시형태에서, 뉴클레이스-불활성 dCas9 도메인은 본 명세서에 기재된 아미노산 서열의 D10A 돌연변이 및 H840A 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 뉴클레이스-불활성 Cas9 도메인은 클로닝 벡터 pPlatTET-gRNA2(등록 번호 BAV54124)에 기재된 아미노산 서열을 포함한다.

[0728] 일부 실시형태에서, dCas9는 dCas9(D10A 및 H840A)의 아미노산 서열을 포함한다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[0729]

[0730] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인).

[0731] 일부 실시형태에서, 예시적인 촉매적으로 불활성인 Cas9(dCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[0732]

[0733] (예를 들어, 문헌[Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell*. 2013; 152(5):1173-83] 참조, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있음).

[0734] 일부 실시형태에서, 예시적인 촉매적으로 불활성인 Cas9(dCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[0735]

[0736] 일부 실시형태에서, Cas9 도메인은 D10A 돌연변이를 포함하는 반면, 840번 위치의 잔기는 위에 제공되는 아미노산 서열에 히스티딘으로 남아 있거나 또는 본 명세서에 제공된 임의의 아미노산 서열의 상응하는 위치에 존재한다.

[0737] 다른 실시형태에서, D10A 및 H840A 이외의 돌연변이를 갖는 dCas9 변이체가 제공되며, 이는, 예를 들어, 뉴클레이스 불활성화된 Cas9(dCas9)를 생성한다. 이러한 돌연변이는, 예로서, D10 및 H840에서의 다른 아미노산 치환 또는 Cas9의 뉴클레이스 도메인 내의 다른 치환(예를 들어, HNH 뉴클레이스 하위도메인 및/또는 RuvC1 하위도메인에서의 치환)을 포함한다. 일부 실시형태에서, 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일한 dCas9의 변이체 또는 상동체가 제공된다. 일부 실시형태에서, 약 5개의 아미노산, 약 10개의 아미노산, 약 15개의 아미노산, 약 20개의 아미노산, 약 25개의 아미노산, 약 30개의 아미노산, 약 40개의 아미노산, 약 50개의 아미노산, 약 75개의 아미노산, 약 100개 이상의 아미노산만큼 더 짧거나 또는 더 긴 아미노산 서열을 갖는 dCas9의 변이체가 제공된다.

[0738] 추가적인 적합한 뉴클레이스-불활성 dCas9 도메인은 본 개시내용 및 당업계의 지식에 기초하여 당업자에게 명백할 것이며 본 개시내용의 범위 내에 있다. 이러한 추가적인 예시적인 적합한 뉴클레이스-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하지만, 이들로 제한되지 않는다(예를 들어, 문헌[Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838] 참조, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있음).

[0739] 일부 실시형태에서, Cas9 뉴클레이스는 불활성(예를 들어, 불활성화된) DNA 절단 도메인을 가지며, 즉, Cas9는 "nCas9" 단백질("닉케이스" Cas9의 경우)로 지칭되는 닉케이스이다. Cas9 닉케이스는 이중나선 핵산 분자(예를 들어, 이중나선 DNA 분자)의 하나의 가닥만을 절단할 수 있는 Cas9 단백질일 수 있다. 일부 실시형태에서, Cas9 닉케이스는 이중나선 핵산 분자의 표적 가닥을 절단하며, 이는 Cas9 닉케이스가 Cas9에 결합된 gRNA(예를 들어, sgRNA)와 염기쌍을 이루는(이에 상보적인) 가닥을 절단함을 의미한다. 일부 실시형태에서, Cas9 닉케이스는 D10A 돌연변이를 포함하며, 840번 위치에 히스티딘을 갖는다. 일부 실시형태에서, Cas9 닉케이스는 이중나선 핵산 분자의 비-표적, 비-염기-편집된 가닥을 절단하며, 이는 Cas9 닉케이스가 Cas9에 결합된 gRNA(예를 들어,

sgRNA)와 염기쌍을 이루지 않는 가닥을 절단함을 의미한다. 일부 실시형태에서, Cas9 닉케이스는 H840A 돌연변이를 포함하며, 위치 10번 위치에 아스파르트산 잔기 또는 상응하는 돌연변이를 갖는다. 일부 실시형태에서, Cas9 닉케이스는 본 명세서에 제공된 Cas9 닉케이스 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 추가적인 적합한 Cas9 닉케이스는 본 개시내용 및 당업계의 지식에 기초하여 당업자에게 명백할 것이며 본 개시내용의 범위 내에 있다.

[0740] 예시적인 촉매적으로 Cas9 닉케이스(nCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA

[0741]

HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[0742]

[0743] 일부 실시형태에서, Cas9는 단일-세포 원핵 미생물의 역(domain) 및 계(kingdom)를 구성하는 고세균(예를 들어, 나노고세균)으로부터의 Cas9를 지칭한다. 일부 실시형태에서, 핵산 프로그램 가능 DNA 결합 단백질은 CasX 또는 CasY 단백질일 수 있으며, 이는, 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌 [Burstein *et al.*, "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21]에 기술되어 있다. 게놈-분해 메타게노믹스를 사용하여, 고대 생명체역(archaeal domain)에서 처음 보고된 Cas9를 포함하여 다수의 CRISPR-Cas 시스템이 확인되었다. 이러한 분기형(divergent) Cas9 단백질은 활성 CRISPR-Cas 시스템의 일부로 거의 연구되지 않은 나노고세균에서 발견되었다. 세균에서, 이전에 알려지지 않은 2개의 시스템인 CRISPR-CasX 및 CRISPR-CasY이 발견되었으며, 이는 아직 발견된 가장 컴팩트한 시스템 중 하나이다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 시스템에서, Cas9는 CasX 또는 CasX의 변이체로 대체된다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 시스템에서 Cas9는 CasY 또는 CasY의 변이체로 대체된다. 다른 RNA-가이드된 DNA 결합 단백질은 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)로 사용될 수 있으며, 이는 본 개시내용의 범위 내에 있음을 이해하여야 한다.

[0744] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)은 CasX 또는 CasY 단백질일 수 있다. 일부 실시형태에서, napDNAbp는 CasX 단백질이다. 일부 실시형태에서, napDNAbp는 CasY 단백질이다. 일부 실시형태에서, 프로그램 가능한 뉴클레오타이드 결합 단백질은 본 명세서에 기재된 임의의 CasX 또는 CasY 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, 프로그램 가능한 뉴클레오타이드 결합 단백질은 자연-발생적 CasX 또는 CasY 단백질이다. 일부 실시형태에서, napDNAbp는 자연-발생적 CasX 또는 CasY 단백질과 적어도 85%, 적어도 90%, 적어

도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종으로부터의 CasX 및 CasY가 또한 본 개시내용에 따라 사용될 수 있음을 이해하여야 한다.

[0745] 예시적인 CasX((uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53) tr|F0NN87|F0NN87_SULIHCRISPR-관련 Casx 단백질 OS = 설포로부스 아이슬랜디쿠스(*Sulfolobus islandicus*)(균주 HVE10/4) GN = SiH_0402 PE=4 SV=1) 아미노산 서열은 다음과 같다:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAK
KKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVK
PEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILY
SLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLT
KLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNL
[0746] NSDDGKVRDLKLISAYVNGELIRGEG.

[0747] 예시적인 CasX(>tr|F0NH53|F0NH53_SULIR CRISPR 관련 단백질, Casx OS = 설포로부스 아이슬랜디쿠스(균주 REY15A) GN=SiRe_0771 PE=4 SV=1) 아미노산 서열은 다음과 같다:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAK
KKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVK
PEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILY
SLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLT
KLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLN
[0748] SDDGKVRDLKLISAYVNGELIRGEG.

[0749] 델타프로테오박테리아(Deltaproteobacteria) CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANN
LRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFAC
SQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRA
LDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKG
NQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDfAYNEVIARVRMWVNLNLWQKLKLSRDDA
KPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYL
PNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIERE
EARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRV
VDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLY
GGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKI
GRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTD
ILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVLV
FANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTIT
YADMDVMLVRLKKTSDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELDRLSEESGNNDI
SKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSY
[0750] KSGKQPFVGAWQAFYKRRLKEVWKPNA

[0751] 예시적인 CasY((ncbi.nlm.nih.gov/protein/APG80656.1) >APG80656.1 CRISPR-관련 단백질 CasY [배양되지 않은 파르쿠박테리아(Parcubacteria) 그룹 세균]) 아미노산 서열은 다음과 같다:

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYG
LSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSH
LYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAG
ASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQK
LDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELE
KRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEM
INRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDV
QEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKY
KNAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVK
NSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWK
DLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFL
EMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLD
LAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKRE
IKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYD
ALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQN
FISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELE
VSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWR
AEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMRG
NSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI.

[0752]

[0753] Cas9 뉴클레이스는 2개 기능적 엔도뉴클레이스 도메인 RuvC 및 HNH를 갖는다. Cas9는 표적 DNA의 반대편 가닥을 절단하도록 뉴클레이스 도메인을 위치시키는 표적 결합시 형태 변화를 겪는다. Cas9-매개성 DNA 절단의 최종 결과는 표적 DNA(PAM 서열의 대략 3개 내지 4개 뉴클레오타이드 상류) 내의 이중-가닥 파손(DSB)이다. 그런 다음, 생성된 DSB는 다음 두 가지 일반적인 복구 경로 중 하나에 의해 복구된다: (1) 효율적이지만 오류가 발생하기 쉬운 비-상동 말단 연결(non-homologous end joining: NHEJ) 경로; 또는 (2) 덜 효율적이지만 정확도가 높은 상동성 유도 복구(homology directed repair: HDR) 경로.

[0754] 비-상동 말단 연결(NHEJ) 및/또는 상동성 유도 복구(HDR)의 "효율성"은 임의의 편리한 방법에 의해 계산될 수 있다. 예를 들어, 일부 실시형태에서, 효율성은 성공적인 HDR의 백분율로 환산하여 표현될 수 있다. 예를 들어, 측량자 뉴클레이스 검정(surveyor nuclease assay)이 절단 생성물을 생성하는데 사용될 수 있고, 생성물 대 기질의 비는 백분율을 계산하는데 사용될 수 있다. 예를 들어, 성공적인 HDR의 결과로서 새로 통합된 제한 서열을 포함하는 DNA를 직접 절단하는 측량자 뉴클레이스 효소가 사용될 수 있다. 더 많이 절단된 기질은 더 높은 비율의 HDR(HDR의 더 높은 효율성)을 나타낸다. 예시적인 예로서, HDR의 분율(백분율)은 다음 방정식 [(절단 생성물)/(기질 + 절단 생성물)](예를 들어, (b+c)/(a+b+c)을 사용하여 계산될 수 있으며, 여기서 "a"는 DNA 기질의 밴드 강도이고, "b" 및 "c"는 절단 생성물임).

[0755] 일부 실시형태에서, 효율성은 성공적인 NHEJ의 백분율로 환산하여 표현될 수 있다. 예를 들어, T7 엔도뉴클레이스 I 검정이 절단 생성물을 생성하는데 사용될 수 있으며, 생성물 대 기질의 비는 NHEJ 백분율을 계산하는데 사용될 수 있다. T7 엔도뉴클레이스 I는 야생형 및 돌연변이체 DNA 가닥의 혼성화로부터 발생하는 불일치 이종이중체를 절단한다(NHEJ는 원래의 파손 부위에 작은 무작위 삽입 또는 결실(indel)을 생성함). 더 많은 절단은 더 높은 비율의 NHEJ(NHEJ의 더 높은 효율성)를 나타낸다. 예시적인 예로서, NHEJ의 분율(백분율)은 다음 방정식을 사용하여 계산될 수 있으며: $(1-(1-(b+c)/(a+b+c))^{1/2})\times 100$, 여기서 "a"는 DNA 기질의 밴드 강도이고, "b" 및 "c"는 절단 생성물이다(문헌[Ran *et. al.*, Cell. 2013 Sep. 12; 154(6):1380-9; 및 Ran *et al.*, Nat Protoc. 2013 Nov.; 8(11): 2281-2308]).

[0756] NHEJ 복구 경로는 가장 활발한 복구 메커니즘이며, 이는 흔히 DSB 부위에서 작은 뉴클레오타이드 삽입 또는 결실(indel)을 유발한다. NHEJ-매개성 DSB 복구의 무작위성은 Cas9 및 gRNA 또는 가이드 폴리뉴클레오타이드를 발현하는 세포의 집단이 다양한 종류의 돌연변이를 초래할 수 있기 때문에 중요한 실용적인 의미를 갖는다. 대부분의 실시형태에서, NHEJ는 표적화된 유전자의 오픈 리딩 프레임(ORF) 내에서 조기 종결 코돈을 초래하는 아미노산 결실, 삽입 또는 프레임시프트 돌연변이를 생성하는 표적 DNA에서 작은 indel을 발생시킨다. 이상적인 최종 결과는 표적화된 유전자 내의 기능 상실 돌연변이이다.

[0757] NHEJ-매개성 DSB 복구가 종종 유전자의 오픈 리딩 프레임을 파괴하지만, 상동성 유도 복구(HDR)가 단일 뉴클레

오타이드 변화에서 형광단 또는 태그의 추가와 같은 큰 삽입에 이르는 특정 뉴클레오타이드 변화를 생성하는데 사용될 수 있다.

[0758] 유전자 편집을 위해 HDR을 이용하기 위하여, 목적하는 서열을 포함하는 DNA 복구 주형이 gRNA(들) 및 Cas9 또는 Cas9 닉케이스와 함께 관심 세포 유형으로 전달될 수 있다. 복구 주형은 목적하는 편집뿐만 아니라 표적의 바로 상류와 하류에 추가적인 상동 서열(왼쪽 및 오른쪽 상동성 암(arm)이라고 함)을 포함할 수 있다. 각 상동성 암의 길이는 도입되는 변화의 크기에 따라 달라질 수 있으며, 더 큰 삽입은 더 긴 상동성 암을 필요로 한다. 복구 주형은 단일-가닥 올리고뉴클레오타이드, 이중-가닥 올리고뉴클레오타이드 또는 이중-가닥 DNA 플라스미드일 수 있다. HDR의 효율성은 Cas9, gRNA 및 외인성 복구 주형을 발현하는 세포에서도 일반적으로 낮다(변형된 대립유전자의 10% 미만). HDR의 효율성은 HDR이 세포 주기의 S 및 G2 단계 동안 발생하므로 세포를 동기화함으로써 향상될 수 있다. NHEJ에 관여하는 유전자를 화학적으로 또는 유전적으로 저해하면 HDR 빈도도 증가할 수 있다.

[0759] 일부 실시형태에서, Cas9는 변형된 Cas9이다. 주어진 gRNA 표적화 서열은 부분적인 상동성이 존재하는 게놈 전체에 걸쳐 추가적인 부위를 가질 수 있다. 이들 부위는 표적외라고도 하며, gRNA를 설계할 때 고려되어야 한다. gRNA 설계를 최적화하는 것 이외에도, CRISPR 특이성은 또한 Cas9에 대한 변형을 통해서도 증가될 수 있다. Cas9는 2개의 뉴클레이스 도메인인 RuvC 및 HNH의 조합된 활성을 통해 이중-가닥 파손(DSB)을 생성한다. SpCas9의 D10A 돌연변이체인 Cas9 닉케이스는 하나의 뉴클레이스 도메인을 보유하며, DSB가 아닌 DNA 닉을 생성한다. 닉케이스 시스템은 또한 특정 유전자 편집을 위해 HDR-매개성 유전자 편집과 조합될 수 있다.

[0760] 일부 실시형태에서, Cas9는 변이체 Cas9 단백질이다. 변이체 Cas9 폴리펩타이드는 야생형 Cas9 단백질의 아미노산 서열과 비교할 때 하나의 아미노산이 상이한(예를 들어, 결실, 삽입, 치환, 융합을 가짐) 아미노산 서열을 갖는다. 일부 예에서, 변이체 Cas9 폴리펩타이드는 Cas9 폴리펩타이드의 뉴클레이스 활성을 감소시키는 아미노산 변화(예를 들어, 결실, 삽입 또는 치환)를 갖는다. 예를 들어, 일부 예에서, 변이체 Cas9 폴리펩타이드는 상응하는 야생형 Cas9 단백질의 뉴클레이스 활성의 50% 미만, 40% 미만, 30% 미만, 20% 미만, 10% 미만, 5% 미만 또는 1% 미만을 갖는다. 일부 실시형태에서, 변이체 Cas9 단백질은 실질적인 뉴클레이스 활성이 없다. 대상체 Cas9 단백질이 실질적인 뉴클레이스 활성이 없는 변이체 Cas9 단백질인 경우, 이는 "dCas9"로 지칭될 수 있다.

[0761] 일부 실시형태에서, 변이체 Cas9 단백질은 감소된 뉴클레이스 활성을 갖는다. 예를 들어, 변이체 Cas9 단백질은 야생형 Cas9 단백질, 예를 들어, 야생형 Cas9 단백질의 엔도뉴클레이스 활성의 약 20% 미만, 약 15% 미만, 약 10% 미만, 약 5% 미만, 약 1% 미만 또는 약 0.1% 미만을 나타낸다.

[0762] 일부 실시형태에서, 변이체 Cas9 단백질은 가이드 표적 서열의 상보적 가닥을 절단할 수 있지만 이중 가닥 가이드 표적 서열의 이중 가닥의 비-상보적 가닥을 절단하는 능력이 감소되어 있다. 예를 들어, 변이체 Cas9 단백질은 RuvC 도메인의 기능을 감소시키는 돌연변이(아미노산 치환)를 가질 수 있다. 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 D10(아미노산 10번 위치에서 아스파테이트에서 알라닌으로)을 가지므로, 이중 가닥 가이드 표적 서열의 상보적 가닥을 절단할 수 있지만 이중 가닥 가이드 표적 서열의 비-상보적 가닥을 절단하는 능력이 감소되어 있다(따라서, 변이체 Cas9 단백질이 이중 가닥 표적 핵산을 절단할 때 이중 가닥 파손(DSB) 대신 단일 가닥 파손(SSB)이 생성됨)(예를 들어, 문헌[Jinek *et al.*, Science. 2012 Aug. 17; 337(6096):816-21] 참조).

[0763] 일부 실시형태에서, 변이체 Cas9 단백질은 이중 가닥 가이드 표적 서열의 비-상보적 가닥을 절단할 수 있지만, 가이드 표적 서열의 상보적 가닥을 절단하는 능력이 감소되어 있다. 예를 들어, 변이체 Cas9 단백질은 HNH 도메인(RuvC/HNH/RuvC 도메인 모티프)의 기능을 감소시키는 돌연변이(아미노산 치환)를 가질 수 있다. 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 H840(아미노산 840번 위치에서 히스티딘에서 알라닌으로) 돌연변이를 가지므로, 가이드 표적 서열의 비-상보적 가닥을 절단할 수 있지만 가이드 표적 서열의 상보적 가닥을 절단하는 능력이 감소되어 있다(따라서, 변이체 Cas9 단백질이 이중 가닥 가이드 표적 서열을 절단할 때 DSB 대신에 SSB를 생성함). 이러한 Cas9 단백질은 가이드 표적 서열(예를 들어, 단일 가닥 가이드 표적 서열)을 절단하는 능력이 감소되어 있지만, 가이드 표적 서열(예를 들어, 단일 가닥 가이드 표적 서열)에 결합하는 능력은 보유한다.

[0764] 일부 실시형태에서, 변이체 Cas9 단백질은 이중 가닥 표적 DNA의 상보적 및 비-상보적 가닥 둘 다를 절단하는 능력이 감소되어 있다. 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 이중 가닥 표적 DNA의 상보적 및 비-상보적 가닥 둘 다를 절단하는 능력이 감소되도록 D10A 및 H840A 돌연변이 둘 다를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력은 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다.

[0765] 또 다른 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA를 절단하는 능력이 감소되도록 W476A 및 W1126A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다.

[0766] 또 다른 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA를 절단하는 능력이 감소되도록 P475A, W476A, N477A, D1125A, W1126A 및 D1127A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다.

[0767] 또 다른 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA를 절단하는 능력이 감소되도록 H840A, W476A 및 W1126A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다. 또 다른 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA를 절단하는 능력이 감소되도록 H840A, D10A, W476A 및 W1126A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다. 일부 실시형태에서, 변이체 Cas9는 Cas9 HNH 도메인(A840H)의 840번 위치에서 촉매성 His 잔기를 복원하였다.

[0768] 또 다른 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA를 절단하는 능력이 감소되도록 H840A, P475A, W476A, N477A, D1125A, W1126A 및 D1127A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다. 또 다른 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA를 절단하는 능력이 감소되도록 D10A, H840A, P475A, W476A, N477A, D1125A, W1126A 및 D1127A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다. 일부 실시형태에서, 변이체 Cas9 단백질이 W476A 및 W1126A 돌연변이를 보유하는 경우 또는 변이체 Cas9 단백질이 P475A, W476A, N477A, D1125A, W1126A 및 D1127A 돌연변이를 보유하는 경우, 변이체 Cas9 단백질은 PAM 서열에 효율적으로 결합하지 않는다. 따라서, 일부 이러한 실시형태에서, 이러한 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 방법은 PAM 서열을 필요로 하지 않는다. 다시 말해서, 일부 실시형태에서, 이러한 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 방법은 가이드 RNA를 포함할 수 있지만, 방법은 PAM 서열의 부재하에 수행될 수 있다(그리고, 따라서 결합의 특이성이 가이드 RNA의 표적화 분절에 의해 제공됨). 다른 잔기가 위의 효과를 달성하도록 돌연변이될 수 있다(즉, 하나 또는 또 다른 뉴클레이스 부분을 불활성화함). 비제한적인 예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 및/또는 A987이 변경(즉, 치환)될 수 있다. 또한, 알라닌 치환 이외의 돌연변이가 적합하다.

[0769] 일부 실시형태에서, 촉매 활성이 감소된 변이체 Cas9 단백질(예를 들어, Cas9 단백질이 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 및/또는 A987 돌연변이, 예를 들어, D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A 및/또는 D986A를 가지는 경우), 변이체 Cas9 단백질은 가이드 RNA와 상호작용하는 능력을 보유하는 한 여전히 부위-특이적 방식으로 표적 DNA에 결합할 수 있다(이는 여전히 가이드 RNA에 의해 표적 DNA 서열로 가이드되기 때문임).

[0770] 일부 실시형태에서, 변이체 Cas 단백질은 spCas9, spCas9-VRQR, spCas9-VRER, xCas9(sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK 또는 spCas9-LRVSQL일 수 있다.

[0771] 일부 실시형태에서, 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E 및 T1337R(SpCas9-MQKFRAER)을 포함하며 변경된 PAM 5'-NGC-3'에 대한 특이성을 갖는 변형된 SpCas9가 사용되었다.

[0772] S. 피오게네스 Cas9에 대한 대안은 포유동물 세포에서 절단 활성을 나타내는 Cpf1 패밀리로부터의 RNA-가이드된 엔도뉴클레이스를 포함할 수 있다. 프레보텔라 및 프란시셀라 1(*Francisella 1*)로부터의 CRISPR(CRISPR/Cpf1)은 CRISPR/Cas9 시스템과 유사한 DNA-편집 기술이다. Cpf1은 클래스 II CRISPR/Cas 시스템의 RNA-가이드된 엔도뉴클레이스이다. 이 후천적 면역 메커니즘은 프레보텔라 및 프란시셀라 세균에서 발견된다. Cpf1 유전자는 가이드 RNA를 사용하여 바이러스 DNA를 찾고 절단하는 엔도뉴클레이스를 코딩하는 CRISPR 유전자좌와 관련이 있다. Cpf1은 Cas9보다 더 작고 단순한 엔도뉴클레이스이며, CRISPR/Cas9 시스템 제한 중 일부를 극복한다. Cas9 뉴클

레이스와 달리, Cpf1-매개성 DNA 절단의 결과는 짧은 3' 돌출부(overhang)가 있는 이중-가닥 파손이다. Cpf1의 엇갈린 절단 패턴은 전통적인 제한 효소 클로닝과 유사한 방향성 유전자 전달 가능성을 열어 유전자 편집의 효율성을 증가시킬 수 있다. 위에 기재된 Cas9 변이체 및 오쏘로그와 마찬가지로, Cpf1은 또한 SpCas9가 선호하는 NGG PAM 부위가 없는 AT-풍부 영역 또는 AT-풍부 게놈으로 CRISPR에 의해 표적화될 수 있는 부위의 수를 확장할 수 있다. Cpf1 유전자좌는 혼합된 알파/베타 도메인, RuvC-I 다음에 나선 영역, RuvC-II 및 아연 핑거-유사 도메인을 포함한다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레이스 도메인을 갖는다.

[0773] 또한, Cpf1은 Cas9와 달리 HNH 엔도뉴클레이스 도메인을 갖지 않고, Cpf1의 N-말단은 Cas9의 알파-나선 인식 로브(lobe)를 갖지 않는다. Cpf1 CRISPR-Cas 도메인 구조는 Cpf1이 기능적으로 고유하며, 클래스 2, 유형 V CRISPR 시스템으로 분류됨을 보여준다. Cpf1 유전자좌는 유형 II 시스템보다 유형 I 및 III에 더 유사한 Cas1, Cas2 및 Cas4 단백질을 암호화한다. 기능적 Cpf1은 트랜스-활성화 CRISPR RNA(tracrRNA)를 필요로 하지 않으므로, CRISPR(crRNA)만이 필요하다. 이는 Cpf1이 Cas9보다 더 작을 뿐만 아니라 더 작은 sgRNA 분자(Cas9의 대략 절반의 뉴클레오타이드)를 갖고 있기 때문에 게놈 편집에 도움이 된다. Cpf1-crRNA 복합체는 Cas9에 의해 표적화된 G-풍부 PAM과 대조적으로 프로토스페이서 인접 모티프 5'-YTN-3' 또는 5'-TTN-3'를 식별함으로써 표적 DNA 또는 RNA를 절단한다. PAM의 식별 후, Cpf1은 4개 또는 5개의 뉴클레오타이드의 돌출부를 갖는 점착성-말단-유사 DNA 이중-가닥 파손을 도입한다.

[0774] 일부 실시형태에서, Cas9는 변경된 PAM 서열에 대한 특이성을 갖는 Cas9 변이체이다. 일부 실시형태에서, 추가적인 Cas9 변이체 및 PAM 서열은 그 전문이 참조에 의해 본 명세서에 원용되어 있는 문헌[Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020)]에 기술되어 있다. 일부 실시형태에서, Cas9 변이체는 특정 PAM 요건이 없다. 일부 실시형태에서, Cas9 변이체, 예를 들어, SpCas9 변이체는 NRNH PAM에 대한 특이성을 갖되, R은 A 또는 G이고, H는 A, C 또는 T이다. 일부 실시형태에서, SpCas9 변이체는 PAM 서열 AAA, TAA, CAA, GAA, TAT, GAT 또는 CAC에 대한 특이성을 갖는다. 일부 실시형태에서, SpCas9 변이체는 1114번, 1134번, 1135번, 1137번, 1139번, 1151번, 1180번, 1188번, 1211번, 1218번, 1219번, 1221번, 1249번, 1256번, 1264번, 1290번, 1318번, 1317번, 1320번, 1321번, 1323번, 1332번, 1333번, 1335번, 1337번 또는 1339번 위치 또는 이의 상응하는 위치에 아미노산 치환을 포함한다. 일부 실시형태에서, SpCas9 변이체는 1114번, 1135번, 1218번, 1219번, 1221번, 1249번, 1320번, 1321번, 1323번, 1332번, 1333번, 1335번 또는 1337번 위치 또는 이의 상응하는 위치에 아미노산 치환을 포함한다. 일부 실시형태에서, SpCas9 변이체는 1114번, 1134번, 1135번, 1137번, 1139번, 1151번, 1180번, 1188번, 1211번, 1219번, 1221번, 1256번, 1264번, 1290번, 1318번, 1317번, 1320번, 1323번, 1333번 위치 또는 이의 상응하는 위치에 아미노산 치환을 포함한다. 일부 실시형태에서, SpCas9 변이체는 1114번, 1131번, 1135번, 1150번, 1156번, 1180번, 1191번, 1218번, 1219번, 1221번, 1227번, 1249번, 1253번, 1286번, 1293번, 1320번, 1321번, 1332번, 1335번, 1339번 위치 또는 이의 상응하는 위치에 아미노산 치환을 포함한다. 일부 실시형태에서, SpCas9 변이체는 1114번, 1127번, 1135번, 1180번, 1207번, 1219번, 1234번, 1286번, 1301번, 1332번, 1335번, 1337번, 1338번, 1349번 위치 또는 이의 상응하는 위치에 아미노산 치환을 포함한다. SpCas9 변이체의 예시적인 아미노산 치환 및 PAM 특이성은 **표 3A** 내지 **표 3D**에 나타나 있다.

표 3

[표 3A]

SpCas9 변이체

| SpCas9 | SpCas9 아미노산 위치 | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1114 | 1135 | 1218 | 1219 | 1221 | 1249 | 1320 | 1321 | 1323 | 1332 | 1333 | 1335 | 1337 |
| | R | D | G | E | Q | P | A | P | A | D | R | R | T |
| AAA | | N | | V | H | | | | | | G | | |
| AAA | | N | | V | H | | | | | | G | | |
| AAA | | | | V | | | | | | | G | | |
| TAA | G | N | | V | | | | | | | I | | |
| TAA | | N | | V | | | | | | | I | | A |
| TAA | G | N | | V | | | | | | | I | | A |
| CAA | | | | V | | | | | | | K | | |
| CAA | | N | | V | | | | | | | K | | |
| CAA | | N | | V | | | | | | | K | | |
| GAA | | | | V | H | | V | | | | K | | |
| GAA | | N | | V | | | V | | | | K | | |
| GAA | | | | V | H | | V | | | | K | | |
| TAT | | | S | V | H | S | | S | | | | L | |
| TAT | | | S | V | H | S | | S | | | | L | |
| TAT | | | S | V | H | S | | S | | | | L | |
| GAT | | | | V | | | | | | | I | | |
| GAT | | | | V | | | | | D | | | Q | |
| GAT | | | | V | | | | | D | | | Q | |
| CAC | | | | V | | | | | | N | | Q | N |
| CAC | | N | | V | | | | | | | | Q | N |
| CAC | | | | V | | | | | | N | | Q | N |

[0775]

[표 3B]

| SpCas9 | SpCas9 아미노산 위치 | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1114 | 1134 | 1135 | 1137 | 1139 | 1151 | 1180 | 1188 | 1211 | 1219 | 1221 | 1256 | 1264 | 1290 | 1318 | 1317 | 1320 | 1323 | 1333 |
| | R | F | D | P | V | K | D | K | K | E | Q | Q | H | V | L | N | A | A | R |
| GAA | | | | | | | | | | V | H | | | | | | V | | K |
| GAA | | | N | S | | | | | | V | | | | | | | V | D | K |
| GAA | | | N | | | | | | | V | H | | Y | | | | V | | K |
| CAA | | | N | | | | | | | V | H | | Y | | | | V | | K |
| CAA | G | | N | S | | | | | | V | H | | Y | | | | V | | K |
| CAA | | | N | | | | | R | | V | H | | | | | | V | | K |
| CAA | | | N | | | | G | | R | V | H | | Y | | | | V | | K |
| CAA | | | N | | | | | | | V | H | | Y | | | | V | | K |
| AAA | | | N | | | | G | | | V | H | R | Y | | | | V | D | K |
| CAA | G | | N | | | | G | | | V | H | | Y | | | | V | D | K |
| CAA | | L | N | | | | G | | | V | H | | Y | | | T | V | D | K |
| TAA | G | | N | | | | G | | | V | H | | Y | G | S | | V | D | K |
| TAA | G | | N | | | E | G | | | V | H | | Y | | S | | V | | K |
| TAA | G | | N | | | | G | | | V | H | | Y | | S | | V | D | K |
| TAA | G | | N | | | | G | | R | V | H | | | | | | V | | K |
| TAA | | | N | | | | G | | R | V | H | | Y | | | | V | | K |
| TAA | G | | N | | A | | G | | | V | H | | | | | | V | | K |
| TAA | G | | N | | | | | | | V | H | | | | | | V | | K |

[0776]

[표 3C]

| SpCas9 | SpCas9 아미노산 위치 | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1114 | 1131 | 1135 | 1150 | 1156 | 1180 | 1191 | 1218 | 1219 | 1221 | 1227 | 1249 | 1253 | 1286 | 1293 | 1320 | 1321 | 1332 | 1335 | 1339 |
| | R | Y | D | E | K | D | K | G | E | Q | A | P | E | N | A | A | P | D | R | T |
| SacB.TAT | | | N | | | | N | | V | H | | | | | | V | S | | L | |
| SacB.TAT | | | N | | | | | S | V | H | | S | | | | | S | G | L | |
| AAT | | | N | | | | | S | V | H | V | S | | K | T | | S | G | L | I |
| TAT | G | | N | | | G | | S | V | H | | S | K | | | | S | G | L | |
| TAT | G | | N | | | G | | S | V | H | | S | | | | | S | G | L | |
| TAT | G | C | N | | | G | | S | V | H | | S | | | | | S | G | L | |
| TAT | G | C | N | | | G | | S | V | H | | S | | | | | S | G | L | |
| TAT | G | C | N | | | G | | S | V | H | | S | | | | | S | G | L | |
| TAT | G | C | N | | E | G | | S | V | H | | S | | | | | S | G | L | |
| TAT | G | C | N | V | | G | | S | V | H | | S | | | | | S | G | L | |
| TAT | | C | N | | | G | | S | V | H | | S | | | | | S | G | L | |
| TAT | G | C | N | | | G | | S | V | H | | S | | | | | S | G | L | |

[0777]

[표 3D]

| SpCas9 | SpCas9 아미노산 위치 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1114 | 1127 | 1135 | 1180 | 1207 | 1219 | 1234 | 1286 | 1301 | 1332 | 1335 | 1337 | 1338 | 1349 |
| | R | D | D | D | E | E | N | N | P | D | R | T | S | H |
| SacB.CAC | | | N | | | V | | | | N | Q | N | | |
| AAC | G | | N | | | V | | | | N | Q | N | | |
| AAC | G | | N | | | V | | | | N | Q | N | | |
| TAC | G | | N | | | V | | | | N | Q | N | | |
| TAC | G | | N | | | V | | H | | N | Q | N | | |
| TAC | G | | N | | G | V | D | H | | N | Q | N | | |
| TAC | G | | N | | | V | | | | N | Q | N | | |
| TAC | G | G | N | E | | V | | H | | N | Q | N | | |
| TAC | G | | N | | | V | | H | | N | Q | N | | |
| TAC | G | | N | | | V | | | | N | Q | N | T | R |

[0778]

[0779] 일부 실시형태에서, Cas9는 네이세리아 메닌지티디스 Cas9(NmeCas9) 또는 이의 변이체이다. 일부 실시형태에서,

NmeCas9는 NNNNGAYW PAM에 대한 특이성을 갖되, Y는 C 또는 T이고, W는 A 또는 T이다. 일부 실시형태에서, NmeCas9는 NNNNGYTT PAM에 대한 특이성을 갖되, Y는 C 또는 T이다. 일부 실시형태에서, NmeCas9는 NNNNGTCT PAM에 대한 특이성을 갖는다. 일부 실시형태에서, NmeCas9는 Nme1 Cas9이다. 일부 실시형태에서, NmeCas9는 NNNNGATT PAM, NNNNCCTA PAM, NNNNCCTC PAM, NNNNCCTT PAM, NNNNCCTG PAM, NNNNCCGT PAM, NNNNCCGGPAM, NNNNCCCA PAM, NNNNCCCT PAM, NNNNCCCC PAM, NNNNCCAT PAM, NNNNCCAG PAM, NNNNCCAT PAM 또는 NNNGATT PAM에 대한 특이성을 갖는다. 일부 실시형태에서, Nme1Cas9는 NNNNGATT PAM, NNNNCCTA PAM, NNNNCCTC PAM, NNNNCCTT PAM 또는 NNNNCCTG PAM에 대한 특이성을 갖는다. 일부 실시형태에서, NmeCas9는 CAA PAM, CAAA PAM 또는 CCA PAM에 대한 특이성을 갖는다. 일부 실시형태에서, NmeCas9는 Nme2 Cas9이다. 일부 실시형태에서, NmeCas9는 NNNNCC(N4CC) PAM에 대한 특이성을 갖되, N은 A, G, C 또는 T 중 어느 하나이다. 일부 실시형태에서, NmeCas9는 NNNNCCGT PAM, NNNNCCGGPAM, NNNNCCCA PAM, NNNNCCCT PAM, NNNNCCCC PAM, NNNNCCAT PAM, NNNNCCAG PAM, NNNNCCAT PAM 또는 NNNGATT PAM에 대한 특이성을 갖는다. 일부 실시형태에서, NmeCas9는 Nme3Cas9이다. 일부 실시형태에서, NmeCas9는 NNNNCAAA PAM, NNNNCC PAM 또는 NNNNCNNN PAM에 대한 특이성을 갖는다. 문헌[Edraki *et al.* Mol. Cell. (2019) 73(4): 714-726]에 기술되어 있는 바와 같은 추가적인 NmeCas9 특징 및 PAM 서열은 그 전체가 참조에 의해 본 명세서에 원용된다.

[0780] 예시적인 네이세리아 메닌지티디스 Cas9 단백질, Nme1Cas9(NCBI 참조: WP_002235162.1; 유형 II CRISPR RNA-가이드된 엔도뉴클레이스 Cas9)는 다음 아미노산 서열을 갖는다:

[0781]
```
   1 MAAFKPNPIN YILGLDIGIA SVGWAMVEID EDENPICLID LGVRVFERAE VPKTGDSLAM
  61 ARRLARSVRR LTRRRAHRLL RARRLLKREG VLQAADFDEN GLIKSLPNTP WQLRAAALDR
 121 KLTPLEWSAV LLHLIKHRGY LSQRKNEGET ADKELGALLK GVADNAHALQ TGDFRTPAEL
 181 ALNKFEKESG HIRNQRGDYS HTFSRKDLQA ELILLFEKQK EFGNPHVSGG LKEGIETLLM
 241 TQRPALSGDA VQKMLGHCTF EPAEPKAAKN TYTAERFIWL TKLNNLRILE QGSERPLTDT
 301 ERATLMDEPY RKSKLTYAQA RKLLGLEDTA FFKGLRYGKD NAEASTLMEM KAYHAISRAL
 361 EKEGLKDKKS PLNLSPELQD EIGTAFSLFK TDEDITGRLK DRIQPEILEA LLKHISFDKF
 421 VQISLKALRR IVPLMEQGKR YDEACAEIYG DHYGKKNTEE KIYLPPIPAD EIRNPVVLRA
 481 LSQARKVING VVRRYGSPAR IHIETAREVG KSFKDRKEIE KRQEENRKDR EKAAAKFREY
 541 FPNFVGEPKS KDILKLRLYE QQHGKCLYSG KEINLGRLNE KGYVEIDHAL PFSRTWDDSF
 601 NNKVLVLGSE NQNKGNQTPY EYFNGKDNSR EWQEFKARVE TSRFPRSKKQ RILLQKFDED
```

[0782]
```
 661 GFKERNLNDT RYVNRFLCQF VADRMRLTGK GKKRVFASNG QITNLLRGFW GLRKVRAEND
 721 RHHALDAVVV ACSTVAMQQK ITRFVRYKEM NAFDGKTIDK ETGEVLHQKT HFPQPWEFFA
 781 QEVMIRVFGK PDGKPEFEEA DTPEKLRTLL AEKLSSRPEA VHEYVTPLFV SRAPNRKMSG
 841 QGHMETVKSA KRLDEGVSVL RVPLTQLKLK DLEKMVNRER EPKLYEALKA RLEAHKDDPA
 901 KAFAEPFYKY DKAGNRTQQV KAVRVEQVQK TGVWVRNHNG IADNATMVRV DVFEKGDKYY
 961 LVPIYSWQVA KGILPDRAVV QGKDEEDWQL IDDSFNFKFS LHPNDLVEVI TKKARMFGYF
1021 ASCHRGTGNI NIRIHDLDHK IGKNGILEGI GVKTALSFQK YQIDELGKEI RPCRLKKRPP
1081 VR
```

[0783] 예시적인 네이세리아 메닌지티디스 Cas9 단백질, Nme2Cas9(NCBI 참조: WP_002230835; 유형 II CRISPR RNA-가이드된 엔도뉴클레이스 Cas9)는 다음 아미노산 서열을 갖는다:

```
   1 MAAFKPNPIN YILGLDIGIA SVGWAMVEID EEENPIRLID LGVRVFERAE VPKTGDSLAM
  61 ARRLARSVRR LTRRRAHRLL RARRLLKREG VLQAADFDEN GLIKSLPNTP WQLRAAALDR
 121 KLTPLEWSAV LLHLIKHRGY LSQRKNEGET ADKELGALLK GVANNAHALQ TGDFRTPAEL
 181 ALNKFEKESG HIRNQRGDYS HTFSRKDLQA ELILLFEKQK EFGNPHVSGG LKEGIETLLM
 241 TQRPALSGDA VQKMLGHCTF EPAEPKAAKN TYTAERFIWL TKLNNLRILE QGSERPLTDT
 301 ERATLMDEPY RKSKLTYAQA RKLLGLEDTA FFKGLRYGKD NAEASTLMEM KAYHAISRAL
 361 EKEGLKDKKS PLNLSSELQD EIGTAFSLFK TDEDITGRLK DRVQPEILEA LLKHISFDKF
 421 VQISLKALRR IVPLMEQGKR YDEACAEIYG DHYGKKNTEE KIYLPPIPAD EIRNPVVLRA
 481 LSQARKVING VVRRYGSPAR IHIETAREVG KSFKDRKEIE KRQEENRKDR EKAAAKFREY
 541 FPNFVGEPKS KDILKLRLYE QQHGKCLYSG KEINLVRLNE KGYVEIDHAL PFSRTWDDSF
 601 NNKVLVLGSE NQNKGNQTPY EYFNGKDNSR EWQEFKARVE TSRFPRSKKQ RILLQKFDED
 661 GFKECNLNDT RYVNRFLCQF VADHILLTGK GKRRVFASNG QITNLLRGFW GLRKVRAEND
 721 RHHALDAVVV ACSTVAMQQK ITRFVRYKEM NAFDGKTIDK ETGKVLHQKT HFPQPWEFFA
 781 QEVMIRVFGK PDGKPEFEEA DTPEKLRTLL AEKLSSRPEA VHEYVTPLFV SRAPNRKMSG
 841 AHKDTLRSAK RFVKHNEKIS VKRVWLTEIK LADLENMVNY KNGREIELYE ALKARLEAYG
 901 GNAKQAFDPK DNPFYKKGGQ LVKAVRVEKT QESGVLLNKK NAYTIADNGD MVRVDVFCKV
 961 DKKGKNQYFI VPIYAWQVAE NILPDIDCKG YRIDDSYTFC FSLHKYDLIA FQKDEKSKVE
1021 FAYYINCDSS NGRFYLAWHD KGSKEQQFRI STQNLVLIQK YQVNELGKEI RPCRLKKRPP
1081 VR
```

[0784]

[0785] ***핵염기 편집기의 Cas12 도메인***

[0786] 전형적으로, 미생물 CRISPR-Cas 시스템은 클래스 1 및 클래스 2 시스템으로 나뉜다. 클래스 1 시스템은 다중 서브유닛 효과기 복합체를 갖는 반면, 클래스 2 시스템은 단일 단백질 효과기를 갖는다. 예를 들어, Cas9 및 Cpf1은 비록 상이한 유형(각각 유형 II 및 유형 V)이지만 클래스 2 효과기이다. Cpf1에 더하여, 클래스 2, 유형 V CRISPR-Cas 시스템은 또한 Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 및 Cas12j/CasΦ를 포함한다. 예를 들어, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Shmakov *et al.*, "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems," Mol. Cell, 2015 Nov. 5; 60(3): 385-397; Makarova *et al.*, "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" CRISPR Journal, 2018, 1(5): 325-336; 및 Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91] 참조. 유형 V Cas 단백질은 RuvC(또는 RuvC-유사) 엔도뉴클레이스 도메인을 포함한다. 성숙 CRISPR RNA(crRNA)의 생산은 일반적으로 tracrRNA-독립적이지만, 예를 들어, Cas12b/C2c1은 crRNA의 생산을 위해 tracrRNA를 필요로 한다. Cas12b/C2c1은 DNA 절단을 위해 crRNA 및 tracrRNA 둘 다에 의존한다.

[0787] 본 발명에서 상정된 핵산 프로그램 가능 DNA 결합 단백질은 클래스 2, 유형 V(Cas12 단백질)로 분류되는 Cas 단백질을 포함한다. Cas 클래스 2, 유형 V 단백질의 비제한적인 예는 Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 및 Cas12j/CasΦ, 이들의 상동체 또는 이들의 변형된 버전을 포함한다. 본 명세서에서 사용되는 바와 같이, Cas12 단백질은 또한 Cas12 뉴클레이스, Cas12 도메인 또는 Cas12 단백질 도메인으로도 지칭될 수 있다. 일부 실시형태에서, 본 발명의 Cas12 단백질은 데아미네이스 도메인과 같은 내부 융합 단백질 도메인에 의해 중단된 아미노산 서열을 포함한다.

[0788] 일부 실시형태에서, Cas12 도메인은 뉴클레이스 불활성 Cas12 도메인 또는 Cas12 닉케이스이다. 일부 실시형태에서, Cas12 도메인은 뉴클레이스 활성 도메인이다. 예를 들어, Cas12 도메인은 이중나선 핵산(예를 들어, 이중나선 DNA 분자)의 하나의 가닥에 흠을 내는(nick) Cas12 도메인일 수 있다. 일부 실시형태에서, Cas12 도메인은 본 명세서에 기재된 바와 같은 아미노산 서열 중 어느 하나를 포함한다. 일부 실시형태에서, Cas12 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, Cas12 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, Cas12 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200

개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 적어도 900개, 적어도 1000개, 적어도 1100개 또는 적어도 1200개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0789] 일부 실시형태에서, Cas12의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시형태에서, 단백질은 2개의 Cas12 도메인 중 하나를 포함한다: (1) Cas12의 gRNA 결합 도메인; 또는 (2) Cas12의 DNA 절단 도메인. 일부 실시형태에서, Cas12 또는 이들의 단편을 포함하는 단백질은 "Cas12 변이체"로 지칭된다. Cas12 변이체는 Cas12 또는 이의 단편에 대한 상동성을 공유한다. 예를 들어, Cas12 변이체는 야생형 Cas12와 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일하다. 일부 실시형태에서, Cas12 변이체는 야생형 Cas12와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 아미노산 변화를 가질 수 있다. 일부 실시형태에서, Cas12 변이체는 단편이 야생형 Cas12의 상응하는 단편과 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일한, Cas12의 단편(예를 들어, gRNA 결합 도메인 또는 DNA 절단 도메인)을 포함한다. 일부 실시형태에서, 단편은 상응하는 야생형 Cas12의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 동일, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5%이다. 일부 실시형태에서, 단편은 적어도 100개의 아미노산 길이이다. 일부 실시형태에서, 단편은 적어도 100개, 150개, 200개, 250개, 300개, 350개, 400개, 450개, 500개, 550개, 600개, 650개, 700개, 750개, 800개, 850개, 900개, 950개, 1000개, 1050개, 1100개, 1150개, 1200개, 1250개 또는 적어도 1300개의 아미노산 길이이다.

[0790] 일부 실시형태에서, Cas12는 Cas12 뉴클레이스 활성을 변경하는 하나 이상의 돌연변이를 갖는 Cas12 아미노산 서열에 상응하거나 또는 이를 부분적으로 또는 전체적으로 포함한다. 예로서, 이러한 돌연변이는 Cas12의 RuvC 뉴클레이스 도메인 내에 아미노산 치환을 포함한다. 일부 실시형태에서, 야생형 Cas12과 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일 또는 적어도 약 99.9% 동일한 Cas12의 변이체 또는 상동체가 제공된다. 일부 실시형태에서, 약 5개의 아미노산, 약 10개의 아미노산, 약 15개의 아미노산, 약 20개의 아미노산, 약 25개의 아미노산, 약 30개의 아미노산, 약 40개의 아미노산, 약 50개의 아미노산, 약 75개의 아미노산, 약 100개 이상의 아미노산만큼 더 짧거나 또는 긴 아미노산 서열을 갖는 Cas12의 변이체가 제공된다.

[0791] 일부 실시형태에서, 본 명세서에 제공된 바와 같은 Cas12 융합 단백질은 Cas12 단백질의 전장 아미노산 서열, 예를 들어, 본 명세서에 제공된 Cas12 서열 중 하나를 포함한다. 그러나, 다른 실시형태에서, 본 명세서에 제공된 바와 같은 융합 단백질은 전장 Cas12 서열을 포함하지 않고, 이의 하나 이상의 단편만을 포함한다. 적합한 Cas12 도메인의 예시적인 아미노산 서열이 본 명세서에 제공되며, Cas12 도메인 및 단편의 추가적인 적합한 서열은 당업자에게 명백할 것이다.

[0792] 일반적으로, 클래스 2, 유형 V Cas 단백질은 단일 기능적 RuvC 엔도뉴클레이스 도메인을 갖는다(예를 들어, 문헌[Chen *et al.*, "CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity," Science 360:436-439 (2018)] 참조). 일부 경우에, Cas12 단백질은 변이체 Cas12b 단백질이다. (문헌[Strecker *et al.*, Nature Communications, 2019, 10(1): Art. No.: 212] 참조). 일 실시형태에서, 변이체 Cas12 폴리펩타이드는 야생형 Cas12 단백질의 아미노산 서열과 비교할 때 1개, 2개, 3개, 4개, 5개 이상의 아미노산이 상이한 아미노산 서열을 갖는다(예를 들어, 결실, 삽입, 치환, 융합을 가짐). 일부 예에서, 변이체 Cas12 폴리펩타이드는 Cas12 폴리펩타이드의 활성을 감소시키는 아미노산 변화(예를 들어, 결실, 삽입 또는 치환)를 갖는다. 예를 들어, 일부 예에서, 변이체 Cas12는 상응하는 야생형 Cas12b 단백질의 닉케이스 활성의 50% 미만, 40% 미만, 30% 미만, 20% 미만, 10% 미만, 5% 미만 또는 1% 미만을 갖는 Cas12b 폴리펩타이드이다. 일부 경우에, 변이체 Cas12b 단백질은 실질적인 닉케이스 활성이 없다.

[0793] 일부 경우에, 변이체 Cas12b 단백질은 감소된 닉케이스 활성을 갖는다. 예를 들어, 변이체 Cas12b 단백질은 야생형 Cas12b 단백질의 닉케이스 활성의 약 20% 미만, 약 15% 미만, 약 10% 미만, 약 5% 미만, 약 1% 미만 또는

약 0.1% 미만을 나타낸다.

[0794] 일부 실시형태에서, Cas12 단백질은 포유동물 세포에서 활성을 나타내는 Cas12a/Cpf1 패밀리로부터의 RNA-가이드된 엔도뉴클레이스를 포함한다. 프레보텔라 및 프란시셀라 1로부터의 CRISPR(CRISPR/Cpf1)은 CRISPR/Cas9 시스템과 유사한 DNA 편집 기술이다. Cpf1은 클래스 II CRISPR/Cas 시스템의 RNA-가이드된 엔도뉴클레이스이다. 이 후천적 면역 메커니즘은 프레보텔라 및 프란시셀라 세균에서 발견된다. Cpf1 유전자는 가이드 RNA를 사용하여 바이러스 DNA를 찾고 절단하는 엔도뉴클레이스를 코딩하는 CRISPR 유전자좌와 관련이 있다. Cpf1은 Cas9보다 더 작고 단순한 엔도뉴클레이스이며, CRISPR/Cas9 시스템 제한 중 일부를 극복한다. Cas9 뉴클레이스와 달리, Cpf1-매개성 DNA 절단의 결과는 짧은 3' 돌출부가 있는 이중-가닥 파손이다. Cpf1의 엇갈린 절단 패턴은 전통적인 제한 효소 클로닝과 유사한 방향성 유전자 전달 가능성을 열어 유전자 편집의 효율성을 증가시킬 수 있다. 위에 기재된 Cas9 변이체 및 오쏘로그와 마찬가지로, Cpf1은 또한 SpCas9가 선호하는 NGG PAM 부위가 없는 AT-풍부 영역 또는 AT-풍부 게놈으로 CRISPR에 의해 표적화될 수 있는 부위의 수를 확장할 수 있다. Cpf1 유전자좌는 혼합된 알파/베타 도메인, RuvC-I 다음에 나선 영역, RuvC-II 및 아연 핑거-유사 도메인을 포함한다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레이스 도메인을 갖는다. 또한, Cpf1은 Cas9와 달리 HNH 엔도뉴클레이스 도메인을 갖지 않고, Cpf1의 N-말단은 Cas9의 알파-나선 인식 로브를 갖지 않는다. Cpf1 CRISPR-Cas 도메인 구조는 Cpf1이 기능적으로 고유하며, 클래스 2, 유형 V CRISPR 시스템으로 분류됨을 보여준다. Cpf1 유전자좌는 유형 II 시스템보다 유형 I 및 III에 더 유사한 Cas1, Cas2 및 Cas4 단백질을 암호화한다. 기능적 Cpf1은 트랜스-활성화 CRISPR RNA(tracrRNA)를 필요로 하지 않으므로, CRISPR(crRNA)만이 필요하다. 이는 Cpf1이 Cas9보다 더 작을 뿐만 아니라 더 작은 sgRNA 분자(Cas9의 대략 절반의 뉴클레오타이드)를 갖고 있기 때문에 게놈 편집에 도움이 된다. Cpf1-crRNA 복합체는 Cas9에 의해 표적화된 G-풍부 PAM과 대조적으로 프로토스페이서 인접 모티프 5'-YTN-3' 또는 5'-TTTN-3'를 식별함으로써 표적 DNA 또는 RNA를 절단한다. PAM의 식별 후, Cpf1은 4개 또는 5개의 뉴클레오타이드의 돌출부를 갖는 점착성-말단-유사 DNA 이중-가닥 파손을 도입한다.

[0795] 본 발명의 일부 양태에서, 돌연변이된 CRISPR 효소가 표적 서열을 포함하는 표적 폴리뉴클레오타이드의 하나 또는 두 가닥 모두를 절단하는 능력이 결여되도록 상응하는 야생형 효소에 대해 돌연변이된 CRISPR 효소를 암호화하는 벡터가 사용될 수 있다. Cas12는 야생형의 예시적인 Cas12 폴리펩타이드(예를 들어, 바실러스 히사시이로부터의 Cas12)에 대해 적어도 또는 적어도 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%의 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 지칭할 수 있다. Cas12는 야생형의 예시적인 Cas12 폴리펩타이드(예를 들어, 바실러스 히사시이(BhCas12b), 바실러스 종 V3-13(BvCas12b) 및 알리사이클로바실러스 아시디필루스(*Alicyclobacillus acidiphilus*)(AaCas12b) 유래)에 대해 최대 또는 최대 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%의 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 지칭할 수 있다. Cas12는 결실, 삽입, 치환, 변이체, 돌연변이, 융합, 키메라 또는 이들의 임의의 조합과 같은 아미노산 변화를 포함할 수 있는 Cas12 단백질의 야생형 또는 변형된 형태를 지칭할 수 있다.

[0796] 일부 실시형태에서, BhCas12b 가이드 폴리뉴클레오타이드는 다음 서열을 갖는다:

[0797] BhCas12b sgRNA 스캐폴드(밑줄) + 20nt 내지 23nt 가이드 서열($N_n$으로 표시됨)

[0798] 5′GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCU
AUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCACNNNNNNNNNNNNNNNNNNNN-3′

[0799] 일부 실시형태에서, BvCas12b 및 AaCas12b 가이드 폴리뉴클레오타이드는 다음 서열을 갖는다:

[0800] BvCas12b sgRNA 스캐폴드(밑줄) + 20nt 내지 23nt 가이드 서열($N_n$으로 표시됨)

[0801] 5′GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGG
AGCACCUGAAAACAGGUGCUUGGCACNNNNNNNNNNNNNNNNNNNN-3′

[0802] AaCas12b sgRNA 스캐폴드(밑줄) + 20nt 내지 23nt 가이드 서열($N_n$으로 표시됨)

[0803] 5′GUCUAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCCAGGUGGCAAAGCCCG
UUGAACUUCUCAAAAAGAACGAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3′

[0804] ***핵산 프로그램 가능 DNA 결합 단백질***

[0805] 본 개시내용의 일부 양태는 염기 편집기와 같은 단백질을 특정 핵산(예를 들어, DNA 또는 RNA) 서열로 가이드하는데 사용될 수 있는 핵산 프로그램 가능 DNA 결합 단백질로서 작용하는 도메인을 포함하는 융합 단백질을 제공한다. 특정 실시형태에서, 융합 단백질은 핵산 프로그램 가능 DNA 결합 단백질 도메인 및 데아미네이스 도메인을 포함한다. 핵산 프로그램 가능 DNA 결합 단백질의 비제한적인 예는 Cas9(예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i 및 Cas12j/CasΦ를 포함한다. Cas 효소의 비제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9(Csn1 또는 Csx12로도 알려져 있음), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, 유형 II Cas 효과기 단백질, 유형 V Cas 효과기 단백질, 유형 VI Cas 효과기 단백질, CARF, DinG, 이들의 상동체, 또는 이들의 변형된 또는 조작된 버전을 포함한다. 다른 핵산 프로그램 가능 DNA 결합 단백질도 또한 본 개시내용의 범위 내에 있지만, 이들은 본 개시내용에 구체적으로 열거되지 않을 수 있다. 예를 들어, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Makarova *et al.* "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" CRISPR J. 2018 Oct;1:325-336. doi: 10.1089/crispr.2018.0033; Yan *et al.*, "Functionally diverse type V CRISPR-Cas systems" Science. 2019 Jan 4;363(6422):88-91. doi: 10.1126/science.aav7271] 참조.

[0806] Cas9와 다른 PAM 특이성을 갖는 핵산 프로그램 가능 DNA-결합 단백질의 하나의 예는 프레보텔라 및 프란시셀라 1(Cpf1)로부터의 클러스터링된 규칙적 간격의 짧은 회문 반복부이다. Cas9와 마찬가지로, Cpf1도 또한 클래스 2 CRISPR 효과기이다. Cpf1은 Cas9와 구별되는 특징으로 강력한 DNA 간섭을 매개하는 것으로 나타났다. Cpf1은 tracrRNA가 결여되어 있는 단일 RNA-가이드된 엔도뉴클레이스이며, 이는 T-풍부 프로토스페이서-인접 모티프(TTN, TTTN 또는 YTN)를 이용한다. 또한, Cpf1은 엇갈린 DNA 이중-가닥 파손을 통해 DNA를 절단한다. 16개의 Cpf1-패밀리 단백질 중 아시다미노코커스(*Acidaminococcus*) 및 라크노스피라세아에(*Lachnospiraceae*)로부터의 두 가지 효소가 인간 세포에서 효율적인 게놈-편집 활성을 갖는 것으로 나타났다. Cpf1 단백질은 당업계에 공지되어 있으며, 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Yamano *et al.*, "Crystal structure of Cpf1 in complex with guide RNA and target DNA." *Cell* (165) 2016, p. 949-962]에 이전에 기술되었다.

[0807] 본 발명의 조성물 및 방법에서 가이드 뉴클레오타이드 서열-프로그램 가능한 DNA-결합 단백질 도메인으로 사용될 수 있는 뉴클레이스-불활성 Cpf1(dCpf1) 변이체가 유용하다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사하지만 HNH 엔도뉴클레이스 도메인을 갖지 않는 RuvC-유사 엔도뉴클레이스 도메인을 가지며, Cpf1의 N-말단은 Cas9의 알파-나선 인식 로브를 갖지 않는다. 문헌[Zetsche *et al.*, *Cell,* 163, 759-771, 2015](참조에 의해 본 명세서에 원용됨)에서, Cpf1의 RuvC-유사 도메인은 두 DNA 가닥의 절단을 담당하며, RuvC-유사 도메인의 불활성화는 Cpf1 뉴클레이스 활성을 불활성화한다는 것이 밝혀졌다. 예를 들어, 프란시셀라 노비시다(*Francisella novicida*) Cpf1의 D917A, E1006A 또는 D1255A에 상응하는 돌연변이는 Cpf1 뉴클레이스 활성을 불활성화한다. 일부 실시형태에서, 본 개시내용의 dCpf1은 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A 또는 D917A/E1006A/D1255A에 상응하는 돌연변이를 포함한다. 임의의 돌연변이, 예를 들어, Cpf1의 RuvC 도메인을 불활성화하는 치환 돌연변이, 결실 또는 삽입이 본 개시내용에 따라 사용될 수 있음을 이해하여야 한다.

[0808] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)은 Cpf1 단백질일 수 있다. 일부 실시형태에서, Cpf1 단백질은 Cpf1 닉케이스(nCpf1)이다. 일부 실시형태에서, Cpf1 단백질은 뉴클레이스 불활성 Cpf1(dCpf1)이다. 일부 실시형태에서, Cpf1, NCpf1 또는 dCpf1은 본 명세서에 개시된 Cpf1 서열과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, dCpf1은 본 명세서에 개시된 Cpf1 서열과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함하며, D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A 또는 D917A/E1006A/D1255A에 상응하는 돌연변이를 포함한다. 다른 세균 종으로부터의 Cpf1도 또한 본 개시내용에 따

라 사용될 수 있음을 이해하여야 한다.

[0809] 야생형 프란시셀라 노비시다 Cpf1(D917, E1006 및 D1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN

[0810]

[0811] 프란시셀라 노비시다 Cpf1 D917A(A917, E1006 및 D1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN

[0812]

[0813] 프란시셀라 노비시다 Cpf1 E1006A(D917, A1006 및 D1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN

[0814] ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI<u>D</u>RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF<u>A</u>DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA<u>D</u>ANGAYHIGLKGLMLLGRIKNNQEGK

[0815] KLNLVIKNEEYFEFVQNRNN

[0816] 프란시셀라 노비시다 Cpf1 D1255A(D917, E1006 및 A1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI<u>D</u>RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF<u>E</u>DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA<u>A</u>ANGAYHIGLKGLMLLGRIKNNQEGK

[0817] KLNLVIKNEEYFEFVQNRNN

[0818] 프란시셀라 노비시다 Cpf1 D917A/E1006A(A917, A1006 및 D1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN

[0819]

[0820] 프란시셀라 노비시다 Cpf1 D917A/D1255A(A917, E1006 및 A1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI

[0821]

FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN

[0822]

[0823] 프란시셀라 노비시다 Cpf1 E1006A/D1255A(D917, A1006 및 A1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN

[0824]

[0825] 프란시셀라 노비시다 Cpf1 D917A/E1006A/D1255A(A917, A1006 및 A1255는 굵게 표시되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEE
ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLI
DAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSS
NDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQR
VFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK
MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKA
KYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIV
PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKI
FDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQ
KGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFEN
ISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY
RKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVY
QKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGK
KLNLVIKNEEYFEFVQNRNN

[0826]

[0827] 일부 실시형태에서, 융합 단백질에 존재하는 Cas9 도메인 중 하나는 PAM 서열에 대한 요건이 없는 가이드 뉴클레오타이드 서열-프로그램 가능한 DNA-결합 단백질 도메인으로 대체될 수 있다.

[0828] 일부 실시형태에서, Cas9 도메인은 스타필로코커스 아우레우스(SaCas9)로부터의 Cas9 도메인이다. 일부 실시형

태에서, SaCas9 도메인은 뉴클레이스 활성 SaCas9, 뉴클레이스 불활성 SaCas9(SaCas9d) 또는 SaCas9 닉케이스(SaCas9n)이다. 일부 실시형태에서, SaCas9는 N579A 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다.

[0829] 일부 실시형태에서, SaCas9 도메인, SaCas9d 도메인 또는 SaCas9n 도메인은 비-표준 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SaCas9 도메인, SaCas9d 도메인 또는 SaCas9n 도메인은 NNGRRT 또는 NNGRRT PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SaCas9 도메인은 E781X, N967X 및 R1014X 돌연변이 중 하나 이상 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함하되, X는 임의의 아미노산이다. 일부 실시형태에서, SaCas9 도메인은 E781K, N967K 및 R1014H 돌연변이 중 하나 이상 또는 본 명세서에 제공된 임의의 아미노산 서열에 하나 이상의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SaCas9 도메인은 E781K, N967K 또는 R1014H 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다.

[0830] 예시적인 SaCas9 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQ
RVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTG
NELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQL
DQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYN
ALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKP
EFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISN
LKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPV
VKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTG
KENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQ
EE**N**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDF
INRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDA
LIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDY
KYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHD
PQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDY
PNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAE
FIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQ
SIKKYSTDILGNLYEVKSKKHPQIIKKG

[0831]

[0832] 밑줄이 그어져 있고 굵게 표시된 위의 잔기 N579가 SaCas9 닉케이스를 생성하기 위해 돌연변이될 수 있다(예를 들어, A579로).

[0833] 예시적인 SaCas9n 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQ
RVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTG
NELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQL
DQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYN
ALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKP
EFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISN
LKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPV
VKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTG
KENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQ
EEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDF
INRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDA
LIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDY
KYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHD
PQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDY
PNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAE
FIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQ
SIKKYSTDILGNLYEVKSKKHPQIIKKG

[0834]

[0835] SaCas9 닉케이스를 생성하기 위해 N579로부터 돌연변이될 수 있는 위의 잔기 A579가 밑줄이 그어져 있고 굵게 표시되어 있다.

[0836] 예시적인 SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQ
RVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTG
NELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQL
DQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYN
ALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKP
EFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISN
LKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPV
VKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTG
KENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQ
EEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDF
INRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDA
LIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDY
KYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHD
PQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDY
PNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAE
FIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQ
SIKKYSTDILGNLYEVKSKKHPQIIKKG.

[0837]

[0838] SaCas9 닉케이스를 생성하기 위해 N579로부터 돌연변이될 수 있는 위의 잔기 A579가 밑줄이 그어져 있고 굵게 표시되어 있다. SaKKH Cas9를 생성하기 위해 E781, N967 및 R1014로부터 돌연변이될 수 있는 위의 잔기 K781, K967 및 H1014가 밑줄이 그어져 있고 이탤릭체로 표시되어 있다.

[0839] 일부 실시형태에서, napDNAbp는 원형 순열이다. 다음 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래되는 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다. 별표(*)는 종결 코돈을 나타낸다.

[0840] CP5(MSP "NGC = NGG와 같이 규칙적인 Cas9 돌연변이가 있는 Pam 변이체" PID=단백질 상호작용 도메인 및

"D10A" 닉케이스 포함):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ
VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKF
LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATL
IHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITD
EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS
NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL
RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR
LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLL
AQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK
IIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH
AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKR
KV*

[0841]

[0842] 일부 실시형태에서, 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)은 미생물 CRISPR-Cas 시스템의 단일 효과기이다. 미생물 CRISPR-Cas 시스템의 단일 효과기는 제한 없이 Cas9, Cpf1, Cas12b/C2c1 및 Cas12c/C2c3을 포함한다. 전형적으로, 미생물 CRISPR-Cas 시스템은 클래스 1 및 클래스 2 시스템으로 나뉜다. 클래스 1 시스템은 다중 서브유닛 효과기 복합체를 갖는 반면, 클래스 2 시스템은 단일 단백질 효과기를 갖는다. 예를 들어, Cas9 및 Cpf1은 클래스 2 효과기이다. Cas9 및 Cpf1에 더하여, 세 개의 별개의 클래스 2 CRISPR-Cas 시스템(Cas12b/C2c1 및 Cas12c/C2c3)이 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Shmakov *et al.*, "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell,* 2015 Nov. 5; 60(3): 385-397]에 기술되어 있다. 두 시스템 Cas12b/C2c1 및 Cas12c/C2c3의 효과기는 Cpf1과 관련된 RuvC-유사 엔도뉴클레이스 도메인을 포함한다. 제3 시스템은 2개의 예상 HEPN RNase 도메인을 갖는 효과기를 포함한다. 성숙 CRISPR RNA의 생산은 Cas12b/C2c1에 의한 CRISPR RNA의 생산과 달리 tracrRNA-독립적이다. Cas12b/C2c1은 DNA 절단을 위해 CRISPR RNA 및 tracrRNA 둘 다에 의존한다.

[0843] 알리사이클로바실러스 아시도테라스트리스(*Alicyclobaccillus acidoterrastris*) Cas12b/C2c1(AacC2c1)의 결정 구조는 키메라 단일-분자 가이드 RNA(sgRNA)와의 복합체로 보고되었다. 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell,* 2017 Jan. 19; 65(2):310-322] 참조. 결정 구조는 또한 삼원 복합체로 표적 DNA에 결합된 알리사이클로바실러스 아시도테레스트리스 C2c1에서 보고되었다. 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Yang *et al.*, "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec. 15; 167(7):1814-1828] 참조. 표적 및 비-표적 DNA 가닥 둘 다를 포함하는 AacC2c1의 촉매적으로 적격한 형태는 단일 RuvC 촉매 포켓 내에 독립적으로 위치하여 포획되고, Cas12b/C2c1-매개성 절단으로 인해 표적 DNA의 엇갈린 7개-뉴클레오타이드 파손이 생성된다. Cas12b/C2c1 삼원 복합체와 이전에 확인된 Cas9 및 Cpf1 대응물 간의 구조 비교는 CRISPR-Cas9 시스템에 의해 사용되는 메커니즘의 다양성을 보여준다.

[0844] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)은 Cas12b/C2c1 또는 Cas12c/C2c3 단백질일 수 있다. 일부 실시형태에서, napDNAbp는 Cas12b/C2c1 단백질이다.

일부 실시형태에서, napDNAbp는 Cas12c/C2c3 단백질이다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12b/C2c1 또는 Cas12c/C2c3 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12b/C2c1 또는 Cas12c/C2c3 단백질이다. 일부 실시형태에서, napDNAbp는 본 명세서에 제공된 napDNAbp 서열 중 어느 하나와 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종으로부터의 Cas12b/C2c1 또는 Cas12c/C2c3도 또한 본 개시내용에 따라 사용될 수 있음을 이해하여야 한다.

[0845] Cas12b/C2c1((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR-관련 엔도뉴클레이스 C2c1 OS = 알리사이클로바실러스 아시도-테레스트리스(*Alicyclobacillus acido-terrestris*)(균주 ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1) 아미노산 서열은 다음과 같다:

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKT
AEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPL
ADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMR
VYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKN
FVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDA
EIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPD
ATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLD
NLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQS
EARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTS
ASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQR
TLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHG
ICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQY
KFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGK
GKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAA
FSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIF
VSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADS
YSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELLVEADEAREKSVVLMRDPSGIINRG
[0846] NWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQDSACENTGDI

[0847] AacCas12b(알리사이클로바실러스 아시디필루스) - WP_067623834

MAVKSMKVKLRLDNMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECYKT
AEECKAELLERLRARQVENGHCGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPL
ADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKAKAEARKSTDRTADVLRALADFGLKPLMR
VYTDSDMSSVQWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGEAYAKLVEQKSRFEQKN
FVGQEHLVQLVNQLQQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEKWEKLDPDAPFDLYDT
EIKNVQRRNTRRFGSHDLFAKLAEPKYQALWREDASFLTRYAVYNSIVRKLNHAKMFATFTLPD
ATAHPIWTRFDKLGGNLHQYTFLFNEFGEGRHAIRFQKLLTVEDGVAKEVDDVTVPISMSAQLD
DLLPRDPHELVALYFQDYGAEQHLAGEFGGAKIQYRRDQLNHLHARRGARDVYLNLSVRVQSQS
EARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTS
ASISVFRVARKDELKPNSEGRVPFCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIREERQR
TLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPMDANQMTPDWREAFEDELQKLKSLYGI
[0848]

CGDREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKDVVGGNSIEQIEYLERQYK
FLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDDERGK
GKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELLNQAQVHDLLVGTMYAA
FSSRFDARTGAPGIRCRRVPARCAREQNPEPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFF
VSPFSAEEGDFHQIHADLNAAQNLQRRLWSDFDISQIRLRCDWGEVDGEPVLIPRTTGKRTADS
YGNKVFYTKTGVTYYERERGKKRRKVFAQEELSEEEAELLVEADEAREKSVVLMRDPSGIINRG
[0849] DWTRQKEFWSMVNQRIEGYLVKQIRSRVRLQESACENTGDI

[0850] BhCas12b(바실러스 히사시이) NCBI 참조 서열: WP_095142515

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGE
ANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWL
KMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKK
KDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG
GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRY
PHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSG
IESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGET
LVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDE
LIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRK
FLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ
AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSR
FHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDR
KCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFIL
KDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW
[0851] MAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK

[0852] 일부 실시형태에서, Cas12b는 BhCas12b의 변이체인 BvCas12b(V4)이며, BhCas12b와 관련된 다음 변화를 포함한다: S893R, K846R 및 E837G. BhCas12b(V4)는 다음과 같이 표현된다: 5' mRNA 캡---5'UTR---bhCas12b---STOP 서열 --- 3'UTR --- 120폴리A 꼬리.

[0853] 5'UTR:

[0854] GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC

[0855] 3' UTR(TriLink 표준 UTR)

GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCC
[0856] TGCACCCGTACCCCCGTGGTCTTTGAATAAAGTCTGA

[0857] bhCas12b(V4)의 핵산 서열

[0858]
```
ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAGATCCT
TCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGGTGCT
GAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAG
CACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGT
GGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGT
GTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAA
GCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGG
GAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTG
GGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGC
AAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCG
TGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACAT
GTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAA
TACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTC
TGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACAC
CAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTG
AAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGC
ACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCAT
CTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAG
AAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCC
GATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACAC
CGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGC
GGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGA
TCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTT
CCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATAC
CCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGC
CTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAA
CTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGC
ATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTG
```

CCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAAT
CAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACA
CTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGA
AGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAA
GCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAG
CTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGC
TCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGA
CGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACCCGGAAG
TTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAACCCGGCC
AGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGAT
GGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAG
GCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGG
AAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAGACAGGT
TGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGCAGCAGA
TTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGCTGCAGG
ACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGCCGTGCT
GAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGG
AAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAA
GAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACAT
CCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTCATTCTG
AAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCTCCAAGC
AGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTCCGAGCT
GAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGACAAATGG
ATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCAACCAGT
ACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGCCACGAA
AAAGGCCGGCCAGGCAAAAAAGAAAAAG

[0859]

[0860] 일부 실시형태에서, Cas12b는 BvCas12B이다. 일부 실시형태에서, Cas12b는 아래에 제공되는 BvCas12b 예시적인 서열에서 번호가 매겨진 아미노산 치환 S893R, K846R 및 E837G를 포함한다.

[0861] BvCas12b(바실러스 종 V3-13) NCBI 참조 서열: WP_101661451.1

MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQAELI
NIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGK
GTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQKDI
EWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTESYYKEHLTGGEEWIEK
IRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLPESASPEELWKVVAEQQNKMS
EGFGDPKVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLPDAIEHPLWIRY
ESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKG
KQEISFSDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGR
LQSPIGKALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASV
SIFEVVKELPKDQEQKLFYSINDTELFAIHKRSFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQ
IRMLANVLRLETKKTPDERKKAIHKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDEIWKE
SLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRTPGEANR
IETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIILFEN
LNRYLFNLDRSRRENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCH
ALTEEDLKAGSNTLKRLIEDGFINESELAYLKKGDIIPSQGGELFVTLSKRYKKDSDNNELTVI
HADINAAQNLQKRFWQQNSEVYRVPCQLARMGEDKLYIPKSQTETIKKYFGKGSFVKNNTEQEV
YKWEKSEKMKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMFRDPSGYFFNNETWRPQKE
YWSIVNNIIKSCLKKKILSNKVEL

[0862]

[0863] 일부 실시형태에서, Cas12b는 BTCas12b.BTCas12b(바실러스 써모아밀로보란스(*Bacillus thermoamylovorans*))이다. NCBI 참조 서열: WP_041902512

MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKV
SKAEIQAELWDFVLKMQKCNSFTHEVDKDVVFNILRELYEELVPSSVEKKGEANQLSNKF
LYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAE
YGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEHKTLEERIKEDIQAFKSLEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREII
QKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYAT
FCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTV
QLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGT
LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKFVNF
KPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLF
FPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFE
DITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGK
EVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQ
LNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERS
RFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQ
KRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWGNAGK
LKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFG
KLERILISKLTNQYSISTIEDDSSKQSM

[0864]

[0865] 일부 실시형태에서, napDNAbp는 Cas12c를 지칭한다. 일부 실시형태에서, Cas12c 단백질은 Cas12c1 또는 Cas12c1의 변이체이다. 일부 실시형태에서, Cas12 단백질은 Cas12c2 또는 Cas12c2의 변이체이다. 일부 실시형태에서, Cas12 단백질은 올레이필루스 종(*Oleiphilus* sp.) HI0009로부터의 Cas12c 단백질(즉, OspCas12c) 또는 OspCas12c의 변이체이다. 이들 Cas12c 분자는 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91]에 기술되어 있다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12c1, Cas12c2 또는 OspCas12c 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12c1, Cas12c2 또는 OspCas12c 단백질이다. 일부 실시형태에서, napDNAbp는 본 명세서에 기재된 임의의 Cas12c1, Cas12c2 또는 OspCas12c 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종으로부터의 Cas12c1, Cas12c2 또는 OspCas12c도 또한 본 개시내용에 따라 사용될 수 있음을 이해하여야 한다.

[0866] Cas12c1

MQTKKTHLHLISAKASRKYRRTIACLSDTAKKDLERRKQSGAADPAQELSCLKTIKFKLEVPEG
SKLPSFDRISQIYNALETIEKGSLSYLLFALILSGFRIFPNSSAAKTFASSSCYKNDQFASQIK
EIFGEMVKNFIPSELESILKKGRRKNNKDWTEENIKRVLNSEFGRKNSEGSSALFDSFLSKFSQ
ELFRKFDSWNEVNKKYLEAAELLDSMLASYGPFDSVCKMIGDSDSRNSLPDKSTIAFTNNAEIT
VDIESSVMPYMAIAALLREYRQSKSKAAPVAYVQSHLTTTNGNGLSWFFKFGLDLIRKAPVSSK
QSTSDGSKSLQELFSVPDDKLDGLKFIKEACEALPEASLLCGEKGELLGYQDFRTSFAGHIDSW
VANYVNRLFELIELVNQLPESIKLPSILTQKNHNLVASLGLQEAEVSHSLELFEGLVKNVRQTL
KKLAGIDISSSPNEQDIKEFYAFSDVLNRLGSIRNQIENAVQTAKKDKIDLESAIEWKEWKKLK
KLPKLNGLGGGVPKQQELLDKALESVKQIRHYQRIDFERVIQWAVNEHCLETVPKFLVDAEKKK
INKESSTDFAAKENAVRFLLEGIGAAARGKTDSVSKAAYNWFVVNNFLAKKDLNRYFINCQGCI
YKPPYSKRRSLAFALRSDNKDTIEVVWEKFETFYKEISKEIEKFNIFSQEFQTFLHLENLRMKL
LLRRIQKPIPAEIAFFSLPQEYYDSLPPNVAFLALNQEITPSEYITQFNLYSSFLNGNLILLRR
SRSYLRAKFSWVGNSKLIYAAKEARLWKIPNAYWKSDEWKMILDSNVLVFDKAGNVLPAPTLKK
VCEREGDLRLFYPLLRQLPHDWCYRNPFVKSVGREKNVIEVNKEGEPKVASALPGSLFRLIGPA
PFKSLLDDCFFNPLDKDLRECMLIVDQEISQKVEAQKVEASLESCTYSIAVPIRYHLEEPKVSN
QFENVLAIDQGEAGLAYAVFSLKSIGEAETKPIAVGTIRIPSIRRLIHSVSTYRKKKQRLQNFK
QNYDSTAFIMRENVTGDVCAKIVGLMKEFNAFPVLEYDVKNLESGSRQLSAVYKAVNSHFLYFK
EPGRDALRKQLWYGGDSWTIDGIEIVTRERKEDGKEGVEKIVPLKVFPGRSVSARFTSKTCSCC
GRNVFDWLFTEKKAKTNKKFNVNSKGELTTADGVIQLFEADRSKGPKFYARRKERTPLTKPIAK
GSYSLEEIERRVRTNLRRAPKSKQSRDTSQSQYFCVYKDCALHFSGMQADENAAINIGRRFLTA
LRKNRRSDFPSNVKISDRLLDN

[0867]

[0868] Cas12c2

MTKHSIPLHAFRNSGADARKWKGRIALLAKRGKETMRTLQFPLEMSEPEAAAINTTPFAVAYNA
IEGTGKGTLFDYWAKLHLAGFRFFPSGGAATIFRQQAVFEDASWNAAFCQQSGKDWPWLVPSKL
YERFTKAPREVAKKDGSKKSIEFTQENVANESHVSLVGASITDKTPEDQKEFFLKMAGALAEKF
DSWKSANEDRIVAMKVIDEFLKSEGLHLPSLENIAVKCSVETKPDNATVAWHDAPMSGVQNLAI
GVFATCASRIDNIYDLNGGKLSKLIQESATTPNVTALSWLFGKGLEYFRTTDIDTIMQDFNIPA
SAKESIKPLVESAQAIPTMTVLGKKNYAPFRPNFGGKIDSWIANYASRLMLLNDILEQIEPGFE
LPQALLDNETLMSGIDMTGDELKELIEAVYAWVDAAKQGLATLLGRGGNVDDAVQTFEQFSAMM
DTLNGTLNTISARYVRAVEMAGKDEARLEKLIECKFDIPKWCKSVPKLVGISGGLPKVEEEIKV
MNAAFKDVRARMFVRFEEIAAYVASKGAGMDVYDALEKRELEQIKKLKSAVPERAHIQAYRAVL
HRIGRAVQNCSEKTKQLFSSKVIEMGVFKNPSHLNNFIFNQKGAIYRSPFDRSRHAPYQLHADK
LLKNDWLELLAEISATLMASESTEQMEDALRLERTRLQLQLSGLPDWEYPASLAKPDIEVEIQT
ALKMQLAKDTVTSDVLQRAFNLYSSVLSGLTFKLLRRSFSLKMRFSVADTTQLIYVPKVCDWAI
PKQYLQAEGEIGIAARVVTESSPAKMVTEVEMKEPKALGHFMQQAPHDWYFDASLGGTQVAGRI
VEKGKEVGKERKLVGYRMRGNSAYKTVLDKSLVGNTELSQCSMIIEIPYTQTVDADFRAQVQAG
LPKVSINLPVKETITASNKDEQMLFDRFVAIDLGERGLGYAVFDAKTLELQESGHRPIKAITNL
LNRTHHYEQRPNQRQKFQAKFNVNLSELRENTVGDVCHQINRICAYYNAFPVLEYMVPDRLDKQ
LKSVYESVTNRYIWSSTDAHKSARVQFWLGGETWEHPYLKSAKDKKPLVLSPGRGASGKGTSQT
CSCCGRNPFDLIKDMKPRAKIAVVDGKAKLENSELKLFERNLESKDDMLARRHRNERAGMEQPL
TPGNYTVDEIKALLRANLRRAPKNRRTKDTTVSEYHCVFSDCGKTMHADENAAVNIGGKFIADI

[0869] EK

[0870] OspCas12c

MTKLRHRQKKLTHDWAGSKKREVLGSNGKLQNPLLMPVKKGQVTEFRKAFSAYARATKGEMTDG
RKNMFTHSFEPPFKTKPSLHQCELADKAYQSLHSYLPGSLAHFLLSAHALGFRIFSKSGEATAFQ
ASSKIEAYESKLASELACVDLSIQNLTISTLFNALTTSVRGKGEETSADPLIARFYTLLTGKPL
SRDTQGPERDLAEVISRKIASSFGTWKEMTANPLQSLQFFEEELHALDANVSLSPAFDVLIKMN
DLQGDLKNRTIVFDPDAPVFEYNAEDPADIIIKLTARYAKEAVIKNQNVGNYVKNAITTTNANG
LGWLLNKGLSLLPVSTDDELLEFIGVERSHPSCHALIELIAQLEAPELFEKNVFSDTRSEVQGM
IDSAVSNHIARLSSSRNSLSMDSEELERLIKSFQIHTPHCSLFIGAQSLSQQLESLPEALQSGV
NSADILLGSTQYMLTNSLVEESIATYQRTLNRINYLSGVAGQINGAIKRKAIDGEKIHLPAAWS
ELISLPFIGQPVIDVESDLAHLKNQYQTLSNEFDTLISALQKNFDLNFNKALLNRTQHFEAMCR
[0871] STKKNALSKPEIVSYRDLLARLTSCLYRGSLVLRRAGIEVLKKHKIFESNSELREHVHERKHFV

FVSPLDRKAKKLLRLTDSRPDLLHVIDEILQHDNLENKDRESLWLVRSGYLLAGLPDQLSSSFI
NLPIITQKGDRRLIDLIQYDQINRDAFVMLVTSAFKSNLSGLQYRANKQSFVVTRTLSPYLGSK
LVYVPKDKDWLVPSQMFEGRFADILQSDYMVWKDAGRLCVIDTAKHLSNIKKSVFSSEEVLAFL
RELPHRTFIQTEVRGLGVNVDGIAFNNGDIPSLKTFSNCVQVKVSRTNTSLVQTLNRWFEGGKV
SPPSIQFERAYYKKDDQIHEDAAKRKIRFQMPATELVHASDDAGWTPSYLLGIDPGEYGMGLSL
VSINNGEVLDSGFIHINSLINFASKKSNHQTKVVPRQQYKSPYANYLEQSKDSAAGDIAHILDR
LIYKLNALPVFEALSGNSQSAADQVWTKVLSFYTWGDNDAQNSIRKQHWFGASHWDIKGMLRQP
PTEKKPKPYIAFPGSQVSSYGNSQRCSCCGRNPIEQLREMAKDTSIKELKIRNSEIQLFDGTIK
LFNPDPSTVIERRRHNLGPSRIPVADRTFKNISPSSLEFKELITIVSRSIRHSPEFIAKKRGIG
[0872] SEYFCAYSDCNSSLNSEANAAANVAQKFQKQLFFEL

[0873] 일부 실시형태에서, napDNAbp는, 예를 들어, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌 [Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91]에 기술되어 있는 Cas12g, Cas12h 또는 Cas12i를 지칭한다. 10 테라바이트 초과의 서열 데이터를 집계하여, Cas12g, Cas12h 및 Cas12i를 포함하여 이전에 특성화된 클래스 V 단백질과 약한 유사성을 나타내는 유형 V Cas 단백질의 새로운 분류가 확인되었다. 일부 실시형태에서, Cas12 단백질은 Cas12g 또는 Cas12g의 변이체이다. 일부 실시형태에서, Cas12 단백질은 Cas12h 또는 Cas12h의 변이체이다. 일부 실시형태에서, Cas12 단백질은 Cas12i 또는 Cas12i의 변이체이다. 다른 RNA-가이드된 DNA 결합 단백질은 napDNAbp로서 사용될 수 있으며 본 개시내용의 범위 내에 있음을 이해하여야 한다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12g, Cas12h 또는 Cas12i 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12g, Cas12h 또는 Cas12i 단백질이다. 일부 실시형태에서, napDNAbp는 본 명세서에 기재된 임의의 Cas12g, Cas12h 또는 Cas12i 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종으로부터의 Cas12g, Cas12h 또는 Cas12i도 또한 본 개시내용에 따라 사용될 수 있음을 이해하여야 한다. 일부 실시형태에서, Cas12i는 Cas12i1 또는 Cas12i2이다.

[0874] Cas12g1

MAQASSTPAVSPRPRPRYREERTLVRKLLPRPGQSKQEFRENVKKLRKAFLQFNADVSGVCQWA
IQFRPRYGKPAEPTETFWKFFLEPETSLPPNDSRSPEFRRLQAFEAAAGINGAAALDDPAFTNE
LRDSILAVASRPKTKEAQRLFSRLKDYQPAHRMILAKVAAEWIESRYRRAHQNWERNYEEWKKE
KQEWEQNHPELTPEIREAFNQIFQQLEVKEKRVRICPAARLLQNKDNCQYAGKNKHSVLCNQFN
EFKKNHLQGKAIKFFYKDAEKYLRCGLQSLKPNVQGPFREDWNKYLRYMNLKEETLRGKNGGRL
PHCKNLGQECEFNPHTALCKQYQQQLSSRPDLVQHDELYRKWRREYWREPRKPVFRYPSVKRHS
IAKIFGENYFQADFKNSVVGLRLDSMPAGQYLEFAFAPWPRNYRPQPGETEISSVHLHFVGTRP
RIGFRFRVPHKRSRFDCTQEELDELRSRTFPRKAQDQKFLEAARKRLLETFPGNAEQELRLLAV
DLGTDSARAAFFIGKTFQQAFPLKIVKIEKLYEQWPNQKQAGDRRDASSKQPRPGLSRDHVGRH
LQKMRAQASEIAQKRQELTGTPAPETTTDQAAKKATLQPFDLRGLTVHTARMIRDWARLNARQI
IQLAEENQVDLIVLESLRGFRPPGYENLDQEKKRRVAFFAHGRIRRKVTEKAVERGMRVVTVPY
[0875] LASSKVCAECRKKQKDNKQWEKNKKRGLFKCEGCGSQAQVDENAARVLGRVFWGEIELPTAIP

[0876] Cas12h1

MKVHEIPRSQLLKIKQYEGSFVEWYRDLQEDRKKFASLLFRWAAFGYAAREDDGATYISPSQAL
LERRLLLGDAEDVAIKFLDVLFKGGAPSSSCYSLFYEDFALRDKAKYSGAKREFIEGLATMPLD
KIIERIRQDEQLSKIPAEEWLILGAEYSPEEIWEQVAPRIVNVDRSLGKQLRERLGIKCRRPHD
AGYCKILMEVVARQLRSHNETYHEYLNQTHEMKTKVANNLTNEFDLVCEFAEVLEEKNYGLGWY
VLWQGVKQALKEQKKPTKIQIAVDQLRQPKFAGLLTAKWRALKGAYDTWKLKKRLEKRKAFPYM
PNWDNDYQIPVGLTGLGVFTLEVKRTEVVVDLKEHGKLFCSHSHYFGDLTAEKHPSRYHLKFRH
KLKLRKRDSRVEPTIGPWIEAALREITIQKKPNGVFYLGLPYALSHGIDNFQIAKRFFSAAKPD
KEVINGLPSEMVVGAADLNLSNIVAPVKARIGKGLEGPLHALDYGYGELIDGPKILTPDGPRCG
ELISLKRDIVEIKSAIKEFKACQREGLTMSEETTTWLSEVESPSDSPRCMIQSRIADTSRRLNS
FKYQMNKEGYQDLAEALRLLDAMDSYNSLLESYQRMHLSPGEQSPKEAKFDTKRASFRDLLRRR
VAHTIVEYFDDCDIVFFEDLDGPSDSDSRNNALVKLLSPRTLLLYIRQALEKRGIGMVEVAKDG
TSQNNPISGHVGWRNKQNKSEIYFYEDKELLVMDADEVGAMNILCRGLNHSVCPYSFVTKAPEK
KNDEKKEGDYGKRVKRFLKDRYGSSNVRFLVASMGFVTVTTKRPKDALVGKRLYYHGGELVTHD
[0877] LHNRMKDEIKYLVEKEVLARRVSLSDSTIKSYKSFAHV

[0878] Cas12i1

MSNKEKNASETRKAYTTKMIPRSHDRMKLLGNFMDYLMDGTPIFFELWNQFGGGIDRDIISGTA
NKDKISDDLLLAVNWFKVMPINSKPQGVSPSNLANLFQQYSGSEPDIQAQEYFASNFDTEKHQW
KDMRVEYERLLAELQLSRSDMHHDLKLMYKEKCIGLSLSTAHYITSVMFGTGAKNNRQTKHQFY
SKVIQLLEESTQINSVEQLASIILKAGDCDSYRKLRIRCSRKGATPSILKIVQDYELGTNHDDE
VNVPSLIANLKEKLGRFEYECEWKCMEKIKAFLASKVGPYYLGSYSAMLENALSPIKGMTTKNC
KFVLKQIDAKNDIKYENEPFGKIVEGFFDSPYFESDTNVKWVLHPHHIGESNIKTLWEDLNAIH
SKYEEDIASLSEDKKEKRIKVYQGDVCQTINTYCEEVGKEAKTPLVQLLRYLYSRKDDIAVDKI
IDGITFLSKKHKVEKQKINPVIQKYPSFNFGNNSKLLGKIISPKDKLKHNLKCNRNQVDNYIWI
EIKVLNTKTMRWEKHHYALSSTRFLEEVYYPATSENPPDALAARFRTKTNGYEGKPALSAEQIE
QIRSAPVGLRKVKKRQMRLEAARQQNLLPRYTWGKDFNININCKRGNNFEVTLATKVKKKKEKNY
KVVLGYDANIVRKNTYAAIEAHANGDGVIDYNDLPVKPIESGFVTVESQVRDKSYDQLSYNGVK
LLYCKPHVESRRSFLEKYRNGTMKDNRGNNIQIDFMKDFEAIADDETSLYYFNMKYCKLLQSSI
RNHSSQAKEYREEIFELLRDGKLSVLKLSSLSNLSFVMFKVAKSLIGTYFGHLLKKPKNSKSDV
KAPPITDEDKQKADPEMFALRLALEEKRLNKVKSKKEVIANKIVAKALELRDKYGPVLIKGENI
SDTTKKGKKSSTNSFLMDWLARGVANKVKEMVMMHQGLEFVEVNPNFTSHQDPFVHKNPENTFR
ARYSRCTPSELTEKNRKEILSFLSDKPSKRPTNAYYNEGAMAFLATYGLKKNDVLGVSLEKFKQ
IMANILHQRSEDQLLFPSRGGMFYLATYKLDADATSVNWNGKQFWVCNADLVAAYNVGLVDIQK
[0879] DFKKK

[0880] Cas12i2

MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFFKMLQGLFGGITPEIVRFSTEQEKQQQD
IALWCAVNWFRPVSQDSLTHTIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWNDCRQQY
YDLCRELGVEVSDLTHDLEILCREKCLAVATESNQNNSIISVLFGTGEKEDRSVKLRITKKILE
AISNLKEIPKNVAPIQEIILNVAKATKETFRQVYAGNLGAPSTLEKFIAKDGQKEFDLKKLQTD
LKKVIRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQ
FKEIQSLNNLLVVKKLNDFFDSEFFSGEETYTICVHHLGGKDLSKLYKAWEDDPADPENAIVVL
CDDLKNNFKKEPIRNILRYIFTIRQECSAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWTN
AVILPEKAQRNDRPNSLDLRIWLYLKLRHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFR
TPRFGYHLPKLTDQTAIRVNKKHVKAAKTEARIRLAIQQGTLPVSNLKITEISATINSKGQVRI
PVKFDVGRQKGTLQIGDRFCGYDQNQTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITE
NRGNQFDQLSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDAICKYQPRLY
KFNKEYAYLLRDIVRGKSLVELQQIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIIYSYF
STALNASKNNPISDEQRKEFDPELFALLEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRG
EGDLSTTNNATKKKANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLYTSHQDPLVHRNPDKA
MKCRWAAIPVKDIGDWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEELLKWRSDRK
SNIPCWVLQNRLAEKLGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAA
NIALTVKGIGEQSSDEENPDGSRIKLQLTS

[0881]

[0882] 염기 편집기의 대표적인 핵산 및 단백질 서열은 다음과 같다:

[0883] P153에서 BhCas12b GGSGGS-ABE8-Xten20

**GCCACC**ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCA
GATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGA
GGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATC
TACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCG
AGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGA

[0884]

CGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAG
GGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTG
GAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCC
Cggaggctctggaggaagc**TCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTG**
**ACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACA**
**ATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGA**
**AATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTG**
**TACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAG**
**TTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCA**
**TCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTG**
**TTGTGTCGTTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTA**
**CTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGG**
**C**TCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATC
CTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGC
CCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAA
GGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAA
GAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCC
AGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCT
GAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA
TGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGC
GGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCA
CTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAG
AAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGT
GGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCT
GCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA
TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACA
ACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCAT
CAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGA
AGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACA
TCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGT
GGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAG
TCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGG
CCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTT
CCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGC

[0885]

GAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGA
ACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAG
AGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAG
GATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGA
AGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCT
GAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC
CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAAC
CCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAA
GAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAA
TGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCT
ACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAG
ACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGC
AGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGC
TGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGC
CGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCT
GGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGAC**GGCCAGACCG**
**TGTACATCC**CTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACT
TCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAG
CTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCC
TCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCG
ACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGAC
CAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCG
GCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATT*
*ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

[0886]

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGE
ANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPGGSGGSSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYV
TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLC
RFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWL

[0887]

KMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKK
KDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG
GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRY
PHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSG
IESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGET
LVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDE
LIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRK
FLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ
AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSR
FHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDR
KCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFIL
KDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW
MAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAY

[0888]

PYDVPDYAYPYDVPDYA

[0889] K255에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCA
GATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGA
GGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATC
TACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCG
AGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGA
CGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAG
GGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTG
GAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCC
CTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATC
CTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGC
CCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAA
GGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAA
GAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAggaggctctg
gaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAA
GAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATC
GGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCC
TTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTT
TGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGT

[0890]

GTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGA
ACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTT
TTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCT
TCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGAGGACATCC
AGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCT
GAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA
TGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGC
GGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCA
CTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAG
AAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGT
GGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCT
GCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA
TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACA
ACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCAT
CAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGA
AGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACA
TCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGT
GGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAG
TCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGG
CCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTT
CCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGC
GAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGA
ACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAG
AGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAG
GATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGA
AGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCT
GAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC
CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAAC
CCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAA
GAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAA
TGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCT
ACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAG
ACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGC

[0891] AGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGC

TGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGC
CGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCT
GGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGT
GTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTC
ATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCT
CCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC
CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCA
ACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATTAC*
*GCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGE
ANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEYKTLEERIKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGE
GWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSG
SETPGTSESATPESSGEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWL
KMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKK
KDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG
GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRY
PHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSG
IESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGET
LVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDE
LIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRK
FLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ
AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSR
FHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDR
KCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFIL
KDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW

[0892]

TGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGC
CGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCT
GGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGT
GTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTC
ATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCT
CCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC
CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCA
ACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATTAC*
*GCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGE
ANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEYKTLEERIKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGE
GWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSG
SETPGTSESATPESSGEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWL
KMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKK
KDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG
GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRY
PHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSG
IESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGET
LVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDE
LIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRK
FLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ
AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSR
FHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDR
KCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFIL
KDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW
MAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAY
PYDVPDYAYPYDVPDYA

[0893]

[0894] D306에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCA
GATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGA
GGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATC
TACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCG
AGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGA
CGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAG
GGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTG
GAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCC
CTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATC
CTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGC
CCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAA
GGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAA
GAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCC
AGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCT
GAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA
TGGCTGAAAATGGACggaggctctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGA
TGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGT
ACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCC
ACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTT
ATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTC
CCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATG
GACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACG
AATGTGCGGCGCTGTTGTGTCGTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAA
AGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACC
CCTGAGAGCTCTGGCGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGC
GGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCA
CTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAG
AAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGT
GGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCT
GCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA

[0895]

[0896]

TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACA
ACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCAT
CAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGA
AGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACA
TCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGT
GGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAG
TCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGG
CCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTT
CCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGC
GAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGA
ACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAG
AGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAG
GATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGA
AGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCT
GAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC
CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAAC
CCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAA
GAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAA
TGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCT
ACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAG
ACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGC
AGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGC
TGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGC
CGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCT
GGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGT
GTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTC
ATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCT
CCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC
CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCA
ACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATTAC*
*GCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGE
ANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWL
KMDGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTA
HAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPE
SSGENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKK
KDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG
GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRY
PHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSG
IESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGET
LVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDE
LIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRK
FLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ
AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSR
FHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDR
KCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFIL
KDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW
MAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAY
[0897] PYDVPDYAYPYDVPDYA

[0898] D980에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCA
GATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGA
GGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATC
TACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCG
AGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGA
CGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAG
GGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTG
GAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCC
[0899] CTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATC

CTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGC
CCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAA
GGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAA
GAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCC
AGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCT
GAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA
TGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGC
GGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCA
CTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAG
AAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGT
GGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCT
GCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA
TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACA
ACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCAT
CAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGA
AGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACA
TCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGT
GGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAG
TCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGG
CCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTT
CCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGC
GAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGA
ACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAG
AGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAG
GATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGA
AGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCT
GAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC
CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAAC
CCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAA
GAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAA
TGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCT
ACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAG
ACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGC

[0900] AGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGC

TGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGC
CGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCT
GGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACggaggctctgg
aggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAG
AGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCG
GCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCT
TCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTT
GAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTG
TTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAA
CCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTT
TTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTT
CTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGGCCAGACCGT
GTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTC
ATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAGGGCAGCT
CCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC
CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCA
ACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATTAC*
*GCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGE
ANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWL
KMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKK
KDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG
GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRY
PHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSG
IESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGET
LVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDE
LIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRK

[0901]

FLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ
AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSR
FHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDR
KCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGGSGGSSEVEFSHEYWMRHALTLAKRA
RDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEP
CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFR
MPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGGQTVYIPESKDQKQKIIEEFGEGYFIL
KDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW
MAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAY

[0902]

PYDVPDYAYPYDVPDYA

[0903] K1019에서 BhCas12b GGSGGS-ABE8-Xten20

**GCCACC**ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCA
GATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGA
GGTGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATC
TACGAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCG
AGCTGTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGA
CGAGGTGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAG
GGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTG
GAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCC
CTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATC
CTGGGCAAGCTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGC
CCATCGTGAAAGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAA
GGACATGTTCATTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAA
GAGGAATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCC
AGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCT
GAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAA
TGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGC
GGAAGCACCCTAGAGAGGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCA
CTTCATCTGGCGGAATCACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAG
AAAAAGAAGGACGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGT
GGGTCCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCT
GCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAA
TCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACA

[0904]

ACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCAT
CAAGTTCCCTCTGAAGGGCACACTCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGA
AGATACCCTCACAAGGTGGAAAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACA
TCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGT
GGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAG
TCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGG
CCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTT
CCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGC
GAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCCAGAGAGGACAATCTGAAACTGATGA
ACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAG
AGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAG
GATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCTGA
AGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGGCGGAAGTCCCT
GAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACATCGACGAGATCGATCGGACC
CGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGCGAAGTGCGTAGACTGGAAC
CCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCTGAAAGAAGATCGGCTGAA
GAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAA
TGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCT
ACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCCAG
ACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGCTCAGTTCAGC
AGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTGACCAAAGAGAAGC
TGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCTGGACAAAATCGC
CGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAG
GATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCT
GGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGT
GTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACTTC
ATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGgaggctctggaggaagcTCCG
AAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGA
TGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGG
AATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAG
GGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGT
AATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCC
AAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAG
AAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCC

[0905]

CAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAA
ACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCCTGAAAATCAAGAAGGGCAGCT
CCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCTGGCCTC
CGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCAGCGAC
AAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTGACCA
ACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGCGGC
CACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATTAC
GCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGE
ANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG
KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWL
KMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKK
KDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESG
GWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRY
PHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSG
IESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGET
LVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDE
LIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRK
FLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ
AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSR
FHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDR
KCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFIL
KDGVYEWVNAGKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETP
GTSESATPESSGLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW
MAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAY
PYDVPDYAYPYDVPDYA

[0906]

[0907] 위의 서열에 대해, 코작 서열은 굵게 표시되고 밑줄이 그어져 있고; N-말단 핵 국부화 신호(NLS)를 표시하며; 소문자는 GGGSGGS 링커를 나타내고; ABE8을 암호화하는 서열을 표시하고, 변형되지 않은 서열은 BhCas12b를 암호화하며; 이중 밑줄은 Xten20 링커를 나타내고; 단일 밑줄은 C-말단 NLS를 나타내며; GGATCC는 GS 링커를 나타내고; 이탤릭체 문자는 3x 헤마글루티닌(HA) 태그의 코딩 서열을 나타낸다.

[0908] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp)은 Cas12j/CasΦ 단백질일 수 있다. Cas12j/CasΦ는 그 전문이 참조에 의해 본 명세서에 원용되어 있는 문헌 [Pausch *et al.*, "CRISPR-CasΦ from huge phages is a hypercompact genome editor," *Science*, 17 July 2020, Vol. 369, Issue 6501, pp. 333-337]에 기술되어 있다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12j/CasΦ 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 쉽게는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, napDNAbp는 자연-발생적 Cas12j/CasΦ 단백질이다. 일부 실시형태에서, napDNAbp는 뉴클레이스 불활성("활성이 없는") Cas12j/CasΦ 단백질이다. 다른 종으로부터의 Cas12j/CasΦ도 또한 본 개시내용에 따라 사용될 수 있음을 이해하여야 한다.

[0909] 예시적인 Cas12j/CasΦ 아미노산 서열은 다음과 같다:

[0910] >CasΦ-1

MADTPTLFTQFLRHHLPGQRFRKDILKQAGRILANKGEDATIAFLRGKSEESPPDFQPP
VKCPIIACSRPLTEWPIYQASVAIQGYVYGQSLAEFEASDPGCSKDGLLGWFDKTGVCTDYFSV
QGLNLIFQNARKRYIGVQTKVTNRNEKRHKKLKRINAKRIAEGLPELTSDEPESALDETGHLID
PPGLNTNIYCYQQVSPKPLALSEVNQLPTAYAGYSTSGDDPIQPMVTKDRLSISKGQPGYIPEH
QRALLSQKKHRRMRGYGLKARALLVIVRIQDDWAVIDLRSLLRNAYWRRIVQTKEPSTITKLLK
LVTGDPVLDATRMVATFTYKPGIVQVRSAKCLKNKQGSKLFSERYLNETVSVTSIDLGSNNLVA
VATYRLVNGNTPELLQRFTLPSHLVKDFERYKQAHDTLEDSIQKTAVASLPQGQQTEIRMWSMY
GFREAQERVCQELGLADGSIPWNVMTATSTILTDLFLARGGDPKKCMFTSEPKKKKNSKQVLYK
IRDRAWAKMYRTLLSKETREAWNKALWGLKRGSPDYARLSKRKEELARRCVNYTISTAEKRAQC
GRTIVALEDLNIGFFHGRGKQEPGWVGLFTRKKENRWLMQALHKAFLELAHHRGYHVIEVNPAY
TSQTCPVCRHCDPDNRDQHNREAFHCIGCGFRGNADLDVATHNIAMVAITGESLKRARGSVASK
TPQPLAAE*

[0911]

[0912] >CasΦ-2

MPKPAVESEFSKVLKKHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQPPAKC
HVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGVSNHGYSHVQG
LNLIFDHTLGRYDGVLKKVQLRNEKARARLESINASRADEGLPEIKAEEEEVATNETGHLLQPP
GINPSFYVYQTISPQAYRPRDEIVLPPEYAGYVRDPNAPIPLGVVRNRCDIQKGCPGYIPEWQR
EAGTAISPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQDALLVTVRIGTDWVVIDVRGLLRNA
RWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTFTYTLDACGTYARKWTLKGKQTKATLDKLTA
TQTVALVAIDLGQTNPISAGISRVTQENGALQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELR
ARSVEALPEAQQAEVRALDGVSKETARTQLCADFGLDPKRLPWDKMSSNTTFISEALLSNSVSR
DQVFFTPAPKKGAKKKAPVEVMRKDRTWARAYKPRLSVEAQKLKNEALWALKRTSPEYLKLSRR
KEELCRRSINYVIEKTRRRTQCQIVIPVIEDLNVRFFHGSGKRLPGWDNFFTAKKENRWFIQGL
HKAFSDLRTHRSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLT
QVALTGKTMPKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS

[0913]

[0914] >CasΦ-3

MEKEITELTKIRREFPNKKFSSTDMKKAGKLLKAEGPDAVRDFLNSCQEIIGDFKPPVKTNIVS
ISRPFEEWPVSMVGRAIQEYYFSLTKEELESVHPGTSSEDHKSFFNITGLSNYNYTSVQGLNLI
FKNAKAIYDGTLVKANNKNKKLEKKFNEINHKRSLEGLPIITPDFEEPFDENGHLNNPPGINRN
IYGYQGCAAKVFVPSKHKMVSLPKEYEGYNRDPNLSLAGFRNRLEIPEGEPGHVPWFQRMDIPE
GQIGHVNKIQRFNFVHGKNSGKVKFSDKTGRVKRYHHSKYKDATKPYKFLEESKKVSALDSILA
IITIGDDWVVFDIRGLYRNVFYRELAQKGLTAVQLLDLFTGDPVIDPKKGVVTFSYKEGVVPVF
SQKIVPRFKSRDTLEKLTSQGPVALLSVDLGQNEPVAARVCSLKNINDKITLDNSCRISFLDDY
KKQIKDYRDSLDELEIKIRLEAINSLETNQQVEIRDLDVFSADRAKANTVDMFDIDPNLISWDS
MSDARVSTQISDLYLKNGGDESRVYFEINNKRIKRSDYNISQLVRPKLSDSTRKNLNDSIWKLK
RTSEEYLKLSKRKLELSRAVVNYTIRQSKLLSGINDIVIILEDLDVKKKFNGRGIRDIGWDNFF
SSRKENRWFIPAFHKAFSELSSNRGLCVIEVNPAWTSATCPDCGFCSKENRDGINFTCRKCGVS
YHADIDVATLNIARVAVLGKPMSGPADRERLGDTKKPRVARSRKTMKRKDISNSTVEAMVTA*

[0915]

[0916] >CasΦ-4

MYSLEMADLKSEPSLLAKLLRDRFPGKYWLPKYWKLAEKKRLTGGEEAACEYMADKQLDSPPPN
FRPPARCVILAKSRPFEDWPVHRVASKAQSFVIGLSEQGFAALRAAPPSTADARRDWLRSHGAS
EDDLMALEAQLLETIMGNAISLHGGVLKKIDNANVKAAKRLSGRNEARLNKGLQELPPEQEGSA
YGADGLLVNPPGLNLNIYCRKSCCPKPVKNTARFVGHYPGYLRDSDSILISGTMDRLTIIEGMP
GHIPAWQREQGLVKPGGRRRRLSGSESNMRQKVDPSTGPRRSTRSGTVNRSNQRTGRNGDPLLV
EIRMKEDWVLLDARGLLRNLRWRESKRGLSCDHEDLSLSGLLALFSGDPVIDPVRNEVVFLYGE
GIIPVRSTKPVGTRQSKKLLERQASMGPLTLISCDLGQTNLIAGRASAISLTHGSLGVRSSVRI
ELDPEIIKSFERLRKDADRLETEILTAAKETLSDEQRGEVNSHEKDSPQTAKASLCRELGLHPP
SLPWGQMGPSTTFIADMLISHGRDDDAFLSHGEFPTLEKRKKFDKRFCLESRPLLSSETRKALN
ESLWEVKRTSSEYARLSQRKKEMARRAVNFVVEISRRKTGLSNVIVNIEDLNVRIFHGGGKQAP
GWDGFFRPKSENRWFIQAIHKAFSDLAAHHGIPVIESDPQRTSMTCPECGHCDSKNRNGVRFLC
KGCGASMDADFDAACRNLERVALTGKPMPKPSTSCERLLSATTGKVCSDHSLSHDAIEKAS*

[0917]

[0918] >CasΦ-5

MSSLPTPLELLKQKHADLFKGLQFSSKDNKMAGKVLKKDGEEAALAFLSERGVSRGELPNFRPP
AKTLVVAQSRPFEEFPIYRVSEAIQLYVYSLSVKELETVPSGSSTKKEHQRFFQDSSVPDFGYT
SVQGLNKIFGLARGIYLGVITRGENQLQKAKSKHEALNKKRRASGEAETEFDPTPYEYMTPERK
LAKPPGVNHSIMCYVDISVDEFDFRNPDGIVLPSEYAGYCREINTAIEKGTVDRLGHLKGGPGY
IPGHQRKESTTEGPKINFRKGRIRRSYTALYAKRDSRRVRQGKLALPSYRHHMMRLNSNAESAI
LAVIFFGKDWVVFDLRGLLRNVRWRNLFVDGSTPSTLLGMFGDPVIDPKRGVVAFCYKEQIVPV
VSKSITKMVKAPELLNKLYLKSEDPLVLVAIDLGQTNPVGVGVYRVMNASLDYEVVTRFALESE
LLREIESYRQRTNAFEAQIRAETFDAMTSEEQEEITRVRAFSASKAKENVCHRFGMPVDAVDWA
TMGSNTIHIAKWVMRHGDPSLVEVLEYRKDNEIKLDKNGVPKKVKLTDKRIANLTSIRLRFSQE
TSKHYNDTMWELRRKHPVYQKLSKSKADFSRRVVNSIIRRVNHLVPRARIVFIIEDLKNLGKVF
HGSGKRELGWDSYFEPKSENRWFIQVLHKAFSETGKHKGYYIIECWPNWTSCTCPKCSCCDSEN
RHGEVFRCLACGYTCNTDFGTAPDNLVKIATTGKGLPGPKKRCKGSSKGKNPKIARSSETGVSV
TESGAPKVKKSSPTQTSQSSSQSAP*

[0919]

[0920] >CasΦ-6

MNKIEKEKTPLAKLMNENFAGLRFPFAIIKQAGKKLLKEGELKTIEYMTGKGSIEPLPNFKPPV
KCLIVAKRRDLKYFPICKASCEIQSYVYSLNYKDFMDYFSTPMTSQKQHEEFFKKSGLNIEYQN
VAGLNLIFNNVKNTYNGVILKVKNRNEKLKKKAIKNNYEFEEIKTFNDDGCLINKPGINNVIYC
FQSISPKILKNITHLPKEYNDYDCSVDRNIIQKYVSRLDIPESQPGHVPEWQRKLPEFNNTNNP
RRRRKWYSNGRNISKGYSVDQVNQAKIEDSLLAQIKIGEDWIILDIRGLLRDLNRRELISYKNK
LTIKDVLGFFSDYPIIDIKKNLVTFCYKEGVIQVVSQKSIGNKKSKQLLEKLIENKPIALVSID
LGQTNPVSVKISKLNKINNKISIESFTYRFLNEEILKEIEKYRKDYDKLELKLINEA

[0921]

[0922] >CasΦ-7

MSNTAVSTREHMSNKTTPPSPLSLLLRAHFPGLKFESQDYKIAGKKLRDGGPEAVISYLTGKGQ
AKLKDVKPPAKAFVIAQSRPFIEWDLVRVSRQIQEKIFGIPATKGRPKQDGLSETAFNEAVASL
EVDGKSKLNEETRAAFYEVLGLDAPSLHAQAQNALIKSAISIREGVLKKVENRNEKNLSKTKRR
KEAGEEATFVEEKAHDERGYLIHPPGVNQTIPGYQAVVIKSCPSDFIGLPSGCLAKESAEALTD
YLPHDRMTIPKGQPGYVPEWQHPLLNRRKNRRRRDWYSASLNKPKATCSKRSGTPNRKNSRTDQ
IQSGRFKGAIPVLMRFQDEWVIIDIRGLLRNARYRKLLKEKSTIPDLLSLFTGDPSIDMRQGVC
TFIYKAGQACSAKMVKTKNAPEILSELTKSGPVVLVSIDLGQTNPIAAKVSRVTQLSDGQLSHE
TLLRELLSNDSSDGKEIARYRVASDRLRDKLANLAVERLSPEHKSEILRAKNDTPALCKARVCA
ALGLNPEMIAWDKMTPYTEFLATAYLEKGGDRKVATLKPKNRPEMLRRDIKFKGTEGVRIEVSP
EAAEAYREAQWDLQRTSPEYLRLSTWKQELTKRILNQLRHKAAKSSQCEVVVMAFEDLNIKMMH
GNGKWADGGWDAFFIKKRENRWFMQAFHKSLTELGAHKGVPTIEVTPHRTSITCTKCGHCDKAN
RDGERFACQKCGFVAHADLEIATDNIERVALTGKPMPKPESERSGDAKKSVGARKAAFKPEEDA
EAAE*

[0923]

[0924] >CasΦ-8

MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIA
KSREFTEWEIYQSSLAIQEVIFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIK
NAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLNGEQEISFEEIKAFDDKGYLLQKPSPNKSIY
CYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGYVPKWQYTFLSK
KENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPV
IDTKANVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELK
KVNGELTKTLISRHPTPIDFCNKITAYRERYDKLESSIKLDAIKQLTSEQKIEVDNYNNNFTPQ
NTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNKSEIYFTSTDKGKTKDVMKSDYKWFQD
YKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIGIENLVKKNNFF
GGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKN
RNGEKFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCERSGDAKKPVRARKAKAPEFHDK
LAPSYTVVLREAV*

[0925]

[0926] >CasΦ-9

MRSSREIGDKILMRQPAEKTAFQVFRQEVIGTQKLSGGDAKTAGRLYKQGKMEAAREWLLKGAR
DDVPPNFQPPAKCLVVAVSHPFEEWDISKTNHDVQAYIYAQPLQAEGHLNGLSEKWEDTSADQH
KLWFEKTGVPDRGLPVQAINKIAKAAVNRAFGVVRKVENRNEKRRSRDNRIAEHNRENGLTEVV
REAPEVATNADGFLLHPPGIDPSILSYASVSPVPYNSSKHSFVRLPEEYQAYNVEPDAPIPQFV
VEDRFAIPPGQPGYVPEWQRLKCSTNKHRRMRQWSNQDYKPKAGRRAKPLEFQAHLTRERAKGA
LLVVMRIKEDWVVFDVRGLLRNVEWRKVLSEEAREKLTLKGLLDLFTGDPVIDTKRGIVTFLYK
AEITKILSKRTVKTKNARDLLLRLTEPGEDGLRREVGLVAVDLGQTHPIAAAIYRIGRTSAGAL
ESTVLHRQGLREDQKEKLKEYRKRHTALDSRLRKEAFETLSVEQQKEIVTVSGSGAQITKDKVC
NYLGVDPSTLPWEKMGSYTHFISDDFLRRGGDPNIVHFDRQPKKGKVSKKSQRIKRSDSQWVGR
MRPRLSQETAKARMEADWAAQNENEEYKRLARSKQELARWCVNTLLQNTRCITQCDEIVVVIED
LNVKSLHGKGAREPGWDNFFTPKTENRWFIQILHKTFSELPKHRGEHVIEGCPLRTSITCPACS
YCDKNSRNGEKFVCVACGATFHADFEVATYNLVRLATTGMPMPKSLERQGGGEKAGGARKARKK
AKQVEKIVVQANANVTMNGASLHSP*

[0927]

[0928] >CasΦ-10

MDMLDTETNYATETPAQQQDYSPKPPKKAQRAPKGFSKKARPEKKPPKPITLFTQKHFSGVRFL
KRVIRDASKILKLSESRTITFLEQAIERDGSAPPDVTPPVHNTIMAVTRPFEEWPEVILSKALQ
KHCYALTKKIKIKTWPKKGPGKKCLAAWSARTKIPLIPGQVQATNGLFDRIGSIYDGVEKKVTN
RNANKKLEYDEAIKEGRNPAVPEYETAYNIDGTLINKPGYNPNLYITQSRTPRLITEADRPLVE
KILWQMVEKKTQSRNQARRARLEKAAHLQGLPVPKFVPEKVDRSQKIEIRIIDPLDKIEPYMPQ
DRMAIKASQDGHVPYWQRPFLSKRRNRRVRAGWGKQVSSIQAWLTGALLVIVRLGNEAFLADIR
GALRNAQWRKLLKPDATYQSLFNLFTGDPVVNTRTNHLTMAYREGVVNIVKSRSFKGRQTREHL
LTLLGQGKTVAGVSFDLGQKHAAGLLAAHFGLGEDGNPVFTPIQACFLPQRYLDSLTNYRNRYD
ALTLDMRRQSLLALTPAQQQEFADAQRDPGGQAKRACCLKLNLNPDEIRWDLVSGISTMISDLY
IERGGDPRDVHQQVETKPKGKRKSEIRILKIRDGKWAYDFRPKIADETRKAQREQLWKLQKASS
EFERLSRYKINIARAIANWALQWGRELSGCDIVIPVLEDLNVGSKFFDGKGKWLLGWDNRFTPK
KENRWFIKVLHKAVAELAPHRGVPVYEVMPHRTSMTCPACHYCHPTNREGDRFECQSCHVVKNT
DRDVAPYNILRVAVEGKTLDRWQAEKKPQAEPDRPMILIDNQES*

[0929]

[0930] 위의 서열에서 별표(*)는 종결 코돈을 나타낸다. 대안적으로, CasΦ-1은 Cas12j 오쏘로그 1이라고도 한다. 따라서, CasΦ-1 내지 CasΦ-10은 또한 각각Cas12j 오쏘로그 1 내지 오쏘로그 10으로도 지칭될 수 있다.

[0931] ***가이드 폴리뉴클레오타이드***

[0932] 실시형태에서, 가이드 폴리뉴클레오타이드는 가이드 RNA이다. 본 명세서에서 사용되는 용어 "가이드 RNA(gRNA)" 및 이의 문법적 등가물은 표적 DNA에 특이적일 수 있으며 Cas 단백질과 복합체를 형성할 수 있는 RNA를 지칭할 수 있다. RNA/Cas 복합체는 Cas 단백질을 표적 DNA로 "가이드"하는데 도움이 될 수 있다. Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 원형 dsDNA 표적을 핵내 분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥이 먼저 핵내 분해적으로 절단된 다음, 3'-5' 핵산말단 분해적으로(exonucleolytically) 트리밍된다. 사실상, DNA-결합 및 절단에는 전형적으로 단백질 및 두 RNA가 필요하다. 그러나, 단일 가이드 RNA("sgRNA", 또는 간단히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 양태를 단일 RNA 종에 혼입하도록 조작될 수 있다. 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Jinek M. *et al.*, Science 337:816-821(2012)] 참조. Cas9는 CRISPR 반복 서열에서 짧은 모티프(PAM 또는 프로토스페이서 인접 모티프)를 인식하여 자기 대 비-자기를 구별하는데 도움이 된다. Cas9 뉴클레이스 서열 및 구조는 당업자에게 잘 알려져 있다(예를 들어, 문헌 ["Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti, J.J. *et al.*, Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E. *et al.*, Nature 471:602-607(2011); 및 "Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M.*et al*, Science 337:816-821(2012)] 참조, 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있음). Cas9 오쏘로그는 S. 피오게네스 및 S. 써모필루스를 포함하지만 이들로 제한되지 않는 다양한 종에서 기술되어 있다. 추가적인 적합한 Cas9 뉴클레이스 및 서열은 본 개시내용에 기초하여 당업자에게 명백할 수 있으며, 이러한 Cas9 뉴클레이스 및 서열은 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737]에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다. 일부 실시형태에서, Cas9 뉴클레이스는 불활성(예를 들어, 불활성화된) DNA 절단 도메인을 가지며, 즉, Cas9는 닉케이스이다.

[0933] 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 적어도 하나의 단일 가이드 RNA("sgRNA" 또는 "gRNA")이다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 적어도 하나의 tracrRNA이다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 폴리뉴클레오타이드-프로그램 가능한 DNA-결합 도메인(예를 들어, Cas9 또는 Cpf1)을 표적 뉴클레오타이드 서열로 가이드하기 위해 PAM 서열을 필요로 하지 않는다.

[0934] 본 명세서에 개시된 염기 편집기의 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인(예를 들어, CRISPR-유래 도메인)은 가이드 폴리뉴클레오타이드와 회합함으로써 표적 폴리뉴클레오타이드 서열을 인식할 수 있다. 가이드 폴리뉴클레오타이드(예를 들어, gRNA)는 전형적으로 단일-가닥이며, 폴리뉴클레오타이드의 표적 서열에 부위-특이적으로 결합(즉, 상보적 염기 페어링을 통해)하도록 프로그래밍되어 가이드 핵산과 함께 있는 염기 편집기를 표적 서열로 지시할 수 있다. 가이드 폴리뉴클레오타이드는 DNA일 수 있다. 가이드 폴리뉴

클레오타이드는 RNA일 수 있다. 당업자에 의해 이해되는 바와 같이, 가이드 폴리뉴클레오타이드 서열에서, 우라실(U)은 서열에서 티민(T)을 대체한다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 천연 뉴클레오타이드(예를 들어, 아데노신)를 포함한다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 비-천연(또는 비천연) 뉴클레오타이드(예를 들어, 펩타이드 핵산 또는 뉴클레오타이드 유사체)를 포함한다. 일부 실시형태에서, 가이드 핵산 서열의 표적화 영역은 적어도 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 30개의 뉴클레오타이드 길이일 수 있다. 가이드 핵산의 표적화 영역은 10개 내지 30개의 뉴클레오타이드 길이, 또는 15개 내지 25개의 뉴클레오타이드 길이 또는 15개 내지 20개의 뉴클레오타이드 길이일 수 있다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 특히 5' 말단에서 1개, 2개, 3개, 4개의 뉴클레오타이드 등으로 절단될 수 있다. 비제한적인 예로서, 20개의 뉴클레오타이드 길이의 가이드 폴리뉴클레오타이드는 특히 5' 말단에서 1개, 2개, 3개, 4개의 뉴클레오타이드 등으로 절단될 수 있다.

[0935] 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 2개 이상의 개별 폴리뉴클레오타이드를 포함하며, 이는, 예를 들어, 상보적 염기 페어링(예를 들어, 이중 가이드 폴리뉴클레오타이드)을 통해 서로 상호작용할 수 있다. 예를 들어, 가이드 폴리뉴클레오타이드는 CRISPR RNA(crRNA) 및 트랜스-활성화 CRISPR RNA(tracrRNA)을 포함할 수 있다. 예를 들어, 가이드 폴리뉴클레오타이드는 하나 이상의 트랜스-활성화 CRISPR RNA(tracrRNA)를 포함할 수 있다.

[0936] 유형 II CRISPR 시스템에서, CRISPR 단백질(예를 들어, Cas9)에 의한 핵산의 표적화는 전형적으로 표적 서열을 인식하는 서열을 포함하는 제1 RNA 분자(crRNA)와 가이드 RNA-CRISPR 단백질 복합체를 안정화하는 스캐폴드 영역을 형성하는 반복 서열을 포함하는 제2 RNA 분자(trRNA) 사이의 상보적 염기 페어링을 필요로 한다. 이러한 이중 가이드 RNA 시스템은 본 명세서에 개시된 염기 편집기를 표적 폴리뉴클레오타이드 서열로 지시하기 위한 가이드 폴리뉴클레오타이드로서 이용될 수 있다.

[0937] 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 단일 가이드 폴리뉴클레오타이드(예를 들어, sgRNA)를 이용한다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 이중 가이드 폴리뉴클레오타이드(예를 들어, 이중 gRNA)를 이용한다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 하나 이상의 가이드 폴리뉴클레오타이드(예를 들어, 다중 gRNA)를 이용한다. 일부 실시형태에서, 단일 가이드 폴리뉴클레오타이드는 본 명세서에 기재된 상이한 염기 편집기에 이용된다. 예를 들어, 단일 가이드 폴리뉴클레오타이드는 사이티딘 염기 편집기 및 아데노신 염기 편집기에 이용될 수 있다.

[0938] 다른 실시형태에서, 가이드 폴리뉴클레오타이드는 단일 분자(즉, 단일-분자 가이드 핵산)에서 핵산의 폴리뉴클레오타이드 표적화 부분 및 핵산의 스캐폴드 부분 둘 다를 포함할 수 있다. 예를 들어, 단일-분자 가이드 폴리뉴클레오타이드는 단일 가이드 RNA(sgRNA 또는 gRNA)일 수 있다. 본 명세서에서 용어 가이드 폴리뉴클레오타이드 서열은 표적 폴리뉴클레오타이드 서열과 상호작용할 수 있고 염기 편집기를 표적 폴리뉴클레오타이드 서열로 지시할 수 있는 임의의 단일, 이중 또는 다중-분자 핵산을 상정한다.

[0939] 전형적으로, 가이드 폴리뉴클레오타이드(예를 들어, crRNA/trRNA 복합체 또는 gRNA)는 표적 폴리뉴클레오타이드 서열을 인식하고 이에 결합할 수 있는 서열을 포함하는 "폴리뉴클레오타이드-표적화 분절" 및 염기 편집기의 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 성분 내의 가이드 폴리뉴클레오타이드를 안정화하는 "단백질-결합 분절"을 포함한다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드의 폴리뉴클레오타이드 표적화 분절은 DNA 폴리뉴클레오타이드를 인식하고 이에 결합하여 DNA에서 염기의 편집을 용이하게 한다. 다른 실시형태에서, 가이드 폴리뉴클레오타이드의 폴리뉴클레오타이드 표적화 분절은 RNA 폴리뉴클레오타이드를 인식하고 이에 결합하여 RNA에서 염기의 편집을 용이하게 한다. 본 명세서에서 "분절"은 분자의 부문 또는 영역, 예를 들어, 가이드 폴리뉴클레오타이드에서 뉴클레오타이드의 연속적인 신장부를 지칭한다. 분절은 또한 하나 이상의 분자의 영역을 포함할 수 있도록 하는 복합체의 영역/부문을 지칭할 수 있다. 예를 들어, 가이드 폴리뉴클레오타이드가 다중 핵산 분자를 포함하는 경우, 이의 단백질-결합 분절은, 예를 들어, 상보성의 영역을 따라 혼성화되는 다중 개별 분자의 전부 또는 일부를 포함할 수 있다. 일부 실시형태에서, 2개 개별 분자를 포함하는 DNA-표적화 RNA의 단백질-결합 분절은 (i) 100개의 염기쌍 길이인 제1 RNA 분자의 염기쌍 40개 내지 75개; 및 (ii) 50개의 염기쌍 길이인 제2 RNA 분자의 염기쌍 10개 내지 25개를 포함할 수 있다. 특정 맥락에서 달리 구체적으로 정의되지 않는 한 "분절"의 정의는 총 염기쌍의 특정 수로 제한되지 않고, 주어진 RNA 분자로부터의 염기쌍의 임의의 특정 수로 제한되지 않으며, 복합체 내의 개별 분자의 특정 수로 제한되지 않고, 임의의 전체 길이의 RNA 분자의 영역을 포함할 수 있으며, 다른 분자에 대해 상보성을 갖는 영역을 포함할 수 있다.

[0940] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 2개 이상의 RNA, 예를 들어, CRISPR RNA(crRNA) 및 전사 촉진

crRNA(tracrRNA)를 포함할 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 때때로 crRNA 및 tracrRNA의 일부(예를 들어, 기능적 부분)의 융합에 의해 형성된 단일-사슬 RNA 또는 단일 가이드 RNA(sgRNA)를 포함할 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 crRNA 및 tracrRNA를 포함하는 이중 RNA일 수 있다. 또한, crRNA는 표적 DNA와 혼성화할 수 있다.

[0941] 위에 논의된 바와 같이, 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 발현 생성물일 수 있다. 예를 들어, 가이드 RNA를 암호화하는 DNA는 가이드 RNA를 코딩하는 서열을 포함하는 벡터일 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 가이드 RNA와 프로모터를 코딩하는 서열을 포함하는 단리된 가이드 RNA 또는 플라스미드 DNA로 세포를 형질감염시킴으로써 세포 내로 전달될 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 바이러스-매개성 유전자 전달을 사용하는 것과 같은 다른 방식으로 세포에 전달될 수 있다.

[0942] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 단리될 수 있다. 예를 들어, 가이드 RNA는 단리된 RNA의 형태로 세포 또는 유기체에 형질감염될 수 있다. 가이드 RNA는 당업계에 공지된 임의의 시험관내 전사 시스템을 사용하여 시험관내 전사에 의해 제조될 수 있다. 가이드 RNA는 가이드 RNA에 대한 암호화 서열을 포함하는 플라스미드의 형태보다는 단리된 RNA의 형태로 세포에 전달될 수 있다.

[0943] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 다음 3개의 영역을 포함할 수 있다: 염색체 서열의 표적 부위에 상보적일 수 있는 5' 말단에 제1 영역, 줄기 루프 구조를 형성할 수 있는 제2 내부 영역 및 단일-가닥일 수 있는 3' 영역에 제3 영역. 각 가이드 RNA의 제1 영역은 또한 각각의 가이드 RNA가 가이드 융합 단백질을 특정 표적 부위로 가이드하도록 상이할 수 있다. 또한, 각 가이드 RNA의 제2 및 제3 영역은 모든 가이드 RNA에서 동일할 수 있다.

[0944] 가이드 RNA 또는 가이드 폴리뉴클레오타이드의 제1 영역은 가이드 RNA의 제1 영역이 표적 부위와 염기쌍을 이룰 수 있도록 염색체 서열의 표적 부위에서 서열에 상보적일 수 있다. 일부 실시형태에서, 가이드 RNA의 제1 영역은 약 10개 뉴클레오타이드 내지 25개의 뉴클레오타이드(즉, 10개 뉴클레오타이드 내지 25개 뉴클레오타이드; 또는 약 10개 뉴클레오타이드 내지 약 25개 뉴클레오타이드; 또는 10개 뉴클레오타이드 내지 약 25개 뉴클레오타이드; 또는 약 10개 뉴클레오타이드 내지 25개의 뉴클레오타이드) 또는 그 이상을 포함할 수 있다. 예를 들어, 가이드 RNA의 제1 영역과 염색체 서열의 표적 부위 사이의 염기 페어링의 영역은 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 22개, 23개, 24개, 25개 이상 또는 약 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 22개, 23개, 24개, 25개 이상의 뉴클레오타이드 길이일 수 있다. 일부 실시형태에서, 가이드 RNA의 제1 영역은 19개, 20개 또는 21개 또는 약 19개, 20개 또는 21개의 뉴클레오타이드 길이일 수 있다.

[0945] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 2차 구조를 형성하는 제2 영역을 포함할 수 있다. 예를 들어, 가이드 RNA에 의해 형성된 2차 구조는 줄기(또는 헤어핀) 및 루프를 포함할 수 있다. 루프 및 줄기의 길이는 다양할 수 있다. 예를 들어, 루프는 약 3개 내지 10개의 뉴클레오타이드 길이의 범위일 수 있고, 줄기는 약 6개 내지 20개의 염기쌍 길이의 범위일 수 있다. 줄기는 1개 내지 10개 또는 약 10개의 뉴클레오타이드의 하나 이상의 벌지(bulge)를 포함할 수 있다. 제2 영역의 전체 길이는 약 16개 내지 60개의 뉴클레오타이드 길이의 범위일 수 있다. 예를 들어, 루프는 4개 또는 약 4개의 뉴클레오타이드 길이일 수 있고, 줄기는 12개 또는 약 12개의 염기쌍일 수 있다.

[0946] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 본질적으로 단일-가닥일 수 있는 3' 말단에 제3 영역을 포함할 수 있다. 예를 들어, 제3 영역은 때때로 관심 세포의 임의의 염색체 서열에 상보적이지 않으며, 때때로 가이드 RNA의 나머지 부분에 대해서도 상보적이지 않다. 또한, 제3 영역의 길이는 다양할 수 있다. 제3 영역은 4개 초과 또는 약 4개 초과의 뉴클레오타이드 길이일 수 있다. 예를 들어, 제3 영역의 길이는 5개 내지 60개 또는 약 5개 내지 60개의 뉴클레오타이드 길이의 범위일 수 있다.

[0947] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 표적 유전자 표적의 임의의 엑손 또는 인트론을 표적으로 할 수 있다. 일부 실시형태에서, 가이드는 유전자의 엑손 1 또는 2를 표적으로 할 수 있고; 다른 실시형태에서, 가이드는 유전자의 엑손 3 또는 4를 표적으로 할 수 있다. 조성물은 모두 동일한 엑손을 표적으로 하는 다중 가이드 RNA 또는 일부 실시형태에서 상이한 엑손을 표적으로 할 수 있는 다중 가이드 RNA를 포함할 수 있다. 유전자의 엑손 및 인트론이 표적화될 수 있다.

[0948] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 20개 또는 약 20개의 뉴클레오타이드의 핵산 서열을 표적으로 할 수 있다. 표적 핵산은 20개 또는 약 20개 미만의 뉴클레오타이드일 수 있다. 표적 핵산은 적어도 5개, 10개, 15

개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 30개, 또는 적어도 약 5개, 10개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 30개, 또는 1개 내지 100개의 뉴클레오타이드 길이 사이일 수 있다. 표적 핵산은 최대 5개, 10개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 30개, 40개, 50개, 또는 최대 5개, 10개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 30개, 40개, 50개 또는 1개 내지 100개의 뉴클레오타이드 길이 사이일 수 있다. 표적 핵산 서열은 PAM의 제1 뉴클레오타이드의 5' 바로 옆에 있는 20개 또는 약 20개의 염기일 수 있다. 가이드 RNA는 표적 핵산 서열일 수 있다. 표적 핵산은 적어도 1개 내지 10개, 1개 내지 20개, 1개 내지 30개, 1개 내지 40개, 1개 내지 50개, 1개 내지 60개, 1개 내지 70개, 1개 내지 80개, 1개 내지 90개 또는 1개 내지 100개, 또는 적어도 약 1개 내지 10개, 1개 내지 20개, 1개 내지 30개, 1개 내지 40개, 1개 내지 50개, 1개 내지 60개, 1개 내지 70개, 1개 내지 80개, 1개 내지 90개 또는 1개 내지 100개의 뉴클레오타이드일 수 있다.

[0949] 가이드 폴리뉴클레오타이드, 예를 들어, 가이드 RNA는 또 다른 핵산, 예를 들어, 세포의 게놈의 표적 핵산 또는 프로토스페이서에 혼성화할 수 있는 핵산을 지칭한다. 가이드 폴리뉴클레오타이드는 RNA일 수 있다. 가이드 폴리뉴클레오타이드는 DNA일 수 있다. 가이드 폴리뉴클레오타이드는 핵산의 서열에 부위-특이적으로 결합하도록 프로그래밍되거나 또는 설계될 수 있다. 가이드 폴리뉴클레오타이드는 폴리뉴클레오타이드 사슬을 포함할 수 있으며, 단일 가이드 폴리뉴클레오타이드로 불릴 수 있다. 가이드 폴리뉴클레오타이드는 2개 폴리뉴클레오타이드 사슬을 포함할 수 있으며, 이중 가이드 폴리뉴클레오타이드로 불릴 수 있다. 가이드 RNA는 RNA 분자로 세포 또는 배아에 도입될 수 있다. 예를 들어, RNA 분자는 시험관내에서 전사될 수 있고/있거나 화학적으로 합성될 수 있다. RNA는 합성 DNA 분자, 예를 들어, gBlocks® 유전자 단편으로부터 전사될 수 있다. 그런 다음, 가이드 RNA는 RNA 분자로 세포 또는 배아에 도입될 수 있다. 가이드 RNA는 또한 비-RNA 핵산 분자, 예를 들어, DNA 분자의 형태로 세포 또는 배아에 도입될 수 있다. 예를 들어, 가이드 RNA를 암호화하는 DNA는 관심 세포 또는 배아에서 가이드 RNA의 발현을 위한 프로모터 제어 서열에 작동 가능하게 연결될 수 있다. RNA 코딩 서열은 RNA 폴리머레이스 III(Pol III)에 의해 인식되는 프로모터 서열에 작동 가능하게 연결될 수 있다. 가이드 RNA를 발현하는데 사용될 수 있는 플라스미드 벡터는 px330 벡터 및 px333 벡터를 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 플라스미드 벡터(예를 들어, px333 벡터)는 적어도 2개의 가이드 RNA-암호화 DNA 서열을 포함할 수 있다.

[0950] 가이드 폴리뉴클레오타이드, 예를 들어, 가이드 RNA 및 표적화 서열을 선택, 설계 및 검증하는 방법이 본 명세서에 기재되어 있으며 당업자에게 공지되어 있다. 예를 들어, 핵염기 편집 시스템(예를 들어, AID 도메인)에서 데아미네이스 도메인의 잠재적 기질 난잡성(promiscuity)의 영향을 최소화하기 위해, 의도하지 않게 탈아미노화에 대해 표적화될 수 있는 잔기(예를 들어, 표적 핵산 유전자좌 내에서 ssDNA에 잠재적으로 상주할 수 있는 표적외 C 잔기)의 수가 최소화될 수 있다. 또한, 소프트웨어 도구가 표적 핵산 서열에 상응하는 gRNA를 최적화하기 위해, 예를 들어, 게놈 전반에 걸쳐 총 표적외 활성을 최소화하기 위해 사용될 수 있다. 예를 들어, S. 피오게네스 Cas9를 사용한 각각의 가능한 표적화 도메인 선택에 대해, 모든 표적외 서열(선택된 PAM, 예를 들어, NAG 또는 NGG에 선행함)은 최대 소정의 수(예를 들어, 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개)의 불일치 염기쌍을 포함하는 게놈 전반에 걸쳐 확인될 수 있다. 표적 부위에 상보적인 gRNA의 제1 영역이 확인될 수 있고, 모든 제1 영역(예를 들어, crRNA)은 예측된 이의 총 표적외 점수에 따라 순위가 매겨질 수 있고; 최상위의 표적화 도메인은 표적에 가장 크게 작용하며 최소 표적외 활성을 가질 가능성이 있는 도메인을 나타낸다. gRNA를 표적화하는 후보물질은 당업계에 공지된 방법 및/또는 본 명세서에 기재된 바와 같은 방법을 사용함으로써 기능적으로 평가될 수 있다.

[0951] 비제한적인 예로서, Cas9와 함께 사용하기 위한 가이드 RNA의 crRNA에서 표적 DNA 혼성화 서열은 DNA 서열 검색 알고리즘을 사용하여 확인될 수 있다. gRNA 설계는 문헌[Bae S., Park J., & Kim J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473-1475 (2014)]에 기술되어 있는 바와 같이 공개 도구 cas-offinder에 기반한 맞춤형 gRNA 설계 소프트웨어를 사용하여 수행될 수 있다. 이 소프트웨어는 전장유전체 표적외 성향(genome-wide off-target propensity)을 계산한 후 가이드를 점수화한다. 전형적으로 17개 내지 24개 길이의 범위인 가이드에 대해 완벽한 일치에서 7개의 불일치에 이르는 일치가 고려된다. 일단 표적외 부위가 계산적으로 결정되면, 각 가이드에 대해 집계 점수가 계산되고, 웹-인터페이스를 사용하여 표 형식의 결과로 요약된다. PAM 서열에 인접한 잠재적 표적 부위를 식별하는 것 외에도, 소프트웨어는 또한 선택된 표적 부위와 1개, 2개, 3개 또는 3개 이상의 뉴클레오타이드가 다른 모든 PAM 인접 서열을 식별한다. 예를 들어, 표적 유전자에 대한 게놈 DNA 서열이 얻어질 수 있고, 반복 요소는 공개적으로 이용 가능한 도구, 예를 들어, RepeatMasker 프로그램을 사용하여 스크리닝될 수 있다. RepeatMasker는 반복되는 요소 및 낮은 복잡성이 낮은 영역에 대해 입력 DNA 서열을 검색한다.

결과는 주어진 쿼리 서열에 있는 반복의 자세한 주석이다.

[0952] 식별 후, 가이드 RNA, 예를 들어, crRNA의 제1 영역은 표적 부위까지의 거리, 이의 직교성 및 관련 PAM 서열과의 근접한 일치를 위한 5' 뉴클레오타이드의 존재에 기초하여(예를 들어, 5' G는 관련 PAM, 예를 들어, S. 피오게네스의 경우 NGG PAM, S. 아우레우스의 경우 NNGRRT 또는 NNGRRV PAM을 포함하는 인간 게놈에서 근접한 일치의 식별에 기초함) 단계적으로 순위가 매겨질 수 있다. 본 명세서에서 사용되는 바와 같이, 직교성은 표적 서열에 대한 최소 수의 불일치를 포함하는 인간 게놈의 서열의 수를 지칭한다. "높은 수준의 직교성" 또는 "우수한 직교성"은, 예를 들어, 의도된 표적 외에 인간 게놈의 동일한 서열 또는 표적 서열에 1개 또는 2개의 불일치를 포함하는 임의의 서열을 갖지 않는 20량체 표적화 도메인을 지칭할 수 있다. 우수한 직교성을 갖는 표적화 도메인은 표적외 DNA 절단을 최소화하도록 선택될 수 있다.

[0953] 일부 실시형태에서, 리포터 시스템은 염기-편집 활성을 검출하고 후보 가이드 폴리뉴클레오타이드를 테스트하는데 사용될 수 있다. 일부 실시형태에서, 리포터 시스템은 염기 편집 활성이 리포터 유전자의 발현을 유도하는 리포터 유전자 기반 검정을 포함할 수 있다. 예를 들어, 리포터 시스템은 비활성화된(deactivated) 개시 코돈, 예를 들어, 3'-TAC-5'에서 3'-CAC-5'로의 주형 가닥의 돌연변이를 포함하는 리포터 유전자를 포함할 수 있다. 표적 C의 성공적인 탈아미노화시, 해당 mRNA는 5'-GUG-3' 대신 5'-AUG-3'로 전사되어 리포터 유전자의 번역을 가능하게 할 것이다. 적합한 리포터 유전자 당업자에게 명백할 것이다. 리포터 유전자의 비제한적인 예는 녹색 형광 단백질(green fluorescence protein: GFP), 적색 형광 단백질(red fluorescence protein: RFP), 루시퍼레이스, 분비성 알칼리 포스파테이스(secreted alkaline 포스파테이스: SEAP) 또는 발현의 검출이 가능하고 당업자에게 명백한 임의의 다른 유전자를 암호화하는 유전자를 포함한다. 리포터 시스템은, 예를 들어, 각 데아미네이스가 표적으로 할 표적 DNA 서열과 관련한 어떤 뉴클레오타이드 잔기(들)를 결정하기 위해 많은 상이한 gRNA를 테스트하는데 사용될 수 있다. 비-주형 가닥 뉴클레오타이드 잔기를 표적으로 하는 sgRNA는 또한 특정 염기 편집 단백질, 예를 들어, Cas9 데아미네이스 융합 단백질의 표적외 효과를 평가하기 위해 테스트될 수 있다. 일부 실시형태에서, 이러한 gRNA는 돌연변이된 개시 코돈이 gRNA와 염기쌍을 이루지 않도록 설계될 수 있다. 가이드 폴리뉴클레오타이드는 표준 뉴클레오타이드, 변형된 뉴클레오타이드(예를 들어, 슈도우리딘), 뉴클레오타이드는 이성질체 및/또는 뉴클레오타이드 유사체를 포함할 수 있다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드는 적어도 하나의 검출 가능한 표지를 포함할 수 있다. 검출 가능한 표지는 형광단(예를 들어, FAM, TMR, Cy3, Cy5, 텍사스 레드, 오레곤 그린, 알렉사 플루오르, 할로 태그 또는 임의의 다른 적합한 형광 염료), 검출 태그(예를 들어, 바이오틴, 디곡시게닌 등), 양자점 또는 금 입자일 수 있다.

[0954] 가이드 폴리뉴클레오타이드는 화학적으로 및/또는 효소적으로 합성될 수 있다. 예를 들어, 가이드 RNA는 표준 포스포라미다이트-기반 고상 합성 방법을 사용하여 합성될 수 있다. 대안적으로, 가이드 RNA는 RNA를 암호화하는 DNA를 파지 RNA 폴리머레이스에 의해 인식되는 프로모터 제어 서열에 작동 가능하게 연결함으로써 시험관내에서 합성될 수 있다. 적합한 파지 프로모터 서열의 예는 T7, T3, SP6 프로모터 서열 또는 이들의 변이를 포함한다. 가이드 RNA가 2개 개별 분자(예를 들어, crRNA 및 tracrRNA)를 포함하는 실시형태에서, crRNA는 화학적으로 합성될 수 있고, tracrRNA는 효소적으로 합성될 수 있다.

[0955] 일부 실시형태에서, 염기 편집 시스템은 다중 가이드 폴리뉴클레오타이드, 예를 들어, gRNA를 포함할 수 있다. 예를 들어, gRNA는 표적 염기 편집기를 하나 이상의 표적 유전자좌(예를 들어, 적어도 하나(1)의 gRNA, 적어도 2개의 gRNA, 적어도 5개의 gRNA, 적어도 10개의 gRNA, 적어도 20개의 gRNA, 적어도 30개의 gRNA 또는 적어도 50개의 gRNA)로 표적화할 수 있다. 일부 실시형태에서, 다중 gRNA 서열은 직렬로 배열될 수 있고, 바람직하게는 직접 반복에 의해 분리된다.

[0956] 가이드 RNA 또는 가이드 폴리뉴클레오타이드를 암호화하는 DNA 서열은 또한 벡터의 일부일 수 있다. 일부 실시형태에서, 벡터는 추가적인 발현 제어 서열(예를 들어, 인핸서 서열, 코작 서열, 폴리아데닐화 서열, 전사 종결 서열 등), 선별 마커 서열(예를 들어, GFP 또는 항생제 저항성 유전자, 예컨대, 퓨로마이신), 복제 원점 등을 포함한다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드를 암호화하는 DNA 분자는 원형 또는 선형일 수 있다.

[0957] 일부 실시형태에서, 염기 편집 시스템의 하나 이상의 성분은 DNA 서열에 의해 암호화될 수 있다. 이러한 DNA 서열은 함께 또는 별도로 발현 시스템, 예를 들어, 세포에 도입될 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 및 가이드 RNA를 암호화하는 DNA 서열이 세포에 도입될 수 있으며, 각 DNA 서열은 개별 분자의 일부(예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 코딩 서열을 포함하는 하나의 벡터 및 가이드 RNA 코딩 서열을 포함하는 제2 벡터)일 수 있거나, 또는 둘 다 동일한 분자의 일부(예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 및 가이드 RNA 둘

다에 대한 코딩(및 조절) 서열을 포함하는 하나의 벡터)일 수 있다.

[0958] 가이드 폴리뉴클레오타이드는 새로운 또는 향상된 특징을 갖는 핵산을 제공하기 위해 하나 이상의 변형을 포함할 수 있다. 가이드 폴리뉴클레오타이드는 핵산 친화성 태그를 포함할 수 있다. 가이드 폴리뉴클레오타이드는 합성 뉴클레오타이드, 합성 뉴클레오타이드 유사체, 뉴클레오타이드 유도체 및/또는 변형된 뉴클레오타이드를 포함할 수 있다.

[0959] 일부 실시형태에서, gRNA 또는 가이드 폴리뉴클레오타이드는 변형을 포함할 수 있다. 변형은 gRNA 또는 가이드 폴리뉴클레오타이드의 임의의 위치에서 이루어질 수 있다. 하나 이상의 변형이 단일 gRNA 또는 가이드 폴리뉴클레오타이드에 대해 이루어질 수 있다. gRNA 또는 가이드 폴리뉴클레오타이드는 변형 후 품질 관리를 거칠 수 있다. 일부 실시형태에서, 품질 관리는 PAGE, HPLC, MS 또는 이들의 임의의 조합을 포함할 수 있다.

[0960] gRNA 또는 가이드 폴리뉴클레오타이드의 변형은 치환, 삽입, 결실, 화학 변형, 물리적 변형, 안정화, 정제 또는 이들의 임의의 조합일 수 있다.

[0961] gRNA 또는 가이드 폴리뉴클레오타이드는 또한 5' 아데닐레이트, 5' 구아노신-트라이포스페이트 캡, 5'N7-메틸구아노신-트라이포스페이트 캡, 5'트라이포스페이트 캡, 3'포스페이트, 3'티오포스페이트, 5'포스페이트, 5'티오포스페이트, Cis-Syn 티미딘 이량체, 삼량체, C12 스페이서, C3 스페이서, C6 스페이서, d스페이서, PC 스페이서, r스페이서, 스페이서 18, 스페이서 9,3'-3' 변형, 5'-5' 변형, 무염기, 아크리딘, 아조벤젠, 바이오틴, 바이오틴 BB, 바이오틴 TEG, 콜레스테릴 TEG, 데스티오바이오틴 TEG, DNP TEG, DNP-X, DOTA, dT-바이오틴, 이중 바이오틴, PC 바이오틴, 소랄렌 C2, 소랄렌 C6, TINA, 3'DABCYL, 블랙홀 소광제 1, 블랙홀 소광제 2, DABCYL SE, dT-DABCYL, IRDye QC-1, QSY-21, QSY-35, QSY-7, QSY-9, 카복실 링커, 티올 링커, 2'-데옥시리보뉴클레오사이드 유사체 퓨린, 2'-데옥시리보뉴클레오사이드 유사체 피리미딘, 리보뉴클레오사이드 유사체, 2'-O-메틸 리보뉴클레오사이드 유사체, 당 변형된 유사체, 워블(wobble)/범용 염기, 형광 염료 표지, 2'-플루오로 RNA, 2'-O-메틸 RNA, 메틸포스포네이트, 포스포다이에스터 DNA, 포스포다이에스터 RNA, 포스포티오에이트 DNA, 포스포로티오에이트 RNA, UNA, 슈도우리딘-5'-트라이포스페이트, 5'-메틸사이티딘-5'-트라이포스페이트 또는 이들의 임의의 조합에 의해 변형될 수 있다.

[0962] 일부 실시형태에서, 변형은 영구적이다. 다른 실시형태에서, 변형은 일시적이다. 일부 실시형태에서, 여러 변형이 gRNA 또는 가이드 폴리뉴클레오타이드에 대해 이루어질 수 있다. gRNA 또는 가이드 폴리뉴클레오타이드 변형은 형태, 극성, 소수성, 화학 반응성, 염기-페어링 상호작용 또는 이들의 임의의 조합과 같은 뉴클레오타이드의 물리화학적 특성을 변경할 수 있다.

[0963] 변형은 또한 포스포로티오에이트 대체물일 수 있다. 일부 실시형태에서, 천연 포스포다이에스터 결합은 세포 뉴클레이스에 의한 빠른 분해에 민감할 수 있고; 포스포로티오에이트(PS) 결합 대체물을 이용한 뉴클레오타이드간 연결의 변형은 세포 분해에 의해 가수분해에 더 안정적일 수 있다. 변형은 gRNA 또는 가이드 폴리뉴클레오타이드에서 안정성을 증가시킬 수 있다. 변형은 또한 생물학적 활성을 향상시킬 수 있다. 일부 실시형태에서, 포스포로티오에이트 강화 RNA gRNA는 RNase A, RNase T1, 송아지 혈청 뉴클레이스 또는 이들의 임의의 조합을 저해할 수 있다. 이러한 특성은 PS-RNA gRNA를 뉴클레이스에 대한 노출이 생체내 또는 시험관내에서 높은 확률로 적용되는 응용 분야에서 사용될 수 있도록 할 수 있다. 예를 들어, 포스포로티오에이트(PS) 결합은 엑소뉴클레이스 분해를 저해할 수 있는 gRNA의 5'- 또는 "-말단에 마지막 3개 내지 5개의 뉴클레오타이드 사이에 도입될 수 있다. 일부 실시형태에서, 포스포로티오에이트 결합은 엔도뉴클레이스에 의한 공격을 감소시키기 위해 전체 gRNA에 걸쳐 첨가될 수 있다.

[0964] 일부 실시형태에서, 가이드 RNA는 스플라이싱 부위(즉, 스플라이스 수용체(SA) 또는 스플라이스 공여체(SD)를 파괴하도록 설계된다. 일부 실시형태에서, 가이드 RNA는 염기 편집이 조기 종결 코돈을 생성하도록 설계된다. **표 4A** 내지 **표 4B**는 스플라이싱 부위를 파괴하거나 또는 조기 종결 코돈을 생성하도록 설계된 gRNA 표적 서열의 비-포괄적인 목록을 제공한다.

표 4

[표 4A]

gRNA: 스플라이스 부위 및 STOP 코돈

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| VISTA | 엑손 1 SD (pos6) | CCTTACCTAGGGACGCAGCC | CCUUACCUAGGGACGCAGCC |
| | 엑손 1 STOP (pos7) | GGATCCCCAGCGCCAGCTGC | GGAUCCCCAGCGCCAGCUGC |
| | 엑손 1 STOP (pos5) | AGCGCCAGCTGCCGGCCTCC | AGCGCCAGCUGCCGGCCUCC |
| | 엑손 1 STOP (pos4) | GCGCCAGCTGCCGGCCTCCA | GCGCCAGCUGCCGGCCUCCA |
| | 엑손 2 STOP (pos8) | CCTGGCTCAGCGCCACGGGC | CCUGGCUCAGCGCCACGGGC |
| | 엑손 2 STOP (pos5) | GCTGCAGGTGCAGACAGGTG | GCUGCAGGUGCAGACAGGUG |
| | 엑손 2 STOP (pos7) | GCGGTACCACGTCTTGTAGA | GCGGUACCACGUCUUGUAGA |
| | 엑손 3 SA (pos4) | TGCCTGTGGGAACAAACAGA | UGCCUGUGGGAACAAACAGA |
| | 엑손 3 SD (pos5) | CTTACTTTCACTATCCTGGG | CUUACUUUCACUAUCCUGGG |
| | 엑손 3 SD (pos8) | TCCCTTACTTTCACTATCCT | UCCCUUACUUUCACUAUCCU |
| | 엑손 3 STOP (pos5) | CTCCCAGGATAGTGAAAGTA | CUCCCAGGAUAGUGAAAGUA |
| | 엑손 4 SA (pos7) | TGATGTCTGAAAGGGCAGAG | UGAUGUCUGAAAGGGCAGAG |
| | 엑손 5 STOP (pos5) | TGCCCAGGAGCTGGTGCGGA | UGCCCAGGAGCUGGUGCGGA |
| | 엑손 6 SA (pos4) | TTGCTGCCACAGAACCAGAA | UUGCUGCCACAGAACCAGAA |
| | 엑손 6 STOP (pos4) | ATTCAAGGGATTGAAAACCC | AUUCAAGGGAUUGAAAACCC |
| | 엑손 6 STOP (pos8) | ACCTGCCCAGGGGATACCCG | ACCUGCCCAGGGGAUACCCG |
| | 엑손 6 STOP (pos7) | CAGCGGCAGCCTTCTGAGTC | CAGCGGCAGCCUUCUGAGUC |

[0965]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| TRAC | 엑손 1 STOP 1 (pos5) | GCTACAAACAAGCTCATCTT | GCUACAAACAAGCUCAUCUU |
| | 엑손 1 STOP 2 (pos6) | CCAGCCAAGTACGTAAGTAG | CCAGCCAAGUACGUAAGUAG |
| | 엑손 2 SA (pos9) | CTGGATATCTGTGGGACAAG | CUGGAUAUCUGUGGGACAAG |
| | 엑손 2 SD | CTTACCTGGGCTGGGGAAGA | CUUACCUGGGCUGGGGAAGA |
| | 엑손 4 SA | TTCGTATCTGTAAAACCAAG | UUCGUAUCUGUAAAACCAAG |
| | 엑손 4 STOP | TTTCAAAACCTGTCAGTGAT | UUUCAAAACCUGUCAGUGAU |
| | 엑손 4 STOP | TTCAAAACCTGTCAGTGATT | UUCAAAACCUGUCAGUGAUU |
| Tim-3 | 엑손 2 SA (pos6) | GGACCCTGCATAGAGAGAGA | GGACCCUGCAUAGAGAGAGA |
| | 엑손 2 STOP (pos5) | TGCCCCAGCAGACGGGCACG | UGCCCCAGCAGACGGGCACG |
| | 엑손 3 SD (pos5) | GTTACCTGGGCCATGTCCCC | GUUACCUGGGCCAUGUCCCC |
| | 엑손 4 SD (pos5) | CTTACTGTTAGATTTATATC | CUUACUGUUAGAUUUAUAUC |
| | 엑손 4 SD (pos4) | TTACTGTTAGATTTATATCA | UUACUGUUAGAUUUAUAUCA |
| | 엑손 5 SA (pos5) | TTTGCTATGGAAACACAAAC | UUUGCUAUGGAAACACAAAC |
| | 엑손 5 STOP (pos8) | TCCATAGCAAATATCCACAT | UCCAUAGCAAAUAUCCACAU |
| | 엑손 7 STOP (pos5) | GCAGCAACCCTCACAACCTT | GCAGCAACCCUCACAACCUU |
| | 엑손 7 STOP (pos 4) | CAGCAACCCTCACAACCTTT | CAGCAACCCUCACAACCUUU |
| TIGIT | 엑손 1 STOP (pos4) | AGGCAGGCTCCCCTCGCCTC | AGGCAGGCUCCCCUCGCCUC |
| | 엑손 2 STOP (5&8) | GGAGCAGCAGGACCAGCTTC | GGAGCAGCAGGACCAGCUUC |
| | 엑손 2 SD (pos9) | CAGGAATACCTGAGCTTTCT | CAGGAAUACCUGAGCUUUCU |

[0966]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 3 STOP (pos7) | AGGTTCCAGATTCCATTGCT | AGGUUCCAGAUUCCAUUGCU |
| | 엑손 1 STOP | CTGGGCCCAGGGGCTGAGGC | CUGGGCCCAGGGGCUGAGGC |
| | 엑손 2 STOP | GATCGAGTGGCCCCAGGTCC | GAUCGAGUGGCCCCAGGUCC |
| TGFbRII | 엑손 1 SD (JMG79) | TCACCCGACTTCTGAACGTG | UCACCCGACUUCUGAACGUG |
| | 엑손 3 SD (JMG83) | TTACCTGCCCACTGTTAGCC | UUACCUGCCCACUGUUAGCC |
| | 엑손 2 STOP (JMG80) | GAAGCCACAGGAAGTCTGTG | GAAGCCACAGGAAGUCUGUG |
| | 엑손 3 STOP (JMG81) | ACTCCAGTTCCTGACGGCTG | ACUCCAGUUCCUGACGGCUG |
| | 엑손 3 STOP (JMG82) | ACCTACAGGAGTACCTGACG | ACCUACAGGAGUACCUGACG |
| | 엑손 4 STOP (JMG84) | TTCCCAGAGCACCAGAGCCA | UUCCCAGAGCACCAGAGCCA |
| | 엑손 1 STOP (JMG85) | ACGTTCAGAAGTCGGGTGAG | ACGUUCAGAAGUCGGGUGAG |
| | 엑손 3 STOP (pos8) | TTCAGAGCAGTTTGAGACAG | UUCAGAGCAGUUUGAGACAG |
| RFXANK | 엑손 2 SA (JMG8) | CCTGCTGGGAAACAGACAAC | CCUGCUGGGAAACAGACAAC |
| | 엑손 2 SD (JMG9) | CACTCACAGTCTAGGGTGGC | CACUCACAGUCUAGGGUGGC |
| | 엑손 2 STOP (pos8) | CAACCGGCAGCGAGGGAACG | CAACCGGCAGCGAGGGAACG |
| | 엑손 3 SA (pos7) | ACAGGGCTGGGGCAGGACAG | ACAGGGCUGGGGCAGGACAG |
| | 엑손 3 STOP (pos8) | CATCCACCAGCTCGCAGCAC | CAUCCACCAGCUCGCAGCAC |
| | 엑손 3 STOP (pos7) | ATCCACCAGCTCGCAGCACA | AUCCACCAGCUCGCAGCACA |
| | 엑손 3 STOP (pos6) | TCCACCAGCTCGCAGCACAG | UCCACCAGCUCGCAGCACAG |
| | 엑손 3 STOP (pos5) | CCACCAGCTCGCAGCACAGG | CCACCAGCUCGCAGCACAGG |

[0967]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 4 SA (JMG10) | TGTCACCTGGCAGGAGGAGG | UGUCACCUGGCAGGAGGAGG |
| | 엑손 4 SA (pos6) | GTCACCTGGCAGGAGGAGGC | GUCACCUGGCAGGAGGAGGC |
| | 엑손 5 SA (pos7) | GGCACCCTGCAGGGAGAAGA | GGCACCCUGCAGGGAGAAGA |
| | 엑손 5 SA (JMG11) | GCACCCTGCAGGGAGAAGAA | GCACCCUGCAGGGAGAAGAA |
| | 엑손 6 SA (pos4) | ATTCTGTCGTGGGTAGGGGC | AUUCUGUCGUGGGUAGGGGC |
| | 엑손 6 SA (JMG12) | CTCCATTCTGTCGTGGGTAG | CUCCAUUCUGUCGUGGGUAG |
| | 엑손 7 SA (pos8) | CCTCGGGCTGCAAAGGAGAG | CCUCGGGCUGCAAAGGAGAG |
| | 엑손 7 SA (pos5) | CGGGCTGCAAAGGAGAGGGG | CGGGCUGCAAAGGAGAGGGG |
| | 엑손 7 SD (pos6) | GCTGACCTTTCCGGTATCCC | GCUGACCUUUCCGGUAUCCC |
| | 엑손 7 SD (pos5) | CTGACCTTTCCGGTATCCCA | CUGACCUUUCCGGUAUCCCA |
| | 엑손 8 SA (pos8) | TGTTGCACTGAGATGGGGCA | UGUUGCACUGAGAUGGGGCA |
| | 엑손 8 SA (pos9) | CTGTTGCACTGAGATGGGGC | CUGUUGCACUGAGAUGGGGC |
| PVRIG (CD112R) | 엑손 1 STOP (pos7) | GCCCTGCAGCCCCCAGAACC | GCCCUGCAGCCCCCAGAACC |
| | 엑손 1 SD (pos5) | CTCACCCGCAGTGACACACA | CUCACCCGCAGUGACACACA |
| | 엑손 1 STOP (pos8) | GCAGCACCCAGGGCAGGACC | GCAGCACCCAGGGCAGGACC |
| | 엑손 1 STOP (pos7) | CAGCACCCAGGGCAGGACCA | CAGCACCCAGGGCAGGACCA |
| | 엑손 2 SA (pos5) | GTCCCTGTGGAACAGCAGCA | GUCCCUGUGGAACAGCAGCA |
| | 엑손 2 STOP (pos8) | GTGGGTTCAAGTTCGGATGG | GUGGGUUCAAGUUCGGAUGG |

[0968]

[0969]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 2 SD (pos 7) | GCCCCACCTGGGTCTGAGCT | GCCCCACCUGGGUCUGAGCU |
| | 엑손 2 SD (pos8) | GGCCCCACCTGGGTCTGAGC | GGCCCCACCUGGGUCUGAGC |
| | 엑손 2 SD (pos4) | CCACCTGGGTCTGAGCTGGG | CCACCUGGGUCUGAGCUGGG |
| | 엑손 2 STOP (pos8) | AGGCCTCCCAGGAGCCCTCA | AGGCCUCCCAGGAGCCCUCA |
| | 엑손 2 STOP (pos4) | CTCCCAGGAGCCCTCAGGGA | CUCCCAGGAGCCCUCAGGGA |
| | 엑손 2 STOP (pos4) | CCCCCAGCTCACAGTCACCA | CCCCCAGCUCACAGUCACCA |
| | 엑손 3 SD (pos8) | GGTCTCACCGGTGCTTATGT | GGUCUCACCGGUGCUUAUGU |
| | 엑손 3 STOP (pos9) | TGCTGCGCCGACATAAGCAC | UGCUGCGCCGACAUAAGCAC |
| | 엑손 4 SA (pos8) | GGCAGGGCTGGGAGAGAGCA | GGCAGGGCUGGGAGAGAGCA |
| | 엑손 4 STOP (pos9) | CGAGAGCACGAGCATGGGTG | CGAGAGCACGAGCAUGGGUG |
| | 엑손 4 STOP (pos6) | GAGCACGAGCATGGGTGAGG | GAGCACGAGCAUGGGUGAGG |
| | 엑손 4 STOP (pos5) | AGCACGAGCATGGGTGAGGA | AGCACGAGCAUGGGUGAGGA |
| | 엑손 4 STOP (pos4) | GCACGAGCATGGGTGAGGAG | GCACGAGCAUGGGUGAGGAG |
| | 엑손 4 SD (pos5) | CTCACCCATGCTCGTGCTCT | CUCACCCAUGCUCGUGCUCU |
| | 엑손 5 SA (pos6) | GGTGCCTGCGCGGGGGAAGG | GGUGCCUGCGCGGGGGAAGG |
| | 엑손 5 SA (pos5) | GTGCCTGCGCGGGGGAAGGA | GUGCCUGCGCGGGGGAAGGA |
| | 엑손 5 SA (pos9) | CTTGGTGCCTGCGCGGGGGA | CUUGGUGCCUGCGCGGGGGA |
| | 엑손 5 STOP (pos6) | GGCCCCAGGGCCCTGCCGCC | GGCCCCAGGGCCCUGCCGCC |
| | 엑손 5 STOP (pos9) | TCTACGCTCAGGCAGGGGAG | UCUACGCUCAGGCAGGGGAG |

[0970]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 5 STOP (pos4) | CCACCAGGACGGCCCCCCAT | CCACCAGGACGGCCCCCCAU |
| | 엑손 5 STOP (pos5) | AGGCCCAGGCGGCAGGGCCC | AGGCCCAGGCGGCAGGGCCC |
| | 엑손 5 STOP (pos4) | GGCCCAGGCGGCAGGGCCCT | GGCCCAGGCGGCAGGGCCCU |
| PDCD1 | 엑손 1 STOP 2 (pos9) | ACGACTGGCCAGGGCGCCTG | ACGACUGGCCAGGGCGCCUG |
| | 엑손 1 STOP 4 (pos7) | CACCGCCCAGACGACTGGCC | CACCGCCCAGACGACUGGCC |
| | 엑손 1 STOP (pos4) | CTACAACTGGGCTGGCGGCC | CUACAACUGGGCUGGCGGCC |
| | 엑손 1 SD | CACCTACCTAAGAACCATCC | CACCUACCUAAGAACCAUCC |
| | 엑손 2 SA | GGAGTCTGAGAGATGGAGAG | GGAGUCUGAGAGAUGGAGAG |
| | 엑손 2 STOP 1 (pos8) | CAGCAACCAGACGGACAAGC | CAGCAACCAGACGGACAAGC |
| | 엑손 2 STOP 2 (pos9) | GTGTCACACAACTGCCCAAC | GUGUCACACAACUGCCCAAC |
| | 엑손 3 STOP 1 (pos8) | AGCCGGCCAGTTCCAAACCC | AGCCGGCCAGUUCCAAACCC |
| | 엑손 3 STOP (pos7) | CAGTTCCAAACCCTGGTGGT | CAGUUCCAAACCCUGGUGGU |
| | 엑손 3 STOP 2 (pos5) | CGGCCAGTTCCAAACCCTGG | CGGCCAGUUCCAAACCCUGG |
| | 엑손 3 STOP (pos5) | GGACCCAGACTAGCAGCACC | GGACCCAGACUAGCAGCACC |
| | 엑손 3 SD | GACGTTACCTCGTGCGGCCC | GACGUUACCUCGUGCGGCCC |
| | 엑손 4 SA | TCCCTGCAGAGAAACACACT | UCCCUGCAGAGAAACACACU |
| | 엑손 4 SD | GAGACTCACCAGGGGCTGGC | GAGACUCACCAGGGGCUGGC |
| | 엑손 5 SA | CCTCCTTCTTTGAGGAGAAA | CCUCCUUCUUUGAGGAGAAA |
| | 엑손 2 STOP (pos 7) | GGGGTTCCAGGGCCTGTCTG | GGGGUUCCAGGGCCUGUCUG |

[0971]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 3 SA | TTCTCTCTGGAAGGGCACAA | UUCUCUCUGGAAGGGCACAA |
| | 엑손 5 STOP 1 (pos 8) | CCAGTGGCGAGAGAAGACCC | CCAGUGGCGAGAGAAGACCC |
| | 엑손 5 STOP 2 (pos 5) | TGCCCAGCCACTGAGGCCTG | UGCCCAGCCACUGAGGCCUG |
| | 엑손 1 STOP 1 (pos8) | CGACTGGCCAGGGCGCCTGT | CGACUGGCCAGGGCGCCUGU |
| | 엑손 1 STOP 3 (pos6) | ACCGCCCAGACGACTGGCCA | ACCGCCCAGACGACUGGCCA |
| Lag3 | 엑손 1 STOP (pos8) | GTTTCTGCAGCCGCTTTGGG | GUUUCUGCAGCCGCUUUGGG |
| | 엑손 1 SD (pos4) | TTACCTGGAGCCACCCAAAG | UUACCUGGAGCCACCCAAAG |
| | 엑손 2 SA (pos4) | TCACTAGGTGAGCAAAAGAG | UCACUAGGUGAGCAAAAGAG |
| | 엑손 2 STOP (pos8) | GCCTCTCCAGCCAGGGGCTG | GCCUCUCCAGCCAGGGGCUG |
| | 엑손 2 STOP (pos 6) | CTTGGCAGCATCAGCCAGAC | CUUGGCAGCAUCAGCCAGAC |
| | 엑손 3 SA (pos4) | CCACTGGGCGGGAAAGAGAA | CCACUGGGCGGGAAAGAGAA |
| | 엑손 3 SD (pos6) | ACATACTCGAGGCCTGGCCC | ACAUACUCGAGGCCUGGCCC |
| | 엑손 3 STOP (pos5) | CCTGCAGCCCCGCGTCCAGC | CCUGCAGCCCCGCGUCCAGC |
| | 엑손 3 STOP (pos7) | CGCGTCCAGCTGGATGAGCG | CGCGUCCAGCUGGAUGAGCG |
| | 엑손 3 STOP (pos6) | TGGGCCAGGCCTCGAGTATG | UGGGCCAGGCCUCGAGUAUG |
| | 엑손 4 SD (pos4) | GGGAGTTACCCAGAACAGTG | GGGAGUUACCCAGAACAGUG |
| | 엑손 4 STOP (pos8) | CCTGCCCCAAGTCAGCCCCA | CCUGCCCCAAGUCAGCCCCA |
| | 엑손 4 STOP (pos9) | GCCAGGGCCGAGTCCCTGTC | GCCAGGGCCGAGUCCCUGUC |
| | 엑손 4 STOP (pos8) | CCAGGGCCGAGTCCCTGTCC | CCAGGGCCGAGUCCCUGUCC |

[0972]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 4 STOP (pos4) | GCCCCAGGGCCCAGAGTCCA | GCCCCAGGGCCCAGAGUCCA |
| | 엑손 5 STOP (pos9) | ATGTGAGCCAGGCCCAGGCT | AUGUGAGCCAGGCCCAGGCU |
| | 엑손 5 STOP (pos 8) | GAGGAGTCCACTTGGCAGTG | GAGGAGUCCACUUGGCAGUG |
| | 엑손 6 SA (pos7) | GAGTCACTGAAAAGAGTAGA | GAGUCACUGAAAAGAGUAGA |
| | 엑손 6 STOP (pos6) | CTGGACAAGAACGCTTTGTG | CUGGACAAGAACGCUUUGUG |
| | 엑손 6 STOP (pos7) | CCATCCCAGAGGAGTTTCTC | CCAUCCCAGAGGAGUUUCUC |
| | 엑손 6 STOP (pos4) | TGGCAATGCCAGCTGTACCA | UGGCAAUGCCAGCUGUACCA |
| | 엑손 6 STOP (pos4) | TACCAGGGGGAGAGGCTTCT | UACCAGGGGGAGAGGCUUCU |
| | 엑손 6 STOP (pos8) | GGCATTGCCAAGGCTGGGAA | GGCAUUGCCAAGGCUGGGAA |
| | 엑손 7 SA (pos6) | GGCACCTATGGAGAAAGTAC | GGCACCUAUGGAGAAAGUAC |
| | 엑손 7 STOP (pos4) | AGACAGGTGAGCCAGGGACA | AGACAGGUGAGCCAGGGACA |
| | 엑손 7 SD (pos7) | GGCTCACCTGTCTTCTCCAA | GGCUCACCUGUCUUCUCCAA |
| | 엑손 8 SA (pos8) | GTCGCCACTGTGAGAAGAGA | GUCGCCACUGUGAGAAGAGA |
| | 엑손 8 STOP (pos8) | GCAGGCTCAGAGCAAGATAG | GCAGGCUCAGAGCAAGAUAG |
| | 엑손 8 STOP (pos8) | GCTGGAGCAAGAACCGGAGC | GCUGGAGCAAGAACCGGAGC |
| CTLA-4 | 엑손 1 SD (pos 6) | ACTCACCTTTGCAGAAGACA | ACUCACCUUUGCAGAAGACA |
| | 엑손 1 SD | CACTCACCTTTGCAGAAGAC | CACUCACCUUUGCAGAAGAC |
| | 엑손 1 STOP (pos5) | AGGGCCAGGTCCTGGTAGCC | AGGGCCAGGUCCUGGUAGCC |
| | 엑손 2 STOP | GGCCCAGCCTGCTGTGGTAC | GGCCCAGCCUGCUGUGGUAC |

[0973]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 2 STOP (pos 8) | GCTTCGGCAGGCTGACAGCC | GCUUCGGCAGGCUGACAGCC |
| | 엑손 2 STOP | TATCCAAGGACTGAGGGCCA | UAUCCAAGGACUGAGGGCCA |
| | 엑손 2 STOP | GGAACCCAGATTTATGTAAT | GGAACCCAGAUUUAUGUAAU |
| | 엑손 2 SD | GCTCACCAATTACATAAATC | GCUCACCAAUUACAUAAAUC |
| | 엑손 2 SD | CTCACCAATTACATAAATCT | CUCACCAAUUACAUAAAUCU |
| | 엑손 1 STOP | CTCAGCTGAACCTGGCTACC | CUCAGCUGAACCUGGCUACC |
| Chi3l1 | 엑손 1 STOP (pos8) | GGCGTCTCAAACAGGTATCT | GGCGUCUCAAACAGGUAUCU |
| | 엑손 1 SA (pos7) | CAAAGCCTGAAGAGAAATCC | CAAAGCCUGAAGAGAAAUCC |
| | 엑손 3 SA (pos6) | AGAGCCTGAAGGAGAAGTCT | AGAGCCUGAAGGAGAAGUCU |
| | 엑손 3 STOP (pos4) | TCCCAGTACCGGGAAGGCGA | UCCCAGUACCGGGAAGGCGA |
| | 엑손 4 SA (pos6) | GGTTCCTGTGGAGCACAGGG | GGUUCCUGUGGAGCACAGGG |
| | 엑손 4 SA (pos9) | TGGGGTTCCTGTGGAGCACA | UGGGGUUCCUGUGGAGCACA |
| | 엑손 6 SA (pos8) | TCATTTCCTAGATGGGAGAC | UCAUUUCCUAGAUGGGAGAC |
| | 엑손 6 SA (pos4) | TTCCTAGATGGGAGACAGGC | UUCCUAGAUGGGAGACAGGC |
| | 엑손 8 SA (pos9) | CCAGGTGTCTGAGGAGGAAG | CCAGGUGUCUGAGGAGGAAG |
| | 엑손 8 SA (pos5) | GTGTCTGAGGAGGAAGGGGA | GUGUCUGAGGAGGAAGGGGA |
| | 엑손 9 SA (pos6) | TAGTCCTGGGTGGGGTAGGG | UAGUCCUGGGUGGGGUAGGG |
| | 엑손 9 SA (pos5) | AGTCCTGGGTGGGGTAGGGT | AGUCCUGGGUGGGGUAGGGU |
| | 엑손 9 SD (pos6) | CATTACCTCATAGTAGGCAA | CAUUACCUCAUAGUAGGCAA |

[0974]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 9 SD (pos7) | CCATTACCTCATAGTAGGCA | CCAUUACCUCAUAGUAGGCA |
| | 엑손 10 SA (pos7) | ACAGATCTGAGCAGATAACA | ACAGAUCUGAGCAGAUAACA |
| | 엑손 10 STOP (pos 7) | TCCTACCCACTGGTTGCCCT | UCCUACCCACUGGUUGCCCU |
| | 엑손 11 STOP (pos7) | AGGTGCAGTACCTGAAGGAC | AGGUGCAGUACCUGAAGGAC |
| | 엑손 11 STOP (pos5) | CAGGCAGCTGGCGGGCGCCA | CAGGCAGCUGGCGGGCGCCA |
| | 엑손 11 STOP (pos7) | GACTTCCAGGGCTCCTTCTG | GACUUCCAGGGCUCCUUCUG |
| CD96 | 엑손 1 STOP (pos5) | CATCCAGATACATTTTGTCA | CAUCCAGAUACAUUUUGUCA |
| | 엑손 2 STOP (pos5) | ACCTGCCAAACACAGACAGT | ACCUGCCAAACACAGACAGU |
| | 엑손 2 STOP (pos7) | CGTGCAGATGCAATGGTCCA | CGUGCAGAUGCAAUGGUCCA |
| | 엑손 3 SA (pos6) | TGTAACTGTAACAAAACATA | UGUAACUGUAACAAAACAUA |
| | 엑손 3 SD (pos6) | ACTTACCACCGACCATGCAT | ACUUACCACCGACCAUGCAU |
| | 엑손 5 SD (pos5) | CTTACCAAAAACCTTGACTG | CUUACCAAAAACCUUGACUG |
| | 엑손 5 STOP (pos6) | CCAGTCCAAATCTTCGATGA | CCAGUCCAAAUCUUCGAUGA |
| | 엑손 5 STOP (pos7) | CAGTCCAAATCTTCGATGAT | CAGUCCAAAUCUUCGAUGAU |
| | 엑손 7 STOP (pos4) | AAACCATGTGATATTTGCTT | AAACCAUGUGAUAUUUGCUU |
| | 엑손 8 STOP (pos6) | ATGTTCCACACTTTATTTCC | AUGUUCCACACUUUAUUUCC |
| | 엑손 10 SD (pos4) | TCACGTTGAGGAGTGGTGTT | UCACGUUGAGGAGUGGUGUU |
| | 엑손 13 SA (pos7) | CATTGTCTAGGGATATAAAG | CAUUGUCUAGGGAUAUAAAG |
| | 엑손 13 SA (pos8) | ACATTGTCTAGGGATATAAA | ACAUUGUCUAGGGAUAUAAA |

[0975]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 13 SA (pos9) | GACATTGTCTAGGGATATAA | GACAUUGUCUAGGGAUAUAA |
| | 엑손 14 STOP (pos4) | TGGCCAGGACATTCCATCTT | UGGCCAGGACAUUCCAUCUU |
| | 엑손 15 SA (pos6) | CCATTCTAGGAACAAAATAT | CCAUUCUAGGAACAAAAUAU |
| Cblb | 엑손 1 STOP | GAGCTTCCAAGTCTTCTCCA | GAGCUUCCAAGUCUUCUCCA |
| | 엑손 1 STOP (JMG44) | TCCCCGAAAAGGTCGAATTT | UCCCCGAAAAGGUCGAAUUU |
| | 엑손 2 STOP | ATGAAGAACAGTCACAGGAC | AUGAAGAACAGUCACAGGAC |
| | 엑손 3 SA | GATTTCGTCTGTAGGCACAA | GAUUUCGUCUGUAGGCACAA |
| | 엑손 4 SD | TAAACTTACCTGAAACAGCC | UAAACUUACCUGAAACAGCC |
| | 엑손 4 STOP | ATTCAGACAGTGCCTTCATG | AUUCAGACAGUGCCUUCAUG |
| | 엑손 6 STOP | GTTGCACTCGATTGGGACAG | GUUGCACUCGAUUGGGACAG |
| | 엑손 6 STOP | TTATTTCAAGCCCTGATTGA | UUAUUUCAAGCCCUGAUUGA |
| | 엑손 7 SD | TTACCTGTGTAACTTTTATA | UUACCUGUGUAACUUUUAUA |
| | 엑손 8 SA (pos8) | ATTGTTCCTGGAATTTGGGG | AUUGUUCCUGGAAUUUGGGG |
| | 엑손 8 SD (JMG48) | ATTATACCTGCCATGCCGTA | AUUAUACCUGCCAUGCCGUA |
| | 엑손 8 SA (pos 5) (JMG46) | GTTCCTGGAATTTGGGGAGG | GUUCCUGGAAUUUGGGGAGG |
| | 엑손 8 STOP (JMG47) | CTGCCATGCCGTAAGGCAAG | CUGCCAUGCCGUAAGGCAAG |
| | 엑손 10 SD (JMG49) | TCTACCTTTGGTGAACCCGT | UCUACCUUUGGUGAACCCGU |
| | 엑손 11 SD (JMG50) | CTTACCTTAGCTCCTTCTAA | CUUACCUUAGCUCCUUCUAA |
| | 엑손 11 STOP | GGGATGTCGACTCCTAGGGG | GGGAUGUCGACUCCUAGGGG |

[0976]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 11 STOP | CGAGGGCACCATGCTTCAAG | CGAGGGCACCAUGCUUCAAG |
| | 엑손 12 SD | AAACTCACTTTATGCTAGGG | AAACUCACUUUAUGCUAGGG |
| | 엑손 12 SD (JMG51) | CTCACTTTATGCTAGGGAGG | CUCACUUUAUGCUAGGGAGG |
| | 엑손 16 SA (JMG52) | CTTCACCTGCATTTAAAGAA | CUUCACCUGCAUUUAAAGAA |
| | 엑손 4 STOP (JMG45) | CCACCAGATTAGCTCTGGCC | CCACCAGAUUAGCUCUGGCC |
| | 엑손 10 SD (pos4) | CTACCTTTGGTGAACCCGTT | CUACCUUUGGUGAACCCGUU |
| BTLA | 엑손 1 STOP (pos6) | ATGTTCCAGATGTCCAGATA | AUGUUCCAGAUGUCCAGAUA |
| | 엑손 1 STOP (pos5) | TGTTCCAGATGTCCAGATAT | UGUUCCAGAUGUCCAGAUAU |
| | 엑손 2 STOP (pos8) | AGATAGACAAACAAGTTGGA | AGAUAGACAAACAAGUUGGA |
| | 엑손 2 STOP (pos9) | AGCTTGCACCAAGTCACATG | AGCUUGCACCAAGUCACAUG |
| | 엑손 3 SD (pos6) | ACCCACCTTGGTGCCTTCTC | ACCCACCUUGGUGCCUUCUC |
| B2M (BE) | 엑손 1 SD | ACTCACGCTGGATAGCCTCC | ACUCACGCUGGAUAGCCUCC |
| | 엑손 2 SA (pos9) | TGGAGTACCTGAGGAATATC | UGGAGUACCUGAGGAAUAUC |
| | 엑손 2 STOP (pos6) | TTACCCCACTTAACTATCTT | UUACCCCACUUAACUAUCUU |
| | 엑손 3 SA | TCGATCTATGAAAAAGACAG | UCGAUCUAUGAAAAAGACAG |
| | 엑손 2 STOP | TACCCCACTTAACTATCT | UACCCCACUUAACUAUCU |
| B2M (ABE) | 엑손 1 SD 1 (pos 5) | ACTCACGCTGGATAGCCTCC | ACUCACGCUGGAUAGCCUCC |
| | 엑손 2 SA (pos 4) | CTCAGGTACTCCAAAGATTC | CUCAGGUACUCCAAAGAUUC |
| | 엑손 2 SD (pos 4) | CTTACCCCACTTAACTATCT | CUUACCCCACUUAACUAUCU |

[0977]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| TET2 | 엑손 1 STOP 1 (pos 8) | CATTTGCCAGACAGAACCTC | CAUUUGCCAGACAGAACCUC |
| | 엑손 1 STOP 2 (pos 4) | AAACAAGACCAAAAGGCTAA | AAACAAGACCAAAAGGCUAA |
| | 엑손 1 STOP 3 (pos 7) | GTAAGCCAAGAAAGAAATCC | GUAAGCCAAGAAAGAAAUCC |
| | 엑손 1 STOP 4 (pos 5) | GCTTCAGATTCTGAATGAGC | GCUUCAGAUUCUGAAUGAGC |
| | 엑손 1 STOP 5 (pos 7) | TTAAAACAAAATGAAATGAA | UUAAAACAAAAUGAAAUGAA |
| | 엑손 1 STOP 6 (pos 7) | GTTCCTCAGCTTCCTTCAGA | GUUCCUCAGCUUCCUUCAGA |
| | 엑손 1 STOP 7 (pos 8) | CAAAGAGCAAGAGATTCTGA | CAAAGAGCAAGAGAUUCUGA |
| | 엑손 1 STOP 8 (pos 7) | AAAGAGCAAGAGATTCTGAA | AAAGAGCAAGAGAUUCUGAA |
| | 엑손 1 STOP 9 (pos 4) | ACACAGCACTATCTGAAACC | ACACAGCACUAUCUGAAACC |
| | 엑손 1 STOP 10 (pos 5) | CACCCAGAAAACAACACAGC | CACCCAGAAAACAACACAGC |
| | 엑손 1 STOP 11 (pos 4) | TACCAAGTTGAAATGAATCA | UACCAAGUUGAAAUGAAUCA |
| | 엑손 1 STOP 12 (pos 7) | ATGAATCAAGGGCAGTCCCA | AUGAAUCAAGGGCAGUCCCA |
| | 엑손 1 STOP 13 (pos 5) | AGGGCAGTCCCAAGGTACAG | AGGGCAGUCCCAAGGUACAG |
| | 엑손 1 STOP 14 (pos 5) | GTTCCAAAAACCCTCACACC | GUUCCAAAAACCCUCACACC |
| | 엑손 1 STOP 15 (pos 5) | GAAACAGCACTTGAATCAAC | GAAACAGCACUUGAAUCAAC |
| | 엑손 1 STOP 16 (pos 5) | ATTACAAATAAAGAATAAAG | AUUACAAAUAAAGAAUAAAG |
| | 엑손 1 STOP 17 (pos 8) | TAATGTCCAAATGGGACTGG | UAAUGUCCAAAUGGGACUGG |
| | 엑손 1 STOP 18 (pos 6) | CAAAGCAAGATCTTCTTCAC | CAAAGCAAGAUCUUCUUCAC |
| | 엑손 1 STOP 19 (pos 5) | ACAACAAGCTTCAGTTCTAC | ACAACAAGCUUCAGUUCUAC |

[0978]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 1 STOP 20 (pos 6) | CTGCGCAACTTGCTCAGCAA | CUGCGCAACUUGCUCAGCAA |
| | 엑손 1 STOP 21 (pos 5) | CACTCAGACCCCTCCCCAGA | CACUCAGACCCCUCCCCAGA |
| | 엑손 1 STOP 22 (pos 6) | TTTTTCCATGTTTTGTTTTC | UUUUUCCAUGUUUUGUUUUC |
| | 엑손 1 SD (pos 4) | TTACCTACACATCTGCAAGA | UUACCUACACAUCUGCAAGA |
| | 엑손 3 SD (pos 8) | ACACTTACCCACTTAGCAAT | ACACUUACCCACUUAGCAAU |
| | 엑손 7 STOP (pos 5) | CATGCAGAATGGCAGCACAT | CAUGCAGAAUGGCAGCACAU |
| | 엑손 8 STOP 1 (pos 6) | AAGCTCAGGAGGAGAAAAAA | AAGCUCAGGAGGAGAAAAAA |
| | 엑손 8 STOP 2 (pos 8) | CGCAAGCCAGGCTAAACAGT | CGCAAGCCAGGCUAAACAGU |
| | 엑손 9 STOP 1 (pos 8) | TTCTCCCCAGTCTCAGCCGA | UUCUCCCCAGUCUCAGCCGA |
| | 엑손 9 STOP 2 (pos 5) | TGGTCAGGAAAAGCAGCCAT | UGGUCAGGAAAAGCAGCCAU |
| | 엑손 9 STOP 3 (pos 7) | CTAGTCCAGGGTGTGGCTTC | CUAGUCCAGGGUGUGGCUUC |
| Spry1 | 엑손 1 STOP 1 | CCCCAAAATCAACATGGCAG | CCCCAAAAUCAACAUGGCAG |
| | 엑손 1 STOP 2 | TGTGATCCAGCAGCCTTCTT | UGUGAUCCAGCAGCCUUCUU |
| | 엑손 1 STOP 3 | GACCAGATCAAGGCCATAAG | GACCAGAUCAAGGCCAUAAG |
| | 엑손 1 STOP 4 | CAAGACAAGAAAAGCATGAA | CAAGACAAGAAAAGCAUGAA |
| | 엑손 1 STOP 5 | CTGAACAGGGACTGTTAGGA | CUGAACAGGGACUGUUAGGA |
| Spry2 | 엑손 1 STOP 1 | CCAGAGCTCAGAGTGGCAAC | CCAGAGCUCAGAGUGGCAAC |
| | 엑손 1 STOP 2 | TTGCTGCAGACGCCCCGTGA | UUGCUGCAGACGCCCCGUGA |
| | 엑손 1 STOP 3 | CTGCAGACGCCCCGTGACGG | CUGCAGACGCCCCGUGACGG |

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 1 STOP 4 | CGACAAGCAGTGCCTTTGCT | CGACAAGCAGUGCCUUUGCU |
| | 엑손 1 STOP 5 | GCCCAGAACGTGATTGACTA | GCCCAGAACGUGAUUGACUA |
| | 엑손 1 STOP 6 | TGTGCCAGGGGTGTTATGAC | UGUGCCAGGGGUGUUAUGAC |
| | 엑손 1 STOP 7 | CAGATCCAGTCTGATGGCAG | CAGAUCCAGUCUGAUGGCAG |
| | 엑손 1 STOP 8 | TGTACACGATGGTCAGCCAT | UGUACACGAUGGUCAGCCAU |
| CIITA | 엑손 1 SD (pos 6) | TTTTACCTTGGGGCTCTGAC | UUUUACCUUGGGGCUCUGAC |
| | 엑손 1 STOP 1 (pos 6) | AGCCCCAAGGTAAAAAGGCC | AGCCCCAAGGUAAAAAGGCC |
| | 엑손 1 STOP 2 (pos 7) | GAGCCCCAAGGTAAAAAGGC | GAGCCCCAAGGUAAAAAGGC |
| | 엑손 2 STOP 1 (pos 8) | CAGCTCACAGTGTGCCACCA | CAGCUCACAGUGUGCCACCA |
| | 엑손 2 STOP 2 (pos 7) | TATGACCAGATGGACCTGGC | UAUGACCAGAUGGACCUGGC |
| | 엑손 4 STOP 1 (pos 8) | ACTGGACCAGTATGTCTTCC | ACUGGACCAGUAUGUCUUCC |
| | 엑손 4 STOP 2 (pos 8) | TGTCTTCCAGGACTCCCAGC | UGUCUUCCAGGACUCCCAGC |
| | 엑손 7 STOP 1 (pos 7) | TTCAACCAGGAGCCAGCCTC | UUCAACCAGGAGCCAGCCUC |
| | 엑손 7 STOP 2 (pos 4) | GACCAGATTCCCAGTATGTT | GACCAGAUUCCCAGUAUGUU |
| | 엑손 7 SD (pos 8) | TAACATACTGGGAATCTGGT | UAACAUACUGGGAAUCUGGU |
| | 엑손 8 SA (pos 8) | AAAGGCACTGCAAGAGACAA | AAAGGCACUGCAAGAGACAA |
| | 엑손 8 STOP (pos 8) | CTCTGGCAAATCTCTGAGGC | CUCUGGCAAAUCUCUGAGGC |
| | 엑손 9 STOP 1 (pos 4) | AGCCAAGTACCCCCTCCCAG | AGCCAAGUACCCCCUCCCAG |
| | 엑손 9 STOP 2 (pos 7) | ACCTCCCGAGCAAACATGAC | ACCUCCCGAGCAAACAUGAC |

[0979]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 9 SD (pos 6) | CCTTACCTGTCATGTTTGCT | CCUUACCUGUCAUGUUUGCU |
| | 엑손 10 SA (pos 5) | TGCTCTGGAGATGGAGAAGC | UGCUCUGGAGAUGGAGAAGC |
| | 엑손 10 STOP 1 (pos 7) | CCCACCCAATGCCCGGCAGC | CCCACCCAAUGCCCGGCAGC |
| | 엑손 10 STOP 2 (pos 4) | AGGCCATTTTGGAAGCTTGT | AGGCCAUUUUGGAAGCUUGU |
| | 엑손 11 SA (pos 8) | ACCGGCTCTGCAAAGGCCAG | ACCGGCUCUGCAAAGGCCAG |
| | 엑손 11 STOP 1 (pos 6) | TGGTGCAGGCCAGGCTGGAG | UGGUGCAGGCCAGGCUGGAG |
| | 엑손 11 STOP 3 (pos 7) | GAACGGCAGCTGGCCCAAGG | GAACGGCAGCUGGCCCAAGG |
| | 엑손 11 STOP 4 (pos 5) | GGCCCAAGGAGGCCTGGCTG | GGCCCAAGGAGGCCUGGCUG |
| | 엑손 11 STOP 5 (pos 5) | GACACGAGTGATTGCTGTGC | GACACGAGUGAUUGCUGUGC |
| | 엑손 11 STOP 5 (pos 6) | CTGGTCAGGGCAAGAGCTAT | CUGGUCAGGGCAAGAGCUAU |
| | 엑손 11 STOP 5 (pos 8) | GGGCCCACAGCCACTCGTGG | GGGCCCACAGCCACUCGUGG |
| | 엑손 11 STOP 6 (pos 4) | TTCCAGAAGAAGCTGCTCCG | UUCCAGAAGAAGCUGCUCCG |
| | 엑손 11 STOP 7 (pos 8) | CCTGGTCCAGAGCCTGAGCA | CCUGGUCCAGAGCCUGAGCA |
| | 엑손 11 STOP 8 (pos 8) | CAGACATCAAAGTACCCTAC | CAGACAUCAAAGUACCCUAC |
| | 엑손 11 STOP 9 (pos 5) | ACATCAAAGTACCCTACAGG | ACAUCAAAGUACCCUACAGG |
| | 엑손 11 STOP 10 (pos 4) | CGCCCAGGTCCTCACGTCTG | CGCCCAGGUCCUCACGUCUG |
| | 엑손 11 STOP 11 (pos 8) | CTTAGTCCAACACCCACCGC | CUUAGUCCAACACCCACCGC |
| | 엑손 11 STOP 12 (pos 8) | CCTCCTGCAATGCTTCCTGG | CCUCCUGCAAUGCUUCCUGG |
| | 엑손 11 STOP 13 (pos 8) | GAGCCAGCCACAGGGCCCCC | GAGCCAGCCACAGGGCCCCC |

[0980]

[0981]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 11 STOP 14 (pos 6) | GGAAGCAGAAGGTGCTTGCG | GGAAGCAGAAGGUGCUUGCG |
| | 엑손 11 STOP 15 (pos 6) | GGCTGCAGCCGGGGACACTG | GGCUGCAGCCGGGGACACUG |
| | 엑손 11 STOP 16 (pos 4) | CTGCCAAATTCCAGCCTCCT | CUGCCAAAUUCCAGCCUCCU |
| | 엑손 11 STOP 17 (pos 8) | GGCGGGCCAAGACTTCTCCC | GGCGGGCCAAGACUUCUCCC |
| | 엑손 12 STOP 1 (pos 6) | AGACTCAGAGGTGAGAGGAG | AGACUCAGAGGUGAGAGGAG |
| | 엑손 14 SA (pos 4) | AGCCTAGGAGGCAAAGAGCA | AGCCUAGGAGGCAAAGAGCA |
| | 엑손 14 STOP 1 (pos 5) | CCCCCAGGCTTTCCCCAAAC | CCCCCAGGCUUUCCCCAAAC |
| | 엑손 14 SD (pos 4) | TCACTCCAGATGCTGCAGGG | UCACUCCAGAUGCUGCAGGG |
| | 엑손 15 SA (pos 4) | AGGCTGCAGGTGGAATCAGA | AGGCUGCAGGUGGAAUCAGA |
| | 엑손 15 STOP 1 (pos 8) | CTTCCCCCAGCTGAAGTCCT | CUUCCCCCAGCUGAAGUCCU |
| | 엑손 15 SD (pos 7) | CACTCACTTGAGGGTTTCCA | CACUCACUUGAGGGUUUCCA |
| | 엑손 16 SA (pos 5) | CAGACTGCGGGGACACAGTG | CAGACUGCGGGGACACAGUG |
| | 엑손 16 SD 1 (pos 8) | CCACTCACCTTAGCCTGAGC | CCACUCACCUUAGCCUGAGC |
| | 엑손 16 SD 2 (pos 7) | CACTCACCTTAGCCTGAGCA | CACUCACCUUAGCCUGAGCA |
| | 엑손 17 SA (pos 8) | GTACAAGCTGTCGGAAACAG | GUACAAGCUGUCGGAAACAG |
| | 엑손 17 SD 1 (pos 8) | ACACTCACTCCATCACCCGG | ACACUCACUCCAUCACCCGG |
| | 엑손 17 SD 2 (pos 7) | CACTCACTCCATCACCCGGA | CACUCACUCCAUCACCCGGA |
| | 엑손 18 STOP (pos 5) | CGTCCAGTACAACAAGTTCA | CGUCCAGUACAACAAGUUCA |
| | 엑손 19 SA 1 (pos 8) | CCACATCCTGCAAGGGGGGA | CCACAUCCUGCAAGGGGGGA |

[0982]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 19 SA 2 (pos 7) | CACATCCTGCAAGGGGGGAT | CACAUCCUGCAAGGGGGGAU |
| | 엑손 19 STOP 1 (pos 8) | TGGGCGTCCACATCCTGCAA | UGGGCGUCCACAUCCUGCAA |
| | 엑손 19 STOP 2 (pos 7) | GGGCGTCCACATCCTGCAAG | GGGCGUCCACAUCCUGCAAG |
| | 엑손 19 STOP 3 (pos 6) | GGCGTCCACATCCTGCAAGG | GGCGUCCACAUCCUGCAAGG |
| | 엑손 19 STOP 4 (pos 5) | GCGTCCACATCCTGCAAGGG | GCGUCCACAUCCUGCAAGGG |
| CD7 | 엑손 1 STOP (pos 4) | GCCCAAGGTAAGAGCTTCCC | GCCCAAGGUAAGAGCUUCCC |
| | 엑손 1 SD 1 (pos 8) | GCTCTTACCTTGGGCAGCCA | GCUCUUACCUUGGGCAGCCA |
| | 엑손 1 SD 2 (pos 9) | AGCTCTTACCTTGGGCAGCC | AGCUCUUACCUUGGGCAGCC |
| | 엑손 2 SA 1 (pos 8) | TGCACCTCTGGGGAGGACCT | UGCACCUCUGGGGAGGACCU |
| | 엑손 2 SA 2 (pos 9) | CTGCACCTCTGGGGAGGACC | CUGCACCUCUGGGGAGGACC |
| | 엑손 2 STOP 1 (pos 7) | CGCCTGCAGCTGTCGGACAC | CGCCUGCAGCUGUCGGACAC |
| | 엑손 2 STOP 2 (pos 8) | CACCTGCCAGGCCATCACGG | CACCUGCCAGGCCAUCACGG |
| | 엑손 2 SD 1 (pos 6) | CCCTACCTGTCACCAGGACC | CCCUACCUGUCACCAGGACC |
| | 엑손 2 SD 2 (pos 5) | CCTACCTGTCACCAGGACCA | CCUACCUGUCACCAGGACCA |
| | 엑손 3 SA (pos 4) | CCTCTGAGAAGGAAAAAAGA | CCUCUGAGAAGGAAAAAAGA |
| | 엑손 3 STOP 1 (pos9) | CAGAGGAACAGTCCCAAGGA | CAGAGGAACAGUCCCAAGGA |
| CD33 | 엑손 1 SD 1 (pos 7) | CACTCACCTGCCCACAGCAG | CACUCACCUGCCCACAGCAG |
| | 엑손 1 SD 2 (pos 8) | CCACTCACCTGCCCACAGCA | CCACUCACCUGCCCACAGCA |
| | 엑손 1 SD (pos 9) | GCCACTCACCTGCCCACAGC | GCCACUCACCUGCCCACAGC |

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 2 SA 1 (pos 8) | AGGGCCCCTGTGGGGAAACG | AGGGCCCCUGUGGGGAAACG |
| | 엑손 2 SA 2 (pos 7) | GGGCCCCTGTGGGGAAACGA | GGGCCCCUGUGGGGAAACGA |
| | 엑손 2 STOP 1 (pos 8) | GCAAGTGCAGGAGTCAGTGA | GCAAGUGCAGGAGUCAGUGA |
| | 엑손 2 STOP 2 (pos 6) | CGGAACCAGTAACCATGAAC | CGGAACCAGUAACCAUGAAC |
| | 엑손 2 STOP 3 (pos 5) | GGAACCAGTAACCATGAACT | GGAACCAGUAACCAUGAACU |
| | 엑손 2 STOP 4 (pos 4) | GAACCAGTAACCATGAACTG | GAACCAGUAACCAUGAACUG |
| | 엑손 2 STOP 5 (pos 8) | GCTAGATCAAGAAGTACAGG | GCUAGAUCAAGAAGUACAGG |
| | 엑손 2 STOP 6 (pos 8) | AGAAGTACAGGAGGAGACTC | AGAAGUACAGGAGGAGACUC |
| | 엑손 3 SA 1 (pos 6) | CAAGTCTAGTGAGGAGAAAG | CAAGUCUAGUGAGGAGAAAG |
| | 엑손 3 SA 2 (pos 5) | AAGTCTAGTGAGGAGAAAGA | AAGUCUAGUGAGGAGAAAGA |
| | 엑손 3 SA 3 (pos 4) | AGTCTAGTGAGGAGAAAGAG | AGUCUAGUGAGGAGAAAGAG |
| | 엑손 3 STOP 1 (pos 7) | ACAGGCCCAGGACACAGAGC | ACAGGCCCAGGACACAGAGC |
| | 엑손 3 STOP 2 (pos 7) | ACCTGTCAGGTGAAGTTCGC | ACCUGUCAGGUGAAGUUCGC |
| | 엑손 3 SD 1 (pos 6) | ACTTACAGGTGACGTTGAGC | ACUUACAGGUGACGUUGAGC |
| | 엑손 4 SA 1 (pos 6) | AACATCTAGGAGAGGAAGAG | AACAUCUAGGAGAGGAAGAG |
| | 엑손 4 STOP 1 (pos 7) | GTTCCACAGAACCCAACAAC | GUUCCACAGAACCCAACAAC |
| | 엑손 4 SD 1 (pos 7) | TTCCTACCTGAGCCATCTCC | UUCCUACCUGAGCCAUCUCC |
| | 엑손 5 SD (pos 8) | ATGCTCACATGAAGAAGATG | AUGCUCACAUGAAGAAGAUG |
| | 엑손 5 STOP 1 (pos 7) | GGGAAACAAGAGACCAGAGC | GGGAAACAAGAGACCAGAGC |

[0983]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 6 SA 1 (pos 6) | TCACTCTGATGGGAGACACC | UCACUCUGAUGGGAGACACC |
| | 엑손 6 SA 2 (pos 5) | CACTCTGATGGGAGACACCA | CACUCUGAUGGGAGACACCA |
| | 엑손 6 SA 1 (pos 4) | TTTCTTATGGAGAGGAAAGA | UUUCUUAUGGAGAGGAAAGA |
| CD52 | 엑손 1 STOP (pos 4) | GTACAGGTAAGAGCAACGCC | GUACAGGUAAGAGCAACGCC |
| | 엑손 1 SD (pos7) | CTCTTACCTGTACCATAACC | CUCUUACCUGUACCAUAACC |
| | 엑손 1 SD (pos 4) | TTACCTGTACCATAACCAGG | UUACCUGUACCAUAACCAGG |
| | 엑손 2 SA (pos 6) | TGTATCTGTAGGAGGAGAAG | UGUAUCUGUAGGAGGAGAAG |
| | 엑손 2 SA (pos 5) | GTATCTGTAGGAGGAGAAGT | GUAUCUGUAGGAGGAGAAGU |
| | 엑손 2 STOP (pos 7) | CAGATACAAACTGGACTCTC | CAGAUACAAACUGGACUCUC |
| CD123 | 엑손 1 SD (pos 6) | TCTTACCTTCCTTCGTTTGC | UCUUACCUUCCUUCGUUUGC |
| | 엑손 2 SA 1 (pos 8) | TTTGGATCTAAAACGGTGAC | UUUGGAUCUAAAACGGUGAC |
| | 엑손 2 SA 2 (pos 4) | GATCTAAAACGGTGACAGGT | GAUCUAAAACGGUGACAGGU |
| | 엑손 2 STOP 1 (pos 8) | AAAGGCTCAGCAGTTGACCT | AAAGGCUCAGCAGUUGACCU |
| | 엑손 2 SD (pos 6) | ATTTACCGGCATAGAATAGT | AUUUACCGGCAUAGAAUAGU |
| | 엑손 3 SA (pos 8) | TCACTGCCTAAGAGAGACAT | UCACUGCCUAAGAGAGACAU |
| | 엑손 3 STOP 1 (pos 6) | AGGATCCACGTGGAGAATGG | AGGAUCCACGUGGAGAAUGG |
| | 엑손 3 STOP 2 (pos 5) | GGATCCACGTGGAGAATGGT | GGAUCCACGUGGAGAAUGGU |
| | 엑손 3 SD (pos 6) | TCTCACTGTTCTCAGGGAAG | UCUCACUGUUCUCAGGGAAG |
| | 엑손 4 STOP 1 (pos 6) | CCTGCCCAAGGCTTCCCACC | CCUGCCCAAGGCUUCCCACC |

[0984]

[0985]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 4 STOP 2 (pos 5) | CTGCCCAAGGCTTCCCACCT | CUGCCCAAGGCUUCCCACCU |
| | 엑손 5 SA 1 (pos 6) | GCCTGCTGCGGTAAGCGGTA | GCCUGCUGCGGUAAGCGGUA |
| | 엑손 5 STOP 1 (pos 7) | GATGCTCAGGGAACACGTAT | GAUGCUCAGGGAACACGUAU |
| | 엑손 5 STOP 2 (pos 5) | TTCTCAAAGTTCCCACATCC | UUCUCAAAGUUCCCACAUCC |
| | 엑손 5 STOP 3 (pos 4) | TCACAGATTGGTGAGTAGCC | UCACAGAUUGGUGAGUAGCC |
| | 엑손 7 SD (pos 5) | CTCACCTGTTCTGTGATTAC | CUCACCUGUUCUGUGAUUAC |
| | 엑손 8 STOP 1 (pos 7) | TCCTTCCAGCTACTCAATCC | UCCUUCCAGCUACUCAAUCC |
| | 엑손 8 STOP 2 (pos 8) | CACAGTACAAATAAGAGCCC | CACAGUACAAAUAAGAGCCC |
| | 엑손 8 STOP 3 (pos 6) | CCCCCCAGCGCTTCGGTGAG | CCCCCCAGCGCUUCGGUGAG |
| | 엑손 8 STOP 4 (pos 5) | CCCCCAGCGCTTCGGTGAGT | CCCCCAGCGCUUCGGUGAGU |
| | 엑손 8 SD (pos 8) | CCACTCACCGAAGCGCTGGG | CCACUCACCGAAGCGCUGGG |
| | 엑손 10 SA (pos 4) | TACCTCGGAGGAAAGAGAAA | UACCUCGGAGGAAAGAGAAA |
| | 엑손 10 STOP (pos 8) | CAGCTTCCAAAACGACAAGC | CAGCUUCCAAAACGACAAGC |
| | 엑손 10 SD (pos 7) | AACATACCAGCTTGTCGTTT | AACAUACCAGCUUGUCGUUU |
| | 엑손 11 SA 1 (pos 8) | AGACCACCTGCAGAGACGAG | AGACCACCUGCAGAGACGAG |
| | 엑손 11 SA 2 (pos 5) | CCACCTGCAGAGACGAGAGG | CCACCUGCAGAGACGAGAGG |
| TRBC1 | 엑손 1 STOP 1 (pos 8) | CCACACCCAAAAGGCCACAC | CCACACCCAAAAGGCCACAC |
| | 엑손 1 STOP 2 (pos 5) | CCCACCAGCTCAGCTCCACG | CCCACCAGCUCAGCUCCACG |
| | 엑손 1 STOP 3 (pos 7) | CGCTGTCAAGTCCAGTTCTA | CGCUGUCAAGUCCAGUUCUA |

[0986]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 1 STOP 4 (pos 6) | GCTGTCAAGTCCAGTTCTAC | GCUGUCAAGUCCAGUUCUAC |
| | 엑손 1 STOP 5 (pos 5) | CACCCAGATCGTCAGCGCCG | CACCCAGAUCGUCAGCGCCG |
| | 엑손 1 SD (pos 8) | CCACTCACCTGCTCTACCCC | CCACUCACCUGCUCUACCCC |
| | 엑손 2 SA (pos 8) | CCACAGTCTGAAAGAAAGCA | CCACAGUCUGAAAGAAAGCA |
| | 엑손 3 SA (pos 5) | GACACTGTTGGCACGGAGGA | GACACUGUUGGCACGGAGGA |
| | 엑손 3 SD (pos 4) | TTACCATGGCCATCAACACA | UUACCAUGGCCAUCAACACA |
| TRBC2 | 엑손 1 STOP 1 (pos 8) | CCACACCCAAAAGGCCACAC | CCACACCCAAAAGGCCACAC |
| | 엑손 1 STOP 2 (pos 5) | CCCACCAGCTCAGCTCCACG | CCCACCAGCUCAGCUCCACG |
| | 엑손 1 STOP 3 (pos 7) | CGCTGTCAAGTCCAGTTCTA | CGCUGUCAAGUCCAGUUCUA |
| | 엑손 1 STOP 4 (pos 6) | GCTGTCAAGTCCAGTTCTAC | GCUGUCAAGUCCAGUUCUAC |
| | 엑손 1 STOP 5 (pos 5) | CACCCAGATCGTCAGCGCCG | CACCCAGAUCGUCAGCGCCG |
| | 엑손 2 SA (pos 8) | CCACAGTCTGAAAGAAAACA | CCACAGUCUGAAAGAAAACA |
| | 엑손 2 SA (pos 7) | CACAGTCTGAAAGAAAACAG | CACAGUCUGAAAGAAAACAG |
| | 엑손 3 SD (pos 4) | TTACCATGGCCATCAGCACG | UUACCAUGGCCAUCAGCACG |
| | 엑손 1 SD (pos 8) | CCACTCACCTGCTCTACCCC | CCACUCACCUGCUCUACCCC |
| CISH | 엑손 1 STOP | TCTGCGTTCAGGGGTAAGCG | UCUGCGUUCAGGGGUAAGCG |
| | 엑손 1 SD | GCGCTTACCCCTGAACGCAG | GCGCUUACCCCUGAACGCAG |
| | 엑손 2 STOP 2 | GACTGGGCAGCGGCCCCTGT | GACUGGGCAGCGGCCCCUGU |
| | 엑손 2 STOP 1 | GGACTGGGCAGCGGCCCCTG | GGACUGGGCAGCGGCCCCUG |

[0987]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 2 STOP 3 | GTCATGCAGCCCTTGCCTGC | GUCAUGCAGCCCUUGCCUGC |
| | 엑손 2 STOP 4 | TCATGCAGCCCTTGCCTGCT | UCAUGCAGCCCUUGCCUGCU |
| | 엑손 2 STOP 5 | CATGCAGCCCTTGCCTGCTG | CAUGCAGCCCUUGCCUGCUG |
| | 엑손 2 SD 1 | CTCACCAGATTCCCGAAGGT | CUCACCAGAUUCCCGAAGGU |
| | 엑손 2 SD 2 | CAGACTCACCAGATTCCCGA | CAGACUCACCAGAUUCCCGA |
| | 엑손 3 SA 1 (pos 4) | AGCCTAGGCAAGTGCAGAGG | AGCCUAGGCAAGUGCAGAGG |
| | 엑손 3 SA 2 (pos 5) | CAGCCTAGGCAAGTGCAGAG | CAGCCUAGGCAAGUGCAGAG |
| | 엑손 3 SA 3 (pos 7) | ACCAGCCTAGGCAAGTGCAG | ACCAGCCUAGGCAAGUGCAG |
| | 엑손 3 STOP 1 (pos 8) | TGGAACCCCAATACCAGCCT | UGGAACCCCAAUACCAGCCU |
| | 엑손 3 STOP 2 (pos 7) | CACCTGCAGAAGATGCCAGA | CACCUGCAGAAGAUGCCAGA |
| ACAT1 | 엑손 1 SD 1 (pos 7) | CGCTCACCTGCACCAGCCTC | CGCUCACCUGCACCAGCCUC |
| | 엑손 3 SA (pos 5) | CTTCCTGGCAAGACACAAGA | CUUCCUGGCAAGACACAAGA |
| | 엑손 3 STOP (pos 5) | AATTCAGGGAGCCATTGAAA | AAUUCAGGGAGCCAUUGAAA |
| | 엑손 3 SD (pos 8) | CTACTGACCTGCCTTTTCAA | CUACUGACCUGCCUUUUCAA |
| | 엑손 5 STOP (pos 7) | GCCTCTCAAAGTCTTATGTG | GCCUCUCAAAGUCUUAUGUG |
| | 엑손 7 STOP (pos 4) | TTCCCATGCTGCTTTACTTC | UUCCCAUGCUGCUUUACUUC |
| | 엑손 8 STOP (pos 8) | TTTAGGTCAACCAGATGTAG | UUUAGGUCAACCAGAUGUAG |
| | 엑손 9 SA (pos 9) | TGTGCCTGAAAGCAAAAATG | UGUGCCUGAAAGCAAAAAUG |
| | 엑손 9 SD (pos 4) | TTACCTACTATTCTTGCCAG | UUACCUACUAUUCUUGCCAG |

[0988]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 10 SA (pos 6) | AAATGCTGTTTAAAAAAAGG | AAAUGCUGUUUAAAAAAAGG |
| | 엑손 11 STOP (pos 4) | CCCCAAAAAGTGAATATCAA | CCCCAAAAAGUGAAUAUCAA |
| Cyp11a1 | 엑손 1 STOP 1 (pos 4) | GTCCAGAATTTCCAGAAGTA | GUCCAGAAUUUCCAGAAGUA |
| | 엑손 2 SA 1 (pos 4) | TCCCTGGAGGGGTGGGGGAG | UCCCUGGAGGGGUGGGGGAG |
| | 엑손 2 SD 1 (pos 4) | TCACTTCAACAGGACTCCTA | UCACUUCAACAGGACUCCUA |
| | 엑손 3 SD 1 (pos 6) | CCTTACACTCAAAGGCAAAG | CCUUACACUCAAAGGCAAAG |
| | 엑손 4 SA (pos 5) | ATGGCTGCAGGGAGAGGAAG | AUGGCUGCAGGGAGAGGAAG |
| | 엑손 4 STOP 1 (pos 8) | GGAGCGCCAGGGGATGCTGG | GGAGCGCCAGGGGAUGCUGG |
| | 엑손 4 STOP 2 (pos 8) | TCACGTCCCATGCAGCCACA | UCACGUCCCAUGCAGCCACA |
| | 엑손 6 SA (pos 8) | TGGACGTCTGGTGGGGAGTA | UGGACGUCUGGUGGGGAGUA |
| | 엑손 8 STOP 1(pos 6) | ACTCACATTGATGAGGAAGA | ACUCACAUUGAUGAGGAAGA |
| | 엑손 9 SA (pos 7) | CAGCATCTGAGAAAGGCAGA | CAGCAUCUGAGAAAGGCAGA |
| | 엑손 9 STOP 1 (pos 5) | AATCCAACACCTCAGCGATG | AAUCCAACACCUCAGCGAUG |
| | 엑손 9 STOP 2 (pos 4) | ATCCAACACCTCAGCGATGT | AUCCAACACCUCAGCGAUGU |
| GATA3 | 엑손 1 STOP 1 (pos 8) | CGCGGCGCAGTACCCGCTGC | CGCGGCGCAGUACCCGCUGC |
| | 엑손 1 SD 1 (pos 7) | CACTCACCGTGGTGGGTCGG | CACUCACCGUGGUGGGUCGG |
| | 엑손 1 SD 2 (pos 6) | ACTCACCGTGGTGGGTCGGA | ACUCACCGUGGUGGGUCGGA |
| | 엑손 2 SA 1 (pos 8) | TGGCTCCCTGTGGGGCAACG | UGGCUCCCUGUGGGGCAACG |
| | 엑손 2 STOP 2 (pos 5) | GATTCCAGGGGGAGGCGGTG | GAUUCCAGGGGGAGGCGGUG |

[0989]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 2 SD 1 (pos 8) | GCTCCTACCTGTGCTGGACC | GCUCCUACCUGUGCUGGACC |
| | 엑손 3 STOP 1 (pos 7) | TCGCCGCCACAGTGGGGTCG | UCGCCGCCACAGUGGGGUCG |
| | 엑손 4 SA (pos 5) | CAGACTGAGAGTGGGGAGAG | CAGACUGAGAGUGGGGAGAG |
| | 엑손 4 STOP 1 (pos 7) | CCTCCTCCAGAGTGTGGTTG | CCUCCUCCAGAGUGUGGUUG |
| NR4A1 | 엑손 1 STOP 1 (pos 8) | AGCCATCCCAGGGAGAGAGC | AGCCAUCCCAGGGAGAGAGC |
| | 엑손 1 STOP 2 (pos 7) | GCCATCCCAGGGAGAGAGCT | GCCAUCCCAGGGAGAGAGCU |
| | 엑손 1 STOP 3 (pos 6) | CCATCCCAGGGAGAGAGCTG | CCAUCCCAGGGAGAGAGCUG |
| | 엑손 1 STOP 4 (pos 5) | CTCACAGGCCACCCACCAGC | CUCACAGGCCACCCACCAGC |
| | 엑손 2 STOP 1 (pos 8) | CCGCTTCCAGAAGTGCCTGG | CCGCUUCCAGAAGUGCCUGG |
| | 엑손 2 STOP 2 (pos 5) | CTTCCAGAAGTGCCTGGCGG | CUUCCAGAAGUGCCUGGCGG |
| | 엑손 3 SA 1 (pos 5) | ACAACTGCAAAGGAATGGGT | ACAACUGCAAAGGAAUGGGU |
| | 엑손 3 SA 2 (pos 4) | CAACTGCAAAGGAATGGGTA | CAACUGCAAAGGAAUGGGUA |
| | 엑손 4 SA (pos 4) | GAACTAGGAAGACGGTCCAG | GAACUAGGAAGACGGUCCAG |
| | 엑손 4 STOP 1 (pos 8) | GGCTGACCAGGACCTGTTGC | GGCUGACCAGGACCUGUUGC |
| | 엑손 4 SD 1 (pos 5) | CTCACCTGTACGCCAGGCGG | CUCACCUGUACGCCAGGCGG |
| | 엑손 4 SD 2 (pos 8) | GCTCTCACCTGTACGCCAGG | GCUCUCACCUGUACGCCAGG |
| | 엑손 5 SA (pos 8) | CTTAGACCTGGCAGGCAGAT | CUUAGACCUGGCAGGCAGAU |
| | 엑손 5 STOP 1 (pos 5) | CAATCCAGTCCCCGAAGCCA | CAAUCCAGUCCCCGAAGCCA |
| | 엑손 5 STOP 2 (pos 4) | AATCCAGTCCCCGAAGCCAC | AAUCCAGUCCCCGAAGCCAC |

[0990]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 5 SD 1 (pos 6) | ACTCACCGGTGATGAGGACA | ACUCACCGGUGAUGAGGACA |
| | 엑손 5 SD 2 (pos 5) | CTCACCGGTGATGAGGACAA | CUCACCGGUGAUGAGGACAA |
| | 엑손 6 SA (pos 6) | CCGGTCTGCGGGAAGGGTAC | CCGGUCUGCGGGAAGGGUAC |
| | 엑손 6 STOP 1 (pos 8) | TGGGCTGCAGGAGCCGCGGC | UGGGCUGCAGGAGCCGCGGC |
| NR4A2 | 엑손 1 STOP 1 (pos 7) | TTGTACCAAATGCCCCTGTC | UUGUACCAAAUGCCCCUGUC |
| | 엑손 1 STOP 2 (pos 8) | CGGACAGCAGTCCTCCATTA | CGGACAGCAGUCCUCCAUUA |
| | 엑손 1 STOP 3 (pos 6) | AGGTGCAGCACAGCCCCATG | AGGUGCAGCACAGCCCCAUG |
| | 엑손 1 STOP 4 (pos 5) | GGTGCAGCACAGCCCCATGT | GGUGCAGCACAGCCCCAUGU |
| | 엑손 1 STOP 5 (pos 7) | AGTTGCCAGATGCGCTTCGA | AGUUGCCAGAUGCGCUUCGA |
| | 엑손 1 STOP 6 (pos 6) | GTTGCCAGATGCGCTTCGAC | GUUGCCAGAUGCGCUUCGAC |
| | 엑손 1 STOP 7 (pos 5) | GTCTCAGCTGCTCGACACGC | GUCUCAGCUGCUCGACACGC |
| | 엑손 3 SD (pos 7) | TTCTTACCCTGGAATAGTCC | UUCUUACCCUGGAAUAGUCC |
| | 엑손 4 SD (pos 5) | ATTACCTGTATGCTAATCGA | AUUACCUGUAUGCUAAUCGA |
| | 엑손 5 STOP 1 (pos 4) | TTGCAATGCGTTCGTGGCTT | UUGCAAUGCGUUCGUGGCUU |
| | 엑손 5 SD (pos 6) | ACTGACCTGTGACCATAGCC | ACUGACCUGUGACCAUAGCC |
| NR4A3 | 엑손 2 SA (pos 4) | TATCTGCAGGGACAGAGAAA | UAUCUGCAGGGACAGAGAAA |
| | 엑손 2 STOP 1 (pos 8) | TGCGGCGCAGACATACAGCT | UGCGGCGCAGACAUACAGCU |
| | 엑손 2 STOP 2 (pos 6) | CCCCGCAGGCGGGGGCGTTA | CCCCGCAGGCGGGGGCGUUA |
| | 엑손 3 STOP 1 (pos 4) | TTTCAGAAGTGTCTCAGTGT | UUUCAGAAGUGUCUCAGUGU |

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 5 SD (pos 5) | ATTACCTGATGGAAAGTCTG | AUUACCUGAUGGAAAGUCUG |
| | 엑손 6 STOP 1 (pos 4) | CTTCAGTGCCTTCGTGGATT | CUUCAGUGCCUUCGUGGAUU |
| | 엑손 7 SA (pos 4) | TTTCTGCAGAGGGATAGAGA | UUUCUGCAGAGGGAUAGAGA |
| | 엑손 7 STOP 1 (pos 8) | AGACCACCAGAGTAAGGGAC | AGACCACCAGAGUAAGGGAC |
| MCJ | 엑손 1 STOP (pos 6) | ACTTGCAGCCCTCGGCCAAA | ACUUGCAGCCCUCGGCCAAA |
| FAS | 엑손 1 SD (pos 9) | AGGGCTCACCAGAGGTAGGA | AGGGCUCACCAGAGGUAGGA |
| | 엑손 3 SA (pos 6) | TTCACCTGCCCAAGGAAAAA | UUCACCUGCCCAAGGAAAAA |
| | 엑손 4 SA (pos 7) | CTAAGCCTAGAAAATCAGTT | CUAAGCCUAGAAAAUCAGUU |
| | 엑손 5 SA (pos 5) | ACATCTAGAAAAAAAAATAC | ACAUCUAGAAAAAAAAAUAC |
| | 엑손 5 SD (pos 5) | ATTACCTTCCTCTTTGCACT | AUUACCUUCCUCUUUGCACU |
| | 엑손 6 SA (pos 5) | GATCCTGTAGGTTGGAACAT | GAUCCUGUAGGUUGGAACAU |
| | 엑손 6 STOP 1 (pos 4) | AAGCCACCCCAAGTTAGATC | AAGCCACCCCAAGUUAGAUC |
| | 엑손 6 SD (pos 7) | AACTTACCCCAAACAATTAG | AACUUACCCCAAACAAUUAG |
| | 엑손 7 SD (pos 8) | ATACCTACAGGATTTAAAGT | AUACCUACAGGAUUUAAAGU |
| | 엑손 8 SA (pos 8) | GTTTCCTAGAAAGCAAAAAA | GUUUCCUAGAAAGCAAAAAA |
| | 엑손 9 STOP 1 (pos 6) | AAGTTCAACTGCTTCGTAAT | AAGUUCAACUGCUUCGUAAU |
| | 엑손 9 STOP (pos 5) | AATTCAGACTATCATCCTCA | AAUUCAGACUAUCAUCCUCA |
| SELPG/PSGL 1 | 엑손 1 STOP 1 (pos 6) | GCTTGCAGCTGTGGGACACC | GCUUGCAGCUGUGGGACACC |

[0991]

| 유전자 | 설명 | 표적화 서열 | gRNA 스페이서 서열 |
|---|---|---|---|
| | 엑손 1 STOP 2 (pos 8) | GACCACTCAACCAGTGCCCA | GACCACUCAACCAGUGCCCA |
| | 엑손 1 STOP 3 (pos 8) | GGAGGCACAGACCACTCCAC | GGAGGCACAGACCACUCCAC |
| | 엑손 1 STOP 4 (pos 5) | GGCACAGACAACTCGACTGA | GGCACAGACAACUCGACUGA |
| | 엑손 1 STOP 5 (pos 8) | GGAGGCACAGACCACTCCAC | GGAGGCACAGACCACUCCAC |
| | 엑손 1 STOP 6 (pos 4) | GCACAGACCACTCAACCCAC | GCACAGACCACUCAACCCAC |
| | 엑손 1 STOP 7 (pos 8) | GACCACTCAACCCACAGGCC | GACCACUCAACCCACAGGCC |
| | 엑손 1 STOP 8 (pos 8) | GACCACTCAAACCACAGCCA | GACCACUCAAACCACAGCCA |
| | 엑손 1 STOP 9 (pos 8) | GACCACTCAACCCACAGCCA | GACCACUCAACCCACAGCCA |
| | 엑손 1 STOP 10 (pos 8) | GGAGGCACAGACCACTCCAC | GGAGGCACAGACCACUCCAC |
| | 엑손 1 STOP 11 (pos 8) | GACCACTCAACCAGCAGCCA | GACCACUCAACCAGCAGCCA |

[0992]

[0993] ***프로토스페이서 인접 모티프***

[0994] 용어 "프로토스페이서 인접 모티프(PAM)" 또는 PAM-유사 모티프는 CRISPR 세균 적응 면역계에서 Cas9 뉴클레이스에 의해 표적화된 DNA 서열 바로 다음에 오는 2개 내지 6개의 염기쌍 DNA 서열을 지칭한다. 일부 실시형태에서, PAM은 5' PAM(즉, 프로토스페이서의 5' 말단의 상류에 위치함)일 수 있다. 다른 실시형태에서, PAM은 3' PAM(즉, 프로토스페이서의 5' 말단의 하류에 위치함)일 수 있다. PAM 서열은 표적 결합에 필수적이지만, 정확한 서열은 Cas 단백질의 유형에 따라 다르다. PAM 서열은 당업계에 공지된 임의의 PAM 서열일 수 있다. 적합한 PAM 서열은 NGG, NGA, NGC, NGN, NGT, NGTT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW 또는 NAAAAC를 포함하지만, 이들로 제한되지 않는다. Y는 피리미딘이고; N은 임의의 뉴클레오타이드 염기이며; W는 A 또는 T이다.

[0995] 본 명세서에 제공된 염기 편집기는 표준 또는 비-표준 프로토스페이서 인접 모티프(PAM) 서열을 포함하는 뉴클레오타이드 서열에 결합할 수 있는 CRISPR 단백질-유래 도메인을 포함할 수 있다. PAM 부위는 표적 폴리뉴클레오타이드 서열에 근접한 뉴클레오타이드 서열이다. 본 개시내용의 일부 양태는 상이한 PAM 특이성을 갖는 CRISPR 단백질의 전부 또는 일부를 포함하는 염기 편집기를 제공한다.

[0996] 예를 들어, 전형적으로 S. 피오게네스로부터의 Cas9(spCas9)와 같은 Cas9 단백질이 특정 핵산 영역에 결합하기

위해서는 표준 NGG PAM 서열이 필요하며, 여기서 "NGG"의 "N"은 아데닌(A), 티민(T), 구아닌(G) 또는 사이토신(C)이고, G는 구아닌이다. PAM은 CRISPR 단백질-특이적일 수 있으며, 상이한 CRISPR 단백질-유래 도메인을 포함하는 상이한 염기 편집기 간에 상이할 수 있다. PAM은 표적 서열의 5' 또는 3'일 수 있다. PAM은 표적 서열의 상류 또는 하류일 수 있다. PAM은 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 이상의 뉴클레오타이드 길이일 수 있다. 종종, PAM은 2개 내지 6개의 뉴클레오타이드 길이이다.

[0997] 일부 실시형태에서, 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[R.T. Walton *et al.*, 2020, *Science*, 10.1126/science.aba8853 (2020)]에 기술되어 있는 바와 같이, PAM은 "NRN" PAM이며, 여기서 "NRN"의 "N"은 아데닌(A), 티민(T), 구아닌(G) 또는 사이토신(C)이고, R은 아데닌(A) 또는 구아닌(G)이거나; 또는 PAM은 "NYN" PAM이며, 여기서 NYN의 "N"은 아데닌(A), 티민(T), 구아닌(G) 또는 사이토신(C)이고, Y는 사이티딘(C) 또는 티민(T)이다.

[0998] 몇몇 PAM 변이체가 아래 **표 5**에 기재되어 있다.

*표 5*

**Cas9 단백질 및 상응하는 PAM 서열**

| 변이체 | PAM |
|---|---|
| spCas9 | NGG |
| spCas9-VRQR | NGA |
| spCas9-VRER | NGCG |
| xCas9 (sp) | NGN |
| saCas9 | NNGRRT |
| saCas9-KKH | NNNRRT |
| spCas9-MQKSER | NGCG |
| spCas9-MQKSER | NGCN |
| spCas9-LRKIQK | NGTN |
| spCas9-LRVSQK | NGTN |
| spCas9-LRVSQL | NGTN |
| spCas9-MQKFRAER | NGC |
| Cpf1 | 5' (TTTV) |
| SpyMac | 5'-NAA-3' |

[0999]

[1000] 일부 실시형태에서, PAM은 NGC이다. 일부 실시형태에서, NGC PAM은 Cas9 변이체에 의해 인식된다. 일부 실시형태에서, NGC PAM 변이체는 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E 및 T1337R(통칭 "MQKFRAER")로부터 선택되는 하나 이상의 아미노산 치환을 포함한다.

[1001] 일부 실시형태에서, PAM은 NGT이다. 일부 실시형태에서, NGT PAM은 Cas9 변이체에 의해 인식된다. 일부 실시형태에서, NGT PAM 변이체는 하나 이상의 1335번, 1337번, 1135번, 1136번, 1218번 및/또는 1219번 잔기에서의 표적화된 돌연변이를 통해 생성된다. 일부 실시형태에서, NGT PAM 변이체는 하나 이상의 1219번, 1335번, 1337번, 1218번 잔기에서의 표적화된 돌연변이를 통해 생성된다. 일부 실시형태에서, NGT PAM 변이체는 하나 이상의 1135번, 1136번, 1218번, 1219번 및 1335번 잔기에서의 표적화된 돌연변이를 통해 생성된다. 일부 실시형태에서, NGT PAM 변이체는 아래 **표 6A** 및 **표 6B**에 제공되는 표적화된 돌연변이의 세트로부터 선택된다.

표 6

[표 6A]

1219 번, 1335 번, 1337 번, 1218 번 잔기에서의 NGT PAM 변이체 돌연변이

| 변이체 | E1219V | R1335Q | T1337 | G1218 |
|---|---|---|---|---|
| 1 | F | V | T | |
| 2 | F | V | R | |
| 3 | F | V | Q | |
| 4 | F | V | L | |
| 5 | F | V | T | R |
| 6 | F | V | R | R |
| 7 | F | V | Q | R |
| 8 | F | V | L | R |
| 9 | L | L | T | |
| 10 | L | L | R | |
| 11 | L | L | Q | |
| 12 | L | L | L | |
| 13 | F | I | T | |
| 14 | F | I | R | |
| 15 | F | I | Q | |
| 16 | F | I | L | |
| 17 | F | G | C | |
| 18 | H | L | N | |
| 19 | F | G | C | A |
| 20 | H | L | N | V |
| 21 | L | A | W | |
| 22 | L | A | F | |
| 23 | L | A | Y | |
| 24 | I | A | W | |
| 25 | I | A | F | |
| 26 | I | A | Y | |

[1002]

[표 6B]

**1135 번, 1136 번, 1218 번, 1219 번 및 1335 번 잔기에서의**

**NGT PAM 변이체 돌연변이**

| 변이체 | D1135L | S1136R | G1218S | E1219V | R1335Q |
|---|---|---|---|---|---|
| 27 | G | | | | |
| 28 | V | | | | |
| 29 | I | | | | |
| 30 | | A | | | |
| 31 | | W | | | |
| 32 | | H | | | |
| 33 | | K | | | |
| 34 | | | K | | |
| 35 | | | R | | |
| 36 | | | Q | | |
| 37 | | | T | | |
| 38 | | | N | | |
| 39 | | | | I | |
| 40 | | | | A | |
| 41 | | | | N | |
| 42 | | | | Q | |
| 43 | | | | G | |
| 44 | | | | L | |
| 45 | | | | S | |
| 46 | | | | T | |
| 47 | | | | | L |
| 48 | | | | | I |
| 49 | | | | | V |
| 50 | | | | | N |
| 51 | | | | | S |
| 52 | | | | | T |
| 53 | | | | | F |
| 54 | | | | | Y |
| 55 | N1286Q | I1331F | | | |

[1003]

[1004] 일부 실시형태에서, NGT PAM 변이체는 **표 6A** 및 **표 6B**의 변이체 5, 7, 28, 31 또는 36으로부터 선택된다. 일부 실시형태에서, 변이체는 NGT PAM 인식이 개선되었다.

[1005] 일부 실시형태에서, NGT PAM 변이체는 1219번, 1335번, 1337번 및/또는 1218번 잔기에 돌연변이를 갖는다. 일부 실시형태에서, NGT PAM 변이체는 아래 **표 7**에 제공되는 변이체로부터 개선된 인식을 위한 돌연변이로 선택된다.

*표 7*

**1219 번, 1335 번, 1337 번 및 1218 번 잔기에서의**

**NGT PAM 변이체 돌연변이**

| 변이체 | E1219V | R1335Q | T1337 | G1218 |
|---|---|---|---|---|
| 1 | F | V | T | |
| 2 | F | V | R | |
| 3 | F | V | Q | |
| 4 | F | V | L | |
| 5 | F | V | T | R |
| 6 | F | V | R | R |
| 7 | F | V | Q | R |
| 8 | F | V | L | R |

[1006]

[1007] 일부 실시형태에서, NGT PAM은 아래 **표 8**에 제공되는 변이체로부터 선택된다.

표 8

NGT PAM 변이체

| | NGTN 변이체 | D1135 | S1136 | G1218 | E1219 | A1322R | R1335 | T1337 |
|---|---|---|---|---|---|---|---|---|
| 변이체 1 | LRKIQK | L | R | K | I | - | Q | K |
| 변이체 2 | LRSVQK | L | R | S | V | - | Q | K |
| 변이체 3 | LRSVQL | L | R | S | V | - | Q | L |
| 변이체 4 | LRKIRQK | L | R | K | I | R | Q | K |
| 변이체 5 | LRSVRQK | L | R | S | V | R | Q | K |
| 변이체 6 | LRSVRQL | L | R | S | V | R | Q | L |

[1008]

[1009] 일부 실시형태에서, NGTN 변이체는 변이체 1이다. 일부 실시형태에서, NGTN 변이체는 변이체 2이다. 일부 실시형태에서, NGTN 변이체는 변이체 3이다. 일부 실시형태에서, NGTN 변이체는 변이체 4이다. 일부 실시형태에서, NGTN 변이체는 변이체 5이다. 일부 실시형태에서, NGTN 변이체는 변이체 6이다.

[1010] 일부 실시형태에서, Cas9 도메인은 스트렙토코커스 피오게네스로부터의 Cas9 도메인(SpCas9)이다. 일부 실시형태에서, SpCas9 도메인은 뉴클레이스 활성 SpCas9, 뉴클레이스 불활성 SpCas9(SpCas9d) 또는 SpCas9 닉케이스(SpCas9n)이다. 일부 실시형태에서, SpCas9는 D9X 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함하되, X는 D를 제외한 임의의 아미노산이다. 일부 실시형태에서, SpCas9는 D9A 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 비-표준 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SpCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 NGG, NGA 또는 NGCG PAM 서열을 갖는 핵산 서열에 결합할 수 있다.

[1011] 일부 실시형태에서, SpCas9 도메인은 D1135X, R1335X 및 T1337X 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이 중 하나 이상을 포함하되, X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 D1135E, R1335Q 및 T1337R 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135E, R1335Q 및 T1337R 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135X, R1335X 및 T1337X 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이 중 하나 이상을 포함하되, X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 D1135V, R1335Q 및 T1337R 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135V, R1335Q 및 T1337R 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135X, G1218X, R1335X 및 T1337X 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이 중 하나 이상을 포함하되, X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 D1135V, G1218R, R1335Q 및 T1337R 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135V, G1218R, R1335Q 및 T1337R 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다.

[1012] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 Cas9 폴리펩타이드와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 임의의 Cas9 폴리펩타이드의 아미노산 서열을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 임의의 Cas9 폴리펩타이드의 아미노산 서열로 구성된다.

[1013] 일부 예에서, 본 명세서에 개시된 염기 편집기의 CRISPR 단백질-유래 도메인에 의해 인식되는 PAM은 염기 편집기를 암호화하는 인서트(insert)(예를 들어, AAV 인서트)에 대한 별도의 올리고뉴클레오타이드 상의 세포에 제공될 수 있다. 이러한 실시형태에서, 별도의 올리고뉴클레오타이드 상에 PAM을 제공하는 것은 표적 서열과 동일한 폴리뉴클레오타이드 상에 인접한 PAM이 존재하지 않기 때문에, 그렇지 않으면 절단될 수 없는 표적 서열의 절단을 가능하게 할 수 있다.

[1014] 실시형태에서, S. 피오게네스 Cas9(SpCas9)는 게놈 조작을 위한 CRISPR 엔도뉴클레이스로 사용될 수 있다. 그러나, 다른 것도 사용될 수 있다. 일부 실시형태에서, 상이한 엔도뉴클레이스가 소정의 게놈 표적을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 비-NGG PAM 서열을 갖는 합성 SpCas9-유래 변이체가 사용될 수 있다. 추가적으로, 다양한 종으로부터의 다른 Cas9 오쏘로그가 확인되었으며, 이러한 "비-SpCas9"는 본 개시내용에 또한 유용할 수 있는 다양한 PAM 서열에 결합할 수 있다. 예를 들어, 상대적으로 큰 크기의 SpCas9(대략 4kb의 코딩 서열)는 세포에서 효율적으로 발현될 수 없는 SpCas9 cDNA를 운반하는 플라스미드로 이어질 수 있다. 반대로, 스타필로코커스 아우레우스 Cas9(SaCas9)에 대한 코딩 서열은 SpCas9보다 대략 1킬로베이스 짧아 세포에서 효율적으로 발현될 가능성이 있다. SpCas9와 유사하게, SaCas9 엔도뉴클레이스는 시험관내 포유동물 세포 및 마우스 생체내에서 표적 유전자를 변형시킬 수 있다. 일부 실시형태에서, Cas 단백질은 상이한 PAM 서열을 표적으로 할 수 있다. 일부 실시형태에서, 표적 유전자는, 예를 들어, Cas9 PAM, 5'-NGG에 인접할 수 있다. 다른 실시형태에서, 다른 Cas9 오쏘로그는 상이한 PAM 요건을 가질 수 있다. 예를 들어, S. 써모필루스(CRISPR1의 경우 5'-NNAGAA 및 CRISPR3의 경우 5'-NGGNG) 및 네이세리아 메닌지티디스(5'-NNNNGATT)의 PAM과 같은 다른 PAM도 또한 표적 유전자에 인접하여 발견될 수 있다.

[1015] 일부 실시형태에서, S. 피오게네스 시스템의 경우, 표적 유전자 서열은 5'-NGG PAM보다 앞에 있을 수 있고(즉, 5'일 수 있음) 및 20-nt 가이드 RNA 서열은 PAM에 인접한 Cas9 절단을 매개하기 위해 반대편 가닥과 염기쌍을 이룰 수 있다. 일부 실시형태에서, 인접한 절단은 PAM의 상류에 있는 3개 또는 약 3개의 염기쌍일 수 있다. 일부 실시형태에서, 인접한 절단은 PAM의 상류에 있는 10개 또는 약 10개의 염기쌍일 수 있다. 일부 실시형태에서, 인접한 절단은 PAM의 상류에 있는 0개 내지 20개 또는 약 0개 내지 20개의 염기쌍일 수 있다. 예를 들어, 인접한 절단은 PAM의 상류에 있는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 30개의 염기쌍 옆에 있을 수 있다. 인접한 절단은 또는 1개 내지 30개의 염기쌍만큼 PAM의 하류일 수 있다. PAM 서열에 결합할 수 있는 예시적인 SpCas9 단백질의 서열은 다음과 같다:

[1016] 예시적인 PAM-결합 SpCas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[1017]

[1018] 예시적인 PAM-결합 SpCas9n의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[1019]

[1020] 예시적인 PAM-결합 SpEQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESVLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD

[1021] SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[1022]

[1023] 위의 서열에서, D1135, R1335 및 T1337에서 돌연변이되어 SpEQR Cas9를 생성할 수 있는 잔기 E1135, Q1335 및 R1337이 밑줄이 그어져 있고, 굵게 표시되어 있다.

[1024] 예시적인 PAM-결합 SpVQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS

[1025]

KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[1026]

[1027] 위의 서열에서, D1135, R1335 및 T1337에서 돌연변이되어 SpVQR Cas9를 생성할 수 있는 잔기 V1135, Q1335 및 R1337이 밑줄이 그어져 있고, 굵게 표시되어 있다.

[1028] 예시적인 PAM-결합 SpVRER Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
A**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVLD
ATLIHQSITGLYETRIDLSQLGGD.

[1029]

[1030] 위의 서열에서, D1135, G1218, R1335 및 T1337에서 돌연변이되어 SpVRER Cas9를 생성할 수 있는 잔기 V1135, R1218, Q1335 및 R1337이 밑줄이 그어져 있고, 굵게 표시되어 있다.

[1031] 일부 실시형태에서, 조작된 SpCas9 변이체는 3' H(비-G PAM) 측면에 있는 프로토스페이서 인접 모티프(PAM) 서열을 인식할 수 있다(**표 3A** 내지 **표 3D** 참조). 일부 실시형태에서, SpCas9 변이체는 NRNH PAM을 인식한다(여기서, R은 A 또는 G이고, H는 A, C 또는 T임). 일부 실시형태에서, 비-G PAM은 NRRH, NRTH 또는 NRCH이다(예를 들어, 문헌[Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020)] 참조, 이의 내용은 그 전문이 참조에 의해 본 명세서에 원용됨).

[1032] 일부 실시형태에서, Cas9 도메인은 재조합 Cas9 도메인이다. 일부 실시형태에서, 재조합 Cas9 도메인은 SpyMacCas9 도메인이다. 일부 실시형태에서, SpyMacCas9 도메인은 뉴클레이스 활성 SpyMacCas9, 뉴클레이스 불활성 SpyMacCas9(SpyMacCas9d) 또는 SpyMacCas9 닉케이스(SpyMacCas9n)이다. 일부 실시형태에서, SaCas9 도메인, SaCas9d 도메인 또는 SaCas9n 도메인은 비-표준 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SpyMacCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 NAA PAM 서열을 갖는 핵산 서열에 결합할 수 있다.

[1033] 천연 5'-NAAN-3' PAM 특이성을 갖는 스트렙토코커스 마카사에(*Streptococcus macacae*)에서 Spy Cas9의 예시적인 Cas9 상동체의 서열은 당업계에 공지되어 있고, 예를 들어, Jakimo 등

(www.biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf)에 기술되어 있으며, 아래에 제공된다.

[1034] SpyMacCas9

```
MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSL
HEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQ
KPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDI
GDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEII
SFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQ
KQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.
```

[1035]

[1036] 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA 또는 RNA를 절단하는 능력이 감소되도록 H840A, P475A, W476A, N477A, D1125A, W1126A 및 D1218A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다. 또 다른 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 폴리펩타이드가 표적 DNA를 절단하는 능력이 감소되도록 D10A, H840A, P475A, W476A, N477A, D1125A, W1126A 및 D1218A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 능력이 감소되어 있지만, 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력은 보유한다. 일부 실시형태에서, 변이체 Cas9 단백질이 W476A 및 W1126A 돌연변이를 보유하는 경우 또는 변이체 Cas9 단백질이 P475A, W476A, N477A, D1125A, W1126A 및 D1218A 돌연변이를 보유하는 경우, 변이체 Cas9 단백질은 PAM 서열에 효율적으로 결합하지 않는다. 따라서, 일부 이러한 경우에, 이러한 변이체 Cas9 단백질이 결합 방법에 사용될 때, 방법은 PAM 서열이 필요하지 않다. 다시 말해서, 일부 실시형태에서, 이러한 변이체 Cas9 단백질이 결합 방법에 사용될 때, 방법은 가이드 RNA를 포함할 수 있지만, 방법은 PAM 서열의 부재하에 수행될 수 있다(그리고, 따라서 결합의 특이성은 가이드 RNA의 표적화 분절에 의해 제공됨). 위의 효과(즉, 하나 또는 다른 뉴클레이스 부분의 불활성화)를 달성하기 위해 다른 잔기가 돌연변이될 수 있다. 비제한적인 예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 및/또는 A987이 변경(즉, 치환)될 수 있다. 또한, 알라닌 치환 이외의 돌연변이가 적합하다.

[1037] 일부 실시형태에서, 염기 편집기의 CRISPR 단백질-유래 도메인은 표준 PAM 서열(NGG)을 갖는 Cas9 단백질의 전부 또는 일부를 포함할 수 있다. 다른 실시형태에서, Cas9-유래 염기 편집기의 도메인은 비-표준 PAM 서열을 이용할 수 있다. 이러한 서열은 당업계에 기술되어 있으며, 당업자에게 명백할 것이다. 예를 들어, 비-표준 PAM 서열에 결합하는 Cas9 도메인은 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" Nature 523, 481-485 (2015); 및 Kleinstiver, B. P., *et al.*, "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition" Nature Biotechnology 33, 1293-1298 (2015); R.T. Walton *et*

*al.* "Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants" *Science* 10.1126/science.aba8853 (2020); Hu *et al.* "Evolved Cas9 variants with broad PAM compatibility and high DNA specificity," *Nature*, 2018 Apr. 5, 556(7699), 57-63; Miller *et al.*, "Continuous evolution of SpCas9 variants compatible with non-G PAMs" *Nat. Biotechnol.*, 2020 Apr;38(4):471-481]에 기술되어 있다.

[1038] ***배타성이 감소된 Cas9 도메인***

[1039] 전형적으로, S. 피오게네스로부터의 Cas9(spCas9)와 같은 Cas9 단백질은 특정 핵산 영역에 결합하기 위해 표준 NGG PAM 서열이 필요하며, 여기서 "NGG"의 "N"은 아데노신(A), 티미딘(T) 또는 사이토신(C)이고, G는 구아노신이다. 이는 게놈 내에서 목적하는 염기를 편집하는 능력을 제한할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집 융합 단백질은 정확한 위치, 예를 들어, PAM의 상류에 있는 표적 염기를 포함하는 영역에 배치될 필요가 있을 수 있다. 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016)] 참조. 따라서, 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 표준(예를 들어, NGG) PAM 서열을 포함하지 않는 뉴클레오타이드 서열에 결합할 수 있는 Cas9 도메인을 포함할 수 있다. 비-표준 PAM 서열에 결합하는 Cas9 도메인이 당업계에 기술되어 있으며 당업자에게 명백할 것이다. 예를 들어, 비-표준 PAM 서열에 결합하는 Cas9 도메인은 각각 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); 및 Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015)]에 기술되어 있다.

[1040] ***내부 삽입을 갖는 융합 단백질***

[1041] 핵산 프로그램 가능한 핵산 결합 단백질, 예를 들어, napDNAbp에 융합된 이종성 폴리펩타이드를 포함하는 융합 단백질이 본 명세서에 제공된다. 이종성 폴리펩타이드는 천연 또는 야생형 napDNAbp 폴리펩타이드 서열에서 발견되지 않는 폴리펩타이드일 수 있다. 이종성 폴리펩타이드는 napDNAbp의 C-말단, napDNAbp의 N-말단에서 napDNAbp에 융합되거나 또는 napDNAbp의 내부 위치에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩타이드는 napDNAbp의 내부 위치에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩타이드는 데아미네이스 또는 기능적 이의 단편이다. 예를 들어, 융합 단백질은 폴리펩타이드인 Cas9 또는 Cas12(예를 들어, Cas12b/C2c1)의 N-말단 단편 및 C-말단 단편의 측면에 있는 데아미네이스를 포함할 수 있다. 융합 단백질의 데아미네이스는 사이티딘 데아미네이스일 수 있다. 일부 실시형태에서, 사이티딘 데아미네이스는 APOBEC 데아미네이스(예를 들어, APOBEC1)이다. 융합 단백질의 데아미네이스는 아데노신 데아미네이스일 수 있다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA(예를 들어, TadA*7.10 또는 TadA*8)이다. 일부 실시형태에서, TadA는 TadA*8 또는 TadA*9이다. 본 명세서에 기재된 바와 같은 TadA 서열(예를 들어, TadA7.10 또는 TadA*8)은 위에 기재된 융합 단백질에 적합한 데아미네이스이다.

[1042] 일부 실시형태에서, 융합 단백질은 다음 구조를 포함한다:

[1043] NH2-[napDNAbp의 N-말단 단편]-[데아미네이스]-[napDNAbp의 C-말단 단편]-COOH;

[1044] NH2-[N-말단 Cas9의 단편]-[아데노신 데아미네이스]-[C-말단 Cas9의 단편]-COOH;

[1045] NH2-[N-말단 Cas12의 단편]-[아데노신 데아미네이스]-[C-말단 Cas12의 단편]-COOH;

[1046] NH2-[N-말단 Cas9의 단편]-[사이티딘 데아미네이스]-[C-말단 Cas9의 단편]-COOH;

[1047] NH2-[N-말단 Cas12의 단편]-[사이티딘 데아미네이스]-[C-말단 Cas12의 단편]-COOH;

[1048] 여기에서, "]-["의 각 개체는 선택적 링커이다.

[1049] 데아미네이스는 원형 순열 데아미네이스일 수 있다. 예를 들어, 데아미네이스는 원형 순열 아데노신 데아미네이스일 수 있다. 일부 실시형태에서, 데아미네이스는 TadA 참조 서열에서 넘버링된 116번 아미노산 잔기에서 원형으로 순열된 원형 순열 TadA이다. 일부 실시형태에서, 데아미네이스는 TadA 참조 서열에서 넘버링된 136번 아미노산 잔기에서 원형으로 순열된 원형 순열 TadA이다. 일부 실시형태에서, 데아미네이스는 TadA 참조 서열에서 넘버링된 65번 아미노산 잔기에서 원형으로 순열된 원형 순열 TadA이다.

[1050] 융합 단백질은 하나 이상의 데아미네이스를 포함할 수 있다. 융합 단백질은, 예를 들어, 1개, 2개, 3개, 4개, 5

개 이상의 데아미네이스를 포함할 수 있다. 일부 실시형태에서, 융합 단백질은 1개의 데아미네이스를 포함한다. 일부 실시형태에서, 융합 단백질은 2개 데아미네이스를 포함한다. 융합 단백질 내 2개 이상의 데아미네이스는 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 이들의 조합일 수 있다. 2개 이상의 데아미네이스는 동종이량체일 수 있다. 2개 이상의 데아미네이스는 이종이량체일 수 있다. 2개 이상의 데아미네이스는 napDNAbp에 나란히 삽입될 수 있다. 일부 실시형태에서, 2개 이상의 데아미네이스는 napDNAbp에 나란히 있지 않을 수 있다.

[1051] 일부 실시형태에서, 융합 단백질의 napDNAbp는 Cas9 폴리펩타이드 또는 이의 단편이다. Cas9 폴리펩타이드는 변이체 Cas9 폴리펩타이드일 수 있다. 일부 실시형태에서, Cas9 폴리펩타이드는 Cas9 닉케이스(nCas9) 폴리펩타이드 또는 이의 단편이다. 일부 실시형태에서, Cas9 폴리펩타이드는 뉴클레이스 활성이 없는 Cas9(dCas9) 폴리펩타이드 또는 이의 단편이다. 융합 단백질의 Cas9 폴리펩타이드는 전장 Cas9 폴리펩타이드일 수 있다. 일부 경우에, 융합 단백질의 Cas9 폴리펩타이드는 전장 Cas9 폴리펩타이드가 아닐 수 있다. Cas9 폴리펩타이드는, 예를 들어, 자연-발생적 Cas9 단백질과 비교하여 N-말단 또는 C-말단 말단에서 절단될 수 있다. Cas9 폴리펩타이드는 원형으로 순열된 Cas9 단백질일 수 있다. Cas9 폴리펩타이드는 표적 폴리뉴클레오타이드 및 가이드 핵산 서열에 여전히 결합할 수 있는 Cas9 폴리펩타이드의 단편, 부분 또는 도메인일 수 있다.

[1052] 일부 실시형태에서, Cas9 폴리펩타이드는 스트렙토코커스 피오게네스 Cas9(SpCas9), 스타필로코커스 아우레우스 Cas9(SaCas9), 스트렙토코커스 써모필루스 1 Cas9(St1Cas9) 또는 이들의 단편 또는 변이체이다.

[1053] 융합 단백질의 Cas9 폴리펩타이드는 자연-발생적 Cas9 폴리펩타이드와 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함할 수 있다.

[1054] 융합 단백질의 Cas9 폴리펩타이드는 아래에 기재된 Cas9 아미노산 서열(아래에서 "Cas9 참조 서열"이라 함)과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함할 수 있다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD

[1055]

[1056] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인).

[1057] Cas9 폴리펩타이드의 N- 및 C-말단 단편의 측면에 있는 이종성 촉매 도메인을 포함하는 융합 단백질은 또한 본 명세서에 기재된 바와 같은 방법에서 염기 편집에 유용하다. Cas9 및 하나 이상의 데아미네이스 도메인, 예를 들어, 아데노신 데아미네이스를 포함하거나 또는 Cas9 서열의 측면에 있는 아데노신 데아미네이스 도메인을 포

함하는 융합 단백질은 또한 표적 서열의 고도로 특이적이고 효율적인 염기 편집에 유용하다. 실시형태에서, 키메라 Cas9 융합 단백질은 Cas9 폴리펩타이드 내에 삽입된 이종성 촉매 도메인(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)을 포함한다. 일부 실시형태에서, 융합 단백질은 Cas9 내에 삽입된 아데노신 데아미네이스 도메인 및 사이티딘 데아미네이스 도메인을 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 Cas9 내에서 융합되고, 사이티딘 데아미네이스는 C-말단에 융합된다. 일부 실시형태에서, 아데노신 데아미네이스는 Cas9 내에서 융합되고, 사이티딘 데아미네이스 N-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas9 내에서 융합되고, 아데노신 데아미네이스는 C-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas9 내에서 융합되고, 아데노신 데아미네이스는 N-말단에 융합된다.

[1058] 아데노신 데아미네이스와 사이티딘 데아미네이스 및 Cas9와의 융합 단백질의 예시적인 구조가 다음과 같이 제공된다:

[1059] $NH_2$-[Cas9(아데노신 데아미네이스)]-[사이티딘 데아미네이스]-COOH;

[1060] $NH_2$-[사이티딘 데아미네이스]-[Cas9(아데노신 데아미네이스)]-COOH;

[1061] $NH_2$-[Cas9(사이티딘 데아미네이스)]-[아데노신 데아미네이스]-COOH; 또는

[1062] $NH_2$-[아데노신 데아미네이스]-[Cas9(사이티딘 데아미네이스)]-COOH.

[1063] 일부 실시형태에서, 위의 일반 구조에서 사용된 "-"는 선택적 링커의 존재를 나타낸다.

[1064] 다양한 실시형태에서, 촉매 도메인은 아데노신 데아미네이스 활성과 같은 DNA 변형 활성(예를 들어, 데아미네이스 활성)을 갖는다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA(예를 들어, TadA*7.10)이다. 일부 실시형태에서, TadA는 TadA*8이다. 일부 실시형태에서, TadA*8은 Cas9 내에서 융합되고, 사이티딘 데아미네이스는 C-말단에 융합된다. 일부 실시형태에서, TadA*8은 Cas9 내에서 융합되고, 사이티딘 데아미네이스는 N-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas9 내에서 융합되고, TadA*8은 C-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas9 내에서 융합되고, TadA*8은 N-말단에 융합된다. TadA*8과 사이티딘 데아미네이스 및 Cas9와의 융합 단백질의 예시적인 구조는 다음과 같이 제공된다:

[1065] $NH_2$-[Cas9(TadA*8)]-[사이티딘 데아미네이스]-COOH;

[1066] $NH_2$-[사이티딘 데아미네이스]-[Cas9(TadA*8)]-COOH;

[1067] $NH_2$-[Cas9(사이티딘 데아미네이스)]-[TadA*8]-COOH; 또는

[1068] $NH_2$-[TadA*8]-[Cas9(사이티딘 데아미네이스)]-COOH.

[1069] 일부 실시형태에서, 위의 일반 구조에서 사용된 "-"는 선택적 링커의 존재를 나타낸다.

[1070] 이종성 폴리펩타이드(예를 들어, 데아미네이스)는, 예를 들어, napDNAbp가 표적 폴리뉴클레오타이드 및 가이드 핵산에 결합하는 능력을 보유하도록 적합한 위치에서 napDNAbp(예를 들어, Cas9 또는 Cas12(예를 들어, Cas12b/C2c1))에 삽입될 수 있다. 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 데아미네이스의 기능(예를 들어, 염기 편집 활성) 또는 napDNAbp의 기능(예를 들어, 표적 핵산 및 가이드 핵산에 결합하는 능력)을 손상시키지 않고 napDNAbp에 삽입될 수 있다. 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는, 예를 들어, 결정학 연구에 의해 나타낸 바와 같이 고은 인자 또는 B-인자를 포함하는 무질서한 영역 또는 영역에서 napDNAbp에 삽입될 수 있다. 덜 질서정연하거나, 무질서하거나 또는 구조화되지 않은 단백질의 영역, 예를 들어, 용매 노출된 영역 및 루프가 구조 또는 기능을 손상시키지 않고 삽입하는 데 사용될 수 있다. 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 가요성 루프 영역 또는 용매-노출 영역에서 napDNAbp에 삽입될 수 있다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노

신 데아미네이스 및 사이티딘 데아미네이스)는 Cas9 또는 Cas12b/C2c1 폴리펩타이드의 가요성 루프에 삽입된다.

[1071] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)의 삽입 위치는 Cas9 폴리펩타이드의 결정 구조의 B-인자 분석에 의해 결정된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 평균보다 더 높은 B-인자(예를 들어, 무질서한 영역을 포함하는 총 단백질 또는 단백질 도메인과 비교하여 더 높은 B 인자)를 포함하는 Cas9 폴리펩타이드의 영역에 삽입된다. B-인자 또는 온도 인자는 평균 위치에서 원자의 변동을 나타낼 수 있다(예를 들어, 온도-의존적 원자 진동 또는 결정 격자의 정적 무질서의 결과로서). 백본 원자에 대한 높은 B-인자(예를 들어, 평균 B-인자보다 높음)는 상대적으로 높은 국부 이동성을 갖는 영역을 나타낼 수 있다. 이러한 영역은 구조 또는 기능을 손상시키지 않고 데아미네이스를 삽입하는데 사용될 수 있다. 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 총 단백질에 대한 평균 B-인자보다 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200% 또는 200% 초과인 B-인자를 갖는 Cα 원자를 갖는 잔기가 있는 위치에 삽입될 수 있다. 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 잔기를 포함하는 Cas9 단백질 도메인의 평균 B-인자보다 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200% 또는 200% 초과인 B-인자를 갖는 Cα 원자를 갖는 잔기가 있는 위치에 삽입될 수 있다. 평균보다 높은 B-인자를 포함하는 Cas9 폴리펩타이드 위치는, 예를 들어, 위의 Cas9 참조 서열에서 넘버링된 768번, 792번, 1052번, 1015번, 1022번, 1026번, 1029번, 1067번, 1040번, 1054번, 1068번, 1246번, 1247번 및 1248번 잔기를 포함할 수 있다. 평균보다 높은 B-인자를 포함하는 Cas9 폴리펩타이드 영역은, 예를 들어, 위의 Cas9 참조 서열에서 넘버링된 792번 내지 872번, 792번 내지 906번 및 2번 내지 791번 잔기를 포함할 수 있다.

[1072] 이종성 폴리펩타이드(예를 들어, 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 768번, 791번, 792번, 1015번, 1016번, 1022번, 1023번, 1026번, 1029번, 1040번, 1052번, 1054번, 1067번, 1068번, 1069번, 1246번, 1247번 및 1248번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩타이드는 위의 Cas9 참조 서열에서 넘버링된 768번 내지 769번, 791번 내지 792번, 792번 내지 793번, 1015번 내지 1016번, 1022번 내지 1023번, 1026번 내지 1027번, 1029번 내지 1030번, 1040번 내지 1041번, 1052번 내지 1053번, 1054번 내지 1055번, 1067번 내지 1068번, 1068번 내지 1069번, 1247번 내지 1248번 또는 1248번 내지 1249번 아미노산 위치, 또는 이들의 상응하는 아미노산 위치에 삽입된다. 일부 실시형태에서, 이종성 폴리펩타이드는 위의 Cas9 참조 서열에서 넘버링된 769번 내지 770번, 792번 내지 793번, 793번 내지 794번, 1016번 내지 1017번, 1023번 내지 1024번, 1027번 내지 1028번, 1030번 내지 1031번, 1041번 내지 1042번, 1053번 내지 1054번, 1055번 내지 1056번, 1068번 내지 1069번, 1069번 내지 1070번, 1248번 내지 1249번 또는 1249번 내지 1250번 아미노산 위치, 또는 이들의 상응하는 아미노산 위치에 삽입된다. 일부 실시형태에서, 이종성 폴리펩타이드는 위의 Cas9 참조 서열에서 넘버링된 768번, 791번, 792번, 1015번, 1016번, 1022번, 1023번, 1026번, 1029번, 1040번, 1052번, 1054번, 1067번, 1068번, 1069번, 1246번, 1247번 및 1248번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체한다. 삽입 위치와 관련하여 위의 Cas9 참조 서열에 대한 언급은 예시를 위한 것임을 이해하여야 한다. 본 명세서에 논의된 바와 같은 삽입은 위의 Cas9 참조 서열의 Cas9 폴리펩타이드 서열로 제한되지 않지만, 변이체 Cas9 폴리펩타이드, 예를 들어, Cas9 닉케이스(nCas9), 뉴클레이스 활성이 없는 Cas9(dCas9), 뉴클레이스 도메인이 결여되어 있는 Cas9 변이체, 절단된 Cas9 또는, 부분 또는 완전 HNH 도메인이 결여되어 있는 Cas9 도메인의 상응하는 위치에서의 삽입을 포함한다.

[1073] 이종성 폴리펩타이드(예를 들어, 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 768번, 792번, 1022번, 1026번, 1040번, 1068번 및 1247번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩타이드는 위의 Cas9 참조 서열에서 넘버링된 768번 내지 769번, 792번 내지 793번, 1022번 내지 1023번, 1026번 내지 1027번, 1029번 내지 1030번, 1040번 내지 1041번, 1068번 내지 1069번 또는 1247번 내지 1248번 아미노산 위치, 또는 이들의 상응하는 아미노산 위치에 삽입된다. 일부 실시형태에서, 이종성 폴리펩타이드는 위의 Cas9 참조 서열에서 넘버링된 769번 내지 770번, 793번 내지 794번, 1023번 내지 1024번, 1027번 내지 1028번, 1030번 내지 1031번, 1041번 내지 1042번, 1069번 내지 1070번 또는 1248번 내지 1249번 아미노산 위치, 또는 이들의 상응하는 아미노산 위치에 삽입된다. 일부 실시형태에서, 이종성 폴리펩타이드는 위의 Cas9 참조 서열에서 넘버

링된 768, 792, 1022, 1026, 1040, 1068 및 1247번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체한다.

[1074] 이종성 폴리펩타이드(예를 들어, 데아미네이스)는 본 명세서에 기재된 바와 같은 아미노산 잔기 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다. 실시형태에서, 이종성 폴리펩타이드(예를 들어, 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1002번, 1003번, 1025번, 1052번 내지 1056번, 1242번 내지 1247번, 1061번 내지 1077번, 943번 내지 947번, 686번 내지 691번, 569번 내지 578번, 530번 내지 539번 및 1060번 내지 1077번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다. 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 잔기의 N-말단 또는 C-말단에 삽입되거나 또는 잔기를 대체할 수 있다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 잔기의 C-말단에 삽입된다.

[1075] 일부 실시형태에서, 아데노신 데아미네이스(예를 들어, TadA)는 위의 Cas9 참조 서열에서 넘버링된 1015번, 1022번, 1029번, 1040번, 1068번, 1247번, 1054번, 1026번, 768번, 1067번, 1248번, 1052번 및 1246번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 아데노신 데아미네이스(예를 들어, TadA)는 위의 Cas9 참조 서열에서 넘버링된 792번 내지 872번, 792번 내지 906번 또는 2번 내지 791번 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기 대신에 삽입된다. 일부 실시형태에서, 아데노신 데아미네이스는 위의 Cas9 참조 서열에서 넘버링된 1015번, 1022번, 1029번, 1040번, 1068번, 1247번, 1054번, 1026번, 768번, 1067번, 1248번, 1052번 및 1246번으로 이루어진 군으로부터 선택되는 아미노산의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 아데노신 데아미네이스는 위의 Cas9 참조 서열에서 넘버링된 1015번, 1022번, 1029번, 1040번, 1068번, 1247번, 1054번, 1026번, 768번, 1067번, 1248번, 1052번 및 1246번으로 이루어진 군으로부터 선택되는 아미노산의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 아데노신 데아미네이스는 위의 Cas9 참조 서열에서 넘버링된 1015번, 1022번, 1029번, 1040번, 1068번, 1247번, 1054번, 1026번, 768번, 1067번, 1248번, 1052번 및 1246번으로 이루어진 군으로부터 선택되는 아미노산, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1076] 일부 실시형태에서, 사이티딘 데아미네이스(예를 들어, APOBEC1)는 위의 Cas9 참조 서열에서 넘버링된 1016번, 1023번, 1029번, 1040번, 1069번 및 1247번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 사이티딘 데아미네이스는 위의 Cas9 참조 서열에서 넘버링된 1016번, 1023번, 1029번, 1040번, 1069번 및 1247번으로 이루어진 군으로부터 선택되는 아미노산의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 사이티딘 데아미네이스는 위의 Cas9 참조 서열에서 넘버링된 1016번, 1023번, 1029번, 1040번, 1069번 및 1247번으로 이루어진 군으로부터 선택되는 아미노산의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 사이티딘 데아미네이스는 위의 Cas9 참조 서열에서 넘버링된 1016번, 1023번, 1029번, 1040번, 1069번 및 1247번으로 이루어진 군으로부터 선택되는 아미노산, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1077] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 768번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 768번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 768번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 768번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1078] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데

아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 791번 아미노산 잔기 또는 792번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 791번 아미노산 잔기의 N-말단 또는 792번 아미노산의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 아미노산 791번 아미노산의 C-말단 또는 792번 아미노산의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 791번 아미노산 또는 792번 아미노산, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1079] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1016번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1016번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1016번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1016번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1080] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1022번 아미노산 잔기 또는 1023번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1022번 아미노산 잔기의 N-말단 또는 1023번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1022번 아미노산 잔기의 C-말단 또는 1023번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1022번 아미노산 잔기 또는 1023번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1081] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1026번 아미노산 잔기 또는 1029번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1026번 아미노산 잔기의 N-말단 또는 1029번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1026번 아미노산 잔기의 C-말단 또는 1029번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1026번 아미노산 잔기 또는 1029번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1082] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1040번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아

데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1040번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1040번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1040번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1083] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1052번 아미노산 잔기 또는 1054번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1052번 아미노산 잔기의 N-말단 또는 1054번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1052번 아미노산 잔기의 C-말단 또는 1054번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1052번 아미노산 잔기 또는 1054번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1084] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1067번 아미노산 잔기 또는 1068번 아미노산 잔기 또는 1069번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1067번 아미노산 잔기의 N-말단 또는 1068번 아미노산 잔기의 N-말단 또는 1069번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1067번 아미노산 잔기의 C-말단 또는 1068번 아미노산 잔기의 C-말단 또는 1069번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1067번 아미노산 잔기 또는 1068번 아미노산 잔기 또는 1069번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1085] 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1246번 아미노산 잔기 또는 1247번 아미노산 잔기 또는 1248번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1246번 아미노산 잔기의 N-말단 또는 1247번 아미노산 잔기의 N-말단 또는 1248번 아미노산 잔기의 N-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1246번 아미노산 잔기의 C-말단 또는 1247번 아미노산 잔기의 C-말단 또는 1248번 아미노산 잔기의 C-말단, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1246번 아미노산 잔기 또는1247번 아미노산 잔기 또는 1248번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[1086] 일부 실시형태에서, 이종성 폴리펩타이드(예를 들어, 데아미네이스)는 Cas9 폴리펩타이드의 가요성 루프에 삽입된다. 가요성 루프 부분은 위의 Cas9 참조 서열에서 넘버링된 530번 내지 537번, 569번 내지 570번, 686번 내지 691번, 943번 내지 947번, 1002번 내지 1025번, 1052번 내지 1077번, 1232번 내지 1247번 또는 1298번 내지

1300번, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기로 이루어진 군으로부터 선택될 수 있다. 가요성 루프 부분은 위의 Cas9 참조 서열에서 넘버링된 1번 내지 529번, 538번 내지 568번, 580번 내지 685번, 692번 내지 942번, 948번 내지 1001번, 1026번 내지 1051번, 1078번 내지 1231번 또는 1248번 내지 1297번, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기로 이루어진 군으로부터 선택될 수 있다.

[1087] 이종성 폴리펩타이드(예를 들어, 아데닌 데아미네이스)는 위의 Cas9 참조 서열에서 넘버링된 1017번 내지 1069번, 1242번 내지 1247번, 1052번 내지 1056번, 1060번 내지 1077번, 1002 번 내지 1003번, 943번 내지 947번, 530번 내지 537번, 568번 내지 579번, 686번 내지 691번, 1242번 내지 1247번, 1298번 내지 1300번, 1066번 내지 1077번, 1052번 내지 1056번 또는 1060번 내지 1077번의 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 상응하는 Cas9 폴리펩타이드 영역에 삽입될 수 있다.

[1088] 이종성 폴리펩타이드(예를 들어, 아데닌 데아미네이스)는 Cas9 폴리펩타이드의 결실된 영역 대신에 삽입될 수 있다. 결실된 영역은 Cas9 폴리펩타이드의 N-말단 또는 C-말단 부분에 해당할 수 있다. 일부 실시형태에서, 결실된 영역은 위의 Cas9 참조 서열에서 넘버링된 792번 내지 872번 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 해당한다. 일부 실시형태에서, 결실된 영역은 위의 Cas9 참조 서열에서 넘버링된 792번 내지 906번 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 해당한다. 일부 실시형태에서, 결실된 영역은 위의 Cas9 참조 서열에서 넘버링된 2번 내지 791번 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 해당한다. 일부 실시형태에서, 결실된 영역은 위의 Cas9 참조 서열에서 넘버링된 1017번 내지 1069번 잔기, 또는 이의 상응하는 아미노산 잔기에 해당한다.

[1089] 예시적인 내부 융합 염기 편집기가 아래 **표 9**에 제공되어 있다.

*표 9*

**Cas9 단백질에서의 삽입 유전자좌**

| BE ID | 변형 | 다른 ID |
|---|---|---|
| IBE001 | Cas9 TadA ins 1015 | ISLAY01 |
| IBE002 | Cas9 TadA ins 1022 | ISLAY02 |
| IBE003 | Cas9 TadA ins 1029 | ISLAY03 |
| IBE004 | Cas9 TadA ins 1040 | ISLAY04 |
| IBE005 | Cas9 TadA ins 1068 | ISLAY05 |
| IBE006 | Cas9 TadA ins 1247 | ISLAY06 |
| IBE007 | Cas9 TadA ins 1054 | ISLAY07 |
| IBE008 | Cas9 TadA ins 1026 | ISLAY08 |
| IBE009 | Cas9 TadA ins 768 | ISLAY09 |
| IBE020 | 델타 HNH TadA 792 | ISLAY20 |
| IBE021 | N-말단 융합 단일 TadA 나선 절단된 165-말단 | ISLAY21 |
| IBE029 | TadA-원형 순열 116 ins1067 | ISLAY29 |
| IBE031 | TadA- 원형 순열 136 ins1248 | ISLAY31 |
| IBE032 | TadA- 원형 순열 136ins 1052 | ISLAY32 |
| IBE035 | 델타 792-872 TadA ins | ISLAY35 |
| IBE036 | 델타 792-906 TadA ins | ISLAY36 |
| IBE043 | TadA-원형 순열 65 ins1246 | ISLAY43 |
| IBE044 | TadA ins C-말단 절단 2 791 | ISLAY44 |

[1090]

[1091] 이종성 폴리펩타이드(예를 들어, 데아미네이스)는 Cas9 폴리펩타이드의 구조적 또는 기능적 도메인 내에 삽입될 수 있다. 이종성 폴리펩타이드(예를 들어, 데아미네이스)는 Cas9 폴리펩타이드의 2개의 구조적 또는 기능적 도

메인 사이에 삽입될 수 있다. 이종성 폴리펩타이드(예를 들어, 데아미네이스)는, 예를 들어,Cas9 폴리펩타이드로부터 도메인을 결실시킨 후 Cas9 폴리펩타이드의 구조적 또는 기능적 도메인 대신에 삽입될 수 있다. Cas9 폴리펩타이드의 구조적 또는 기능적 도메인은, 예를 들어, RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI 또는 HNH를 포함할 수 있다.

[1092] 일부 실시형태에서, Cas9 폴리펩타이드는 RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI 또는 HNH 도메인으로 이루어진 군으로부터 선택되는 하나 이상의 도메인이 결여되어 있다. 일부 실시형태에서, Cas9 폴리펩타이드는 뉴클레이스 도메인이 결여되어 있다. 일부 실시형태에서, Cas9 폴리펩타이드는 HNH 도메인이 결여되어 있다. 일부 실시형태에서, Cas9 폴리펩타이드는 Cas9 폴리펩타이드가 HNH 활성이 감소 또는 폐지되도록 HNH 도메인의 일부가 결여되어 있다. 일부 실시형태에서, Cas9 폴리펩타이드는 뉴클레이스 도메인의 결실을 포함하고, 데아미네이스는 뉴클레이스 도메인을 대체하도록 삽입된다. 일부 실시형태에서, HNH 도메인이 결실되고, 그 자리에 데아미네이스가 삽입된다. 일부 실시형태에서, 하나 이상의 RuvC 도메인이 결실되고, 그 자리에 데아미네이스가 삽입된다.

[1093] 이종성 폴리펩타이드를 포함하는 융합 단백질은 napDNAbp의 N-말단 및 C-말단 단편의 측면에 있을 수 있다. 일부 실시형태에서, 융합 단백질은 Cas9 폴리펩타이드의 N-말단 단편 및 C-말단 단편의 측면에 데아미네이스를 포함한다. N 말단 단편 또는 C 말단 단편은 표적 폴리뉴클레오타이드 서열에 결합할 수 있다. N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 Cas9 폴리펩타이드의 가요성 루프의 일부를 포함할 수 있다. N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 Cas9 폴리펩타이드의 알파-나선 구조의 일부를 포함할 수 있다. N-말단 단편 또는 C-말단 단편은 DNA 결합 도메인을 포함할 수 있다. N-말단 단편 또는 C-말단 단편은 RuvC 도메인을 포함할 수 있다. N-말단 단편 또는 C-말단 단편은 HNH 도메인을 포함할 수 있다. 일부 실시형태에서, N-말단 단편 및 C-말단 단편은 모두 HNH 도메인을 포함하지 않는다.

[1094] 일부 실시형태에서, N 말단 Cas9 단편의 C-말단은 융합 단백질이 표적 핵염기를 탈아미노화할 때 표적 핵염기에 근접한 아미노산을 포함한다. 일부 실시형태에서, C 말단 Cas9 단편의 N-말단은 융합 단백질이 표적 핵염기를 탈아미노화할 때 표적 핵염기에 근접한 아미노산을 포함한다. 상이한 데아미네이스의 삽입 위치는 표적 핵염기와 N 말단 Cas9 단편의 C-말단 또는 C 말단 Cas9 단편의 N-말단의 아미노산 사이에 근접성을 갖기 위해 상이할 수 있다. 예를 들어, 데아미네이스의 삽입 위치는 위의 Cas9 참조 서열에서 넘버링된 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052 및 1246번으로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기일 수 있다.

[1095] 융합 단백질의 N-말단 Cas9 단편(즉, 융합 단백질에서 데아미네이스 측면에 있는 N-말단 Cas9 단편)은 Cas9 폴리펩타이드의 N-말단을 포함할 수 있다. 융합 단백질의 N-말단 Cas9 단편은 적어도 약 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개, 900개, 1000개, 1100개, 1200개 또는 1300개의 아미노산 길이를 포함할 수 있다. 융합 단백질의 N-말단 Cas9 단편은 위의 Cas9 참조 서열에서 넘버링된 1번 내지 56번, 1번 내지 95번, 1번 내지 200번, 1번 내지 300번, 1번 내지 400번, 1번 내지 500번, 1번 내지 600번, 1번 내지 700번, 1번 내지 718번, 1번 내지 765번, 1번 내지 780번, 1번 내지 906번, 1번 내지 918번 또는 1번 내지 1100번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 상응하는 서열을 포함할 수 있다. N-말단 Cas9 단편은 위의 Cas9 참조 서열에서 넘버링된 1번 내지 56번, 1번 내지 95번, 1번 내지 200번, 1번 내지 300번, 1번 내지 400번, 1번 내지 500번, 1번 내지 600번, 1번 내지 700번, 1번 내지 718번, 1번 내지 765번, 1번 내지 780번, 1번 내지 906번, 1번 내지 918번 또는 1번 내지 1100번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5%의 서열 동일성을 포함하는 서열을 포함할 수 있다.

[1096] 융합 단백질의 C-말단 Cas9 단편(즉, 융합 단백질에서 데아미네이스 측면에 있는 C-말단 Cas9 단편)은 Cas9 폴리펩타이드의 C-말단을 포함할 수 있다. 융합 단백질의 C-말단 Cas9 단편은 적어도 약 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개, 900개, 1000개, 1100개, 1200개 또는 1300개의 아미노산 길이를 포함할 수 있다. 융합 단백질의 C-말단 Cas9 단편은 위의 Cas9 참조 서열에서 넘버링된 1099번 내지 1368번, 918번 내지 1368번, 906번 내지 1368번, 780번 내지 1368번, 765번 내지 1368번, 718번 내지 1368번, 94번 내지 1368번 또는 56번 내지 1368번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 상응하는 서열을 포함할 수 있다. N-말단 Cas9 단편은 위의 Cas9 참조 서열에서 넘버링된 1099번 내지 1368번, 918번 내지 1368번, 906번 내지 1368번, 780번 내지 1368번, 765번 내지 1368번, 718번 내지 1368번, 94번 내지 1368번 또는 56번 내지 1368번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드의 상응하는 아미노산 잔기에 대해 적어

도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5%의 서열 동일성을 포함하는 서열을 포함할 수 있다.

[1097] 융합 단백질의 N-말단 Cas9 단편 및 C-말단 Cas9 단편은 함께, 예를 들어, 위의 Cas9 참조 서열에 기재된 바와 같은 전장의 자연 발생적 Cas9 폴리펩타이드 서열에 상응하지 않을 수 있다.

[1098] 본 명세서에 기재된 융합 단백질은 감소된 게놈 전반의 거짓(spurious) 탈아미노화와 같은 비-표적 부위(예를 들어, 표적외 부위)에서의 감소된 탈아미노화로 표적화된 탈아미노화에 영향을 미칠 수 있다. 본 명세서에 기재된 융합 단백질은 비-표적 부위에서 감소된 방관자(bystander) 탈아미노화로 표적화된 탈아미노화에 영향을 미칠 수 있다. 목적하지 않은 탈아미노화 또는 표적외 탈아미노화는, 예를 들어, Cas9 폴리펩타이드의 N 말단 또는 C 말단에 융합된 데아미네이스를 포함하는 말단 융합 단백질과 비교하여 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99%의 감소될 수 있다. 목적하지 않은 탈아미노화 또는 표적외 탈아미노화는, 예를 들어, Cas9 폴리펩타이드의 N 말단 또는 C 말단에 융합된 데아미네이스를 포함하는 말단 융합 단백질과 비교하여 적어도 1-배, 적어도 2-배, 적어도 3-배, 적어도 4-배, 적어도 5-배, 적어도 10배, 적어도 15배, 적어도 20배, 적어도 30배, 적어도 40배, 적어도 50배, 적어도 60배, 적어도 70배, 적어도 80배, 적어도 90배 또는 적어도 100배 감소될 수 있다.

[1099] 일부 실시형태에서, 융합 단백질의 데아미네이스(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)는 R-루프의 범위 내에서 2개 이하의 핵염기를 탈아미노화한다. 일부 실시형태에서, 데아미네이스의 융합 단백질은 R-루프의 범위 내에서 3개 이하의 핵염기를 탈아미노화한다. 일부 실시형태에서, 데아미네이스의 융합 단백질은 R-루프의 범위 내에서 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개 이하의 핵염기를 탈아미노화한다. R-루프는 DNA:RNA 하이브리드, DNA:DNA 또는 RNA:RNA 상보적 구조를 포함하는 3-가닥의 핵산 구조이며, 단일-가닥 DNA와 관련이 있다. 본 명세서에서 사용되는 바와 같이, R-루프는 표적 폴리뉴클레오타이드가 CRISPR 복합체 또는 염기 편집 복합체와 접촉할 때 형성될 수 있되, 가이드 폴리뉴클레오타이드, 예를 들어, 가이드 RNA의 일부는 표적 폴리뉴클레오타이드, 예를 들어, 표적 DNA의 일부와 혼성화하고 이를 대체한다. 일부 실시형태에서, R-루프는 스페이서 서열과 표적 DNA 상보적 서열의 혼성화된 영역을 포함한다. R-루프 영역은 약 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개 또는 50개의 핵염기쌍 길이일 수 있다. 일부 실시형태에서, R-루프 영역은 약 20개의 핵염기쌍 길이이다. 본 명세서에서 사용되는 바와 같이, R-루프 영역은 가이드 폴리뉴클레오타이드와 혼성화하는 표적 DNA 가닥으로 제한되지 않음을 이해하여야 한다. 예를 들어, R-루프 영역 내의 표적 핵염기의 편집은 가이드 RNA에 대한 상보적 가닥을 포함하는 DNA 가닥에 대한 것일 수 있거나, 또는 가이드 RNA에 대한 상보적인 가닥의 반대 가닥인 DNA 가닥에 대한 것일 수 있다. 일부 실시형태에서, R-루프의 영역에서의 편집은 비-상보적 가닥(프로토스페이서 가닥) 상의 핵염기를 표적 DNA 서열의 가이드 RNA로 편집하는 것을 포함한다.

[1100] 본 명세서에 기재된 융합 단백질은 표준 염기 편집과 상이한 편집 창(window)에서 표적 탈아미노화에 영향을 미칠 수 있다. 일부 실시형태에서, 표적 핵염기는 표적 폴리뉴클레오타이드 서열에서 PAM 서열의 상류에 있는 약 1개 내지 약 20개의 염기이다. 일부 실시형태에서, 표적 핵염기는 표적 폴리뉴클레오타이드 서열에서 PAM 서열의 상류에 있는 약 2개 내지 약 12개의 염기이다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류로부터 떨어져 있거나 또는 이의 상류에 있는 약 1개 내지 9개의 염기쌍, 약 2개 내지 10개의 염기쌍, 약 3개 내지 11개의 염기쌍, 약 4개 내지 12개의 염기쌍, 약 5개 내지 13개의 염기쌍, 약 6개 내지 14개의 염기쌍, 약 7개 내지 15개의 염기쌍, 약 8개 내지 16개의 염기쌍, 약 9개 내지 17개의 염기쌍, 약 10개 내지 18개의 염기쌍, 약 11개 내지 19개의 염기쌍, 약 12개 내지 20개의 염기쌍, 약 1개 내지 7개의 염기쌍, 약 2개 내지 8개의 염기쌍, 약 3개 내지 9개의 염기쌍, 약 4개 내지 10개의 염기쌍, 약 5개 내지 11개의 염기쌍, 약 6개 내지 12개의 염기쌍, 약 7개 내지 13개의 염기쌍, 약 8개 내지 14개의 염기쌍, 약 9개 내지 15개의 염기쌍, 약 10개 내지 16개의 염기쌍, 약 11개 내지 17개의 염기쌍, 약 12개 내지 18개의 염기쌍, 약 13개 내지 19개의 염기쌍, 약 14개 내지 20개의 염기쌍, 약 1개 내지 5개의 염기쌍, 약 2개 내지 6개의 염기쌍, 약 3개 내지 7개의 염기쌍, 약 4개 내지 8개의 염기쌍, 약 5개 내지 9개의 염기쌍, 약 6개 내지 10개의 염기쌍, 약 7개 내지 11개의 염기쌍, 약 8개 내지 12개의 염기쌍, 약 9개 내지 13개의 염기쌍, 약 10개 내지 14개의 염기쌍, 약 11개 내지 15개의 염기쌍, 약 12개 내지 16개의 염기쌍, 약 13개 내지 17개의 염기쌍, 약 14개 내지 18개의 염기쌍, 약 15개 내지 19개의 염기쌍, 약 16개 내지 20개의 염기쌍, 약 1개 내지 3개의 염기쌍, 약 2개 내지 4개의 염기쌍, 약 3개 내지 5개의 염기쌍, 약 4개 내지 6개의 염기쌍, 약 5개 내지 7개의 염기쌍, 약 6개 내지 8개의 염기쌍, 약 7개 내지 9개의

염기쌍, 약 8개 내지 10개의 염기쌍, 약 9개 내지 11개의 염기쌍, 약 10개 내지 12개의 염기쌍, 약 11개 내지 13개의 염기쌍, 약 12개 내지 14개의 염기쌍, 약 13개 내지 15개의 염기쌍, 약 14개 내지 16개의 염기쌍, 약 15개 내지 17개의 염기쌍, 약 16개 내지 18개의 염기쌍, 약 17개 내지 19개의 염기쌍, 약 18개 내지 20개의 염기쌍이다. 일부 실시형태에서, 표적 핵염기는 PAM 서열로부터 떨어져 있거나 또는 이의 상류에 있는 약 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개 이상의 염기쌍이다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류에 있는 약 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개 또는 9개의 염기쌍이다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류에 있는 약 2개, 3개, 4개 또는 6개의 염기쌍이다.

[1101] 융합 단백질은 하나 이상의 이종성 폴리펩타이드를 포함할 수 있다. 예를 들어, 융합 단백질은 하나 이상의 UGI 도메인 및/또는 하나 이상의 핵 국부화 신호를 추가적으로 포함할 수 있다. 2개 이상의 이종성 도메인은 나란히 삽입될 수 있다. 2개 이상의 이종성 도메인은 napDNAbp에서 나란히 있지 않은 위치에 삽입될 수 있다.

[1102] 융합 단백질은 데아미네이스와 napDNAbp 폴리펩타이드 사이에 링커를 포함할 수 있다. 링커는 펩타이드 또는 비-펩타이드 링커일 수 있다. 예를 들어, 링커는 XTEN, (GGGS)n, (GGGGS)n, (G)n, (EAAAK)n, (GGS)n, SGSETPGTSESATPES일 수 있다. 일부 실시형태에서, 융합 단백질은 N-말단 Cas9 단편과 데아미네이스 사이에 링커를 포함한다. 일부 실시형태에서, 융합 단백질은 C-말단 Cas9 단편과 데아미네이스 사이에 링커를 포함한다. 일부 실시형태에서, napDNAbp의 N-말단 및 C-말단 단편은 링커로 데아미네이스에 연결된다. 일부 실시형태에서, N-말단 및 C-말단 단편은 링커 없이 데아미네이스 도메인에 연결된다. 일부 실시형태에서, 융합 단백질은 N-말단 Cas9 단편과 데아미네이스 사이에 링커를 포함하지만, C-말단 Cas9 단편과 데아미네이스 사이에는 링커를 포함하지 않는다. 일부 실시형태에서, 융합 단백질은 C-말단 Cas9 단편과 데아미네이스 사이에 링커를 포함하지만, N-말단 Cas9 단편과 데아미네이스 사이에는 링커를 포함하지 않는다.

[1103] 일부 실시형태에서, 융합 단백질의 napDNAbp는 Cas12 폴리펩타이드, 예를 들어, Cas12b/C2c1 또는 이의 단편이다. Cas12 폴리펩타이드는 변이체 Cas12 폴리펩타이드일 수 있다. 다른 실시형태에서, N- 또는 C-말단 Cas12의 단편 폴리펩타이드는 핵산 프로그램 가능 DNA 결합 도메인 또는 RuvC 도메인을 포함한다. 다른 실시형태에서, 융합 단백질은 Cas12 폴리펩타이드와 촉매 도메인 사이에 링커를 포함한다. 다른 실시형태에서, 링커의 아미노산 서열은 GGSGGS 또는 GSSGSETPGTSESATPESSG이다. 다른 실시형태에서, 링커는 강성 링커이다. 위의 양태의 다른 실시형태에서, 링커는 GGAGGCTCTGGAGGAAGC 또는 GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC에 의해 암호화된다.

[1104] Cas12 폴리펩타이드의 N- 및 C-말단 단편의 측면에 있는 이종성 촉매 도메인을 포함하는 융합 단백질은 또한 본 명세서에 기재된 방법 바와 같은 염기 편집에 유용하다. Cas12 및 하나 이상의 데아미네이스 도메인, 예를 들어, 아데노신 데아미네이스를 포함하거나 또는 Cas12 서열의 측면에 있는 아데노신 데아미네이스 도메인을 포함하는 융합 단백질은 또한 표적 서열의 고도로 특이적이고 효율적인 염기 편집에 유용하다. 실시형태에서, 키메라 Cas12 융합 단백질은 Cas12 폴리펩타이드 내에 삽입된 이종성 촉매 도메인(예를 들어, 아데노신 데아미네이스, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 사이티딘 데아미네이스)을 포함한다. 일부 실시형태에서, 융합 단백질은 Cas12 내에 삽입된 아데노신 데아미네이스 도메인 및 사이티딘 데아미네이스 도메인을 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 Cas12 내에서 융합되고, 사이티딘 데아미네이스는 C-말단에 융합된다. 일부 실시형태에서, 아데노신 데아미네이스는 Cas12 내에서 융합되고, 사이티딘 데아미네이스는 N-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas12 내에서 융합되고, 아데노신 데아미네이스는 C-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas12 내에서 융합되고, 아데노신 데아미네이스는 N-말단에 융합된다. 아데노신 데아미네이스와 사이티딘 데아미네이스 및 Cas12와의 융합 단백질의 예시적인 구조가 다음과 같이 제공된다:

[1105] $NH_2$-[Cas12(아데노신 데아미네이스)]-[사이티딘 데아미네이스]-COOH;

[1106] $NH_2$-[사이티딘 데아미네이스]-[Cas12(아데노신 데아미네이스)]-COOH;

[1107] $NH_2$-[Cas12(사이티딘 데아미네이스)]-[아데노신 데아미네이스]-COOH; 또는

[1108] $NH_2$-[아데노신 데아미네이스]-[Cas12(사이티딘 데아미네이스)]-COOH.

[1109] 일부 실시형태에서, 위의 일반 구조에서 사용된 "-"는 선택적 링커의 존재를 나타낸다.

[1110] 다양한 실시형태에서, 촉매 도메인은 아데노신 데아미네이스 활성과 같은 DNA 변형 활성(예를 들어, 데아미네이

스 활성)을 갖는다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA(예를 들어, TadA*7.10)이다. 일부 실시형태에서, TadA는 TadA*8이다. 일부 실시형태에서, TadA*8은 Cas12 내에서 융합되고, 사이티딘 데아미네이스는 C-말단에 융합된다. 일부 실시형태에서, TadA*8은 Cas12 내에서 융합되고, 사이티딘 데아미네이스는 N-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas12 내에서 융합되고, TadA*8은 C-말단에 융합된다. 일부 실시형태에서, 사이티딘 데아미네이스는 Cas12 내에서 융합되고, TadA*8은 N-말단에 융합된다. TadA*8과 사이티딘 데아미네이스 및 Cas12와의 융합 단백질의 예시적인 구조는 다음과 같이 제공된다:

[1111] N-[Cas12(TadA*8)]-[사이티딘 데아미네이스]-C;

[1112] N-[사이티딘 데아미네이스]-[Cas12(TadA*8)]-C;

[1113] N-[Cas12(사이티딘 데아미네이스)]-[TadA*8]-C; 또는

[1114] N-[TadA*8]-[Cas12(사이티딘 데아미네이스)]-C.

[1115] 일부 실시형태에서, 위의 일반 구조에서 사용된 "-"는 선택적 링커의 존재를 나타낸다.

[1116] 다른 실시형태에서, 융합 단백질은 하나 이상의 촉매 도메인을 포함한다. 다른 실시형태에서, 하나 이상의 촉매 도메인 중 적어도 하나는 Cas12 폴리펩타이드 내에 삽입되거나 또는 Cas12 N-말단 또는 C-말단에서 융합된다. 다른 실시형태에서, 하나 이상의 촉매 도메인 중 적어도 하나는 Cas12 폴리펩타이드의 루프, 알파 나선 영역, 구조화되지 않은 부분 또는 용매 접근 가능한 부분 내에 삽입된다. 다른 실시형태에서, Cas12 폴리펩타이드는 Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i 또는 Cas12j/CasΦ이다. 다른 실시형태에서, Cas12 폴리펩타이드는 바실러스 히사시이 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 V3-13 Cas12b 또는 알리사이클로바실러스 아시디필루스 Cas12b에 대해 적어도 약 85%의 아미노산 서열 동일성을 갖는다. 다른 실시형태에서, Cas12 폴리펩타이드는 바실러스 히사시이 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 V3-13 Cas12b 또는 알리사이클로바실러스 아시디필루스 Cas12b에 대해 적어도 약 90% 아미노산 서열 동일성을 갖는다. 다른 실시형태에서, Cas12 폴리펩타이드는 바실러스 히사시이 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 V3-13 Cas12b 또는 알리사이클로바실러스 아시디필루스 Cas12b에 대해 적어도 약 95% 아미노산 서열 동일성을 갖는다. 다른 실시형태에서, Cas12 폴리펩타이드는 바실러스 히사시이 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 V3-13 Cas12b 또는 알리사이클로바실러스 아시디필루스 Cas12b의 단편을 포함하거나 또는 이로 본질적으로 구성된다.

[1117] 다른 실시형태에서, 촉매 도메인은 BhCas12b의 153번 내지 154번, 255번 내지 256번, 306번 내지 307번, 980번 내지 981번, 1019번 내지 1020번, 534번 내지 535번, 604번 내지 605번 또는 344번 내지 345번 아미노산 위치 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i 또는 Cas12j/CasΦ의 상응하는 아미노산 잔기 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BhCas12b의 아미노산 P153과 S154 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BhCas12b의 아미노산 K255와 E256 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BhCas12b의 아미노산 D980과 G981의 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BhCas12b의 아미노산 K1019와 L1020 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BhCas12b의 아미노산 F534와 P535 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BhCas12b의 아미노산 K604와 G605 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BhCas12b의 아미노산 H344 및 F345 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BvCas12b의 147번과 148번, 248번과 249번, 299번과 300번, 991번과 992번 또는 1031번과 1032번 아미노산 위치 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i 또는 Cas12j/CasΦ의 상응하는 아미노산 잔기 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BvCas12b의 아미노산 P147과 D148 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BvCas12b의 아미노산 G248과 G249 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BvCas12b의 아미노산 P299와 E300 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BvCas12b의 아미노산 G991과 E992 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 BvCas12b의 아미노산 K1031과 M1032 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 AaCas12b의 157번과 158번, 258번과 259번, 310번과 311번, 1008번과 1009번 또는 1044번과 1045번 아미노산 위치 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i 또는 Cas12j/CasΦ의 상응하는 아미노산 잔기 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 AaCas12b의 아미노산 P157과 G158 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 AaCas12b의 아미노산 V258과 G259 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 AaCas12b의 아미노산

D310과 P311 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 AaCas12b의 아미노산 G1008과 E1009 사이에 삽입된다. 다른 실시형태에서, 촉매 도메인은 AaCas12b의 아미노산 G1044와 K1045 사이에 삽입된다.

[1118] 다른 실시형태에서, 융합 단백질은 핵 국부화 신호(예를 들어, 이분 핵 국부화 신호)를 포함한다. 다른 실시형태에서, 핵 국부화 신호의 아미노산 서열은 MAPKKKRKVGIHGVPAA이다. 위의 양태의 다른 실시형태에서, 핵 국부화 신호는 다음 서열에 의해 암호화된다: ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC.

[1119] 다른 실시형태에서, Cas12b 폴리펩타이드는 RuvC 도메인의 촉매 활성을 침묵시키는 돌연변이를 포함한다. 다른 실시형태에서, Cas12b 폴리펩타이드는 D574A, D829A 및/또는 D952A 돌연변이를 포함한다. 다른 실시형태에서, 융합 단백질은 태그(예를 들어, 인플루엔자 헤마글루티닌 태그)를 추가로 포함한다.

[1120] 일부 실시형태에서, 융합 단백질은 내부 융합된 핵염기 편집 도메인(예를 들어, 데아미네이스 도메인, 예를 들어, 아데노신 데아미네이스 도메인의 전부 또는 일부)을 갖는 napDNAbp 도메인(예를 들어, Cas12-유래 도메인)을 포함한다. 일부 실시형태에서, napDNAbp는 Cas12b이다. 일부 실시형태에서, 염기 편집기는 아래 **표 10**에 제공되는 유전자좌에 삽입된 내부 융합된 TadA*8 도메인을 갖는 BhCas12b 도메인을 포함한다.

*표 10*

**Cas12b 단백질에서의 삽입 유전자좌**

[1121]

| BhCas12b | 삽입 부위 | aa 사이에 삽입됨 |
|---|---|---|
| 위치 1 | 153 | PS |
| 위치 2 | 255 | KE |
| 위치 3 | 306 | DE |
| 위치 4 | 980 | DG |
| 위치 5 | 1019 | KL |
| 위치 6 | 534 | FP |
| 위치 7 | 604 | KG |
| 위치 8 | 344 | HF |
| | | |
| BvCas12b | 삽입 부위 | Aa 사이에 삽입됨 |
| 위치 1 | 147 | PD |
| 위치 2 | 248 | GG |
| 위치 3 | 299 | PE |
| 위치 4 | 991 | GE |
| 위치 5 | 1031 | KM |
| | | |
| AaCas12b | 삽입 부위 | Aa 사이에 삽입됨 |
| 위치 1 | 157 | PG |
| 위치 2 | 258 | VG |
| 위치 3 | 310 | DP |
| 위치 4 | 1008 | GE |
| 위치 5 | 1044 | GK |

[1122] 비제한적인 예로서, 핵산 서열을 효과적으로 편집하는 융합 단백질(예를 들어, TadA*8.13-BhCas12b)을 생산하기 위해 아데노신 데아미네이스(예를 들어, TadA*8.13)가 BhCas12b에 삽입된다.

[1123] 일부 실시형태에서, 본 명세서에 기재된 염기 편집 시스템은 Cas9에 TadA가 삽입된 ABE이다. Cas9에 TadA가 삽입된 관련 ABE의 서열이 제공된다.

[1124] 101 Cas9 TadAins 1015

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHAL
TLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQG
GLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGS
LMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSST
DYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1125]

[1126] 102 Cas9 TadAins 1022

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[1127] LGGD

[1128] 103 Cas9 TadAins 1029

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGS
EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMP
RQVFNAQKKAQSSTDGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1129]

[1130] 103 Cas9 TadAins 1040

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC
AALLCYFFRMPRQVFNAQKKAQSSTDNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1131]

[1132] 105 Cas9 TadAins 1068

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGEGSSGSETPGTSESATPESSGSEVEFSHEYWMR
HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL
RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQ
SSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1133]

[1134] 106 Cas9 TadAins 1247

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGGSS
GSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL
VLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTF
EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1135]

[1136] 107 Cas9 TadAins 1054

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[1137] LGGD

[1138] 108 Cas9 TadAins 1026

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVE
FSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQV
FNAQKKAQSSTDQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1139] LGGD

[1140] 109 Cas9 TadAins 768

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMR
HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL
RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE
IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1141]

[1142] 110.1 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV
TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
TEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1143]

[1144] 110.1 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT
LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
[1145] YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1146] 110.2 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT
LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
[1147] YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1148] 110.3 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
NHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1149]

[1150] 110.4 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
NHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1151]

[1152] 110.5 Cas9 TadAins 1249

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSGS
SGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1153]

[1154] 110.5 Cas9 TadAins 델타 59-66 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFE
PCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEG
ILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
D

[1155]

[1156] 110.6 Cas9 TadAins 1251

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
GSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMRRDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1157]

[1158] 110.7 Cas9 TadAins 1252

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARD
EREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMRRNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1159]

[1160] 110.8 Cas9 TadAins 델타 59-66 C-절단 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
VLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC
AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1161]

[1162] 111.1 Cas9 TadAins 997

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM
ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE
LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG
APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1163]

[1164] 111.2 Cas9 TadAins 997

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGSSGSSGSETPGTSESATPESSGGSSIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS
QLGGD

[1165]

[1166] 112 델타 HNH TadA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
[1167] IHQSITGLYETRIDLSQLGGD

[1168] 113 N-말단 단일 TadA 나선 절단 165-말단

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG
RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
MPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP
LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL
INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF
RKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVY
DVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL
DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD

[1169]

[1170] 114 N-말단 단일 TadA 나선 절단 165-말단 델타 59-65

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRSGGS
SGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITD
EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK
RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH
IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY
VDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE
EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA
FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1171]

[1172] 115.1 Cas9 TadAins1004

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDA
TLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1173]

[1174] 115.2 Cas9 TadAins1005

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1175]

[1176] 115.3 Cas9 TadAins1006

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
NHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1177]

[1178] 115.4 Cas9 TadAins1007

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1179]

[1180] 116.1 Cas9 TadAins C-말단 절단2 792

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETP
GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR
VIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLCYFFRMPRQSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1181]

[1182] 116.2 Cas9 TadAins C-말단 절단2 791

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPG
TSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV
IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC
AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1183]

[1184] 116.3 Cas9 TadAins C-말단 절단2 790

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC
AALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1185]

[1186] 117 Cas9 델타 1017-1069

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV
TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVN
FLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1187]

[1188] 118 Cas9 TadA-CP116ins 1067

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRAR
DEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1189]

[1190] 119 Cas9 TadAins 701

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATL
YVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDLTFKEDIQKAQVS
GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA
RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY
YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1191]

[1192] 120 Cas9 TadACP136ins 1248

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSMN
HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1193]

[1194] 121 Cas9 TadACP136ins 1052

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLAMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS
ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEP
CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGNGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1195]

[1196] 122 Cas9 TadACP136ins 1041

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSMNHRVEITEG
ILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES
SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI
GLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI
GRVVFGVRNAKTGAAGSLMDVLHYPGNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1197] LGGD

[1198] 123 Cas9 TadACP139ins 1299

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRMN
HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1199]

[1200] 124 Cas9 델타 792-872 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1201]

[1202] 125 Cas9 델타 792-906 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI
KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL
PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
HQSITGLYETRIDLSQLGGD

[1203]

[1204] 126 TadA CP65ins 1003

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGR
VVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1205]

[1206] 127 TadA CP65ins 1016

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
PYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[1207] LGGD

[1208] 128 TadA CP65ins 1022

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYV
TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESAT
PESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1209]

[1210] 129 TadA CP65ins 1029

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEITAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETP
GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR
VIGEGWNRAIGLHDPGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1211]

[1212] 130 TadA CP65ins 1041

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAA
GSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQS
STDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1213]

[1214] 131 TadA CP65ins 1054

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG
RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
MPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRH
ALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1215]

[1216] 132 TadA CP65ins 1246

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN
AQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKR
ARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1217]

[1218] 일부 실시형태에서, 확인된 위치에서 Cas9 폴리펩타이드 내로 TadA 또는 이의 변이체를 삽입하기 위해 아데노신 데아미네이스 염기 편집기를 생성하였다.

[1219] 예시적이지만 비제한적인 융합 단백질은 PCT 특허 PCT/US2020/016285 및 미국 특허 가출원 제62/852,228호 및 제62/852,224호에 기술되어 있으며, 이들의 내용은 그 전문이 참조에 의해 본 명세서에 원용된다.

[1220] ***A에서 G로의 편집***

[1221] 일부 실시형태에서, 본 명세서에 기재된 염기 편집기는 아데노신 데아미네이스 도메인을 포함하는 융합 단백질을 포함한다. 염기 편집기의 이러한 아데노신 데아미네이스 도메인은 G의 염기 페어링 특성을 나타내는 이노신(I)을 형성함으로써 아데닌(A) 핵염기를 구아닌(G) 핵염기로 편집하는 것을 용이하게 할 수 있다. 아데노신 데아미네이스는 데옥시리보핵산(DNA)에서 데옥시아데노신 잔기의 아데닌을 탈아미노화(즉, 아민기를 제거함)할 수 있다.

[1222] 일부 실시형태에서, 본 명세서에 제공된 핵염기 편집기는 하나 이상의 단백질 도메인을 함께 융합하여 융합 단백질을 생성함으로써 제조될 수 있다. 소정의 실시형태에서, 본 명세서에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성(예를 들어, 효율성, 선택성 및 특이성)을 개선하는 하나 이상의 특징을 포함한다. 예를 들어, 본 명세서에 제공된 융합 단백질은 감소된 뉴클레이스 활성을 갖는 Cas9 도메인을 포함할 수 있다. 일부 실시형

태에서, 본 명세서에 제공된 융합 단백질은 뉴클레이스 활성(dCas9)이 없는 Cas9 도메인 또는 Cas9 닉케이스(nCas9)로 지칭되는 이중나선 DNA 분자의 한 가닥을 절단하는 Cas9 도메인을 가질 수 있다. 임의의 특정 이론에 얽매이지 않고, 촉매 잔기(예를 들어, H840)의 존재는 표적화된 A의 반대편에 있는 T를 포함하는 편집되지 않은(예를 들어, 탈아미노화되지 않은) 가닥을 절단하는 Cas9의 활성을 유지한다. Cas9의 촉매 잔기(예를 들어, D10에서 A10으로)의 돌연변이는 표적화된 A 잔기를 포함하는 편집된 가닥의 절단을 방지한다. 이러한 Cas9 변이체는 또한 gRNA-정의된 표적 서열을 기반으로 특정 위치에서 단일-가닥 DNA 파손(닉)을 생성할 수 있으며, 이는 편집되지 않은 가닥의 복구를 유도하여 궁극적으로 편집되지 않은 가닥의 T에서 C로의 변화를 초래한다. 일부 실시형태에서, A에서 G로의 염기 편집기는 이노신 염기 절단 복구 저해제, 예를 들어, 우라실 글리코실레이스 저해제(UGI) 도메인 또는 촉매적으로 불활성인 이노신 특이적 뉴클레이스를 추가로 포함한다. 임의의 특정 이론에 얽매이지 않고, UGI 도메인 또는 촉매적으로 불활성인 이노신 특이적 뉴클레이스는 탈아미노화된 아데노신 잔기(예를 들어, 이노신)의 염기 절단 복구를 저해 또는 방지할 수 있으며, 염기 편집기의 활성 또는 효율성을 개선할 수 있다.

[1223] 아데노신 데아미네이스를 포함하는 염기 편집기는 DNA, RNA 및 DNA-RNA 하이브리드를 포함하는 임의의 폴리뉴클레오타이드에 작용할 수 있다. 소정의 실시형태에서, 아데노신 데아미네이스를 포함하는 염기 편집기는 RNA를 포함하는 폴리뉴클레오타이드의 표적 A를 탈아미노화할 수 있다. 예를 들어, 염기 편집기는 RNA 폴리뉴클레오타이드 및/또는 DNA-RNA 하이브리드 폴리뉴클레오타이드의 표적 A를 탈아미노화할 수 있는 아데노신 데아미네이스 도메인을 포함할 수 있다. 실시형태에서, 염기 편집기에 혼입된 아데노신 데아미네이스는 RNA에 작용하는 아데노신 데아미네이스(ADAR, 예를 들어, ADAR1 또는 ADAR2)의 전부 또는 일부를 포함한다. 또 다른 실시형태에서, 염기 편집기에 혼입된 아데노신 데아미네이스는 tRNA에 작용하는 아데노신 데아미네이스(ADAT)의 전부 또는 일부를 포함한다. 아데노신 데아미네이스를 포함하는 염기 편집기 도메인은 또한 DNA 폴리뉴클레오타이드의 핵염기를 탈아미노화할 수 있다. 실시형태에서, 염기 편집기의 아데노신 데아미네이스 도메인은 ADAT가 DNA에서 표적 A를 탈아미노화하도록 하는 하나 이상의 돌연변이를 포함하는 ADAT의 전부 또는 일부를 포함한다. 예를 들어, 염기 편집기는 다음 돌연변이: D108N, A106V, D147Y, E155V, L84F, H123Y, I156F 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하는 대장균(EcTadA)으로부터의 ADAT의 전부 또는 일부를 포함할 수 있다.

[1224] 아데노신 데아미네이스는 임의의 적합한 유기체(예를 들어, 대장균)로부터 유래될 수 있다. 일부 실시형태에서, 아데노신 데아미네이스는 원핵생물로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 세균으로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균, 스타필로코커스 아우레우스, 살모넬라 티피, 쉬와넬라 푸트레파시엔스, 헤모필루스 인플루엔자에, 카울로박터 크레센투스 또는 바실러스 서브틸리스로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균으로부터 유래한다. 일부 실시형태에서, 아데닌 데아미네이스는 본 명세서에 제공된 임의의 돌연변이(예를 들어, ecTadA의 돌연변이)에 상응하는 하나 이상의 돌연변이를 포함하는 자연-발생적 아데노신 데아미네이스이다. 임의의 상동 단백질에서 상응하는 잔기는, 예를 들어, 서열 정렬 및 상동 잔기의 결정에 의해 확인될 수 있다. 본 명세서에 기재된 임의의 돌연변이(예를 들어, ecTadA에서 확인된 임의의 돌연변이)에 상응하는 임의의 자연-발생적 아데노신 데아미네이스(예를 들어, ecTadA와 상동성을 가짐)의 돌연변이가 그에 따라 생성될 수 있다.

[1225] ***아데노신 데아미네이스***

[1226] 일부 실시형태에서, 본 발명의 융합 단백질은 하나 이상의 아데노신 데아미네이스 도메인을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 아데노신 데아미네이스는 아데닌을 탈아미노화할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 아데노신 데아미네이스는 DNA의 데옥시아데노신 잔기에서 아데닌을 탈아미노화할 수 있다. 아데노신 데아미네이스는 의의 적합한 유기체(예를 들어, 대장균)로부터 유래될 수 있다. 일부 실시형태에서, 아데닌 데아미네이스는 본 명세서에 제공된 임의의 돌연변이(예를 들어, ecTadA의 돌연변이)에 상응하는 하나 이상의 돌연변이를 포함하는 자연-발생적 아데노신 데아미네이스이다. 당업자는, 예를 들어, 서열 정렬 및 상동 잔기의 결정에 의해 임의의 상동 단백질에서 상응하는 잔기를 확인할 수 있을 것이다. 따라서, 당업자는 본 명세서에 기재된 임의의 돌연변이, 예를 들어, ecTadA에서 확인된 임의의 돌연변이에 상응하는 임의의 자연-발생적 아데노신 데아미네이스(예를 들어, ecTadA와 상동성을 가짐)에서 돌연변이를 생성할 수 있을 것이다. 일부 실시형태에서, 아데노신 데아미네이스는 원핵생물로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 세균으로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균, 스타필로코커스 아우레우스, 살모넬라 티피, 쉬와넬라 푸트레파시엔스, 헤모필루스 인플루엔자에, 카울로박터 크레센투스 또는 바실러스 서브틸리스로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균으로부터 유래한다.

[1227] 본 발명은 증가된 효율성(50% 내지 60% 미만) 및 특이성을 갖는 아데노신 데아미네이스 변이체를 제공한다. 특히, 본 명세서에 기재된 아데노신 데아미네이스 변이체는 폴리뉴클레오타이드 내에서 목적하는 염기를 편집할 가능성이 더 높고, 변경되도록 의도되지 않은 염기(즉, "방관자")를 편집할 가능성이 적다.

[1228] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 데아미네이스이다. 특정 실시형태에서, TadA는 이는 그 전문이 참조에 의해 본 명세서에 원용되어 있는 PCT/US2017/045381(WO 2018/027078)에 기술되어 있는 TadA 중 어느 하나이다.

[1229] 야생형 TadA(wt) 아데노신 데아미네이스는 다음 서열(TadA 참조 서열이라고도 함)을 갖는다:

[1230]
```
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD
```

[1231] 일부 실시형태에서, 아데노신 데아미네이스는 전장 대장균 TadA 데아미네이스이다. 예를 들어, 소정의 실시형태에서, 아데노신 데아미네이스는 다음 아미노산 서열을 포함한다:

[1232]
```
MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHD
PTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSL
MDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD.
```

[1233] 일부 실시형태에서, 아데노신 데아미네이스는 원핵생물로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 세균으로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균(E. 콜라이), 스타필로코커스 아우레우스(S. 아우레우스), 살모넬라 티피무리움(S. 티피무리움), 쉬와넬라 푸트레파시엔스(S. 푸트레파시엔스), 헤모필루스 인플루엔자에(H. 인플루엔자에), 카울로박터 크레센투스(C. 크레센투스), 지오박터 설퍼레듀센스(*Geobacter sulfurreducens*, G. 설퍼레듀센스) 또는 바실러스 서브틸리스로부터 유래한다. 일부 실시형태에서, 아데노신 데아미네이스는 대장균으로부터 유래한다.

[1234] 그러나, 본 출원에서 유용한 추가적인 아데노신 데아미네이스는 당업자에게 명백할 것이며, 본 개시내용의 범위 내에 있음을 이해하여야 한다. 예를 들어, 아데노신 데아미네이스는 tRNA에 작용하는 아데노신 데아미네이스(ADAT)의 상동체일 수 있다. 제한 없이, 예시적인 ADAT 상동체의 아미노산 서열은 다음을 포함한다:

[1235] 스타필로코커스 아우레우스(S. 아우레우스) TadA:

[1236]
```
MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIE
RAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNH
RAIVDKGVLKEACSTLLTTFFKNLRANKKSTN
```

[1237] 바실러스 서브틸리스(B. 서브틸리스) TadA:

[1238]
```
MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACK
ALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEV
VSGVLEEECGGMLSAFFRELRKKKKAARKNLSE
```

[1239] 살모넬라 티피무리움(S. 티피무리움) TadA:

[1240]
```
MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHD
PTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSL
IDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV
```

[1241] 쉬와넬라 푸트레파시엔스(S. 푸트레파시엔스) TadA:

[1242]
```
MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKL
ENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTS
GVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE
```

[1243] 헤모필루스 인플루엔자에 F3031(H. 인플루엔자에) TadA:

[1244]
```
MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEI
IALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDY
KMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK
```

[1245] 카울로박터 크레센투스(C. 크레센투스) TadA:

[1246]
```
MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEI
AAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQP
TCHWRPEVTGGVLADESADLLRGFFRARRKAKI
```

[1247] 지오박터 설퍼레듀센스(G. 설퍼레듀센스) TadA:

[1248]
```
MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEM
IAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADP
RLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP
```

[1249] 대장균 TadA(ecTadA)의 실시형태는 다음을 포함한다:

[1250]
```
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD
```

[1251] 일부 실시형태에서, 아데노신 데아미네이스는 본 명세서에 제공된 임의의 아데노신 데아미네이스에 기재된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 본 명세서에 제공된 아데노신 데아미네이스는 하나 이상의 돌연변이(예를 들어, 본 명세서에 제공된 임의의 돌연변이)를 포함할 수 있음을 이해하여야 한다. 본 개시내용은 소정의 동일성 백분율을 갖는 임의의 데아미네이스 도메인과 본 명세서에 기재된 임의의 돌연변이 또는 이의 조합을 제공한다. 일부 실시형태에서, 아데노신 데아미네이스는 참조 서열 또는 본 명세서에 제공된 임의의 아데노신 데아미네이스와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 당업계에 공지되거나 또는 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 5개, 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개, 적어도 30개, 적어도 35개, 적어도 40개, 적어도 45개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 110개, 적어도 120개, 적어도 130개, 적어도 140개, 적어도 150개, 적어도 160개 또는 적어도 170개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[1252] 본 명세서에 제공된 임의의 돌연변이(예를 들어, TadA 참조 서열을 기반으로 함)는 대장균 TadA(ecTadA), S. 아우레우스 TadA(saTadA) 또는 다른 아데노신 데아미네이스(예를 들어, 세균 아데노신 데아미네이스)와 같은 다른 아데노신 데아미네이스에 도입될 수 있음을 이해하여야 한다. 추가적인 데아미네이스가 본 명세서에 제공된 바와 같이 돌연변이될 수 있는 상동 아미노산 잔기를 확인하기 위해 유사하게 정렬될 수 있다는 것이 당업자에게 명백할 것이다. 따라서, TadA 참조 서열에서 확인된 임의의 돌연변이는 상동 아미노산 잔기를 갖는 다른 아데노신 데아미네이스(예를 들어, ecTada)에서 만들어질 수 있다. 또한, 본 명세서에 제공된 임의의 돌연변이는 TadA 참조 서열 또는 또 다른 아데노신 데아미네이스에서 개별적으로 또는 임의의 조합으로 만들어질 수 있음을 이해하여야 한다.

[1253] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 D108X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 08G, D108N, D108V, D108A 또는 D108Y 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 그러나, 추가적인 데아미네이스는 본 명세서에 제공된 바와 같이 돌연변이될 수 있는 상동 아미노산 잔기를 확인하기 위해

유사하게 정렬될 수 있음을 이해하여야 한다.

[1254] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A106X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A106V 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1255] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 E155X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X의 존재는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 E155D, E155G 또는 E155V 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1256] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 D147X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X의 존재는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 D147Y 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1257] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A106X, E155X 또는 D147X 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 E155D, E155G 또는 E155V 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 D147Y를 포함한다.

[1258] 본 명세서에 제공된 임의의 돌연변이(예를 들어, TadA 참조 서열의 ecTadA 아미노산 서열을 기반으로 함)는 S. 아우레우스 TadA(saTadA)와 같은 다른 아데노신 데아미네이스, 또는 다른 아데노신 데아미네이스(예를 들어, 세균 아데노신 데아미네이스)에 도입될 수 있음을 이해하여야 한다. ecTadA에서 돌연변이된 잔기와 어떻게 상동성을 갖는지는 당업자에게 명백할 것이다. 따라서, ecTadA에서 확인된 임의의 돌연변이는 상동 아미노산 잔기를 갖는 다른 아데노신 데아미네이스에서 만들어질 수 있다. 또한, 본 명세서에 제공된 임의의 돌연변이는 ecTadA 또는 또 다른 아데노신 데아미네이스에서 개별적으로 또는 임의의 조합으로 만들어질 수 있음을 이해하여야 한다.

[1259] 예를 들어, 아데노신 데아미네이스는 TadA 참조 서열의 D108N, A106V, E155V 및/또는 D147Y 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함할 수 있다. 일부 실시형태에서, 아데노신 데아미네이스는 다음의 TadA 참조 서열의 돌연변이(돌연변이 그룹은 ";"로 구분됨), 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이의 그룹을 포함한다: D108N 및 A106V; D108N 및 E155V; D108N 및 D147Y; A106V 및 E155V; A106V 및 D147Y; E155V 및 D147Y; D108N, A106V 및 E155V; D108N, A106V 및 D147Y; D108N, E155V 및 D147Y; A106V, E155V 및 D147Y; 및 D108N, A106V, E155V 및 D147Y. 그러나, 본 명세서에 제공된 상응하는 돌연변이의 임의의 조합이 아데노신 데아미네이스(예를 들어, ecTadA)에서 만들어질 수 있음을 이해하여야 한다.

[1260] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, K110X, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X 및/또는 K157X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하되, X의 존재는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, T17S, L18E, W23L, L34S, W45L, R51H, A56E 또는 A56S, E59G, E85K 또는 E85G, M94L, I95L, V102A, F104L, A106V, R107C 또는 R107H 또는 R107P, D108G 또는 D108N 또는 D108V 또는 D108A 또는 D108Y, K110I, M118K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D 및/또는 K157R 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함한다.

[1261] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, D108X 및/또는 N127X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하되, X는 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, D108N 및/또는 N127S 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함한다.

[1262] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, R26X, M61X, L68X, M70X, A106X, D108X, A109X, N127X, D147X, R152X, Q154X, E155X, K161X, Q163X 및/또는 T166X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, R26W, M61I, L68Q, M70V, A106T, D108N, A109T, N127S, D147Y, R152C, Q154H 또는 Q154R, E155G 또는 E155V 또는 E155D, K161Q, Q163H 및/또는 T166P 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함한다.

[1263] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, D108X, N127X, D147X, R152X 및 Q154X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, M61X, M70X, D108X, N127X, Q154X, E155X 및 Q163X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, D108X, N127X, E155X 및 T166X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다.

[1264] 일부 실시형태에서, 아데노신 데아미네이스는 H8X, A106X 및 D108X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 H8X, R26X, L68X, D108X, N127X, D147X 및 E155X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다.

[1265] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, R126X, L68X, D108X, N127X, D147X 및 E155X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개 또는 7개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, D108X, A109X, N127X 및 E155X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다.

[1266] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, D108N, N127S, D147Y, R152C 및 Q154H로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, M61I, M70V, D108N, N127S, Q154R, E155G 및 Q163H로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, D108N, N127S, E155V 및 T166P로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, A106T, D108N, N127S, E155D 및 K161Q로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, R26W, L68Q, D108N, N127S, D147Y 및 E155V로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, D108N, A109T, N127S 및 E155G로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이 또는 돌연변이를 포함한다.

[1267] 일부 실시형태에서, 아데노신 데아미네이스는 또 다른 아데노신 데아미네이스의 하나 이상의 또는 하나 이상의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 D108N, D108G 또는 D108V 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A106V 및 D108N 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R107C 및 D108N 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, D108N, N127S, D147Y 및 Q154H 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, D108N, N127S, D147Y 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 D108N, D147Y 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, D108N 및 N127S 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A106V, D108N, D147Y 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1268] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 S2X, H8X, I49X, L84X, H123X, N127X, I156X 및/또는 K160X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하되, X의 존재는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 S2A, H8Y, I49F, L84F, H123Y, N127S, I156F 및/또는 K160S 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 하나 이상의 상응하는 돌연변이를 포함한다.

[1269] 일부 실시형태에서, 아데노신 데아미네이스는 L84X 돌연변이 아데노신 데아미네이스를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 L84F 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1270] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H123X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H123Y 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1271] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 I156X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 I156F 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1272] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 L84X, A106X, D108X, H123X, D147X, E155X 및 I156X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개 또는 7개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 S2X, I49X, A106X, D108X, D147X 및 E155X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8X, A106X, D108X, N127X 및 K160X로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다.

[1273] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 L84F, A106V, D108N, H123Y, D147Y, E155V 및 I156F로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개 또는 7개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 S2A, I49F, A106V, D108N, D147Y 및 E155V로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개의 돌연변이를 포함한다.

[1274] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H8Y, A106T, D108N, N127S 및 K160S로 이루어진 군으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이 또는 돌연변이를 포함한다.

[1275] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 E25X, R26X, R107X, A142X 및/또는 A143X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하되, X의 존재는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 E25M, E25D, E25A, E25R, E25V, E25S, E25Y, R26G, R26N, R26Q, R26C, R26L, R26K, R107P, R107K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q 및/또는 A143R 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열에 상응하는 본 명세서에 기재된 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함한다.

[1276] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 E25X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 E25M, E25D, E25A, E25R, E25V, E25S 또는 E25Y 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1277] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R26X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R26G, R26N, R26Q, R26C, R26L 또는 R26K 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1278] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R107X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R107P, R107K, R107A, R107N, R107W, R107H 또는 R107S 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1279] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A142X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A142N, A142D, A142G, 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1280] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A143X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q 및/또는 A143R 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 상응하는 돌연변이를 포함한다.

[1281] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X 및/또는 K161X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하되, X의 존재는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N 및/또는 K161T 돌연변이, 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)의 하나 이상의 상응하는 돌연변이를 포함한다.

[1282] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H36X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 H36L 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1283] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 N37X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 N37T 또는 N37S 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1284] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 P48X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 P48T 또는 P48L 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1285] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R51X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R51H 또는 R51L 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1286] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 S146X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 S146R 또는 S146C 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1287] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 K157X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 K157N 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1288] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 P48X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 P48S, P48T 또는 P48A 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1289] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A142X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 A142N 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1290] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 W23X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, where X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 W23R 또는 W23L 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1291] 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R152X 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함하되, X는 야생형 아데노신 데아미네이스의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열의 R152P 또는 R52H 돌연변이, 또는 또 다른 아데노신 데아미네이스의 상응하는 돌연변이를 포함한다.

[1292] 일 실시형태에서, 아데노신 데아미네이스는 돌연변이 H36L, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F 및 K157N을 포함할 수 있다. 일부 실시형태에서, 아데노신 데아미네이스는 TadA 참조 서열과 관련된 다음의 돌연변이의 조합을 포함하되, 조합의 각 돌연변이는 "_"로 구분되며, 각 돌연변이의 조합은 괄호 사이에 있다:

[1293] (A106V_D108N),

[1294] (R107C_D108N),

[1295] (H8Y_D108N_N127S_D147Y_Q154H),

[1296] (H8Y_D108N_N127S_D147Y_E155V),

[1297] (D108N_D147Y_E155V),

[1298] (H8Y_D108N_N127S),

[1299] (H8Y_D108N_N127S_D147Y_Q154H),

[1300] (A106V_D108N_D147Y_E155V),

[1301] (D108Q_D147Y_E155V),

[1302] (D108M_D147Y_E155V),

[1303] (D108L_D147Y_E155V),

[1304] (D108K_D147Y_E155V),

[1305] (D108I_D147Y_E155V),

[1306] (D108F_D147Y_E155V),

[1307] (A106V_D108N_D147Y),

[1308] (A106V_D108M_D147Y_E155V),

[1309] (E59A_A106V_D108N_D147Y_E155V),

[1310] (E59A cat dead_A106V_D108N_D147Y_E155V),

[1311] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),

[1312] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1313] (D103A_D104N),

[1314] (G22P_D103A_D104N),

[1315] (D103A_D104N_S138A),

[1316] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[1317] (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[1318] (E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F),

[1319] (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[1320] (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[1321] (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),

[1322] (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),

[1323] (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[1324] (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),

[1325] (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

[1326] (A106V_D108N_A142N_D147Y_E155V),

[1327] (R26G_A106V_D108N_A142N_D147Y_E155V),

[1328] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),

[1329] (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),

[1330] (E25D_R26G_A106V_D108N_A142N_D147Y_E155V),

[1331] (A106V_R107K_D108N_A142N_D147Y_E155V),

[1332] (A106V_D108N_A142N_A143G_D147Y_E155V),

[1333] (A106V_D108N_A142N_A143L_D147Y_E155V),

[1334] (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[1335] (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),

[1336] (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),

[1337] (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),

[1338] (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),

[1339] (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),

[1340] (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)

[1341] (H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),

[1342] (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),

[1343] (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1344] (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),

[1345] (D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),

[1346] (H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),

[1347] (Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),

[1348] (E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),

[1349] (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1350] (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),

[1351] (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1352] (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1353] (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),

[1354] (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[1355] (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),

[1356] (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),

[1357] (L84F_A106V_D108N_D147Y_E155V_I156F),

[1358] (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),

[1359] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),

[1360] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),

[1361] (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E),

[1362] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1363] (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1364] (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1365] (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1366] (L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),

[1367] (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),

[1368] (P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

[1369] (P48S_A142N),

[1370] (P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),

[1371] (P48T_I49V_A142N),

[1372] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[1373] (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F)

[1374] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[1375] (H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_ I156F_K157N),

[1376] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[1377] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),

[1378] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N),

[1379] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[1380] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

[1381] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),

[1382] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N),

[1383] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),

[1384] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),

[1385] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F_K157N),

[1386] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P_E155V_I156F_K157N),

[1387] (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),

[1388] (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),

[1389] (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155 V_I156F_K157N).

[1390] 일부 실시형태에서, TadA 데아미네이스는 TadA 변이체이다. 일부 실시형태에서, TadA 변이체는 TadA*7.10이다. 특정 실시형태에서, 융합 단백질은 단일 TadA*7.10 도메인(예를 들어, 단량체로 제공됨)을 포함한다. 다른 실시형태에서, 융합 단백질은 TadA*7.10 및 TadA(wt)를 포함하며, 이는 이종이량체를 형성할 수 있다. 일 실시형태에서, 본 발명의 융합 단백질은 Cas9 닉케이스에 연결된 TadA*7.10에 연결된 야생형 TadA를 포함한다.

[1391] 일부 실시형태에서, TadA*7.10은 적어도 하나의 변경을 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 다음 서열의 변경을 포함한다:

[1392] TadA*7.10

[1393] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[1394] 일부 실시형태에서, TadA*7.10은 아미노산 82번 및/또는 166번 아미노산에 변경을 포함한다. 특정 실시형태에서, TadA*7.10은 다음 변경 하나 이상을 포함한다: Y147T, Y147R, Q154S, Y123H, V82S, T166R 및/또는 Q154R. 다른 실시형태에서, TadA*7.10의 변이체는 다음으로 이루어진 군으로부터 선택되는 변경의 조합을 포함한다: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R

+Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R.

[1395] 일부 실시형태에서, 아데노신 데아미네이스 변이체(예를 들어, TadA*8)는 결실을 포함한다. 일부 실시형태에서, 아데노신 데아미네이스 변이체는 C 말단의 결실을 포함한다. 특정 실시형태에서, 아데노신 데아미네이스 변이체는 TadA 참조 서열인 TadA*7.10에 대해 149번, 150번, 151번, 152번, 153번, 154번, 155번, 156번 및 157번 잔기에서 시작하는 C 말단의 결실, 또는 또 다른 TadA의 상응하는 돌연변이를 포함한다.

[1396] 다른 실시형태에서, 아데노신 데아미네이스 변이체(예를 들어, TadA*8)는 TadA 참조 서열인 TadA*7.10에 대해 다음 변경: Y147T, Y147R, Q154S, Y123H, V82S, T166R 및/또는 Q154R 중 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 단량체이다. 다른 실시형태에서, 아데노신 데아미네이스 변이체(TadA*8)는 TadA 참조 서열인 TadA*7.10에 대해 Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R로 이루어진 군으로부터 선택되는 변경의 조합, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 단량체이다.

[1397] 다른 실시형태에서, 아데노신 데아미네이스 변이체는 각각 TadA 참조 서열인 TadA*7.10에 대해 다음 변경 Y147T, Y147R, Q154S, Y123H, V82S, T166R 및/또는 Q154R 중 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이를 갖는 2개의 아데노신 데아미네이스 도메인(예를 들어, TadA*8)를 포함하는 동종이량체이다. 다른 실시형태에서, 아데노신 데아미네이스 변이체는 TadA 참조 서열인 TadA*7.10에 대해 Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R로 이루어진 군으로부터 선택되는 변경의 조합, 또는 또 다른 TadA의 상응하는 돌연변이를 갖는 2개의 아데노신 데아미네이스 도메인(예를 들어, TadA*8)을 포함하는 동종이량체이다.

[1398] 다른 실시형태에서, 아데노신 데아미네이스 변이체는 야생형 아데노신 데아미네이스 도메인 및 TadA 참조 서열인 TadA*7.10에 대해 다음 변경 Y147T, Y147R, Q154S, Y123H, V82S, T166R 및/또는 Q154R 중 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체이다. 다른 실시형태에서, 아데노신 데아미네이스 변이체는 TadA 참조 서열인 TadA*7.10에 대해 Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R로 이루어진 군으로부터 선택되는 변경의 조합, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 야생형 아데노신 데아미네이스 도메인 및 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체이다.

[1399] 다른 실시형태에서, 아데노신 데아미네이스 변이체는 TadA*7.10 도메인 및 TadA 참조 서열인 TadA*7.10에 대해 다음 변경 Y147T, Y147R, Q154S, Y123H, V82S, T166R 및/또는 Q154R 중 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체이다. 다른 실시형태에서, 아데노신 데아미네이스 변이체는 TadA 참조 서열인 TadA*7.10에 대해 Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R로 이루어진 군으로부터 선택되는 변경의 조합, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 TadA*7.10 도메인 및 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체이다.

[1400] 특정 실시형태에서, 아데노신 데아미네이스 이종이량체는 TadA*8 도메인 및 스타필로코커스 아우레우스(S. 아우레우스) TadA, 바실러스 서브틸리스(B. 서브틸리스) TadA, 살모넬라 티피무리움(S. 티피무리움) TadA, 쉬와넬라 푸트레파시엔스(S. 푸트레파시엔스) TadA, 헤모필루스 인플루엔자에 F3031(H. 인플루엔자에) TadA, 카울로박터 크레센투스(C. 크레센투스) TadA, 지오박터 설퍼레듀센스(G. 설퍼레듀센스) TadA로부터 선택되는 아데노신 데아미네이스 도메인 또는 TadA*7.10을 포함한다.

[1401] 일부 실시형태에서, 아데노신 데아미네이스는 TadA*8이다. 일 실시형태에서, 아데노신 데아미네이스는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 이로 본질적으로 구성되는 TadA*8이다:

[1402] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[1403] 일부 실시형태에서, TadA*8은 절단된다. 일부 실시형태에서, 절단된 TadA*8은 전장 TadA*8에 비해 1개, 2개, 3개, 4개, 5 개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개 또는 20개의 N-말단 아미노산 잔기가 누락되어 있다. 일부 실시형태에서, 절단된 TadA*8은 전장 TadA*8에 비해 1개, 2개, 3개, 4개, 5 개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개 또는 20개의 C-말단 아미노산 잔기가 누락되어 있다. 일부 실시형태에서, 아데노신 데아미네이스 변이체는 전장 TadA*8이다.

[1404] 일부 실시형태에서, TadA*8은 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23 또는 TadA*8.24이다.

[1405] 다른 실시형태에서, 아데노신 데아미네이스 변이체(예를 들어, TadA*8) 단량체를 포함하는 본 개시내용의 염기 편집기는 TadA 참조 서열인 TadA*7.10에 대해 다음 변경: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I 및/또는 D167N 중 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이를 포함한다. 다른 실시형태에서, 아데노신 데아미네이스 변이체(TadA*8) 단량체는 TadA 참조 서열인 TadA*7.10에 대해 R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; 및 A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N로 이루어진 군으로부터 선택되는 변경의 조합, 또는 또 다른 TadA의 상응하는 돌연변이를 포함한다.

[1406] 다른 실시형태에서, 염기 편집기는 야생형 아데노신 데아미네이스 도메인과 TadA 참조 서열인 TadA*7.10에 대해 다음 변경 R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I 및/또는 D167N 중 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체를 포함한다. 다른 실시형태에서, 염기 편집기는 야생형 아데노신 데아미네이스 도메인 및 TadA 참조 서열인 TadA*7.10에 대해 R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; 및 A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N로 이루어진 군으로부터 선택되는 변경의 조합, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체를 포함한다.

[1407] 다른 실시형태에서, 염기 편집기는 TadA*7.10 도메인 및 TadA 참조 서열인 TadA*7.10에 대해 다음 변경 R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I 및/또는 D167N 중 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체를 포함한다. 다른 실시형태에서, 염기 편집기는 TadA*7.10 도메인 및 TadA 참조 서열인 TadA*7.10에 대해 R26C + A109S

+ T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; 및 A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N으로 이루어진 군으로부터 선택되는 변경의 조합, 또는 또 다른 TadA의 상응하는 돌연변이를 포함하는 아데노신 데아미네이스 변이체 도메인(예를 들어, TadA*8)의 이종이량체를 포함한다.

[1408] 일부 실시형태에서, TadA*8은 표 11에 나타나 있는 변이체이다. 표 11은 TadA 아미노산 서열의 소정의 아미노산 위치 번호 및 TadA-7.10 아데노신 데아미네이스의 해당 위치에 존재하는 아미노산을 보여준다. 표 11은 또한 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[M. Richter *et al.*, 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z]에 기술되어 있는 바와 같이 파지-보조 비-연속 진화(phage-assisted non-continuous evolution: PANCE) 및 파지-보조 연속 진화(phage-assisted continuous evolution: PACE) 후 TadA-7.10에 대한 TadA 변이체의 아미노산 변화를 보여준다. 일부 실시형태에서, TadA*8은 TadA*8a, TadA*8b, TadA*8c, TadA*8d 또는 TadA*8e이다. 일부 실시형태에서, TadA*8은 TadA*8e이다.

*표 11*

**추가적인 TadA*8 변이체**

| | | TadA 아미노산 번호 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | TadA | 26 | 88 | 109 | 111 | 119 | 122 | 147 | 149 | 166 | 167 |
| | TadA-7.10 | R | V | A | T | D | H | Y | F | T | D |
| PANCE 1 | | | | | R | | | | | | |
| PANCE 2 | | | | S/T | R | | | | | | |
| | TadA-8a | C | | S | R | N | N | D | Y | I | N |
| | TadA-8b | | A | S | R | N | N | | Y | I | N |
| PACE | TadA-8c | C | | S | R | N | N | | Y | I | N |
| | TadA-8d | | A | | R | N | | | Y | | |
| | TadA-8e | | | S | R | N | N | D | Y | I | N |

[1409]

[1410] 일 실시형태에서, 본 발명의 융합 단백질은 Cas9 닉케이스에 연결된 본 명세서에 기재된 아데노신 데아미네이스 변이체(예를 들어, TadA*8)에 연결된 야생형 TadA를 포함한다. 특정 실시형태에서, 융합 단백질은 단일 TadA*8 도메인(예를 들어, 단량체로 제공됨)을 포함한다. 다른 실시형태에서, 융합 단백질은 이종이량체를 형성할 수 있는 TadA*8 및 TadA(wt)를 포함한다.

[1411] 일부 실시형태에서, 아데노신 데아미네이스는 본 명세서에 제공된 임의의 아데노신 데아미네이스에 기재된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 본 명세서에 제공된 아데노신 데아미네이스는 하나 이상의 돌연변이(예를 들어, 본 명세서에 제공된 임의의 돌연변이)를 포함할 수 있음을 이해하여야 한다. 본 개시내용은 소정의 동일성 백분율을 갖는 임의의 데아미네이스 도메인과 본 명세서에 기재된 임의의 돌연변이 또는 이들의 조합을 제공한다. 일부 실시형태에서, 아데노신 데아미네이스는 참조 서열 또는 본 명세서에 제공된 임의의 아데노신 데아미네이스와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, 아데노신 데아미네이스는 당업계에 공지되거나 또는 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 5개, 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개, 적어도 30개, 적어도 35개, 적어도 40개, 적어도 45개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 110개, 적어도 120개, 적어도 130개, 적어도 140개, 적어도 150

개, 적어도 160개 또는 적어도 170개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[1412] 특정 실시형태에서, TadA*8은 굵게 표시된 임의의 위치에 하나 이상의 돌연변이를 포함한다. 다른 실시형태에서, TadA*8은 밑줄로 표시된 임의의 위치에 하나 이상의 돌연변이를 포함한다:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG 50
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG 100
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR 150

[1413] MPRQVFNAQK KAQSSTD

[1414] 예를 들어, TadA*8은 TadA 참조 서열인 TadA*7.10에 대해 82번 및/또는 166번 아미노산 위치의 변경(예를 들어, V82S, T166R)을 단독으로 또는 다음 Y147T, Y147R, Q154S, Y123H 및/또는 Q154R 중 임의의 하나 이상, 또는 또 다른 TadA의 상응하는 돌연변이와 조합하여 포함한다. 특정 실시형태에서, 변경의 조합은 TadA 참조 서열인 TadA*7.10에 대해 Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R, 또는 또 다른 TadA의 상응하는 돌연변이로 이루어진 군으로부터 선택된다.

[1415] 일부 실시형태에서, 아데노신 데아미네이스는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 TadA*8이다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[1416]

[1417] 일부 실시형태에서, TadA*8은 절단된다. 일부 실시형태에서, 절단된 TadA*8은 전장 TadA*8에 비해 1개, 2개, 3개, 4개, 5 개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개 또는 20개의 N-말단 아미노산 잔기가 누락되어 있다. 일부 실시형태에서, 절단된 TadA*8은 전장 TadA*8에 비해 1개, 2개, 3개, 4개, 5 개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개 또는 20개의 C-말단 아미노산 잔기가 누락되어 있다. 일부 실시형태에서, 아데노신 데아미네이스 변이체는 전장 TadA*8이다.

[1418] 일 실시형태에서, 본 발명의 융합 단백질은 Cas9 닉케이스에 연결된 본 명세서에 기재된 아데노신 데아미네이스 변이체(예를 들어, TadA*8)에 연결된 야생형 TadA를 포함한다. 특정 실시형태에서, 융합 단백질은 단일 TadA*8 도메인(예를 들어, 단량체로 제공됨)을 포함한다. 다른 실시형태에서, 염기 편집기는 이종이량체를 형성할 수 있는 TadA*8 및 TadA(wt)를 포함한다.

[1419] 특정 실시형태에서, 융합 단백질은 단일(예를 들어, 단량체로 제공됨) TadA*8을 포함한다. 일부 실시형태에서, TadA*8은 Cas9 닉케이스에 연결된다. 일부 실시형태에서, 본 발명의 융합 단백질은 TadA*8에 연결된 야생형 TadA(TadA(wt))의 이종이량체로 구성된다. 다른 실시형태에서, 본 발명의 융합 단백질은 TadA*8에 연결된 TadA*7.10의 이종이량체로 구성된다. 일부 실시형태에서, 염기 편집기는TadA*8 변이체 단량체를 포함하는 ABE8이다. 일부 실시형태에서, 염기 편집기는 TadA*8과 TadA(wt)의 이종이량체를 포함하는 ABE8이다. 일부 실시형태에서, 염기 편집기는 TadA*8과 TadA*7.10의 이종이량체를 포함하는 ABE8이다. 일부 실시형태에서, 염기 편집기는 TadA*8의 이종이량체를 포함하는 ABE8이다. 일부 실시형태에서, TadA*8은 표 11, 표 13 또는 표 14로부터 선택된다. 일부 실시형태에서, ABE8은 표 13, 표 14 또는 표16으로부터 선택된다.

[1420] 일부 실시형태에서, 아데노신 데아미네이스는 TadA*9 변이체이다. 일부 실시형태에서, 아데노신 데아미네이스는

다음 서열(TadA*7.10이라 함)을 참조하여 아래에 기재된 변이체로부터 선택되는 TadA*9 변이체이다:

```
         10         20         30         40         50
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
         60         70         80         90        100
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
        110        120        130        140        150
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
        160
MPRQVFNAQK KAQSSTD
```

[1421]

[1422] 일부 실시형태에서, 아데노신 데아미네이스는 다음 변경 중 하나 이상을 포함한다: R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R 및 A158K. 하나 이상의 대안이 밑줄 및 굵은 글자로 위의 순서대로 표시된다.

[1423] 일부 실시형태에서, 아데노신 데아미네이스는 다음 조합의 변경 중 하나 이상을 포함한다: V82S + Q154R + Y147R; V82S + Q154R + Y123H; V82S + Q154R + Y147R+ Y123H; Q154R + Y147R + Y123H + I76Y+ V82S; V82S + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; Q154R + Y147R + Y123H + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; V82S + Q154R + Y147R; V82S + Q154R + Y147R; Q154R + Y147R + Y123H + I76Y; Q154R + Y147R + Y123H + I76Y + V82S; I76Y_V82S_Y123H_Y147R_Q154R; Y147R + Q154R + H123H; 및 V82S + Q154R.

[1424] 일부 실시형태에서, 아데노신 데아미네이스는 다음 조합의 변경 중 하나 이상을 포함한다: E25F + V82S + Y123H, T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; Q71M + V82S + Y123H + Y147R + Q154R; E25F + V82S + Y123H + T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; V82S + Y123H + P124W + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; R23H + V82S + Y123H + Y147R + Q154R; R21N + V82S + Y123H + Y147R + Q154R; V82S + Y123H + Y147R + Q154R + A158K; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; 및 M70V + V82S + M94V + Y123H + Y147R + Q154R.

[1425] 일부 실시형태에서, 아데노신 데아미네이스는 다음 조합의 변경 중 하나 이상을 포함한다: Q71M + V82S + Y123H + Y147R + Q154R; E25F + I76Y+ V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; E25F + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; 및 V82S + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; M70V +V82S + M94V + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; 및 M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R. 일부 실시형태에서, 아데노신 데아미네이스는 단량체로 발현된다. 다른 실시형태에서, 아데노신 데아미네이스는 이종이량체로 발현된다. 일부 실시형태에서, 데아미네이스 또는 다른 폴리펩타이드 서열은, 예를 들어, 융합 단백질의 성분으로 포함될 때 메티오닌이 결여되어 있다. 이는 위치의 번호를 변경할 수 있다. 그러나, 당업자는 이러한 상응하는 돌연변이가 동일한 돌연변이, 예를 들어, Y73S 및 Y72S와 D139M 및 D138M을 지칭한다는 것을 이해할 것이다.

[1426] 일부 실시형태에서, TadA*9 변이체는 본 명세서에 기재된 바와 같은 표 17에 기재되어 있는 변경을 포함한다.

일부 실시형태에서, TadA*9 변이체는 단량체이다. 일부 실시형태에서, TadA*9 변이체는 야생형 TadA 아데노신 데아미네이스를 갖는 이종이량체이다. 일부 실시형태에서, TadA*9 변이체는 또 다른 TadA 변이체(예를 들어, TadA*8, TadA*9)를 갖는 이종이량체이다. TadA*9 아데노신 데아미네이스에 대한 추가적인 세부사항은 그 전문이 참조에 의해 본 명세서에 원용되어 있는 PCT 출원 PCT/2020/049975에 기술되어 있다.

[1427] 본 명세서에 제공된 임의의 돌연변이 및 임의의 추가적인 돌연변이(예를 들어, ecTadA 아미노산 서열에 기초함)는 임의의 다른 아데노신 데아미네이스에 도입될 수 있다. 본 명세서에 제공된 임의의 돌연변이는 TadA 참조 서열 또는 또 다른 아데노신 데아미네이스(예를 들어, ecTadA)에서 개별적으로 또는 임의의 조합으로 만들어질 수 있다.

[1428] A에서 G로의 핵염기 편집 단백질에 대한 세부사항은 PCT 출원 PCT/2017/045381(WO2018/027078) 및 문헌 [Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature, 551, 464-471 (2017)]에 기술되어 있으며, 이들의 전체 내용은 참조에 의해 본 명세서에 원용된다.

[1429] ***C에서 T로의 편집***

[1430] 일부 실시형태에서, 본 명세서에 개시된 염기 편집기는 탈아미노화 폴리뉴클레오타이드의 표적 사이티딘(C) 염기를 탈아미노화하여 티민의 염기 페어링 특성을 갖는 우리딘(U)을 생성할 수 있는 사이티딘 데아미네이스를 포함하는 융합 단백질을 포함한다. 그런 다음, 일부 실시형태에서, 예를 들어, 폴리뉴클레오타이드가 이중-가닥(예를 들어, DNA)인 경우, 우리딘 염기는 C:G에서 T:A로의 전이를 일으키기 위해 티미딘 염기(예를 들어, 세포 복구 기구에 의해)로 치환될 수 있다. 다른 실시형태에서, 염기 편집기에 의한 핵산에서 C에서 U로의 탈아미노화는 U에서 T로의 치환을 동반할 수 없다.

[1431] U를 생성하기 위한 폴리뉴클레오타이드에서 표적 C의 탈아미노화는 본 명세서에 기재된 염기 편집기에 의해 실행될 수 있는 염기 편집의 유형의 비제한적인 예이다. 또 다른 예에서, 사이티딘 데아미네이스 도메인을 포함하는 염기 편집기는 사이토신(C) 염기에서 구아닌(G) 염기로의 전환을 매개할 수 있다. 예를 들어, 사이티딘 데아미네이스 염기 편집기의 도메인에 의한 사이티딘의 탈아미노화에 의해 생성된 폴리뉴클레오타이드의 U는 염기 절단 복구 메커니즘(예를 들어, 우라실 DNA 글리코실레이스(UDG) 도메인에 의해)에 의해 폴리뉴클레오타이드로부터 절단되어 무염기성 부위를 생성할 수 있다. 그런 다음, 무염기성 부위 반대편의 핵염기는, 예를 들어, 전이 중합효소에 의해 C와 같은 또 다른 염기로 치환(예를 들어, 염기 복구 기구에 의해)될 수 있다. 무염기성 부위 반대편의 핵염기가 C로 대체되는 것이 전형적이지만, 다른 치환(예를 들어, A, G 또는 T)도 또한 발생할 수 있다.

[1432] 따라서, 일부 실시형태에서, 본 명세서에 기재된 염기 편집기는 폴리뉴클레오타이드에서 표적 C을 U로 탈아미노화할 수 있는 탈아미노화 도메인(예를 들어, 사이티딘 데아미네이스 도메인)을 포함한다. 또한, 아래에 기재된 바와 같이, 염기 편집기는 일부 실시형태에서 T 또는 G로의 탈아미노화로 인한 U의 전환을 촉진하는 추가적인 도메인을 포함할 수 있다. 예를 들어, 사이티딘 데아미네이스 도메인을 포함하는 염기 편집기는 U의 T로의 치환을 매개하기 위해 우라실 글리코실레이스 저해제(UGI) 도메인을 추가로 포함하여, C에서 T로의 염기 편집 이벤트를 완료할 수 있다. 또 다른 예에서, 염기 편집기는 전이 중합효소가 무염기성 부위 반대편에 있는 C의 통합을 촉진할 수 있기 때문에(즉, 무염기성 부위에서 G의 혼입을 일으켜 C에서 G로의 염기 편집 이벤트를 완성함), 전이 중합효소를 혼입하여 C에서 G로의 염기 편집의 효율성을 개선시킬 수 있다.

[1433] 도메인으로서 사이티딘 데아미네이스를 포함하는 염기 편집기는 DNA, RNA 및 DNA-RNA 하이브리드를 포함하는 임의의 폴리뉴클레오타이드에서 표적 C를 탈아미노화할 수 있다. 전형적으로, 사이티딘 데아미네이스는 폴리뉴클레오타이드의 단일-가닥 부분의 맥락에서 위치하는 C 핵염기를 촉매한다. 일부 실시형태에서, 표적 C를 포함하는 전체 폴리뉴클레오타이드는 단일-가닥일 수 있다. 예를 들어, 염기 편집기 내로 혼입된 사이티딘 데아미네이스는 단일-가닥 RNA 폴리뉴클레오타이드에서 표적 C를 탈아미노화할 수 있다. 다른 실시형태에서, 사이티딘 데아미네이스 도메인을 포함하는 염기 편집기는 이중-가닥 폴리뉴클레오타이드에 작용할 수 있지만, 표적 C는 탈아미노화 반응시 단일-가닥 상태인 폴리뉴클레오타이드의 일부에 위치할 수 있다. 예를 들어, NAGPB 도메인이 Cas9 도메인을 포함하는 실시형태에서, Cas9-gRNA-표적 DNA 복합체가 형성되는 동안 여러 뉴클레오타이드가 페어링되지 않은 상태로 남아, Cas9 "R-루프 복합체"가 형성될 수 있다. 이러한 페어링되지 않은 뉴클레오타이드는 단일-가닥 특이적 뉴클레오타이드 데아미네이스 효소(예를 들어, 사이티딘 데아미네이스)에 대한 기질 역할

을 할 수 있는 단일-가닥 DNA의 기포를 형성할 수 있다.

[1434] 일부 실시형태에서, 염기 편집기의 사이티딘 데아미네이스는 아포지단백질 B mRNA 편집 복합체(apolipoprotein B mRNA editing complex: APOBEC) 패밀리 데아미네이스의 전부 또는 일부를 포함할 수 있다. APOBEC는 진화적으로 보존된 사이티딘 데아미네이스의 패밀리이다. 이 패밀리의 구성원은 C에서 U로의 편집 효소이다. APOBEC 유사 단백질의 N-말단 도메인은 촉매 도메인인 반면, C-말단 도메인은 유사촉매 도메인이다. 보다 구체적으로, 촉매 도메인은 아연 의존성 사이티딘 데아미네이스 도메인이며, 사이티딘 탈아미노화에 중요하다. APOBEC 패밀리 구성원은 APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D("APOBEC3E"는 이제 이것을 지칭함), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 및 활성화-유도성(사이티딘) 데아미네이스를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC1 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC2 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3A 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3B 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3C 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3D 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3E 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3F 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3G 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC3H 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC4 데아미네이스의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 활성화-유도성 데아미네이스(activation-induced deaminase: AID)의 전부 또는 일부를 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 사이티딘 데아미네이스 1(cytidine deaminase 1: CDA1)의 전부 또는 일부를 포함한다. 염기 편집기는 임의의 적합한 유기체(예를 들어, 인간 또는 래트)로부터의 데아미네이스를 포함할 수 있음을 이해하여야 한다. 일부 실시형태에서, 데아미네이스 염기 편집기의 도메인은 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스로부터 유래한다. 일부 실시형태에서, 데아미네이스 염기 편집기의 도메인은 래트로부터 유래한다(예를 들어, 래트 APOBEC1). 일부 실시형태에서, 데아미네이스 염기 편집기의 도메인은 인간 APOBEC1이다. 일부 실시형태에서, 데아미네이스 염기 편집기의 도메인은 pmCDA1이다.

[1435] PmCDA1의 아미노산 및 핵산 서열은 본 명세서의 아래에 나타나 있다.

[1436] >tr|A5H718|A5H718_PETMA 사이토신 데아미네이스 OS=페트로마이존 마리누스 OX=7757 PE=2 SV=1 아미노산 서열:

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERG
IHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYY
EKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELS

[1437] IMIQVKILHTTKSPAV

[1438] 핵산 서열: >EF094822.1 페트로마이존 마리누스 단리물 PmCDA.21 사이토신 데아미네이스 mRNA, 완전 cds:

TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAG
AGAGGACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACG
CTGAGTACGTGAGAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAA
CAAAAAATCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTGAACGTAGA
GCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGCGGGACAGAACGTGGAATTCACGCCG
AAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACAATTCACGATAAA
TTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAGAATGGTATAACCAG
GAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATG
CGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGGGTTGAATGTAATGGTAAG
TGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATTGAATGAGAAT
AGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATGATTC
AGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGTTTTC

[1439]

[1440] 인간 활성화-유도성 사이티딘 데아미네이스(AID)의 코딩 서열(CDS)의 아미노산 및 핵산 서열이 아래에 나타나 있다.

[1441] >tr|Q6QJ80|Q6QJ80_인간 활성화-유도성 사이티딘 데아미네이스 OS=호모 사피엔스 OX=9606 GN=AICDA PE=2 SV=1 아미노산 서열:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL
FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK
AEPEGLRRLHRAGVQIAIMTFKAPV

[1442]

[1443] 인간 활성화-유도성 사이티딘 데아미네이스(AID)의 코딩 서열(CDS)의 아미노산 및 핵산 서열이 아래에 나타나 있다.

[1444] >tr|Q6QJ80|Q6QJ80_인간 활성화-유도성 사이티딘 데아미네이스 OS=호모 사피엔스 OX=9606 GN=AICDA PE=2 SV=1 아미노산 서열:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL
FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK
AEPEGLRRLHRAGVQIAIMTFKAPV

[1445]

[1446] 핵산 서열: >NG_011588.1:5001-15681 호모 사피엔스 활성화 유도성 사이티딘 데아미네이스(AICDA), 염색체 12의 RefSeqGene(LRG_17):

AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGAC
ACTCTGGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTC
CTCTCAGAGCAAATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGAC
TGATAAGATCAGCTTGATCAATATGCATATATATTTTTTGATCTGTCTCCTTTTCTTCTATTCA
GATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCA
TGTGGCAAAAGAAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTACCATGGTTGAAA
CTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATCATGAAAAGGTAATA
GCAGTACTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTTGTAAATATTCAACAGTAA
AACAACTTGAAGACACACTTTCCTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAA
TTTGTAATTTTAGAAATGCCAAGCATTCTAAATTAATTGCTTGAAAGTCACTATGATTGTGTCC
ATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAAGGCCCAATTGTTAGGCAGT
TAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTACCTCTTA
GGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCAT
AGAACTTATTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTA
TACTTTACCTCAATTTAACTCCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAA
TATATTTTTTTGTAGAGACAGGGTCTTAGCCCAGCCGAGGCTGGTCTCTAAGTCCTGGCCCAAG
CGATCCTCCTGCCTGGGCCTCCTAAAGTGCTGGAATTATAGACATGAGCCATCACATCCAATAT
ACAGAATAAAGATTTTTAATGGAGGATTTAATGTTCTTCAGAAAATTTTCTTGAGGTCAGACAA

[1447]

TGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTG
AAGGGTCCATTGGAAATACTTGTTCAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAA
ATTCAGAGAAAGACAGAAAAGGAGAAAAGATGAAATTCAACAGGACAGAAGGGAAATATATTAT
CATTAAGGAGGACAGTATCTGTAGAGCTCATTAGTGATGGCAAAATGACTTGGTCAGGATTATT
TTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGGTTCTGCCTCAGGGAGCA
CAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCTAC
TCAGGACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAAT
GGATCAACAAAGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATG
TGACAGAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAG
TCCTTTATCTATGCCACATCCTTCTTATCTATACTTCCAGGACACTTTTTCTTCCTTATGATAA
GGCTCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACACACACAAAC
ACACACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCCTCTGCCTTTCTCATCTACAC
AGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTA
ACACTGGGCCTCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAAGCACCTATTATGTGTTGA
GCTTATATATACAAAGGGTTATTATATGCTAATATAGTAATAGTAATGGTGGTTGGTACTATGG
TAATTACCATAAAAATTATTATCCTTTTAAAATAAAGCTAATTATTATTGGATCTTTTTTAGTA
TTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCTGTTACCCAGGCTG
GAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTG
CCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTA
GATTAAAATATGCATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTAT
ACTGGCAATAAATCTAGTTTGCTGCCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGT
GAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTGTGGTGGCTCACGCCTGTAATCCCAGCA
CTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCTGAGCAACAT
GGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCAGC
TACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGGTCAAGGCTGCACTGAGCAGTGC
TTGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCAAAAAAATAAGAAGAAA
AATTAAAAATAAATGGAAACAACTACAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCT
GATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTACATTATTAAAATATCA
ATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACCTTTATTCACAAAA
CCCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATAT
CCATGCAATGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATG
ACGCTAAATGAAAGAGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAA
TGAAAGTAACTTATAGTTACAGAAAGCAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCC
CAGCACTTTGAGAGGCCACGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCA
ACACAGTGAAACTCCATTCTCCACAAAAATGGGAAAAAAAGAAAGCAAATCAGTGGTTGTCCTG

[1448]

TGGGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATTCAGGTT
CTGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGCA
TCTTAAATGGGTGGAGTTTACTGTATGTAAATTATACCTCAATGTAAGAAAAAATAATGTGTAA
GAAAACTTTCAATTCTCTTGCCAGCAAACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAAT
TCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGGATAAATGCAT
TTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCATGCTGTAC
AGCTTGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTAC
CAGAGTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAG
CCAGAAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGAT
CATTTTGACTAGTTAAAAAAGCAGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAAAT
CATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACACAAGAATCTTAACAGTCATTGTC
ATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTTA
ATATTTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTATTCAAGTTTGACAATGATCAAGTAT
TACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTTGGTCTTGTTGCCCATGC
TGGAGTGGAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAGCTGT
CGCCTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCCACCACCACACTCGGCTAATGTTTGTA
TTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAGG
ATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGCCCGGCCAAG
TATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCCAGCCAGGTATTGCTCTTATA
CATTAAAAAATAGGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGC
GGGCAGAACACCCGAGGTCAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCT
ATTAAAAATACAAACATTACCTGGGCATGATGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGG
CTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTGCCTGAGCCTGGGAGGTTGAGGCTACAGT
AAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCGACAAAGTGAGACCGTAACAAAAAAAAAA
AAATTTAAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAAAGTGGCCTAAACACCACATTAA
AGAGTTTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCTTCAGCATGGGAATGGC
ATGGTGCACCTGGTTTTTGTGAGATCATGGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGA
CTGGAGGCAGACAGACCGGTTAAAAGGCCAGCACAACAGATAAGGAGGAAGAAGATGAGGGCTT
GGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAAT
CAACACATTTAGATGATTAATTAAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATGGTT
CCAGGCTGCTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGG
GCAGTTGAGGAATATTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAG
ACTGGAGGGGAAATCTGAATATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTGTT
TCGTTTTCTTGTTGAAGAACAAATTTAATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTG
GCAGGAGGTGACTGTCTTGTGGGTAAGGGTTTGGGGTCCTTGATGAGTATCTCTCAATTGGCCT

[1449]

TAAATATAAGCAGGAAAAGGAGTTTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAGGGCTCA
GGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCCAAGTAATGACTTCCTTAAAA
AGCTGAAGGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTC
TCACCCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGC
TAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCTTT
TCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTTAA
TGGTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTTGGCAT
TTGTGTCTCTATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCAT
ATTAGACATGGCCCAAAATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCA
CTCCTTCCTTCAGTGCCAAGAACAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGA
ATTGCCTTTGAGATTAATTAAGCTAAAAGCATTTTTATATGGGAGAATATTATCAGCTTGTCCA
AGCAAAAATTTTAAATGTGAAAAACAAATTGTGTCTTAAGCATTTTTGAAAATTAAGGAAGAAG
AATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCCAAATGATTTCTTTTCCCTCCTACT
CACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCAGAAAACTCAGAGAAGCCT
CGGCTGATGATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATTTCCAAGAGACTTC
TTCACCAAAATCCAGATGGGTTTACATAAACTTCTGCCCACGGGTATCTCCTCTCTCCTAACAC
GCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAAGCATCCGTGGGGTGGAAGGTCATCGTCTG
GCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACATTTGTATTGAATACATC
CCAATCTCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCAAGC
ACTTTCATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTA
CCAAAATCTATTTCCCCTTTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCCACC
TAGTGGTCCAGGTATATTTCCACAATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCA
AAAGGTGCAAAAAGCAACTTCATAAACACAAATTAAATCTTCGGTGAGGTAGTGTGATGCTGCT
TCCTCCCAACTCAGCGCACTTCGTCTTCCTCATTCCACAAAAACCCATAGCCTTCCTTCACTCT
GCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGCTT
AGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATG
ACTAAGGCTACCAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTCTGTCTCTCCA
GAACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGC
CGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGG
CCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTG
TGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCC
ATCATGACCTTCAAAGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTCGG
GATTGCGATGCGGAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCT
GGTTCACCTCTGGAGCCGAAATTAAAGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACA
AGGCCCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCTTCTTTCCCCTCGATTTGGAACCTGAA

[1450]

CTGTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTTTTTTTTTTTTTGAAGATTATTTTTACT
GCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTCAAAGCCTGGGAAGGGCTGCATGAAAA
TTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCTTCCTCGCTTTTTAA
ATTTTCTTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTTATTGTT
CAATCACTCTCAGTTTTCATCTGATGAAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTCTG
CTGTTTCACCATTCAGAGCCCTCTGCTAAGGTTCCTTTTCCCTCCCTTTTCTTTCTTTTGTTGT
TTCACATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTCAGAATTCTTTTC
TCCTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCAAAAAAACTCTTT
CCCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCC
CACCGACCCCCAACCTCGTTTTGAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCC
CTGTATGAGGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTTTGATAGCAACTTCCAG
GAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAAGTCT
TCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTTATATACGAC
TCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAGGAACACAGGTCTGGCCAGGGACGTGCTGC
AATTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACTGCAGGACCTGGG
AGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGTAGATCC
TAAAAAGCATGGTGAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATTTGATTCATTTG
AGTTAACAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAA
CACAAACTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTG
ACCCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAA
GCATGTTTTTATGTTTGTACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCA
TGGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACA
CCCTAATAATGGGTTGATGTCTGAAGTAGCAAATCTTCTGGAAACGCAAACTCTTTTAAGGAAG
TCCCTAATTTAGAAACACCCACAAACTTCACATATCATAATTAGCAAACAATTGGAAGGAAGTT
GCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCA
ATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAACTAT
ATAAGAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTA
GCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAA
CATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGG
TGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGA
GTTTGAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCA
GGCGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAA
CCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGA
GCAAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAAAGAGAACAATATTTGG
[1451] GAGAGAAGGATGGGGAAGCATTGCAAGGAAATTGTGCTTTATCCAACAAAATGTAAGGAGCCAA

TAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAACAACTGTCTTTGACAGTGAG
AAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCTTGCAATGAAGATG
AGCAGATCCACAGGAAAACTTGAATGCACAACTGTCTTATTTTAATCTTATTGTACATAAGTTT
GTAAAAGAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATTATTTTGCGTCTAAT
GATTTTTTATTAACATGATTTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAAAT
TTATAACTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAATGGTGCTA
CGAAGCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTT
CAATCAGTTAAATAAATGATAAATAATTTTGGAAGCTGTGAAGATAAAATACCAAATAAAATAA
[1452] TATAAAAGTGATTTATATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

[1453] 본 개시내용의 양태에 따라 Cas9에 융합될 수 있는 다른 예시적인 데아미네이스가 아래에 제공된다. 실시형태에서, 데아미네이스는 활성화-유도성 데아미네이스(AID)이다. 일부 실시형태에서, 각 서열의 활성 도메인은, 예를 들어, 국부화 신호가 없는 도메인(핵 방출 신호, 세포질 국부화 신호가 없는 핵 국부화 서열)이 사용될 수 있음을 이해하여야 한다.

[1454] 인간 AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRY
ISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRR
LHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDA
[1455] FRTLGL

[1456] (밑줄: 핵 국부화 서열; 이중 밑줄: 핵 방출 신호)

[1457] 마우스 AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRY
ISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRR
LHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDA
[1458] FRMLGF

[1459] (밑줄: 핵 국부화 서열; 이중 밑줄: 핵 방출 신호)

[1460] 개 AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRY
ISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRR
LHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDA
[1461] FRTLGL

[1462] (밑줄: 핵 국부화 서열; 이중 밑줄: 핵 방출 신호)

[1463] 소 AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRY
ISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLR
RLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRD
[1464] AFRTLGL

[1465] (밑줄: 핵 국부화 서열; 이중 밑줄: 핵 방출 신호)

[1466] 래트 AID:

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKF
LYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLDPGRC
YRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTF
[1467] VENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL

[1468] (밑줄: 핵 국부화 서열; 이중 밑줄: 핵 방출 신호)

[1469] clAID(카니스 루프스 파밀리아리스(*Canis lupus familiaris*)):

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDL
DPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMT
[1470] FKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1471] btAID(보스 타우루스(*Bos Taurus*)):

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDL
DPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIM
[1472] TFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1473] mAID(무스 무스쿨루스(*Mus musculus*)):

[1474]
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDL
DPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMT
FKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[1475] rAPOBEC-1(라투스 노르베기쿠스(*Rattus norvegicus*)):

[1476]
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVT
IQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIAL
QSCHYQRLPPHILWATGLK

[1477] maAPOBEC-1(메소크리세투스 아우라투스(*Mesocricetus auratus*)):

[1478]
MSSETGPVVVDPTLRRRIEPHEFDAFFDQGELRKETCLLYEIRWGGRHNIWRHTGQNTSRHVEINFIEKF
TSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVTLFIYAARLYHHTDQRNRQGLRDLISRGVT
IRIMTEQEYCYCWRNFVNYPPSNEVYWPRYPNLWMRLYALELYCIHLGLPPCLKIKRRHQYPLTFFRLNL
QSCHYQRIPPHILWATGFI

[1479] ppAPOBEC-1(폰고 피그마에우스(*Pongo pygmaeus*)):

[1480]
MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKF
TSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVT
IQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHL
QNCHYQTIPPHILLATGLIHPSVTWR

[1481] ocAPOBEC1(오릭토라구스 쿠니쿨루스(*Oryctolagus cuniculus*)):

[1482]
MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKNTTNHVEVNFLEKL
TSEGRLGPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVTLIIFVARLFQHMDRRNRQGLKDLVTSGVT
VRVMSVSEYCYCWENFVNYPPGKAAQWPRYPPRWMLMYALELYCIILGLPPCLKISRRHQKQLTFFSLTP
QYCHYKMIPPYILLATGLLQPSVPWR

[1483] mdAPOBEC-1(모노델피스 도메스티카(*Monodelphis domestica*)):

[1484]
MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELRKETCLLYEIKWGNQNIWRHSNQNTSQHAEINFMEKFT
AERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVTLAIFISRLYWHMDQQHRQGLKELVHSGVTI
QIMSYSEYHYCWRNFVDYPQGEEDYWPKYPYLWIMLYVLELHCIILGLPPCLKISGSHSNQLALFSLDLQ
DCHYQKIPYNVLVATGLVQPFVTWR

[1485] ppAPOBEC-2(폰고 피그마에우스):

[1486]
MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFL
CYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLS
KTKNLRLLILVGRLFMWEELEIQDALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQE
NFLYYEEKLADILK

[1487] btAPOBEC-2(보스 타우루스):

[1488]
MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTFL
CYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLN
KTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQE
NFLYYEEKLADILK

[1489] mAPOBEC-3-(1)(무스 무스쿨루스):

MQPQRLGPRAGMGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSS
RLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRP
CYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQ
FNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYT
SRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKES
WGLQDLVNDFGNLQLGPPMS

[1490]

[1491] 마우스 APOBEC-3-(2):

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNI
*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQDPETQ
QNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSS
TLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCL
LSEKGKQ*HAEILFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPF
QKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFG
NLQLGPPMS

[1492]

[1493] (이탤릭체: 핵산 편집 도메인)

[1494] 래트 APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDN
I*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLATHHNLSLDIFSSRLYNIRDPEN
QQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSS
STLSNICLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGC
LLSEKGKQ*HAEILFLDKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRP
FQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDLVNDF
GNLQLGPPMS

[1495]

[1496] (이탤릭체: 핵산 편집 도메인)

[1497] hAPOBEC-3(호모 사피엔스):

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRH
AELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEAL
QMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

[1498]

[1499] hAPOBEC-3F(호모 사피엔스):

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCF
LSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQA
GARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKA
YGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPC
PECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEP
FKPWKGLKYNFLFLDSKLQEILE

[1500]

[1501] 레서스 마카크(Rhesus macaque) APOBEC-3G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFHKW
RQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHA
TMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHE
TYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPC
FSCAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPF
QPWDGLDEHSQALSGRLRAI

[1502]

[1503] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국부화 신호)

[1504] 침팬지 APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKY*HPEMRF*
*FHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQ
KRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNEL
WVRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVT*
*CFTSWSPCFSC*AQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTF
VDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[1505]

[1506] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국부화 신호)

[1507] 녹색 원숭이 APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKD*HPEMKF*
*LHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQ
ERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKP
WVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTC*
*FTSWSPCFSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFV
DRQGRPFQPWDGLDEHSQALSGRLRAI

[1508]

[1509] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국부화 신호)

[1510] 인간 APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKY*HPEMRF*
*FHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQ
KRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEP
WVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVT*
*CFTSWSPCFSC*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTF
VDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[1511]

[1512] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 국부화 신호)

[1513] 인간 APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEH*HAEMCF*
*LSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQA
GARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKA
YGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETH*CHAERCFLSWFCDDILSPNTNYEVTWYTSWSPC*
*PEC*AGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEP
FKPWKGLKYNFLFLDSKLQEILE

[1514]

[1515] (이탤릭체: 핵산 편집 도메인)

[1516] 인간 APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQY*HAEMC*
*FLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQ
AGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLR
RRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFIS*
*WSPCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVY
[1517] RQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[1518] (이탤릭체: 핵산 편집 도메인)

[1519] 래트 APOBEC-3B:

MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCALP
VPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAI
FSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQE
IFSRMNLLREDVFYLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEK
MRSMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIHVDVM
[1520] DLPQFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL

[1521] 소 APOBEC-3B:

DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAPYY
RRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNF
ITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQY
[1522] SASIRRRLQRILTAPI

[1523] 침팬지 APOBEC-3B:

MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMC
FLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQ
AGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLR
RHQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFIS
WSPCFSWGCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVY
RQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPA
[1524] PSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG

[1525] 인간 APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETH*CHAERC*
*FLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQ
[1526] EGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ

[1527] (이탤릭체: 핵산 편집 도메인)

[1528] 고릴라 APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETH*CHAERC*
*FLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGLRSLSQ
[1529] EGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE

[1530] (이탤릭체: 핵산 편집 도메인)

[1531] 인간 APOBEC-3A:

[1532] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRH
AELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEAL
QMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

[1533] (이탤릭체: 핵산 편집 도메인)

[1534] 레서스 마카크 APOBEC-3A:

[1535] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCGDY
GCHVELRFLCEVPSWQLDPAQTYRVTWFISWSPCFRRGCAGQVRVFLQENKHVRLRIFAARIYDYDPLYQ
EALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN

[1536] (이탤릭체: 핵산 편집 도메인)

[1537] 소 APOBEC-3A:

[1538] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSWNL
DRNQHYRLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARITIMT
FEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN

[1539] (이탤릭체: 핵산 편집 도메인)

[1540] 인간 APOBEC-3H:

[1541] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDE
TQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPK
FADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV

[1542] (이탤릭체: 핵산 편집 도메인)

[1543] 레서스 마카크 APOBEC-3H:

[1544] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDE
TQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPE
FTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR

[1545] 인간 APOBEC-3D:

[1546] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQE
VYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYRD
RDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYP
HIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNT
NYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFV
SCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

[1547] (이탤릭체: 핵산 편집 도메인)

[1548] 인간 APOBEC-1:

[1549] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKF
TSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVT
IQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHL
QNCHYQTIPPHILLATGLIHPSVAWR

[1550] 마우스 APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKF
TTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVT
IQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITL

[1551] QTCHYQRIPPHLLWATGLK

[1552] 래트 APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVT
IQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIAL

[1553] QSCHYQRLPPHILWATGLK

[1554] 인간 APOBEC-2:

MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFL
CYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLS
KTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQE

[1555] NFLYYEEKLADILK

[1556] 마우스 APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTFL
CYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLS
KTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQE

[1557] NFLYYEEKLADILK

[1558] 래트 APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTFL
CYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLS
KTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQE

[1559] NFLYYEEKLADILK

[1560] 소 APOBEC-2:

MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTFL
CYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLN
KTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQE

[1561] NFLYYEEKLADILK

[1562] 페트로마이존 마리누스 CDA1(pmCDA1):

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIF
SIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWN

[1563] LRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV

[1564] 인간 APOBEC3G D316R D317R:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRF
FHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQ
KRDGPRATMKFNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHFMLGEILRHSMDPPTFTFNFNNEPW
VRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTC
FTSWSPCFSCAQEMAKFISKKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDH

[1565] QGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[1566] 인간 APOBEC3G 사슬 A:

[1567]
MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIP
FWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISF
TYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[1568] 인간 APOBEC3G 사슬 A D120R D121R:

[1569]
MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVI
PFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKIS
FMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[1570] hAPOBEC-4(호모 사피엔스):

[1571]
MEPIYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTFYE
LKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYSNNSPCNEANHCCISKMYN
FLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFISGVSGSHV
FQPILTGRALADRHNAYEINAITGVKPYFTDVLLQTKRNPNTKAQEALESYPLNNAFPGQFFQMPSGQLQ
PNLPPDLRAPVVFVLVPLRDLPPMHMGQNPNKPRNIVRHLNMPQMSFQETKDLGRLPTGRSVEIVEITEQ
FASSKEADEKKKKKGKK

[1572] mAPOBEC-4(무스 무스쿨루스):

[1573]
MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDL
DPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMT
FKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[1574] rAPOBEC-4(라투스 노르베기쿠스):

[1575]
MEPLYEEYLTHSGTIVKPYYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYPQTKHLTFYEL
RSSSGNLIQKGLASNCTGSHTHPESMLFERDGYLDSLIFHDSNIRHIILYSNNSPCDEANHCCISKMYNF
LMNYPEVTLSVFFSQLYHTENQFPTSAWNREALRGLASLWPQVTLSAISGGIWQSILETFVSGISEGLTA
VRPFTAGRTLTDRYNAYEINCITEVKPYFTDALHSWQKENQDQKVWAASENQPLHNTTPAQWQPDMSQDC
RTPAVFMLVPYRDLPPIHVNPSPQKPRTVVRHLNTLQLSASKVKALRKSPSGRPVKKEEARKGSTRSQEA
NETNKSKWKKQTLFIKSNICHLLEREQKKIGILSSWSV

[1576] mfAPOBEC-4(마카카 파스시쿨라리스(*Macaca fascicularis*)):

[1577]
MEPTYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTYPQTKHLTFYE
LKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHIILYCNNSPCNEANHCCISKVYN
FLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFVSGVSGSHV
FQPILTGRALTDRYNAYEINAITGVKPFFTDVLLHTKRNPNTKAQMALESYPLNNAFPGQSFQMTSGIPP
DLRAPVVFVLLPLRDLPPMHMGQDPNKPRNIIRHLNMPQMSFQETKDLERLPTRRSVETVEITERFASSK
QAEEKTKKKKGKK

[1578] pmCDA-1(페트로마이존 마리누스):

[1579]
MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILM
SMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLTMHFSRIYDRDREGDHRGLRGLK
HVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIPLHLFTL
QTPLLSGRVVWWRV

[1580] pmCDA-2(페트로마이존 마리누스):

MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHAE
VLLLSAVRAALLRRRRCEDGEEATRGCTLHCYSTYSPCRDCVEYIQEFGASTGVRVVIHCCRLYELDVNR
RRSEAEGVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDE
DLHAQVQRNKQIREAYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGP
[1581] EIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLIVVMRKHNYQGP

[1582] pmCDA-5(페트로마이존 마리누스):

MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILM
SMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLMMHFSRIYDRDREGDHRGLRGLK
[1583] HVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGMPLHLFT

[1584] yCD(사카로마이세스 세레비시아에(*Saccharomyces cerevisiae*)):

MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSATLHGEISTLEN
CGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGENVNFKSKGEKYLQTRGHEVVVVDDERCKKI
[1585] MKQFIDERPQDWFEDIGE

[1586] rAPOBEC-1(델타 177-186):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVT
IQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPH
[1587] ILWATGLK

[1588] rAPOBEC-1(델타 202-213):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVT
IQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQHYQRLPPH
[1589] ILWATGLK

[1590] 마우스 APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVF
KNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHHNLSLDIFS
SRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSK
LQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYY
HRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVTITCYLTWSPCPNC*A
WQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKR
[1591] PFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[1592] (이탤릭체: 핵산 편집 도메인)

[1593] 본 개시내용의 일부 양태는, 예를 들어, 데아미네이스 도메인에서 점 돌연변이를 생성함으로써 본 명세서에 기재된 임의의 융합 단백질의 데아미네이스 도메인 촉매 활성을 조절하는 것이 융합 단백질(예를 들어, 염기 편집기)의 진행성(processivity)에 영향을 미친다는 인식에 기초한다. 예를 들어, 염기 편집 융합 단백질 내의 데아미네이스 도메인의 촉매 활성을 감소시키지만 제거하지는 않는 돌연변이는 데아미네이스 도메인이 표적 잔기에 인접한 잔기의 탈아미노화를 촉매할 가능성을 줄여 탈아미노화 창(window)을 좁힐 수 있다. 탈아미노화 창을 좁힐 수 있는 능력은 특정 표적 잔기에 인접한 잔기의 원치 않는 탈아미노화를 방지할 수 있으며, 이는 표적외 효과를 줄이거나 또는 방지할 수 있다.

[1594] 예를 들어, 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 H121X, H122X,

R126X, R126X, R118X, W90X, W90X 및 R132X로 이루어진 군으로부터 선택되는 하나 이상의 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함할 수 있되, X는 임의의 아미노산이다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 H121R, H122R, R126A, R126E, R118A, W90A, W90Y 및 R132E로 이루어진 군으로부터 선택되는 하나 이상의 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함할 수 있다.

[1595] 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 D316X, D317X, R320X, R320X, R313X, W285X, W285X, R326X로 이루어진 군으로부터 선택되는 하나 이상의 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함할 수 있되, X는 임의의 아미노산이다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 hAPOBEC3G의 D316R, D317R, R320A, R320E, R313A, W285A, W285Y, R326E로 이루어진 군으로부터 선택되는 하나 이상의 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함한다.

[1596] 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 H121R 및 H122R 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 R126A 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 R126E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 R118A 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 W90A 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 W90Y 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 R132E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 W90Y 및 R126E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 R126E 및 R132E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 W90Y 및 R132E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 rAPOBEC1의 W90Y, R126E 및 R132E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다.

[1597] 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 D316R 및 D317R 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 hAPOBEC3G의 R320A 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 R320E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 R313A 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 W285A 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 W285Y 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 R326E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 W285Y 및 R320E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 R320E 및 R326E 돌연변이,

또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 W285Y 및 R326E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다. 일부 실시형태에서, APOBEC 염기 편집기에 혼입된 데아미네이스는 hAPOBEC3G의 W285Y, R320E 및 R326E 돌연변이, 또는 또 다른 APOBEC 데아미네이스의 하나 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미네이스를 포함할 수 있다.

[1598] 애드진으로부터 입수 가능한 SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, VRER-BE3, YE1-BE3, EE-BE3, YE2-BE3 및 YEE-BE3(플라스미드 85169, 85170, 85171, 85172, 85173, 85174, 85175, 85176, 85177)을 포함하지만 이들로 제한되지 않는 다수의 변형된 사이티딘 데아미네이스가 상업적으로 이용 가능하다. 일부 실시형태에서, 염기 편집기에 혼입된 데아미네이스는 APOBEC1 데아미네이스의 전부 또는 일부를 포함한다.

[1599] C에서 T로의 핵염기 편집 단백질에 대한 세부사항은 PCT 출원 PCT/US2016/058344(WO2017/070632) 및 문헌 [Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016)]에 기술되어 있으며, 이들의 전체 내용은 참조에 의해 본 명세서에 원용된다.

[1600] ***사이티딘 데아미네이스***

[1601] 일부 실시형태에서, 본 발명의 융합 단백질은 하나 이상의 사이티딘 데아미네이스 도메인을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 사이티딘 데아미네이스는 사이토신 또는 5-메틸사이토신을 우라실 또는 티민으로 탈아미노화할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 사이티딘 데아미네이스는 DNA에서 사이토신을 탈아미노화할 수 있다. 사이티딘 데아미네이스는 임의의 적합한 유기체로부터 유래될 수 있다. 일부 실시형태에서, 사이티딘 데아미네이스는 본 명세서에 제공된 임의의 돌연변이에 상응하는 하나 이상의 돌연변이를 포함하는 자연-발생적 사이티딘 데아미네이스이다. 당업자는, 예를 들어, 서열 정렬 및 상동 잔기의 결정에 의해 임의의 상동 단백질에서 상응하는 잔기를 확인할 수 있을 것이다. 따라서, 당업자는 본 명세서에 기재된 임의의 돌연변이에 상응하는 임의의 자연-발생적 사이티딘 데아미네이스에 돌연변이를 생성할 수 있을 것이다. 일부 실시형태에서, 사이티딘 데아미네이스는 원핵생물로부터 유래한다. 일부 실시형태에서, 사이티딘 데아미네이스는 세균으로부터 유래한다. 일부 실시형태에서, 사이티딘 데아미네이스는 포유동물(예를 들어, 인간)로부터 유래한다.

[1602] 일부 실시형태에서, 사이티딘 데아미네이스는 본 명세서에 기재된 사이티딘 데아미네이스 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 본 명세서에 제공된 사이티딘 데아미네이스는 하나 이상의 돌연변이(예를 들어, 본 명세서에 제공된 임의의 돌연변이)를 포함할 수 있음을 이해하여야 한다. 일부 실시형태는 임의의 이전의 양태의 또는 본 명세서에 상세히 기술된 바와 같은 사이티딘 데아미네이스 핵염기 편집기 폴리펩타이드를 암호화하는 폴리뉴클레오타이드 분자를 제공한다. 일부 실시형태에서, 폴리뉴클레오타이드는 코돈 최적화된다.

[1603] 본 개시내용은 소정의 동일성 백분율과 본 명세서에 기재된 임의의 돌연변이 또는 이들의 조합을 갖는 임의의 데아미네이스 도메인을 제공한다. 일부 실시형태에서, 사이티딘 데아미네이스는 참조 서열 또는 임의의 본 명세서에 제공된 사이티딘 데아미네이스와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, 사이티딘 데아미네이스는 당업계에 공지되거나 또는 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 5개, 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개, 적어도 30개, 적어도 35개, 적어도 40개, 적어도 45개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 110개, 적어도 120개, 적어도 130개, 적어도 140개, 적어도 150개, 적어도 160개 또는 적어도 170개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[1604] 본 발명의 제2 단백질의 융합 단백질은 2개 이상의 핵산 편집 도메인을 포함한다. 일부 실시형태에서, 핵산 편집 도메인은 C에서 U로의 염기 변화를 촉매할 수 있다. 일부 실시형태에서, 핵산 편집 도메인은 데아미네이스 도메인이다. 일부 실시형태에서, 데아미네이스는 사이티딘 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 아포지단백질 B mRNA-편집 복합체(APOBEC) 패밀리 데아미네이스이다. 일부 실시형태에서, 데아미네이스는

APOBEC1 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC2 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3A 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3B 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3C 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3D 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3E 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3F 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3G 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC3H 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 APOBEC4 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 활성화-유도성 데아미네이스(AID)이다. 일부 실시형태에서, 데아미네이스는 척추동물 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 무척추동물 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 인간 데아미네이스이다. 일부 실시형태에서, 데아미네이스는 래트 데아미네이스, 예를 들어, rAPOBEC1이다. 일부 실시형태에서, 데아미네이스는 페트로마이존 마리누스 사이티딘 데아미네이스 1(pmCDA1)이다. 일부 실시형태에서, 데아미네이스는 인간 APOBEC3G이다. 일부 실시형태에서, 데아미네이스는 인간 APOBEC3G의 단편이다. 일부 실시형태에서, 데아미네이스는 D316R D317R 돌연변이를 포함하는 인간 APOBEC3G 변이체이다. 일부 실시형태에서, 데아미네이스는 인간 APOBEC3G의 단편이고, D316R D317R 돌연변이에 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 핵산 편집 도메인은 본 명세서에 기재된 임의의 데아미네이스의 데아미네이스 도메인과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%) 또는 적어도 99.5% 동일하다.

[1605] 소정의 실시형태에서, 본 명세서에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성을 개선하는 하나 이상의 특징을 포함한다. 예를 들어, 본 명세서에 제공된 임의의 융합 단백질은 뉴클레이스 활성이 감소된 Cas9 도메인을 포함할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 뉴클레이스 활성(dCas9)이 없는 Cas9 도메인 또는 Cas9 닉케이스(nCas9)로 지칭되는 이중나선 DNA 분자의 하나의 가닥을 절단하는 Cas9 도메인을 가질 수 있다.

[1606] ***충실도가 높은 Cas9 도메인***

[1607] 본 개시내용의 일부 양태는 충실도(fidelity)가 높은 Cas9 도메인을 제공한다. 일부 실시형태에서, 충실도가 높은 Cas9 도메인은 상응하는 야생형 Cas9 도메인과 비교하여 Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 정전기 상호작용을 감소시키는 하나 이상의 돌연변이를 포함하는 조작된 Cas9 도메인이다. 임의의 특정 이론에 얽매이지 않고, DNA의 당-포스페이트 백본과의 정전기 상호작용이 감소된 충실도가 높은 Cas9 도메인은 표적외 효과가 적을 수 있다. 일부 실시형태에서, Cas9 도메인(예를 들어, 야생형 Cas9 도메인)은 Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 회합을 감소시키는 하나 이상의 돌연변이를 포함한다. 일부 실시형태에서, Cas9 도메인은 Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 회합을 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65% 또는 적어도 70%의 감소시키는 하나 이상의 돌연변이를 포함한다.

[1608] 일부 실시형태에서, 본 명세서에 제공된 임의의 Cas9 융합 단백질은 N497X, R661X, Q695X 및/또는 Q926X 돌연변이 중 하나 이상 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함하되, X는 임의의 아미노산이다. 일부 실시형태에서, 본 명세서에 제공된 임의의 Cas9 융합 단백질은 N497A, R661A, Q695A 및/또는 Q926A 돌연변이 중 하나 이상 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다. 일부 실시형태에서, Cas9 도메인은 D10A 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에 상응하는 돌연변이를 포함한다. 당업계에 공지되어 있는 충실도가 높은 Cas9 도메인은 당업자에게 명백할 것이다. 예를 들어, 충실도가 높은 Cas9 도메인은 문헌[Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); 및 Slaymaker, I.M., *et al.* "Rationally engineered Cas9 nucleases with improved specificity." *Science* 351, 84-88 (2015)]에 기술되어 있으며; 이들 각각의 전체 내용은 참조에 의해 본 명세서에 원용된다.

[1609] 일부 실시형태에서, 변형된 Cas9는 충실도가 높은 Cas9 효소이다. 일부 실시형태에서, 충실도가 높은 Cas9 효소는 SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1 또는 초정밀(hyper accurate) Cas9 변이체(HypaCas9)이다. 변형된 Cas9 eSpCas9(1.1)는 HNH/RuvC 홈과 비-표적 DNA 가닥간의 상호작용을 약화시키는 알라닌 치환을 포함하여, 가닥 분리 및 표적외 부위에서의 절단을 방지한다. 유사하게는, SpCas9-HF1은 DNA 포스페이트 백본과 Cas9와의 상호작용을 방해하는 알라닌 치환을 통해 표적외 편집을 낮춘다. HypaCas9는 Cas9 교정 및 표적 식별을 증가시

키는 REC3 도메인의 돌연변이(SpCas9 N692A/M694A/Q695A/H698A)를 포함한다. 세 가지의 높은 충실도 효소는 모두 야생형 Cas9보다 표적외 편집을 적게 생성한다.

[1610] 예시적인 충실도가 높은 Cas9는 아래에 제공된다. Cas9에 대한 충실도가 높은 Cas9 도메인 돌연변이는 굵은 글씨 및 밑줄로 표시된다.

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD

[1611]

[1612] ***napDNAbp 및 사이티딘 데아미네이스 및/또는 아데노신 데아미네이스를 포함하는 융합 단백질***

[1613] 본 개시내용의 일부 양태는 Cas9 도메인 또는 다른 핵산 프로그램 가능 DNA 결합 단백질(예를 들어, Cas12) 및 하나 이상의 사이티딘 데아미네이스 또는 아데노신 데아미네이스 도메인 및/또는 DNA 글리코실레이스 도메인을 포함하는 융합 단백질을 제공한다. Cas9 도메인은 본 명세서에 제공된 임의의 Cas9 도메인 또는 Cas9 단백질(예를 들어, dCas9 또는 nCas9)일 수 있음을 이해하여야 한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 Cas9 도메인 또는 Cas9 단백질(예를 들어, dCas9 또는 nCas9)은 본 명세서에 제공된 임의의 사이티딘 데아미네이스 및/또는 아데노신 데아미네이스와 융합될 수 있다. 본 명세서에 개시된 염기 편집기의 도메인은 임의의 순서로 배열될 수 있다.

[1614] 일부 실시형태에서, 융합 단백질은 다음 도메인 A 내지 C, A 내지 D 또는 A 내지 E를 포함하되,

[1615] $NH_2$-[A-B-C]-COOH;

[1616] $NH_2$-[A-B-C-D]-COOH; 또는

[1617] $NH_2$-[A-B-C-D-E]-COOH;

[1618] A 및 C 또는 A, C 및 E는 각각 다음 중 하나 이상을 포함하고,

[1619] 아데노신 데아미네이스 도메인 또는 이의 활성 단편,

[1620] 사이티딘 데아미네이스 도메인 또는 이의 활성 단편,

[1621] DNA 글리코실레이스 도메인 또는 이의 활성 단편; 및

[1622] B 또는 B 및 D는 각각 핵산 서열 특이적 결합 활성을 갖는 하나 이상의 도메인을 포함한다.

[1623] 일부 실시형태에서, 융합 단백질은 다음 구조를 포함하되,

[1624] $NH_2$-[$A_n$-$B_o$-$C_n$]-COOH;

[1625] $NH_2$-[$A_n$-$B_o$-$C_n$-$D_o$]-COOH; 또는

[1626] $NH_2$-[$A_n$-$B_o$-$C_p$-$D_o$-$E_q$]-COOH;

[1627] A 및 C 또는 A, C 및 E는 각각 다음 중 하나 이상을 포함하고,

[1628] 아데노신 데아미네이스 도메인 또는 이의 활성 단편,

[1629] 사이티딘 데아미네이스 도메인 또는 이의 활성 단편,

[1630] DNA 글리코실레이스 도메인 또는 이의 활성 단편; 및

[1631] n은 1, 2, 3, 4 또는 5의 정수이고, p는 0, 1, 2, 3, 4 또는 5의 정수이며; q는 0, 1, 2, 3, 4 또는 5의 정수이고; B 또는 B 및 D는 각각 핵산 서열 특이적 결합 활성을 갖는 도메인을 포함하며; o는 1, 2, 3, 4 또는 5의 정수이다.

[1632] 예를 들어 그리고 제한 없이, 일부 실시형태에서, 융합 단백질은 다음 구조를 포함한다:

[1633] $NH_2$-[아데노신 데아미네이스]-[Cas9 도메인]-COOH;

[1634] $NH_2$-[Cas9 도메인]-[아데노신 데아미네이스]-COOH;

[1635] $NH_2$-[사이티딘 데아미네이스]-[Cas9 도메인]-COOH;

[1636] $NH_2$-[Cas9 도메인]-[사이티딘 데아미네이스]-COOH;

[1637] $NH_2$-[사이티딘 데아미네이스]-[Cas9 도메인]-[아데노신 데아미네이스]-COOH;

[1638] $NH_2$-[아데노신 데아미네이스]-[Cas9 도메인]-[사이티딘 데아미네이스]-COOH;

[1639] $NH_2$-[아데노신 데아미네이스]-[사이티딘 데아미네이스]-[Cas9 도메인]-COOH;

[1640] $NH_2$-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-[Cas9 도메인]-COOH;

[1641] $NH_2$-[Cas9 도메인]-[아데노신 데아미네이스]-[사이티딘 데아미네이스]-COOH; 또는

[1642] $NH_2$-[Cas9 도메인]-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-COOH.

[1643] 일부 실시형태에서, 본 명세서에 제공된 임의의 Cas12 도메인 또는 Cas12 단백질은 본 명세서에 제공된 임의의 사이티딘 또는 아데노신 데아미네이스와 융합될 수 있다. 예를 들어 그리고 제한 없이, 일부 실시형태에서, 융합 단백질은 다음 구조를 포함한다:

[1644] $NH_2$-[아데노신 데아미네이스]-[Cas12 도메인]-COOH;

[1645] $NH_2$-[Cas12 도메인]-[아데노신 데아미네이스]-COOH;

[1646] $NH_2$-[사이티딘 데아미네이스]-[Cas12 도메인]-COOH;

[1647] $NH_2$-[Cas12 도메인]-[사이티딘 데아미네이스]-COOH;

[1648] $NH_2$-[사이티딘 데아미네이스]-[Cas12 도메인]-[아데노신 데아미네이스]-COOH;

[1649] $NH_2$-[아데노신 데아미네이스]-[Cas12 도메인]-[사이티딘 데아미네이스]-COOH;

[1650] $NH_2$-[아데노신 데아미네이스]-[사이티딘 데아미네이스]-[Cas12 도메인]-COOH;

[1651] $NH_2$-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-[Cas12 도메인]-COOH;

[1652] $NH_2$-[Cas12 도메인]-[아데노신 데아미네이스]-[사이티딘 데아미네이스]-COOH; 또는

[1653] $NH_2$-[Cas12 도메인]-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-COOH.

[1654] 일부 실시형태에서, 아데노신 데아미네이스는 TadA*8이다. 예시적인 융합 단백질 구조는 다음을 포함한다:

[1655] $NH_2$-[TadA*8]-[Cas9 도메인]-COOH;

[1656] $NH_2$-[Cas9 도메인]-[TadA*8]-COOH;

[1657] $NH_2$-[TadA*8]-[Cas12 도메인]-COOH; 또는

[1658] $NH_2$-[Cas12 도메인]-[TadA*8]-COOH.

[1659] 일부 실시형태에서, 아데노신 데아미네이스의 융합 단백질은 TadA*8 및 사이티딘 데아미네이스 및/또는 아데노신 데아미네이스를 포함한다. 일부 실시형태에서, TadA*8은 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23 또는 TadA*8.24이다.

[1660] 예시적인 융합 단백질 구조는 다음을 포함한다:

[1661] $NH_2$-[TadA*8]-[Cas9/Cas12]-[아데노신 데아미네이스]-COOH;

[1662] $NH_2$-[아데노신 데아미네이스]-[Cas9/Cas12]-[TadA*8]-COOH;

[1663] $NH_2$-[TadA*8]-[Cas9/Cas12]-[사이티딘 데아미네이스]-COOH; 또는

[1664] $NH_2$-[사이티딘 데아미네이스]-[Cas9/Cas12]-[TadA*8]-COOH.

[1665] 일부 실시형태에서, 아데노신 데아미네이스의 융합 단백질은 TadA*9 및 사이티딘 데아미네이스 및/또는 아데노신 데아미네이스를 포함한다. 예시적인 융합 단백질 구조는 다음을 포함한다:

[1666] $NH_2$-[TadA*9]-[Cas9/Cas12]-[아데노신 데아미네이스]-COOH;

[1667] $NH_2$-[아데노신 데아미네이스]-[Cas9/Cas12]-[TadA*9]-COOH;

[1668] $NH_2$-[TadA*9]-[Cas9/Cas12]-[사이티딘 데아미네이스]-COOH; 또는

[1669] $NH_2$-[사이티딘 데아미네이스]-[Cas9/Cas12]-[TadA*9]-COOH.

[1670] 일부 실시형태에서, 융합 단백질은 Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편 및 C-말단 단편의 측면에 있는 데아미네이스를 포함할 수 있다. 일부 실시형태에서, 융합 단백질은 Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편 및 C-말단 단편의 측면에 있는 사이티딘 데아미네이스를 포함한다. 일부 실시형태에서, 융합 단백질은 Cas9 또는 Cas 12 폴리펩타이드의 N-말단 단편 및 C-말단 단편의 측면에 있는 아데노신 데아미네이스를 포함한다.

[1671] 일부 실시형태에서, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 napDNAbp(예를 들어, Cas9 또는 Cas12 도메인)를 포함하는 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시형태에서, 링커는 사이티딘 또는 아데노신 데아미네이스와 napDNAbp 사이에 존재한다. 일부 실시형태에서, 위의 일반 구조에서 사용된 "-"는 선택적 링커의 존재를 나타낸다. 일부 실시형태에서, 사이티딘 또는 아데노신 데아미네이스 및 napDNAbp는 본 명세서에 제공된 임의의 링커를 통해 융합된다. 예를 들어, 일부 실시형태에서, 사이티딘 또는 아데노신 데

아미네이스 및 napDNAbp는 본 명세서에 제공된 임의의 링커를 통해 융합된다.

[1672] 본 개시내용의 융합 단백질은 하나 이상의 추가적인 특징을 포함할 수 있음을 이해하여야 한다. 예를 들어, 일부 실시형태에서, 융합 단백질은 저해제, 세포질 국부화 서열, 방출 서열, 예컨대, 핵 방출 서열 또는 다른 국부화 서열뿐만 아니라 융합 단백질의 가용화, 정제 또는 검출에 유용한 서열 태그를 포함할 수 있다. 본 명세서에 제공된 적합한 단백질 태그는 바이오틴 카복실레이스 운반 단백질(biotin carboxylase carrier protein: BCCP) 태그, myc-태그, 칼모듈린-태그, FLAG-태그, 헤마글루티닌(hemagglutinin: HA)-태그, 히스티딘 태그 또는 His-태그로도 지칭되는 폴리히스티딘 태그, 말토스 결합 단백질(maltose binding protein: MBP)-태그, nus-태그, 글루타티온-S-트랜스퍼레이스(glutathione-S-transferase: GST)-태그, 녹색 형광 단백질(green fluorescent protein: GFP)-태그, 티오레독신-태그, S-태그, 소프태그(예를 들어, 소프태그 1, 소프태그 3), 스트렙-태그, 바이오틴 라이게이스 태그, FlAsH 태그, V5 태그 및 SBP-태그를 포함하지만 이들로 제한되지 않는다. 추가적인 적합한 서열은 당업자에게 명백할 것이다. 일부 실시형태에서, 융합 단백질은 하나 이상의 His 태그를 포함한다.

[1673] 예시적이지만 비제한적인 융합 단백질은 PCT 특허 PCT/2017/044935, PCT/US2019/044935 및 PCT/US2020/016288에 기술되어 있으며, 이들 각각은 그 전문이 참조에 의해 본 명세서에 원용된다.

[1674] ***핵 국부화 서열(NLS)을 포함하는 융합 단백질***

[1675] 일부 실시형태에서, 본 명세서에 제공된 융합 단백질은 하나 이상(예를 들어, 2개, 3개, 4개, 5개)의 핵 표적화 서열, 예를 들어, 핵 국부화 서열(NLS)을 더 포함한다. 일 실시형태에서, 이분 NLS이 사용된다. 일부 실시형태에서, NLS는 (예를 들어, 핵 수송에 의해) NLS를 포함하는 단백질의 세포 핵으로의 유입을 용이하게 하는 아미노산 서열을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 핵 국부화 서열(NLS)을 더 포함한다. 일부 실시형태에서, NLS는 융합 단백질의 N-말단에 융합된다. 일부 실시형태에서, NLS는 융합 단백질의 C-말단에 융합된다. 일부 실시형태에서, NLS는 Cas9 도메인의 N-말단에 융합된다. 일부 실시형태에서, NLS는 Cas9 도메인의 C-말단에 융합된다. 일부 실시형태에서, NLS는 nCas9 도메인 또는 dCas9 도메인의 C-말단에 융합된다. 일부 실시형태에서, NLS는 Cas12 도메인의 N-말단에 융합된다. 일부 실시형태에서, NLS는 Cas12 도메인의 C-말단에 융합된다. 일부 실시형태에서, NLS는 사이티딘 또는 아데노신 데아미네이스의 N-말단에 융합된다. 일부 실시형태에서, NLS는 사이티딘 또는 아데노신 데아미네이스의 C-말단에 융합된다. 일부 실시형태에서, NLS는 하나 이상의 링커를 통해 융합 단백질에 융합된다. 일부 실시형태에서, NLS는 링커 없이 융합 단백질에 융합된다. 일부 실시형태에서, NLS는 본 명세서에서 제공되거나 또는 참조된 NLS 서열 중 어느 하나의 아미노산 서열을 포함한다. 추가적인 핵 국부화 서열은 당업계에 공지되어 있으며, 당업자에게 명백할 것이다. 예를 들어, NLS 서열은 예시적인 핵 국부화 서열의 개시에 대해 그 내용이 참조에 의해 본 명세서에 원용되어 있는 Plank 등의 PCT/EP2000/011690에 기술되어 있다. 일부 실시형태에서, NLS는 아미노산 서열 PKKKRKVEGADKRTADGSEFESPKKKRKV, KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRKPKKKRKV 또는 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC를 포함한다.

[1676] 일부 실시형태에서, 사이티딘 또는 아데노신 데아미네이스, Cas9 도메인 및 NLS를 포함하는 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시형태에서, 하나 이상의 도메인 또는 단백질(예를 들어, 사이티딘 또는 아데노신 데아미네이스, Cas9 도메인 또는 NLS) 사이에 링커 서열이 존재한다. 일부 실시형태에서, 링커는 사이티딘 데아미네이스 및 아데노신 데아미네이스 도메인과 napDNAbp 사이에 존재한다. 일부 실시형태에서, 아래의 일반적인 구조에 사용된 "-"는 선택적 링커의 존재를 나타낸다. 일부 실시형태에서, 사이티딘 데아미네이스 및 아데노신 데아미네이스 및 napDNAbp는 본 명세서에 제공된 임의의 링커를 통해 융합된다. 예를 들어, 일부 실시형태에서, 사이티딘 데아미네이스 및 아데노신 데아미네이스 및 napDNAbp는 본 명세서에 제공된 임의의 링커를 통해 융합된다.

[1677] 일부 실시형태에서, 예시적인 napDNAbp(예를 들어, Cas9 또는 Cas12) 융합 단백질과 사이티딘 또는 아데노신 데아미네이스 및 napDNAbp(예를 들어, Cas9 또는 Cas12) 도메인의 일반적인 구조는 다음 구조 중 어느 하나를 포함하되, NLS는 핵 국부화 서열(예를 들어, 본 명세서에 제공된 임의의 NLS)이고, NH2는 융합 단백질의 N-말단이며, COOH 융합 단백질의 C-말단이다:

[1678] $NH_2$-NLS-[사이티딘 데아미네이스]-[napDNAbp 도메인]-COOH;

[1679] $NH_2$-NLS [napDNAbp 도메인]-[사이티딘 데아미네이스]-COOH;

[1680] $NH_2$-[사이티딘 데아미네이스]-[napDNAbp 도메인]-NLS-COOH;

[1681] $NH_2$-[napDNAbp 도메인]-[사이티딘 데아미네이스]-NLS-COOH;

[1682] $NH_2$-NLS-[아데노신 데아미네이스]-[napDNAbp 도메인]-COOH;

[1683] $NH_2$-NLS [napDNAbp 도메인]-[아데노신 데아미네이스]-COOH;

[1684] $NH_2$-[아데노신 데아미네이스]-[napDNAbp 도메인]-NLS-COOH;

[1685] $NH_2$-[napDNAbp 도메인]-[아데노신 데아미네이스]-NLS-COOH;

[1686] $NH_2$-NLS-[사이티딘 데아미네이스]-[napDNAbp 도메인]-[아데노신 데아미네이스]-COOH;

[1687] $NH_2$-NLS-[아데노신 데아미네이스]-[napDNAbp 도메인]-[사이티딘 데아미네이스]-COOH;

[1688] $NH_2$-NLS-[아데노신 데아미네이스] [사이티딘 데아미네이스]-[napDNAbp 도메인]-COOH;

[1689] $NH_2$-NLS-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-[napDNAbp 도메인]-COOH;

[1690] $NH_2$-NLS-[napDNAbp 도메인]-[아데노신 데아미네이스]-[사이티딘 데아미네이스]-COOH;

[1691] $NH_2$-NLS-[napDNAbp 도메인]-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-COOH;

[1692] $NH_2$-[사이티딘 데아미네이스]-[napDNAbp 도메인]-[아데노신 데아미네이스]-NLS-COOH;

[1693] $NH_2$-[아데노신 데아미네이스]-[napDNAbp 도메인]-[사이티딘 데아미네이스]-NLS-COOH;

[1694] $NH_2$-[아데노신 데아미네이스] [사이티딘 데아미네이스]-[napDNAbp 도메인]-NLS-COOH;

[1695] $NH_2$-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-[napDNAbp 도메인]-NLS-COOH;

[1696] $NH_2$-[napDNAbp 도메인]-[아데노신 데아미네이스]-[사이티딘 데아미네이스]-NLS-COOH; 또는

[1697] $NH_2$-[napDNAbp 도메인]-[사이티딘 데아미네이스]-[아데노신 데아미네이스]-NLS-COOH. 일부 실시형태에서, NLS는 링커에 존재하거나, 또는 NLS는, 예를 들어, 본 명세서에 기재된 링커에 의해 측면에 배치된다. 이분 NLS는 상대적으로 짧은 스페이서 서열에 의해 분리된 2개의 염기성 아미노산 클러스터를 포함한다(그러므로, 이분은 2부분이지만 단분적 NLS는 그렇지 않음). 뉴클레오플라스민의 NLS인 KR[PAATKKAGQA]KKKK는 편재된 이분 신호의 원형(prototype): 약 10개의 아미노산의 스페이서로 분리된 염기성 아미노산의 2개 클러스터이다. 예시적인 이분 NLS의 서열은 다음과 같다:

[1698] PKKKRKVEGADKRTADGSEFESPKKKRKV

[1699] 하나 이상의 핵 국부화 서열(NLS)을 포함하는 CRISPR 효소를 암호화하는 벡터가 사용될 수 있다. 예를 들어, 약 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개의 NLS가 사용되거나 사용될 수 있다. CRISPR 효소는 아미노-말단 또는 그 부근에 NLS, 카복시-말단 또는 그 부근에 약 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 또는 약 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 초과의 NLS, 또는 이들의 임의의 조합(예를 들어, 아미노-말단에 하나 이상의 NLS 및 카복시 말단에 하나 이상의 NLS)을 포함할 수 있다. 하나 이상의 NLS가 존재하는 경우, 각각은 단일 NLS가 하나 이상의 카피로 그리고/또는 하나 이상의 카피에 존재하는 하나 이상의 다른 NLS와 조합하여 존재할 수 있도록 서로 독립적으로 선택될 수 있다.

[1700] 방법에서 사용되는 CRISPR 효소는 약 6개의 NLS를 포함할 수 있다. NLS는 NLS에 가장 가까운 아미노산이 N-말단 또는 C-말단으로부터 폴리펩타이드 사슬을 따라 약 50개의 아미노산 이내, 예를 들어, 1개, 2개, 3개, 4개, 5개, 10개, 15개, 20개, 25개, 30개, 40개 또는 50개의 아미노산 이내인 경우, N-말단 또는 C-말단 근처인 것으로 간주된다.

[1701] ***추가적인 도메인***

[1702] 본 명세서에 기재된 염기 편집기는 폴리뉴클레오타이드의 핵염기의 핵염기 편집, 변형 또는 변경을 용이하게 하는데 도움을 주는 임의의 도메인을 포함할 수 있다. 일부 실시형태에서, 염기 편집기는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인(예를 들어, Cas9), 핵염기 편집 도메인(예를 들어, 데아미네이스 도메인) 및 하나 이상의 추가적인 도메인을 포함한다. 일부 실시형태에서, 추가적인 도메인은 염기 편집기의 효소 또는 촉매 기능, 염기 편집기의 결합 기능을 촉진하거나 또는 목적하는 염기 편집 결과를 방해할 수 있는 세포 기구(예를 들어, 효소)의 저해제일 수 있다. 일부 실시형태에서, 염기 편집기는 뉴클레이스, 닉케이스, 재조합효소, 데아미네이스, 메틸트랜스퍼레이스, 메틸레이스, 아세틸레이스, 아세틸트랜스퍼레이스, 전사 활성인자 또는 전사 억제 도메인을 포함할 수 있다.

[1703] 일부 실시형태에서, 염기 편집기는 우라실 글리코실레이스 저해제(UGI) 도메인을 포함할 수 있다. 일부 실시형태에서, U:G 이종이중체 DNA의 존재에 대한 세포 DNA 복구 반응은 세포에서 핵염기 편집 효율의 감소의 원인이 될 수 있다. 이러한 실시형태에서, 우라실 DNA 글리코실레이스(UDG)는 세포에서 DNA로부터 U의 제거를 촉매할 수 있으며, 이는 염기 절단 복구(BER)를 개시할 수 있으며 대부분 U:G 쌍을 C:G 쌍으로 되돌린다. 이러한 실시형태에서, BER은 단일 가닥에 결합하고, 편집된 염기를 차단하며, UGI를 저해하고, BER을 저해하며, 편집된 염기를 보호하고/하거나 편집되지 않은 가닥의 복구를 촉진하는 하나 이상의 도메인을 포함하는 염기 편집기에서 저해될 수 있다. 따라서, 본 개시내용은 UGI 도메인을 포함하는 염기 편집기 융합 단백질을 상정한다.

[1704] 일부 실시형태에서, 염기 편집기는 이중-가닥 파손(DSB) 결합 단백질의 전부 또는 일부를 도메인으로 포함한다. 예를 들어, DSB 결합 단백질은 DSB의 말단에 결합할 수 있으며 분해로부터 이들을 보호할 수 있는 박테리오파지 Mu의 Gam 단백질을 포함할 수 있다. 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)] 참조.

[1705] 추가적으로, 일부 실시형태에서, Gam 단백질은 염기 편집기의 N 말단에 융합될 수 있다. 일부 실시형태에서, Gam 단백질은 염기 편집기의 C 말단에 융합될 수 있다. 박테리오파지 Mu의 Gam 단백질은 이중 가닥 파손(DSB)의 말단에 결합되어 분해로부터 이들을 보호할 수 있다. 일부 실시형태에서, DSB의 자유 말단에 결합시키기 위해 Gam을 사용하는 것은 염기 편집의 과정 동안 indel 형성을 줄일 수 있다. 일부 실시형태에서, 174개-잔기의 Gam 단백질은 염기 편집기의 N 말단에 융합된다. 문헌[Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)] 참조. 일부 실시형태에서, 돌연변이 또는 돌연변이들은 야생형 도메인에 비해 염기 편집기 도메인의 길이를 변경할 수 있다. 예를 들어, 적어도 하나의 도메인에서 적어도 하나의 아미노산의 결실은 염기 편집기의 길이를 줄일 수 있다. 또 다른 경우에, 돌연변이 또는 돌연변이들은 야생형 도메인에 비해 도메인의 길이를 변경하지 않는다. 예를 들어, 임의의 도메인의 치환은 염기 편집기의 길이를 변경하지 않는다.

[1706] 모든 도메인의 길이가 야생형 도메인과 동일한 이러한 염기 편집기의 비제한적인 예는 다음을 포함할 수 있다:

[1707] NH2-[핵염기 편집 도메인]-링커1-[APOBEC1]-링커2-[핵염기 편집 도메인]-COOH;

[1708] NH2-[핵염기 편집 도메인]-링커1-[APOBEC1]-[핵염기 편집 도메인]-COOH;

[1709] NH2-[핵염기 편집 도메인]-[APOBEC1]-링커2-[핵염기 편집 도메인]-COOH;

[1710] NH2-[핵염기 편집 도메인]-[APOBEC1]-[핵염기 편집 도메인]-COOH;

[1711] NH2-[핵염기 편집 도메인]-링커1-[APOBEC1]-링커2-[핵염기 편집 도메인]-[UGI]-COOH;

[1712] NH2-[핵염기 편집 도메인]-링커1-[APOBEC1]-[핵염기 편집 도메인]-[UGI]-COOH;

[1713] NH2-[핵염기 편집 도메인]-[APOBEC1]-링커2-[핵염기 편집 도메인]-[UGI]-COOH;

[1714] NH2-[핵염기 편집 도메인]-[APOBEC1]-[핵염기 편집 도메인]-[UGI]-COOH;

[1715] NH2-[UGI]-[핵염기 편집 도메인]-링커1-[APOBEC1]-링커2-[핵염기 편집 도메인]-COOH;

[1716] NH2-[UGI]-[핵염기 편집 도메인]-링커1-[APOBEC1]-[핵염기 편집 도메인]-COOH;

[1717] NH2-[UGI]-[핵염기 편집 도메인]-[APOBEC1]-링커2-[핵염기 편집 도메인]-COOH; 또는

[1718] NH2-[UGI]-[핵염기 편집 도메인]-[APOBEC1]-[핵염기 편집 도메인]-COOH.

[1719] 염기 편집 시스템

[1720] 염기 편집 시스템을 사용하여 핵염기를 편집하기 위한 시스템, 조성물 및 방법이 본 명세서에 제공된다. 일부 실시형태에서, 염기 편집 시스템은 (1) 핵염기를 편집하기 위한 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 및 핵염기 편집 도메인(예를 들어, 데아미네이스 도메인)을 포함하는 염기 편집기(BE); 및 (2) 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인과 함께 가이드 폴리뉴클레오타이드(예를 들어, 가이드 RNA)를 포함한다. 일부 실시형태에서, 염기 편집 시스템 사이티딘 염기 편집기(CBE)이다. 일부 실시형태에서, 염기 편집 시스템은 아데노신 염기 편집기(ABE)이다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인이다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그램 가능한 RNA 결합 도메인이다. 일부 실시형태에서, 핵염기 편집 도메인은 데아미네이스 도메인이다. 일부 실시형태에서, 데아미네이스 도메인은 사이티딘 데아미네이스 또는 사이토신 데아미네이스일 수 있다. 일부 실시형태에서, 데아미네이스 도메인은 아데닌 데아미네이스 또는 아데노신 데아미네이스일 수 있다. 일부 실시형태에서, 아데노신 염기 편집기는 DNA에서 아데닌을 탈아미노화할 수 있다. 일부 실시형태에서, 염기 편집기는 DNA에서 사이티딘을 탈아미노화할 수 있다.

[1721] 일부 실시형태에서, 본 명세서에 제공된 바와 같은 염기 편집 시스템은 이중-가닥 DNA 파손을 생성하지 않고, 공여체 DNA 주형을 필요로 하지 않으며, 확률론적 삽입 및 결실을 과도하게 유도하지 않고 DNA에 프로그램 가능한 단일 뉴클레오타이드(C→T 또는 A→G) 변화를 유도하기 위해, 촉매적으로 결함이 있는 스트렙토코커스 피오게네스 Cas9, 데아미네이스(예를 들어, 사이티딘 또는 아데노신 데아미네이스) 및 염기 절단 복구 저해제를 포함하는 융합 단백질을 사용하는 게놈 편집에 대한 새로운 접근법을 제공한다.

[1722] 핵염기 편집 단백질에 대한 세부사항은 PCT 특허 PCT/2017/045381(WO2018/027078) 및 PCT/US2016/058344(WO2017/070632)에 기술되어 있으며, 이들 각각은 그 전문이 참조에 의해 본 명세서에 원용된다. 또한, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); 및 Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)] 참조.

[1723] 본 명세서에 제공된 염기 편집 시스템의 사용은 (a) 대상체의 폴리뉴클레오타이드(예를 들어, 이중-가닥 또는 단일 가닥 DNA 또는 RNA)의 표적 뉴클레오타이드 서열을 핵염기 편집기(예를 들어, 아데노신 염기 편집기 또는 사이티딘 염기 편집기) 및 가이드 다중핵산(예를 들어, gRNA)를 포함하는 염기 편집 시스템과 접촉시키는 단계로서, 표적 뉴클레오타이드 서열은 표적화된 핵염기쌍을 포함하는, 상기 접촉시키는 단계; (b) 상기 표적 영역의 가닥 분리를 유도하는 단계; (c) 표적 영역의 단일 가닥에 있는 상기 표적 핵염기쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계; 및 (d) 제1 핵염기 염기에 상보적인 제3 핵염기가 제2 핵염기에 상보적인 제4 핵염기로 대체되는 상기 표적 영역의 가닥을 1개 이하로 절단하는 단계를 포함한다. 일부 실시형태에서, (b) 단계가 생략됨을 이해하여야 한다. 일부 실시형태에서, 상기 표적화된 핵염기쌍은 하나 이상의 유전자에서 복수의 핵염기쌍이다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집 시스템은 하나 이상의 유전자에서 복수의 핵염기쌍을 다중 편집할 수 있다. 일부 실시형태에서, 복수의 핵염기쌍이 동일한 유전자에 위치한다. 일부 실시형태에서, 복수의 핵염기쌍이 하나 이상의 유전자에 위치하되, 적어도 하나의 유전자는 상이한 유전자좌에 위치한다.

[1724] 일부 실시형태에서, 절단된 단일 가닥(틈이 있는 가닥)은 가이드 핵산에 혼성화된다. 일부 실시형태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥 반대편에 있다. 일부 실시형태에서, 염기 편집기는 Cas9 도메인을 포함한다. 일부 실시형태에서, 제1 염기는 아데닌이고, 제2 염기는 G, C, A 또는 T가 아니다. 일부 실시형태에서, 제2 염기는 이노신이다.

[1725] 일부 실시형태에서, 단일 가이드 폴리뉴클레오타이드는 데아미네이스를 표적 핵산 서열로 표적화하는데 이용될 수 있다. 일부 실시형태에서, 단일 쌍의 가이드 폴리뉴클레오타이드는 상이한 데아미네이스를 표적 핵산 서열로 표적화하는데 이용될 수 있다.

[1726] 염기 편집 시스템의 핵염기 성분 및 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 성분은 공유적으로 또는 비공유적으로 서로 회합될 수 있다. 예를 들어, 일부 실시형태에서, 데아미네이스 도메인은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다.

일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미네이스 도메인에 융합 또는 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미네이스 도메인과 비공유적으로 상호작용하거나 또는 이와 회합함으로써 데아미네이스 도메인을 표적 뉴클레오타이드 서열로 표적화할 수 있다. 예를 들어, 일부 실시형태에서, 핵염기 편집 성분, 예를 들어, 데아미네이스 성분은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 일부인 추가적인 이종성 부분 또는 도메인과 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인을 포함할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드와 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드와 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가적인 이종성 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머레이스 Ku 결합 모티프 및 Ku 단백질, 텔로머레이스 Sm7 결합 모티프 및 Sm7 단백질 또는 RNA 인식 모티프일 수 있다.

[1727] 염기 편집 시스템은 가이드 폴리뉴클레오타이드 성분을 더 포함할 수 있다. 염기 편집 시스템의 성분은 공유 결합, 비공유 상호작용 또는 이들의 회합 및 상호작용의 임의의 조합을 통해 서로 회합될 수 있음을 이해하여야 한다. 일부 실시형태에서, 데아미네이스 도메인은 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다. 예를 들어, 일부 실시형태에서, 염기 편집 시스템의 핵염기 편집 성분, 예를 들어, 데아미네이스 성분은 가이드 폴리뉴클레오타이드의 부분 또는 분절(예를 들어, 폴리뉴클레오타이드 모티프)과 상호작용할 수 있거나, 이와 회합할 수 있거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)은 데아미네이스 도메인에 융합 또는 연결될 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드와 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드와 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가적인 이종성 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머레이스 Ku 결합 모티프 및 Ku 단백질, 텔로머레이스 Sm7 결합 모티프 및 Sm7 단백질 또는 RNA 인식 모티프일 수 있다.

[1728] 일부 실시형태에서, 염기 편집 시스템은 염기 절단 복구 저해제(BER) 성분을 더 포함할 수 있다. 염기 편집 시스템의 성분은 공유 결합, 비공유 상호작용 또는 이들의 회합 및 상호작용의 임의의 조합을 통해 서로 회합될 수 있음을 이해하여야 한다. BER 성분의 저해제는 염기 절단 복구 저해제를 포함할 수 있다. 일부 실시형태에서, 염기 절단 복구 저해제는 우라실 DNA 글리코실레이스 저해제(UGI)일 수 있다. 일부 실시형태에서, 염기 절단 복구 저해제는 이노신 염기 절단 복구 저해제일 수 있다. 일부 실시형태에서, 염기 절단 복구 저해제는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 염기 절단 복구 저해제에 융합 또는 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미네이스 도메인 및 염기 절단 복구 저해제에 융합 또는 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 염기 절단 복구 저해제와 비공유적으로 상호작용하거나 또는 이와 회합함으로써 염기 절단 복구 저해제를 표적 뉴클레오타이드 서열로 표적화할 수 있다. 예를 들어, 일부 실시형태에서, 염기 절단 복구 저해제 성분은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 일부인 추가적인 이종성 부분 또는 도메인과 상호작용할 수 있거나, 이와 회합할 수 있거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인을 포함할 수 있다. 일부 실시형태에서, 염기 절단 복구 저해제 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열로 표적화될 수 있다. 예를 들어, 일부 실시형태에서, 염기 절단 복구 저해제는 가이드 폴리뉴클레오타이드의 부분 또는 분절(예를 들어, 폴리뉴클레오타이드 모티프)과 상호작용할 수 있거나, 이와 회합할 수 있거나 또는 이와 복합체를 형성할 수 있는 추가적인 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과

같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 실시형태에서, 가이드 폴리뉴클레오타이드(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)의 추가적인 이종성 부분 또는 도메인은 염기 절단 복구 저해제에 융합 또는 연결될 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드와 결합하거나, 이와 상호작용하거나, 이와 회합하거나 또는 이와 복합체를 형성할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가적인 이종성 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가적인 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머레이스 Ku 결합 모티프 및 Ku 단백질, 텔로머레이스 Sm7 결합 모티프 및 Sm7 단백질 또는 RNA 인식 모티프일 수 있다.

[1729] 일부 실시형태에서, 염기 편집기는 편집된 가닥의 염기 절단 복구(BER)를 저해한다. 일부 실시형태에서, 염기 편집기는 편집되지 않은 가닥을 보호하거나 또는 이에 결합한다. 일부 실시형태에서, 염기 편집기는 UGI 활성을 포함한다. 일부 실시형태에서, 염기 편집기는 촉매적으로 불활성인 이노신-특이적 뉴클레이스를 포함한다. 일부 실시형태에서, 염기 편집기는 닉케이스 활성을 포함한다. 일부 실시형태에서, 염기쌍의 의도된 편집은 PAM 부위의 상류이다. 일부 실시형태에서, 염기쌍의 의도된 편집은 PAM 부위의 상류에서 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드이다. 일부 실시형태에서, 염기쌍의 의도된 편집은 PAM 부위의 하류이다. 일부 실시형태에서, 염기쌍 의도된 편집은 PAM 부위의 하류에서 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드이다.

[1730] 일부 실시형태에서, 방법은 표준(예를 들어, NGG) PAM 부위를 필요로 하지 않는다. 일부 실시형태에서, 핵염기 편집기는 링커 또는 스페이서를 포함한다. 일부 실시형태에서, 링커 또는 스페이서는 1개 내지 25개의 아미노산 길이이다. 일부 실시형태에서, 링커 또는 스페이서는 5개 내지 20개의 아미노산 길이이다. 일부 실시형태에서, 링커 또는 스페이서는 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 아미노산 길이이다.

[1731] 일부 실시형태에서, 본 명세서에 제공된 염기 편집 융합 단백질은, 예를 들어, 표적 염기가 정의된 영역(예를 들어, "탈아미노화 창") 내에 배치되는 정확한 위치에 위치되어야 한다. 일부 실시형태에서, 표적은 4개의 염기 영역 내에 있을 수 있다. 일부 실시형태에서, 이러한 정의된 표적 영역은 PAM의 상류에서 대략 15개의 염기일 수 있다. 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); 및 Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)] 참조.

[1732] 일부 실시형태에서, 표적 영역은 표적 창(target window)을 포함하되, 표적 창은 표적 핵염기쌍을 포함한다. 일부 실시형태에서, 표적 창은 1개 내지 10개의 뉴클레오타이드이다. 일부 실시형태에서, 표적 창은 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드 길이이다. 일부 실시형태에서, 염기쌍의 의도된 편집은 표적 창 내에 있다. 일부 실시형태에서, 표적 창은 염기쌍의 의도된 편집을 포함한다. 일부 실시형태에서, 방법은 본 명세서에 제공된 임의의 염기 편집기를 사용하여 수행된다. 일부 실시형태에서, 표적 창은 탈아미노화 창이다. 탈아미노화 창은 염기 편집기가 표적 뉴클레오타이드에 작용하고 이를 탈아미노화하는 정의된 영역일 수 있다. 일부 실시형태에서, 탈아미노화 창은 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개의 염기 영역 내에 있다. 일부 실시형태에서, 탈아미노화 창은 PAM의 상류에서 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개 또는 25개의 염기이다.

[1733] 본 개시내용의 염기 편집기는 표적 폴리뉴클레오타이드 서열의 편집을 용이하게 하는 임의의 도메인, 특징 또는 아미노산 서열을 포함할 수 있다. 예를 들어, 일부 실시형태에서, 염기 편집기는 핵 국부화 서열(NLS)을 포함한다. 일부 실시형태에서, 염기 편집기의 NLS는 데아미네이스 도메인과 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인 사이에 국부화된다. 일부 실시형태에서, 염기 편집기의 NLS는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 C-말단에 국부화된다.

[1734] 본 명세서에 개시된 바와 같은 염기 편집기에 존재할 수 있는 다른 예시적인 특징은 국부화 서열, 예컨대, 세포질 국부화 서열, 방출 서열, 예컨대, 핵 방출 서열 또는 다른 국부화 서열뿐만 아니라 융합 단백질의 가용화, 정제 또는 검출에 유용한 서열 태그이다. 본 명세서에 제공된 적합한 단백질 태그는 바이오틴 카복실레이스 운반 단백질(BCCP) 태그, myc-태그, 칼모듈린-태그, FLAG-태그, 헤마글루티닌(HA)-태그, 히스티딘 태그 또는 His-태그로도 지칭되는 폴리히스티딘 태그, 말토스 결합 단백질(MBP)-태그, nus-태그, 글루타티온-S-트랜스퍼레이스(GST)-태그, 녹색 형광 단백질(GFP)-태그, 티오레독신-태그, S-태그, 소프태그(예를 들어, 소프태그 1, 소프태그 3), 스트렙-태그, 바이오틴 라이게이스 태그, FlAsH 태그, V5 태그 및 SBP-태그를 포함하지만, 이들로 제한되지 않는다. 추가적인 적합한 서열은 당업자에게 명백할 것이다. 일부 실시형태에서, 융합 단백질은 하나 이상의 His 태그를 포함한다.

[1735] 일부 실시형태에서, 비제한적인 예시적인 사이티딘 염기 편집기(CBE)는 BE1(APOBEC1-XTEN-dCas9), BE2(APOBEC1-XTEN-dCas9-UGI), BE3(APOBEC1-XTEN-dCas9(A840H)-UGI), BE3-Gam, saBE3, saBE4-Gam, BE4, BE4-Gam, saBE4 또는 saB4E-Gam을 포함한다. BE4는 APOBEC1-Cas9n(D10A) 링커를 32개의 아미노산으로 확장하고, Cas9n-UGI 링커를 9개의 아미노산으로 확장하며, UGI의 제2 카피를 작제물의 C-말단에 추가하고 또 다른 9-아미노산 링커를 단일 염기 편집기 작제물에 추가한다. 염기 편집기 saBE3 및 saBE4에서는 S. 피오게네스 Cas9n(D10A)이 더 작은 S. 아우레우스 Cas9n(D10A)로 대체되었다. BE3-Gam, saBE3-Gam, BE4-Gam 및 saBE4-Gam은 16개의 아미노산 XTEN 링커를 통해 BE3, saBE3, BE4 및 saBE4의 N-말단에 융합된 Gam 단백질의 174개의 잔기를 갖는다.

[1736] 일부 실시형태에서, 아데노신 염기 편집기(ABE)는 DNA에서 아데닌을 탈아미노화할 수 있다. 일부 실시형태에서, ABE는 BE3의 APOBEC1 성분을 천연 또는 조작된 대장균 TadA, 인간 ADAR2, 마우스 ADA 또는 인간 ADAT2로 대체함으로써 생성되었다. 일부 실시형태에서, ABE는 진화된 TadA 변이체를 포함한다. 일부 실시형태에서, ABE는 ABE 1.2(TadA*-XTEN-nCas9-NLS)이다. 일부 실시형태에서, TadA*는 A106V 및 D108N 돌연변이를 포함한다.

[1737] 일부 실시형태에서, ABE는 2세대 ABE이다. 일부 실시형태에서, ABE는 ABE2.1이며, 이는 TadA*(TadA*2.1)에 추가적인 돌연변이 D147Y 및 E155V를 포함한다. 일부 실시형태에서, ABE는 인간 알킬 아데닌 DNA 글리코실레이스의 촉매적으로 불활성화된 버전(E125Q 돌연변이를 갖는 AAG)에 융합된 ABE2.1인 ABE2.2이다. 일부 실시형태에서, ABE는 대장균 Endo V의 촉매적으로 불활성화된 버전(D35A 돌연변이로 불활성화됨)에 융합된 ABE2.1인 ABE2.3이다. 일부 실시형태에서, ABE는 ABE2.1의 링커보다 2배 긴 링커(32개의 아미노산, $(SGGS)_2$-XTEN-$(SGGS)_2$)인 ABE2.6이다. 일부 실시형태에서, ABE는 추가적인 야생형 TadA 단량체로 테더링된 ABE2.1인 ABE2.7이다. 일부 실시형태에서, ABE는 추가적인 TadA*2.1 단량체로 테더링된 ABE2.1인 ABE2.8이다. 일부 실시형태에서, ABE는 ABE2.1의 N-말단에 대한 진화된 TadA(TadA*2.1)의 직접 융합인 ABE2.9다. 일부 실시형태에서, ABE는 ABE2.1의 N-말단에 대한 야생형 TadA의 직접 융합인 ABE2.10이다. 일부 실시형태에서, ABE는 TadA* 단량체의 N-말단에 불활성화 E59A 돌연변이가 있는 ABE2.9인 ABE2.11이다. 일부 실시형태에서, ABE는 내부 TadA* 단량체에 불활성화 E59A 돌연변이가 있는 ABE2.9인 ABE2.12이다.

[1738] 일부 실시형태에서, ABE는 3세대 ABE이다. 일부 실시형태에서, ABE는 3개의 추가적인 TadA 돌연변이(L84F, H123Y 및 I156F)를 갖는 ABE2.3인 ABE3.1이다.

[1739] 일부 실시형태에서, ABE는 4세대 ABE이다. 일부 실시형태에서, ABE는 추가적인 TadA 돌연변이 A142N(TadA*4.3)을 갖는 ABE3.1인 ABE4.3이다.

[1740] 일부 실시형태에서, ABE는 5세대 ABE이다. 일부 실시형태에서, ABE는 살아남은 클론(H36L, R51L, S146C 및 K157N)에서 ABE3.1로 돌연변이의 컨센서스 세트를 수송함으로써 생성된 ABE5.1이다. 일부 실시형태에서, ABE는 내부 진화된 TadA*에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE5.3이다. 일부 실시형태에서, ABE는 아래 표 12에 나타나 있는 바와 같은 ABE5.2, ABE5.4, ABE5.5, ABE5.6, ABE5.7, ABE5.8, ABE5.9, ABE5.10, ABE5.11, ABE5.12, ABE5.13 또는 ABE5.14이다. 일부 실시형태에서, ABE는 6세대 ABE이다. 일부 실시형태에서, ABE는 아래 표 12에 나타나 있는 바와 같은 ABE6.1, ABE6.2, ABE6.3, ABE6.4, ABE6.5 또는 ABE6.6이다. 일부 실시형태에서, ABE는 7세대 ABE이다. 일부 실시형태에서, ABE는 아래 표 12에 나타나 있는 바와 같은 ABE7.1, ABE7.2, ABE7.3, ABE7.4, ABE7.5, ABE7.6, ABE7.7, ABE7.8, ABE 7.9 또는 ABE7.10이다.

표 12

ABE 의 유전자형

| | 23 | 26 | 36 | 37 | 48 | 49 | 51 | 72 | 84 | 87 | 106 | 108 | 123 | 125 | 142 | 146 | 147 | 152 | 155 | 156 | 157 | 161 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABE0.1 | W | R | H | N | P | | R | N | L | S | A | D | H | G | A | S | D | R | E | I | K | K |
| ABE0.2 | W | R | H | N | P | | R | N | L | S | A | D | H | G | A | S | D | R | E | I | K | K |
| ABE1.1 | W | R | H | N | P | | R | N | L | S | A | N | H | G | A | S | D | R | E | I | K | K |
| ABE1.2 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | D | R | E | I | K | K |
| ABE2.1 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.2 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.3 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.4 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.5 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.6 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.7 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.8 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.9 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.10 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.11 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE2.12 | W | R | H | N | P | | R | N | L | S | V | N | H | G | A | S | Y | R | V | I | K | K |
| ABE3.1 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE3.2 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE3.3 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE3.4 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE3.5 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE3.6 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE3.7 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE3.8 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE4.1 | W | R | H | N | P | | R | N | L | S | V | N | H | G | N | S | Y | R | V | I | K | K |
| ABE4.2 | W | G | H | N | P | | R | N | L | S | V | N | H | G | N | S | Y | R | V | I | K | K |
| ABE4.3 | W | R | H | N | P | | R | N | F | S | V | N | Y | G | N | S | Y | R | V | F | K | K |
| ABE5.1 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |

[1741]

| | 23 | 26 | 36 | 37 | 48 | 49 | 51 | 72 | 84 | 87 | 106 | 108 | 123 | 125 | 142 | 146 | 147 | 152 | 155 | 156 | 157 | 161 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABE5.2 | W | R | H | S | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | T |
| ABE5.3 | W | R | L | N | P | | L | N | I | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.4 | W | R | H | S | P | | R | N | F | S | V | N | Y | G | A | S | Y | R | V | F | K | T |
| ABE5.5 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.6 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.7 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.8 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.9 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.10 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.11 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.12 | W | R | L | N | P | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE5.13 | W | R | H | N | P | | L | D | F | S | V | N | Y | A | A | S | Y | R | V | F | K | K |
| ABE5.14 | W | R | H | N | S | | L | N | F | C | V | N | Y | G | A | S | Y | R | V | F | K | K |
| ABE6.1 | W | R | H | N | S | | L | N | F | S | V | N | Y | G | N | S | Y | R | V | F | K | K |
| ABE6.2 | W | R | H | N | T | V | L | N | F | S | V | N | Y | G | N | S | Y | R | V | F | N | K |
| ABE6.3 | W | R | L | N | S | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE6.4 | W | R | L | N | S | | L | N | F | S | V | N | Y | G | N | C | Y | R | V | F | N | K |
| ABE6.5 | W | R | L | N | T | V | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE6.6 | W | R | L | N | T | V | L | N | F | S | V | N | Y | G | N | C | Y | R | V | F | N | K |
| ABE7.1 | W | R | L | N | A | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE7.2 | W | R | L | N | A | | L | N | F | S | V | N | Y | G | N | C | Y | R | V | F | N | K |
| ABE7.3 | L | R | L | N | A | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE7.4 | R | R | L | N | A | | L | N | F | S | V | N | Y | G | A | C | Y | R | V | F | N | K |
| ABE7.5 | W | R | L | N | A | | L | N | F | S | V | N | Y | G | A | C | Y | H | V | F | N | K |
| ABE7.6 | W | R | L | N | A | | L | N | I | S | V | N | Y | G | A | C | Y | P | V | F | N | K |
| ABE7.7 | L | R | L | N | A | | L | N | F | S | V | N | Y | G | A | C | Y | P | V | F | N | K |
| ABE7.8 | L | R | L | N | A | | L | N | F | S | V | N | Y | G | N | C | Y | R | V | F | N | K |
| ABE7.9 | L | R | L | N | A | | L | N | F | S | V | N | Y | G | N | C | Y | P | V | F | N | K |
| ABE7.10 | R | R | L | N | A | | L | N | F | S | V | N | Y | G | A | C | Y | P | V | F | N | K |

[1742]

[1743] 일부 실시형태에서, 염기 편집기는 8세대 ABE(ABE8)이다. 일부 실시형태에서, ABE8은 TadA*8 변이체를 포함한다. 일부 실시형태에서, ABE8은 TadA*8 변이체를 포함하는 단량체 작제물을 갖는다("ABE8.x-m"). 일부 실시형태에서, ABE8은 Y147T 돌연변이(TadA*8.1)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.1-m이다. 일부 실시형태에서, ABE8은 Y147R 돌연변이(TadA*8.2)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.2-m이다. 일부 실시형태에서, ABE8은 Q154S 돌연변이(TadA*8.3)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.3-m이다. 일부 실시형태에서, ABE8은 Y123H 돌연변이(TadA*8.4)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.4-m이다. 일부 실시형태에서, ABE8은 V82S 돌연변이(TadA*8.5)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.5-m이다. 일부 실시형태에서, ABE8은 T166R 돌연변이(TadA*8.6)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.6-m이다. 일부 실시형태에서, ABE8은 Q154R 돌연변이

(TadA*8.7)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.7-m이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 Y123H 돌연변이(TadA*8.8)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.8-m이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 I76Y 돌연변이(TadA*8.9)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.9-m이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 T166R 돌연변이(TadA*8.10)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.10-m이다. 일부 실시형태에서, ABE8은 Y147T 및 Q154R 돌연변이(TadA*8.11)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.11-m이다. 일부 실시형태에서, ABE8은 Y147T 및 Q154S 돌연변이(TadA*8.12)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.12-m이다.

[1744] 일부 실시형태에서, ABE8은 Y123H(H123Y에서 Y123H로 복귀됨), Y147R, Q154R 및 I76Y 돌연변이(TadA*8.13)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.13-m이다. 일부 실시형태에서, ABE8은 I76Y 및 V82S 돌연변이(TadA*8.14)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.14-m이다. 일부 실시형태에서, ABE8은 V82S 및 Y147R 돌연변이(TadA*8.15)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.15-m이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Y147R 돌연변이(TadA*8.16)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.16-m이다. 일부 실시형태에서, ABE8은 V82S 및 Q154R 돌연변이(TadA*8.17)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.17-m이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Q154R 돌연변이(TadA*8.18)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.18-m이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨), Y147R 및 Q154R 돌연변이(TadA*8.19)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.19-m이다. 일부 실시형태에서, ABE8은 I76Y, V82S, Y123H(H123Y에서 Y123H로 복귀됨), Y147R 및 Q154R 돌연변이(TadA*8.20)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.20-m이다. 일부 실시형태에서, ABE8은 Y147R 및 Q154S 돌연변이(TadA*8.21)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.21-m이다. 일부 실시형태에서, ABE8은 V82S 및 Q154S 돌연변이(TadA*8.22)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.22-m이다. 일부 실시형태에서, ABE8은 V82S 및 Y123H(H123Y에서 Y123H로 복귀됨) 돌연변이(TadA*8.23)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.23-m이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Y147T 돌연변이(TadA*8.24)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8.24-m이다.

[1745] 일부 실시형태에서, ABE8은 TadA*8 변이체에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는다("ABE8.x-d"). 일부 실시형태에서, ABE8은 Y147T 돌연변이(TadA*8.1)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.1-d이다. 일부 실시형태에서, ABE8은 Y147R 돌연변이(TadA*8.2)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.2-d이다. 일부 실시형태에서, ABE8은 Q154S 돌연변이(TadA*8.3)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.3-d이다. 일부 실시형태에서, ABE8은 Y123H 돌연변이(TadA*8.4)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.4-d이다. 일부 실시형태에서, ABE8은 V82S 돌연변이(TadA*8.5)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.5-d이다. 일부 실시형태에서, ABE8은 T166R 돌연변이(TadA*8.6)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.6-d이다. 일부 실시형태에서, ABE8은 Q154R 돌연변이(TadA*8.7)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.7-d이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 Y123H 돌연변이(TadA*8.8)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.8-d이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 I76Y 돌연변이(TadA*8.9)가 있는 TadA*7.10에 융합

된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.9-d이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 T166R 돌연변이(TadA*8.10)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.10-d이다. 일부 실시형태에서, ABE8은 Y147T 및 Q154R 돌연변이(TadA*8.11)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.11-d이다. 일부 실시형태에서, ABE8은 Y147T 및 Q154S 돌연변이(TadA*8.12)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.12-d이다. 일부 실시형태에서, ABE8은 Y123H(H123Y에서 Y123H로 복귀됨), Y147R, Q154R 및 I76Y 돌연변이(TadA*8.13)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.13-d이다. 일부 실시형태에서, ABE8은 I76Y 및 V82S 돌연변이(TadA*8.14)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.14-d이다. 일부 실시형태에서, ABE8은 V82S 및 Y147R 돌연변이(TadA*8.15)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.15-d이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Y147R 돌연변이(TadA*8.16)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.16-d이다. 일부 실시형태에서, ABE8은 TadA*7.10에 융합된 야생형 대장균 V82S 및 Q154R 돌연변이(TadA*8.17)가 있는 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.17-d이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Q154R 돌연변이(TadA*8.18)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.18-d이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨), Y147R 및 Q154R 돌연변이(TadA*8.19)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.19-d이다. 일부 실시형태에서, ABE8은 I76Y, V82S, Y123H(H123Y에서 Y123H로 복귀됨), Y147R 및 Q154R 돌연변이(TadA*8.20)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.20-d이다. 일부 실시형태에서, ABE8은 Y147R 및 Q154S 돌연변이(TadA*8.21)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.21-d이다. 일부 실시형태에서, ABE8은 V82S 및 Q154S 돌연변이(TadA*8.22)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.22-d이다. 일부 실시형태에서, ABE8은 V82S 및 Y123H(H123Y에서 Y123H로 복귀됨) 돌연변이(TadA*8.23)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.23-d이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Y147T 돌연변이(TadA*8.24)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8.24-d이다.

[1746] 일부 실시형태에서, ABE8은 TadA*8 변이체에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는다("ABE8.x-7"). 일부 실시형태에서, ABE8은 Y147T 돌연변이(TadA*8.1)가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.1-7이다. 일부 실시형태에서, ABE8은 Y147R 돌연변이(TadA*8.2)가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.2-7이다. 일부 실시형태에서, ABE8은 TadA*7.10에 융합된 Q154S 돌연변이(TadA*8.3)가 있는 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.3-7이다. 일부 실시형태에서, ABE8은 Y123H 돌연변이(TadA*8.4))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.4-7이다. 일부 실시형태에서, ABE8은 V82S 돌연변이(TadA*8.5))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.5-7이다. 일부 실시형태에서, ABE8은 T166R 돌연변이(TadA*8.6))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.6-7이다. 일부 실시형태에서, ABE8은 Q154R 돌연변이(TadA*8.7))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.7-7이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 Y123H 돌연변이(TadA*8.8))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.8-7이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 I76Y 돌연변이(TadA*8.9))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.9-7이다. 일부 실시형태에서, ABE8은 Y147R, Q154R 및 T166R 돌연변이

(TadA*8.10))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.10-7이다. 일부 실시형태에서, ABE8은 Y147T 및 Q154R 돌연변이(TadA*8.11))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.11-7이다. 일부 실시형태에서, ABE8은 Y147T 및 Q154S 돌연변이(TadA*8.12))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.12-7이다. 일부 실시형태에서, ABE8은 Y123H(H123Y에서 Y123H로 복귀됨), Y147R, Q154R 및 I76Y 돌연변이(TadA*8.13))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.13-7이다. 일부 실시형태에서, ABE8은 I76Y 및 V82S 돌연변이(TadA*8.14))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.14-7이다. 일부 실시형태에서, ABE8은 V82S 및 Y147R 돌연변이(TadA*8.15))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.15-7이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Y147R 돌연변이(TadA*8.16))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.16-7이다. 일부 실시형태에서, ABE8은 V82S 및 Q154R 돌연변이(TadA*8.17))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.17-7이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Q154R 돌연변이(TadA*8.18))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.18-7이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨), Y147R 및 Q154R 돌연변이(TadA*8.19))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.19-7이다. 일부 실시형태에서, ABE8은 I76Y, V82S, Y123H(H123Y에서 Y123H로 복귀됨), Y147R 및 Q154R 돌연변이(TadA*8.20))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.20-7이다. 일부 실시형태에서, ABE8은 Y147R 및 Q154S 돌연변이(TadA*8.21))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.21-7이다. 일부 실시형태에서, ABE8은 V82S 및 Q154S 돌연변이(TadA*8.22))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.22-7이다. 일부 실시형태에서, ABE8은 V82S 및 Y123H(H123Y에서 Y123H로 복귀됨) 돌연변이(TadA*8.23))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.23-7이다. 일부 실시형태에서, ABE8은 V82S, Y123H(H123Y에서 Y123H로 복귀됨) 및 Y147T 돌연변이(TadA*8.24))가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8.24-7이다.

[1747] 일부 실시형태에서, ABE는 아래 표 13에 나타나 있는 바와 같은 ABE8.1-m, ABE8.2-m, ABE8.3-m, ABE8.4-m, ABE8.5-m, ABE8.6-m, ABE8.7-m, ABE8.8-m, ABE8.9-m, ABE8.10-m, ABE8.11-m, ABE8.12-m, ABE8.13-m, ABE8.14-m, ABE8.15-m, ABE8.16-m, ABE8.17-m, ABE8.18-m, ABE8.19-m, ABE8.20-m, ABE8.21-m, ABE8.22-m, ABE8.23-m, ABE8.24-m, ABE8.1-d, ABE8.2-d, ABE8.3-d, ABE8.4-d, ABE8.5-d, ABE8.6-d, ABE8.7-d, ABE8.8-d, ABE8.9-d, ABE8.10-d, ABE8.11-d, ABE8.12-d, ABE8.13-d, ABE8.14-d, ABE8.15-d, ABE8.16-d, ABE8.17-d, ABE8.18-d, ABE8.19-d, ABE8.20-d, ABE8.21-d, ABE8.22-d, ABE8.23-d 또는 ABE8.24-d이다.

표 13

아데노신 데아미네이스 염기 편집 8 (ABE8) 변이체

| ABE8 | 아데노신 데아미네이스 | 아데노신 데아미네이스 설명 |
|---|---|---|
| ABE8.1-m | TadA*8.1 | 단량체_TadA*7.10 + Y147T |
| ABE8.2-m | TadA*8.2 | 단량체_TadA*7.10 + Y147R |
| ABE8.3-m | TadA*8.3 | 단량체_TadA*7.10 + Q154S |
| ABE8.4-m | TadA*8.4 | 단량체_TadA*7.10 + Y123H |
| ABE8.5-m | TadA*8.5 | 단량체_TadA*7.10 + V82S |
| ABE8.6-m | TadA*8.6 | 단량체_TadA*7.10 + T166R |
| ABE8.7-m | TadA*8.7 | 단량체_TadA*7.10 + Q154R |
| ABE8.8-m | TadA*8.8 | 단량체_TadA*7.10 + Y147R_Q154R_Y123H |
| ABE8.9-m | TadA*8.9 | 단량체_TadA*7.10 + Y147R_Q154R_I76Y |
| ABE8.10-m | TadA*8.10 | 단량체_TadA*7.10 + Y147R_Q154R_T166R |
| ABE8.11-m | TadA*8.11 | 단량체_TadA*7.10 + Y147T_Q154R |
| ABE8.12-m | TadA*8.12 | 단량체_TadA*7.10 + Y147T_Q154S |
| ABE8.13-m | TadA*8.13 | 단량체_TadA*7.10 + Y123H_Y147R_Q154R_I76Y |
| ABE8.14-m | TadA*8.14 | 단량체_TadA*7.10 + I76Y_V82S |
| ABE8.15-m | TadA*8.15 | 단량체_TadA*7.10 + V82S_Y147R |
| ABE8.16-m | TadA*8.16 | 단량체_TadA*7.10 + V82S_Y123H_Y147R |
| ABE8.17-m | TadA*8.17 | 단량체_TadA*7.10 + V82S_Q154R |
| ABE8.18-m | TadA*8.18 | 단량체_TadA*7.10 + V82S_Y123H_Q154R |
| ABE8.19-m | TadA*8.19 | 단량체_TadA*7.10 + V82S_Y123H_Y147R_Q154R |
| ABE8.20-m | TadA*8.20 | 단량체_TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R |
| ABE8.21-m | TadA*8.21 | 단량체_TadA*7.10 + Y147R_Q154S |
| ABE8.22-m | TadA*8.22 | 단량체_TadA*7.10 + V82S_Q154S |
| ABE8.23-m | TadA*8.23 | 단량체_TadA*7.10 + V82S_Y123H |

[1748]

| | | |
|---|---|---|
| ABE8.24-m | TadA*8.24 | 단량체_TadA*7.10 + V82S_Y123H_Y147T |
| ABE8.1-d | TadA*8.1 | 이종이량체_(WT) + (TadA*7.10 + Y147T) |
| ABE8.2-d | TadA*8.2 | 이종이량체_(WT) + (TadA*7.10 + Y147R) |
| ABE8.3-d | TadA*8.3 | 이종이량체_(WT) + (TadA*7.10 + Q154S) |
| ABE8.4-d | TadA*8.4 | 이종이량체_(WT) + (TadA*7.10 + Y123H) |
| ABE8.5-d | TadA*8.5 | 이종이량체_(WT) + (TadA*7.10 + V82S) |
| ABE8.6-d | TadA*8.6 | 이종이량체_(WT) + (TadA*7.10 + T166R) |
| ABE8.7-d | TadA*8.7 | 이종이량체_(WT) + (TadA*7.10 + Q154R) |
| ABE8.8-d | TadA*8.8 | 이종이량체_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H) |
| ABE8.9-d | TadA*8.9 | 이종이량체_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y) |
| ABE8.10-d | TadA*8.10 | 이종이량체_(WT) + (TadA*7.10 + Y147R_Q154R_T166R) |
| ABE8.11-d | TadA*8.11 | 이종이량체_(WT) + (TadA*7.10 + Y147T_Q154R) |
| ABE8.12-d | TadA*8.12 | 이종이량체_(WT) + (TadA*7.10 + Y147T_Q154S) |
| ABE8.13-d | TadA*8.13 | 이종이량체_(WT) + (TadA*7.10 + Y123H_Y147T_Q154R_I76Y) |
| ABE8.14-d | TadA*8.14 | 이종이량체_(WT) + (TadA*7.10 + I76Y_V82S) |
| ABE8.15-d | TadA*8.15 | 이종이량체_(WT) + (TadA*7.10 + V82S_ Y147R) |
| ABE8.16-d | TadA*8.16 | 이종이량체_(WT) + (TadA*7.10 + V82S_Y123H_Y147R) |
| ABE8.17-d | TadA*8.17 | 이종이량체_(WT) + (TadA*7.10 + V82S_Q154R) |
| ABE8.18-d | TadA*8.18 | 이종이량체_(WT) + (TadA*7.10 + V82S_Y123H_Q154R) |
| ABE8.19-d | TadA*8.19 | 이종이량체_(WT) + (TadA*7.10 + V82S_Y123H_Y147R_Q154R) |
| ABE8.20-d | TadA*8.20 | 이종이량체_(WT) + (TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R) |

[1749]

| | | |
|---|---|---|
| ABE8.21-d | TadA*8.21 | 이종이량체_(WT) + (TadA*7.10 + Y147R_Q154S) |
| ABE8.22-d | TadA*8.22 | 이종이량체_(WT) + (TadA*7.10 + V82S_Q154S) |
| ABE8.23-d | TadA*8.23 | 이종이량체_(WT) + (TadA*7.10 + V82S_Y123H) |
| ABE8.24-d | TadA*8.24 | 이종이량체_(WT) + (TadA*7.10 + V82S_Y123H_Y147T) |

[1750]

[1751] 일부 실시형태에서, ABE8은 R26C, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I 및 D167N 돌연변이(TadA*8a)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8a-m이다. 일부 실시형태에서, ABE8은 V88A,

A109S, T111R, D119N, H122N, F149Y, T166I 및 D167N 돌연변이(TadA*8b)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8b-m이다. 일부 실시형태에서, ABE8은 R26C, A109S, T111R, D119N, H122N, F149Y, T166I 및 D167N 돌연변이(TadA*8c)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8c-m이다. 일부 실시형태에서, ABE8은 V88A, T111R, D119N 및 F149Y 돌연변이(TadA*8d)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8d-m이다. 일부 실시형태에서, ABE8은 A109S, T111R, D119N, H122N, Y147D, F149Y, T166I 및 D167N 돌연변이(TadA*8e)가 있는 TadA*7.10을 포함하는 단량체 작제물을 갖는 ABE8e-m이다.

[1752] 일부 실시형태에서, ABE8은 R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I 및 D167N 돌연변이(TadA*8a)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8a-d이다. 일부 실시형태에서, ABE8은 V88A, A109S, T111R, D119N, H122N, F149Y, T166I 및 D167N 돌연변이(TadA*8b)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8b-d이다. 일부 실시형태에서, ABE8은 R26C, A109S, T111R, D119N, H122N, F149Y, T166I 및 D167N 돌연변이(TadA*8c)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8c-d이다. 일부 실시형태에서, ABE8은 V88A, T111R, D119N 및 F149Y 돌연변이(TadA*8d)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8d-d이다. 일부 실시형태에서, ABE8은 A109S, T111R, D119N, H122N, Y147D, F149Y, T166I 및 D167N 돌연변이(TadA*8e)가 있는 TadA*7.10에 융합된 야생형 대장균 TadA를 포함하는 이종이량체 작제물을 갖는 ABE8e-d이다.

[1753] 일부 실시형태에서, ABE8은 R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I 및 D167N 돌연변이(TadA*8a)가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8a-7이다. 일부 실시형태에서, ABE8은 V88A, A109S, T111R, D119N, H122N, F149Y, T166I 및 D167N 돌연변이(TadA*8b)가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8b-7이다. 일부 실시형태에서, ABE8은 R26C, A109S, T111R, D119N, H122N, F149Y, T166I 및 D167N 돌연변이(TadA*8c)가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8c-7이다. 일부 실시형태에서, ABE8은 V88A, T111R, D119N 및 F149Y 돌연변이(TadA*8d)가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8d-7이다. 일부 실시형태에서, ABE8은 A109S, T111R, D119N, H122N, Y147D, F149Y, T166I 및 D167N 돌연변이(TadA*8e)가 있는 TadA*7.10에 융합된 TadA*7.10을 포함하는 이종이량체 작제물을 갖는 ABE8e-7이다.

[1754] 일부 실시형태에서, ABE는 아래 표 14에 나타나 있는 바와 같은 ABE8a-m, ABE8b-m, ABE8c-m, ABE8d-m, ABE8e-m, ABE8a-d, ABE8b-d, ABE8c-d, ABE8d-d 또는 ABE8e-d이다. 일부 실시형태에서, ABE는 ABE8e-m 또는 ABE8e-d이다. ABE8e는 SpCas9 이외의 Cas 상동체, 예를 들어, SaCas9, SaCas9-KKH, Cas12a 상동체, 예를 들어, LbCas12a, enAs-Cas12a, SpCas9-NG 및 원형으로 순열된 CP1028-SpCas9 및 CP1041-SpCas9와 함께 사용될 때 효율적인 아데닌 염기 편집 활성 및 낮은 indel 형성을 보여준다. 표 14에서 ABE8e에 대해 나타낸 돌연변이에 더하여, V106W 치환을 TadA 도메인에 도입함으로써 표적외 RNA 및 DNA 편집이 감소되었다(그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[M. Richter *et al.*, 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z]에 기술된 바와 같음).

표 14

추가적인 아데노신 데아미네이스 염기 편집 8 변이체

| ABE8 염기 편집 | 아데노신 데아미네이스 | 아데노신 데아미네이스 설명 |
|---|---|---|
| ABE8a-m | TadA*8a | 단량체_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N |
| ABE8b-m | TadA*8b | 단량체_TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N |
| ABE8c-m | TadA*8c | 단량체_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N |
| ABE8d-m | TadA*8d | 단량체_TadA*7.10 + V88A + T111R + D119N + F149Y |
| ABE8e-m | TadA*8e | 단량체_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N |
| ABE8a-d | TadA*8a | 이종이량체_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N) |
| ABE8b-d | TadA*8b | 이종이량체_(WT) + (TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N) |
| ABE8c-d | TadA*8c | 이종이량체_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N) |
| ABE8d-d | TadA*8d | 이종이량체_(WT) + (TadA*7.10 + V88A + T111R + D119N + F149Y) |
| ABE8e-d | TadA*8e | 이종이량체_(WT) + (TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N) |

[1755]

[1756] 일부 실시형태에서, 염기 편집기(예를 들어, ABE8)는 아데노신 데아미네이스 변이체(예를 들어, TadA*8)를 원형 순열 Cas9(예를 들어, CP5 또는 CP6) 및 이분된 핵 국부화 서열을 포함하는 스캐폴드에 클로닝함으로써 생성된다. 일부 실시형태에서, 염기 편집기(예를 들어, ABE7.9, ABE7.10 또는 ABE8)는 NGC PAM CP5 변이체(S. 피오게네스 Cas9 또는 spVRQR Cas9)이다. 일부 실시형태에서, 염기 편집기(예를 들어, ABE7.9, ABE7.10 또는 ABE8)는 AGA PAM CP5 변이체(S. 피오게네스 Cas9 또는 spVRQR Cas9)이다. 일부 실시형태에서, 염기 편집기(예를 들어, ABE7.9, ABE7.10 또는 ABE8)는 NGC PAM CP6 변이체(S. 피오게네스 Cas9 또는 spVRQR Cas9)이다. 일부 실시형태에서, 염기 편집기(예를 들어, ABE7.9, ABE7.10 또는 ABE8)는 AGA PAM CP6 변이체(S. 피오게네스 Cas9 또는 spVRQR Cas9)이다.

[1757] 일부 실시형태에서, ABE는 아래 표 15에 나타낸 바와 같은 유전자형을 갖는다.

표 15

ABE 의 유전자형

| | 23 | 26 | 36 | 37 | 48 | 49 | 51 | 72 | 84 | 87 | 105 | 108 | 123 | 125 | 142 | 145 | 147 | 152 | 155 | 156 | 157 | 161 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABE7.9 | L | R | L | N | A | | L | N | F | S | V | N | Y | G | N | C | Y | P | V | F | N | K |
| ABE7.10 | R | R | L | N | A | | L | N | F | S | V | N | Y | G | A | C | Y | P | V | F | N | K |

[1758]

[1759] 아래 표 16에 나타낸 바와 같이, 40개의 ABE8의 유전자형이 기재되어 있다. ABE의 진화된 대장균 TadA 부분의 잔기 위치가 표시되어 있다. ABE8의 돌연변이 변화는 ABE7.10 돌연변이와 구별되는 경우 나타난다. 일부 실시형태에서, ABE는 아래 표 16에 나타낸 바와 같은 ABE 중 하나의 유전자형을 갖는다.

표 16

진화된 TadA 에서 잔기의 확인

| | 23 | 36 | 48 | 51 | 76 | 82 | 84 | 106 | 108 | 123 | 146 | 147 | 152 | 154 | 155 | 156 | 157 | 166 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABE7.10 | R | L | A | L | I | V | F | V | N | Y | C | Y | P | Q | V | F | N | T |
| ABE8.1-m | | | | | | | | | | | | T | | | | | | |
| ABE8.2-m | | | | | | | | | | | | R | | | | | | |
| ABE8.3-m | | | | | | | | | | | | | | S | | | | |
| ABE8.4-m | | | | | | | | | | H | | | | | | | | |
| ABE8.5-m | | | | | | S | | | | | | | | | | | | |
| ABE8.6-m | | | | | | | | | | | | | | | | | | R |
| ABE8.7-m | | | | | | | | | | | | | | R | | | | |
| ABE8.8-m | | | | | | | | | | H | | R | | R | | | | |
| ABE8.9-m | | | | | Y | | | | | | | R | | R | | | | |
| ABE8.10-m | | | | | | | | | | | | R | | R | | | | R |
| ABE8.11-m | | | | | | | | | | | | T | | R | | | | |
| ABE8.12-m | | | | | | | | | | | | T | | S | | | | |
| ABE8.13-m | | | | | Y | | | | | H | | R | | R | | | | |
| ABE8.14-m | | | | | Y | S | | | | | | | | | | | | |
| ABE8.15-m | | | | | | S | | | | | | R | | | | | | |
| ABE8.16-m | | | | | | S | | | | H | | R | | | | | | |
| ABE8.17-m | | | | | | S | | | | | | | | R | | | | |
| ABE8.18-m | | | | | | S | | | | H | | | | R | | | | |
| ABE8.19-m | | | | | | S | | | | H | | R | | R | | | | |
| ABE8.20-m | | | | | Y | S | | | | H | | R | | R | | | | |
| ABE8.21-m | | | | | | | | | | | | R | | S | | | | |
| ABE8.22-m | | | | | | S | | | | | | | | S | | | | |
| ABE8.23-m | | | | | | S | | | | H | | | | | | | | |
| ABE8.24-m | | | | | | S | | | | H | | T | | | | | | |
| ABE8.1-d | | | | | | | | | | | | T | | | | | | |
| ABE8.2-d | | | | | | | | | | | | R | | | | | | |
| ABE8.3-d | | | | | | | | | | | | | | S | | | | |
| ABE8.4-d | | | | | | | | | | H | | | | | | | | |
| ABE8.5-d | | | | | | S | | | | | | | | | | | | |
| ABE8.6-d | | | | | | | | | | | | | | | | | | R |
| ABE8.7-d | | | | | | | | | | | | | | R | | | | |
| ABE8.8-d | | | | | | | | | | H | | R | | R | | | | |

[1760]

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABE8.9-d | | | | | Y | | | | | | | R | | R | | | | |
| ABE8.10-d | | | | | | | | | | | | R | | R | | | | R |
| ABE8.11-d | | | | | | | | | | | | T | | R | | | | |
| ABE8.12-d | | | | | | | | | | | | T | | S | | | | |
| ABE8.13-d | | | | | Y | | | | | H | | R | | R | | | | |
| ABE8.14-d | | | | | Y | S | | | | | | | | | | | | |
| ABE8.15-d | | | | | | S | | | | | | R | | | | | | |
| ABE8.16-d | | | | | | S | | | | H | | R | | | | | | |
| ABE8.17-d | | | | | | S | | | | | | | | R | | | | |
| ABE8.18-d | | | | | | S | | | | H | | | | R | | | | |
| ABE8.19-d | | | | | | S | | | | H | | R | | R | | | | |
| ABE8.20-d | | | | | Y | S | | | | H | | R | | R | | | | |
| ABE8.21-d | | | | | | | | | | | | R | | S | | | | |
| ABE8.22-d | | | | | | S | | | | | | | | S | | | | |
| ABE8.23-d | | | | | | S | | | | H | | | | | | | | |
| ABE8.24-d | | | | | | S | | | | H | | T | | | | | | |

[1761]

[1762] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.1이다:

[1763] ABE8.1_Y147T_CP5_NGC PAM_단량체

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTK
EVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA

[1764]

KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVK
YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG
TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEF
ESPKKKRKV*

[1765]

[1766] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된

서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타낸다.

[1767] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.1이다:

[1768] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_단량체

[1769]
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSG*GGSSGSETPGTSESATPESS*
*GGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAK
VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTK
EVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSV

[1770]
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVK
YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG
TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEF
ESPKKKRKV*

[1771] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타낸다.

[1772] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.14이다:

[1773] pNMG-357_ABE8.14 with NGC PAM CP5

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDGGSSGGS*SGSETPGTSESATPESSG*
*GSSGGS*MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARK
EYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLA
IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR
RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS
GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRT
ADGSEFESPKKKRKV*

[1774]

[1775] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타낸다.

[1776] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.8-m이다:

[1777] ABE8.8-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1778]

[1779] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1780] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.8-d이다:

[1781] ABE8.8-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGSS*EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
HPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN
REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP
ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD
YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF
ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1782]

[1783] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1784] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.13-m이다:

[1785] ABE8.13-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1786]

[1787] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1788] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.13-d이다:

[1789] ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGSS*EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
HPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN
REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP
ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD

[1790]

YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF
ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1791]

[1792] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1793] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.17-m이다:

[1794] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK

[1795]

VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1796]

[1797] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1798] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.17-d이다:

[1799] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGSS*EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN
REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED

[1800]

REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP
ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD
YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF
ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1801]

[1802] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1803] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.20-m이다:

[1804] ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK

[1805]

ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1806]

[1807] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1808] 일부 실시형태에서, 염기 편집기는 아데노신 데아미네이스 활성을 갖는 다음 서열 또는 이의 단편을 포함하거나 또는 이로 본질적으로 구성되는 ABE8.20-d이다:

[1809] ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
HPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL
SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN
REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP
ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGD
YKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF
ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[1810]

[1811] 위의 서열에서, 일반 텍스트는 아데노신 데아미네이스 서열을 나타내고, 굵은 글씨의 서열은 Cas9로부터 유래된 서열을 나타내며, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄이 그어진 서열은 이분된 핵 국부화 서열을 나타내며, 이중으로 밑줄이 그어진 서열은 돌연변이를 나타낸다.

[1812] 일부 실시형태에서, 본 발명의 ABE8은 다음 서열로부터 선택된다:

[1813] 01. monoABE8.1_bpNLS + Y147T

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1814]

[1815] 02. monoABE8.1_bpNLS + Y147R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCRFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1816]

[1817] 03. monoABE8.1_bpNLS + Q154S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1818]

[1819] 04. monoABE8.1_bpNLS + Y123H

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1820]

[1821] 05. monoABE8.1_bpNLS + V82S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1822]

[1823] 06. monoABE8.1_bpNLS + T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1824]

[1825] 07. monoABE8.1_bpNLS + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1826]

[1827] 08. monoABE8.1_bpNLS + Y147R_Q154R_Y123H

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1828]

[1829] 09. monoABE8.1_bpNLS + Y147R_Q154R_I76Y

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1830]

[1831] 10. monoABE8.1_bpNLS + Y147R_Q154R_T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1832]

[1833] 11. monoABE8.1_bpNLS + Y147T_Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1834]

[1835] 12. monoABE8.1_bpNLS + Y147T_Q154S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCTFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1836]

[1837] 13. monoABE8.1_bpNLS + H123Y123H_Y147R_Q154R_I76Y

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1838]

[1839] 14. monoABE8.1_bpNLS + V82S + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESS
GGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP
ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK
ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK
VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE
MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGR
KRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS
EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRS
TKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[1840]

[1841] 일부 실시형태에서, 염기 편집기는 9세대 ABE(ABE9)이다. 일부 실시형태에서, ABE9는 TadA*9 변이체를 포함한다. ABE9 염기 편집기는 본 명세서에 기재된 바와 같이 ABE 7*10 참조 서열에 대한 변경을 포함하는 아미노산 서열을 포함하는 아데노신 데아미네이스 변이체를 포함한다. 표 19에서 사용되는 용어 "단량체"는 표 19에 기재된 변경을 포함하는 TadA*7.10의 단량체 형태를 지칭한다. 표 19에서 사용되는 용어 "이종이량체"는 본 명세서에 기재된 바와 같은 표 19에 기재된 변경을 포함하는 TadA*7.10에 융합된 지정된 야생형 대장균 TadA 아데노신 데아미네이스를 지칭한다. ABE9 염기 편집기에 대한 세부사항은 그 전문이 참조에 의해 본 명세서에 원용되어 있는 PCT 출원 PCT/2020/049975에 기술되어 있다.

표 17

아데노신 데아미네이스 염기 편집 9 (ABE9) 변이체

| ABE9 설명 | 변경 |
|---|---|
| ABE9.1_단량체 | E25F, V82S, Y123H, T133K, Y147R, Q154R |
| ABE9.2_단량체 | E25F, V82S, Y123H, Y147R, Q154R |
| ABE9.3 _단량체 | V82S, Y123H, P124W, Y147R, Q154R |
| ABE9.4_단량체 | L51W, V82S, Y123H, C146R, Y147R, Q154R |
| ABE9.5_단량체 | P54C, V82S, Y123H, Y147R, Q154R |
| ABE9.6_단량체 | Y73S, V82S, Y123H, Y147R, Q154R |
| ABE9.7_단량체 | N38G, V82T, Y123H, Y147R, Q154R |
| ABE9.8_단량체 | R23H, V82S, Y123H, Y147R, Q154R |
| ABE9.9_단량체 | R21N, V82S, Y123H, Y147R, Q154R |
| ABE9.10_단량체 | V82S, Y123H, Y147R, Q154R, A158K |
| ABE9.11_단량체 | N72K, V82S, Y123H, D139L, Y147R, Q154R, |
| ABE9.12_단량체 | E25F, V82S, Y123H, D139M, Y147R, Q154R |
| ABE9.13_단량체 | M70V, V82S, M94V, Y123H, Y147R, Q154R |
| ABE9.14_단량체 | Q71M, V82S, Y123H, Y147R, Q154R |
| ABE9.15_이종이량체 | E25F, V82S, Y123H, T133K, Y147R, Q154R |
| ABE9.16_이종이량체 | E25F, V82S, Y123H, Y147R, Q154R |
| ABE9.17_이종이량체 | V82S, Y123H, P124W, Y147R, Q154R |
| ABE9.18_이종이량체 | L51W, V82S, Y123H, C146R, Y147R, Q154R |
| ABE9.19_이종이량체 | P54C, V82S, Y123H, Y147R, Q154R |
| ABE9.2 _이종이량체 | Y73S, V82S, Y123H, Y147R, Q154R |
| ABE9.21_이종이량체 | N38G, V82T, Y123H, Y147R, Q154R |
| ABE9.22_이종이량체 | R23H, V82S, Y123H, Y147R, Q154R |
| ABE9.23_이종이량체 | R21N, V82S, Y123H, Y147R, Q154R |
| ABE9.24_이종이량체 | V82S, Y123H, Y147R, Q154R, A158K |

[1842]

| | |
|---|---|
| ABE9.25_이종이량체 | N72K, V82S, Y123H, D139L, Y147R, Q154R, |
| ABE9.26_이종이량체 | E25F, V82S, Y123H, D139M, Y147R, Q154R |
| ABE9.27 _이종이량체 | M70V, V82S, M94V, Y123H, Y147R, Q154R |
| ABE9.28_이종이량체 | Q71M, V82S, Y123H, Y147R, Q154R |
| ABE9.29_단량체 | E25F_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.30_단량체 | I76Y_V82T_Y123H_Y147R_Q154R |
| ABE9.31_단량체 | N38G_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.32_단량체 | N38G_I76Y_V82T_Y123H_Y147R_Q154R |
| ABE9.33_단량체 | R23H_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.34_단량체 | P54C_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.35_단량체 | R21N_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.36_단량체 | I76Y_V82S_Y123H_D138M_Y147R_Q154R |
| ABE9.37_단량체 | Y72S_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.38_이종이량체 | E25F_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.39_이종이량체 | I76Y_V82T_Y123H_Y147R_Q154R |
| ABE9.40_이종이량체 | N38G_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.41_이종이량체 | N38G_I76Y_V82T_Y123H_Y147R_Q154R |
| ABE9.42_이종이량체 | R23H_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.43_이종이량체 | P54C_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.44_이종이량체 | R21N_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.45_이종이량체 | I76Y_V82S_Y123H_D138M_Y147R_Q154R |
| ABE9.46_이종이량체 | Y72S_I76Y_V82S_Y123H_Y147R_Q154R |
| ABE9.47_단량체 | N72K_V82S, Y123H, Y147R, Q154R |
| ABE9.48_단량체 | Q71M_V82S, Y123H, Y147R, Q154R |
| ABE9.49_단량체 | M70V,V82S, M94V, Y123H, Y147R, Q154R |
| ABE9.50_단량체 | V82S, Y123H, T133K, Y147R, Q154R |
| ABE9.51_단량체 | V82S, Y123H, T133K, Y147R, Q154R, A158K |
| ABE9.52_단량체 | M70V,Q71M,N72K,V82S, Y123H, Y147R, Q154R |
| ABE9.53_이종이량체 | N72K_V82S, Y123H, Y147R, Q154R |
| ABE9.54_이종이량체 | Q71M_V82S, Y123H, Y147R, Q154R |

[1843]

| | |
|---|---|
| ABE9.55_이종이량체 | M70V,V82S, M94V, Y123H, Y147R, Q154R |
| ABE9.56_이종이량체 | V82S, Y123H, T133K, Y147R, Q154R |
| ABE9.57_이종이량체 | V82S, Y123H, T133K, Y147R, Q154R, A158K |
| ABE9.58_이종이량체 | M70V, Q71M, N72K, V82S, Y123H, Y147R, Q154R |

[1844]

[1845] 일부 실시형태에서, 염기 편집기는 핵염기 편집 도메인(예를 들어, 데아미네이스 도메인의 전부 또는 일부)에 융합된 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인(예를 들어, Cas9-유래 도메인)를 포함하는 융합 단백질이다. 소정의 실시형태에서, 본 명세서에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성을 개선하는 하나 이상의 특징을 포함한다. 예를 들어, 본 명세서에 제공된 임의의 융합 단백질은 뉴클레이스 활성이 감소된 Cas9 도메인을 포함할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 뉴클레이스 활성(dCas9)이 없는 Cas9 도메인 또는 Cas9 닉케이스(nCas9)로 지칭되는 이중나선 DNA 분자의 하나의 가닥을 절단하는 Cas9 도메인을 가질 수 있다.

[1846] 일부 실시형태에서, 염기 편집기는 우라실 글리코실레이스 저해제(UGI)의 전부 또는 일부를 포함하는 도메인을 더 포함한다. 일부 실시형태에서, 염기 편집기는 우라실 DNA 글리코실레이스(UDG)와 같은 우라실 결합 단백질(UBP)의 전부 또는 일부를 포함하는 도메인을 포함한다. 일부 실시형태에서, 염기 편집기는 핵산 폴리머레이스의 전부 또는 일부를 포함하는 도메인을 포함한다. 일부 실시형태에서, 염기 편집기는 핵산 폴리머레이스(NAP)의 전부 또는 일부를 도메인으로 포함할 수 있다. 예를 들어, 염기 편집기는 진핵생물 NAP의 전부 또는 일부를 포함할 수 있다. 일부 실시형태에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 DNA 폴리머레이스이다. 일부 실시형태에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 트랜스병변 폴리머레이스 활성을 갖는다. 일부 실시형태에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 전이 DNA 폴리머레이스이다. 일부 실시형태에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 Rev7, Rev1 복합체, 폴리머레이스 아이오타, 폴리머레이스 카파 또는 폴리머레이스 에타이다. 일부 실시형태에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 진핵생물 폴리머레이스 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 아이오타, 카파, 람다, 뮤 또는 누 성분이다. 일부 실시형태에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 핵산 폴리머레이스(예를 들어, 전이 DNA 폴리머레이스)와 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, 염기 편집기에 혼입된 핵산 폴리머레이스 또는 이의 일부는 전이 DNA 폴리머레이스이다.

[1847] 일부 실시형태에서, 염기 편집기의 도메인은 다중 도메인을 포함할 수 있다. 예를 들어, Cas9로부터 유래되는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기는 야생형 또는 천연 Cas9의 REC 로브 및 NUC 로브에 상응하는 REC 로브 및 NUC 로브를 포함할 수 있다. 또 다른 예에서, 염기 편집기는 RuvCI 도메인, BH 도메인, REC1 도메인, REC2 도메인, RuvCII 도메인, L1 도메인, HNH 도메인, L2 도메인, RuvCIII 도메인, WED 도메인, TOPO 도메인 또는 CTD 도메인 중 하나 이상을 포함할 수 있다. 일부 실시형태에서, 하나 이상의 염기 편집기의 도메인은 도메인을 포함하는 폴리펩타이드의 야생형 버전에 비해 돌연변이(예를 들어, 치환, 삽입, 결실)를 포함한다. 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인의 HNH 도메인은 H840A 치환을 포함할 수 있다. 또 다른 예에서, 폴리뉴클레오타이드 프로그램 가능한 DNA 결합 도메인의 RuvCI 도메인은 D10A 치환을 포함할 수 있다.

[1848] 본 명세서에 개시된 염기 편집기의 상이한 도메인(예를 들어, 인접 도메인)은 하나 이상의 링커 도메인(예를 들어, XTEN 링커 도메인)을 사용하거나 또는 사용하지 않고 연결될 수 있다. 일부 실시형태에서, 링커 도메인은 결합(예를 들어, 공유 결합), 화학기 또는 2개의 분자 또는 모이어티, 예를 들어, 제1 도메인(예를 들어, Cas9-유래 도메인) 및 제2 도메인(예를 들어, 아데노신 데아미네이스 도메인 또는 사이티딘 데아미네이스 도메인)과 같은, 예를 들어, 융합 단백질의 2개의 도메인을 연결하는 분자일 수 있다. 일부 실시형태에서, 링커는 공유 결합(예를 들어, 탄소-탄소 결합, 이황화물 결합, 탄소-헤테로 원자 결합 등)이다. 소정의 실시형태에서, 링커는 아마이드 연결의 탄소 질소 결합이다. 소정의 실시형태에서, 링커는 고리형 또는 비고리형, 치환 또는 비치환, 분지형 또는 비분지형 지방족 또는 헤테로지방족 링커이다. 소정의 실시형태에서, 링커는 중합체(예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아마이드, 폴리에스터 등)이다. 소정의 실시형태에서, 링커는 아미노알칸산의 단량체, 이량체 또는 중합체를 포함한다. 일부 실시형태에서, 링커는 아미노알칸산(예를 들어, 글리신, 에탄산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노뷰탄산, 5-펜탄산 등)을 포함한다. 일부 실시형태에서, 링커는 아미노헥산산(aminohexanoic acid: Ahx)의 단량체, 이량체 또는 중합체를 포함한다. 소정의 실시형태에서, 링커는 탄소고리형 모이어티(예를 들어, 사이클로펜테인, 사이클로헥세인)를 기반으로 한다. 다른 실시형태에서, 링커는 폴리에틸렌 글리콜 모이어티(PEG)를 포함한다. 소정의 실시형태에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 소정의 실시형태에서, 링커는 페닐 고리를 기반으로 한다. 링커는 펩타이드로부터 링커로의 친핵체(예를 들어, 티올, 아미노)의 부착을 용이하게 하는 기능화된 모이어티를 포함할 수 있다. 임의의 친전자체가 링커의 일부로서 사용될 수 있다. 예시적인 친전자체는 활성화된 에스터, 활성화된 아마이드, Michael 수용체, 알킬 할라이드, 아릴 할라이드, 아실 할라이드 및 아이소티오사이아네이트를 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 링커는 Cas9 뉴클레이스 도메인을 포함하는 RNA-프로그램 가능한 뉴클레이스의 gRNA 결합 도메인과 핵산 편집 단백질의 촉매 도메인을 연결한다. 일부 실시형태에서, 링커는 dCas9와 제2 도메인(예를 들어, UGI 등)을 연결한다.

[1849] ***링커***

[1850] 소정의 실시형태에서, 링커는 본 발명의 임의의 펩타이드 또는 펩타이드 도메인을 연결하는데 사용될 수 있다. 링커는 공유 결합만큼 단순할 수 있거나 또는 많은 원자 길이의 중합체 링커일 수 있다. 소정의 실시형태에서, 링커는 폴리펩타이드이거나 또는 아미노산을 기반으로 한다. 다른 실시형태에서, 링커는 펩타이드와 유사하지 않다. 소정의 실시형태에서, 링커는 공유 결합(예를 들어, 탄소-탄소 결합, 이황화물 결합, 탄소-헤테로원자 결

합 등)이다. 소정의 실시형태에서, 링커는 아마이드 연결의 탄소-질소 결합이다. 소정의 실시형태에서, 링커는 고리형 또는 비고리형, 치환 또는 비치환, 분지형 또는 비분지형 지방족 또는 헤테로지방족 링커이다. 소정의 실시형태에서, 링커는 중합체(예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아마이드, 폴리에스터 등)이다. 소정의 실시형태에서, 링커는 아미노알칸산의 단량체, 이량체 또는 중합체를 포함한다. 소정의 실시형태에서, 링커는 아미노알칸산(예를 들어, 글리신, 에탄산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노뷰탄산, 5-펜탄산 등)을 포함한다. 소정의 실시형태에서, 링커는 아미노헥산산(Ahx)의 단량체, 이량체 또는 중합체를 포함한다. 소정의 실시형태에서, 링커는 탄소고리형 모이어티(예를 들어, 사이클로펜테인, 사이클로헥세인)를 기반으로 한다. 다른 실시형태에서, 링커는 폴리에틸렌 글리콜 모이어티(PEG)를 포함한다. 다른 실시형태에서, 링커는 아미노산을 포함한다. 소정의 실시형태에서, 링커는 펩타이드를 포함한다. 소정의 실시형태에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 소정의 실시형태에서, 링커는 페닐 고리를 기반으로 한다. 링커는 펩타이드로부터 링커로의 친핵체(예를 들어, 티올, 아미노)의 부착을 용이하게 하기 위해 기능화된 모이어티를 포함할 수 있다. 임의의 친전자체는 링커의 일부로 사용될 수 있다. 예시적인 친전자체는 활성화된 에스터, 활성화된 아마이드, Michael 수용체, 알킬 할라이드, 아릴 할라이드, 아실 할라이드 및 아이소티오사이아네이트를 포함하지만, 이들로 제한되지 않는다.

[1851] 전형적으로, 링커는 두 그룹, 분자 또는 다른 모이어티 사이에 또는 그 측면에 위치하며, 공유 결합을 통해 각각 연결되어 둘을 연결한다. 일부 실시형태에서, 링커는 아미노산 또는 복수의 아미노산(예를 들어, 펩타이드 또는 단백질)이다. 일부 실시형태에서, 링커는 유기 분자, 그룹, 중합체 또는 화학 모이어티이다. 일부 실시형태에서, 링커는 2개 내지 100개의 아미노산 길이, 예를 들어, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 30개 내지 35개, 35개 내지 40개, 40개 내지 45개, 45개 내지 50개, 50개 내지 60개, 60개 내지 70개, 70개 내지 80개, 80개 내지 90개, 90개 내지 100개, 100개 내지 150개 또는 150개 내지 200개의 아미노산 길이이다. 일부 실시형태에서, 링커는 약 3개 내지 약 104개(예를 들어, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개, 55개, 60개, 65개, 70개, 75개, 80개, 85개, 90개, 95개 또는 100개)의 아미노산 길이이다. 더 길거나 또는 더 짧은 링커도 또한 상정된다.

[1852] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 링커를 통해 서로 융합된 사이티딘 또는 아데노신 데아미네이스와 Cas9 도메인을 포함한다. 사이티딘 또는 아데노신 데아미네이스 핵염기 편집기에 대한 활성을 위한 최적의 길이를 달성하기 위해 사이티딘 또는 아데노신 데아미네이스와 Cas9 도메인 사이의 다양한 링커 길이 및 가요성이 이용될 수 있다(예를 들어, $(GGGS)_n$, $(GGGGS)_n$, 및 $(G)_n$ 형식의 매우 가요성인 링커에서 $(EAAAK)_n$, $(SGGS)_n$, SGSETPGTSESATPES(예를 들어, 문헌[Guilinger JP, *et al.* Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82] 참조; 그 전체 내용이 참조에 의해 본 명세서에 원용됨) 및 $(XP)_n$ 형식의 매우 강성인 링커에 이르기까지 다양함)). 일부 실시형태에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15이다. 일부 실시형태에서, 링커는 $(GGS)_n$ 모티프를 포함하되, n은 1, 3 또는 7이다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 사이티딘 데아미네이스 또는 아데노신 데아미네이스 및 Cas9 도메인은 XTEN 링커로도 지칭될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다. 일부 실시형태에서, 링커는 복수의 프롤린 잔기를 포함하며, 5개 내지 21개, 5개 내지 14개, 5개 내지 9개, 5개 내지 7개의 아미노산 길이인, 예를 들어, PAPAP, PAPAPA, PAPAPAP, PAPAPAPA, $P(AP)_4$, $P(AP)_7$, $P(AP)_{10}$이다(예를 들어, 문헌[Tan J, Zhang F, Karcher D, Bock R. Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat Commun. 2019 Jan 25;10(1):439] 참조; 전체 내용이 참조에 의해 본 명세서에 원용됨). 이러한 프롤린-풍부 링커는 또한 "강성" 링커라고도 한다.

[1853] 또 다른 실시형태에서, 염기 편집 시스템은 데아미네이스(DNA 데아미네이스), 예를 들어, 아데노신 또는 사이티딘 데아미네이스와 비공유적으로 상호작용하며, 최소한의 또는 감소된 방관자 또는 표적-인접 효과를 갖는 특정 편집을 위해 아데노신 또는 사이티딘 데아미네이스를 표적 폴리뉴클레오타이드 서열의 표적 핵염기로 일시적으로 유인하는 성분(단백질)을 포함한다. 데아미네이스-상호작용 단백질을 포함하는 이러한 비공유 시스템 및 방법은 DNA 데아미네이스를 특정 게놈 표적 핵염기로 유인하는 역할을 하며, 표적 및 표적-인접 편집의 이벤트를 분리하여 보다 정확한 단일 염기 치환 돌연변이의 달성을 향상시킨다. 실시형태에서, 데아미네이스-상호작용 단

백질은 데아미네이스의 활성(촉매) 부위가 표적 핵염기(예를 들어, 각각 아데노신 또는 사이티딘)와 결합하는 것을 차단 또는 방해하지 않고 데아미네이스(예를 들어, 아데노신 데아미네이스 또는 사이티딘 데아미네이스)에 결합한다. "MagnEdit"라고 하는 이러한 시스템은 Cas9 및 gRNA 복합체에 테더링된 단백질 간의 상호작용을 포함하고, 특정 게놈 표적 부위를 편집하기 위해 공동 발현된 아데노신 또는 사이티딘 데아미네이스(외인성 또는 내인성)를 유도할 수 있으며, 이는 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[McCann, J. *et al.*, 2020, "MagnEdit - interacting factors that recruit DNA-editing enzymes to single base targets," Life-Science-Alliance, Vol. 3, No. 4 (e201900606), (doi 10.26508/lsa.201900606)]에 기술되어 있다. 실시형태에서, DNA 데아미네이스는 본 명세서에 기재된 바와 같은 아데노신 데아미네이스 변이체(예를 들어, TadA*8)이다.

[1854] 또 다른 실시형태에서, "Suntag"라고 하는 시스템은, 예를 들어, 이들 각각의 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Tanenbaum, M.E. *et al.*, "A protein tagging system for signal amplification in gene expression and fluorescence imaging," *Cell*. 2014 October 23; 159(3): 635-646. doi:10.1016/j.cell.2014.09.039; 및 Huang, Y.-H. *et al.*, 2017, "DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A," *Genome Biol* 18: 176. doi:10.1186/s13059-017-1306-z]에 기술되어 있는 바와 같이, 인접 표적 편집이 감소된 부위에서 염기 편집을 달성하기 위해 염기 편집기의 단백질(예를 들어, 아데노신 데아미네이스 또는 사이티딘 데아미네이스) 성분 또는 이의 다중 카피를 폴리뉴클레오타이드 표적 부위로 동원하는 데 사용되는 비공유적으로 상호작용하는 성분을 포함한다. 실시형태에서, DNA 데아미네이스는 본 명세서에 기재된 바와 같은 아데노신 데아미네이스 변이체(예를 들어, TadA*8)이다.

[1855] ***가이드 RNA를 갖는 핵산 프로그램 가능 DNA 결합 단백질***

[1856] 숙주 세포, 예를 들어, 면역 세포(예를 들어, T-세포 또는 NK-세포)에서의 염기 편집을 위한 조성물 및 방법이 본 명세서에 제공된다. 가이드 다중핵산 서열, 예를 들어, 가이드 RNA 서열, 또는 본 명세서에 제공된 바와 같은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 가이드 RNA의 조합을 포함하는 조성물이 본 명세서에 추가로 제공된다. 일부 실시형태에서, 본 명세서에 제공된 바와 같은 염기 편집을 위한 조성물은 염기 편집기, 예를 들어, C-염기 편집기 또는 A-염기 편집기를 암호화하는 폴리뉴클레오타이드를 추가로 포함한다. 예를 들어, 염기 편집을 위한 조성물은 BE, BE4, ABE 및 제공된 바와 같은 하나 이상의 가이드 RNA의 조합을 암호화하는 mRNA 서열을 포함할 수 있다. 염기 편집을 위한 조성물은 염기 편집기 폴리펩타이드 및 본 명세서에 제공된 바와 같은 임의의 가이드 RNA 중 하나 이상의 조합을 포함할 수 있다. 이러한 조성물은 상이한 전달 접근법, 예를 들어, 전기천공법, 뉴클레오펙션, 바이러스 형질도입 또는 형질감염을 통해 면역 세포에서 염기 편집을 수행하는데 사용될 수 있다. 일부 실시형태에서, 염기 편집을 위한 조성물은 염기 편집기를 암호화하는 mRNA 서열 및 전기천공법을 위해 본 명세서에 제공된 하나 이상의 가이드 RNA 서열의 조합을 포함한다.

[1857] 본 개시내용의 일부 양태는 본 명세서에 제공된 임의의 융합 단백질, 및 융합 단백질의 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 도메인(예를 들어, Cas9(예를 들어, dCas9, 뉴클레이스 활성이 있는 Cas9 또는 Cas9 닉케이스) 또는 Cas12)에 결합된 가이드 RNA를 포함하는 복합체를 제공한다. 이들 복합체는 리보핵단백질(RNP)이라고도 한다. 일부 실시형태에서, 가이드 핵산(예를 들어, 가이드 RNA)은 15개 내지 100개의 뉴클레오타이드 길이이고, 표적 서열에 상보적인 적어도 10개의 연속 뉴클레오타이드의 서열을 포함한다. 일부 실시형태에서, 가이드 RNA는 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개 또는 50개의 뉴클레오타이드 길이이다. 일부 실시형태에서, 가이드 RNA는 표적 서열에 상보적인 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개 또는 40개의 연속 뉴클레오타이드의 서열을 포함한다. 일부 실시형태에서, 표적 서열은 DNA 서열이다. 일부 실시형태에서, 표적 서열은 RNA 서열이다. 일부 실시형태에서, 표적 서열은 세균, 효모, 균류, 곤충, 식물 또는 동물의 게놈의 서열이다. 일부 실시형태에서, 표적 서열은 인간 게놈의 서열이다. 일부 실시형태에서, 표적 서열의 3' 말단 은 표준 PAM 서열(NGG)에 바로 인접한다. 일부 실시형태에서, 표적 서열의 3' 말단은 비-표준 PAM 서열에 바로 인접한다(예를 들어, 표 5에 열거된 서열 또는 5'-NAA-3'). 일부 실시형태에서, 가이드 핵산(예를 들어, 가이드 RNA)은 관심 유전자(예를 들어, 질환 또는 장애 관련 유전자)의 서열에 상보적이다.

[1858] 본 개시내용의 일부 양태는 본 명세서에 제공된 융합 단백질 또는 복합체를 사용하는 방법을 제공한다. 예를 들

어, 본 개시내용의 일부 양태는 DNA 분자를 본 명세서에 제공된 임의의 융합 단백질 및 적어도 하나의 가이드 RNA와 접촉시키는 단계를 포함하는 방법을 제공하되, 가이드 RNA는 약 15개 내지 100개의 뉴클레오타이드 길이이고, 표적 서열에 상보적인 적어도 10개의 연속 뉴클레오타이드의 서열을 포함한다. 일부 실시형태에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG 또는 CAA 서열에 바로 인접한다. 일부 실시형태에서, 표적 서열의 3' 말단은 NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN 또는 5'(TTTV) 서열에 바로 인접한다. 일부 실시형태에서, 표적 서열의 3' 말단은, 예를 들어, TTN, DTTN, GTTN, ATTN, ATTC, DTTNT, WTTN, HATY, TTTN, TTTV, TTTC, TG, RTR 또는 YTN PAM 부위에 바로 인접한다.

[1859] 각 서열에서 특정 위치 또는 잔기의 넘버링은 사용된 특정 단백질 및 넘버링 방식에 의존한다는 것이 이해될 것이다. 넘버링은, 예를 들어, 성숙 단백질의 전구체 및 성숙 단백질 자체에서 다를 수 있으며, 종마다 서열의 차이가 넘버링에 영향을 미칠 수 있다. 당업자는 당업계에 잘 알려진 방법, 예를 들어, 서열 정렬 및 상동 잔기의 결정에 의해 임의의 상동 단백질 및 각각의 암호화 핵산에서 각 잔기를 확인할 수 있을 것이다.

[1860] 본 명세서에 개시된 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위로 표적화하기 위해, 전형적으로 가이드 RNA와 함께 융합 단백질을 공동 발현하는 것이 필요하다는 것이 당업자에게 명백할 것이다. 본 명세서의 다른 곳에 더 상세히 설명된 바와 같이, 가이드 RNA는 전형적으로 napDNAbp(예를 들어, Cas9 또는 Cas12) 결합을 가능하게 하는 tracrRNA 프레임워크, 및 napDNAbp:핵산 편집 효소/도메인 융합 단백질에 서열 특이성을 부여하는 가이드 서열을 포함한다. 대안적으로, 가이드 RNA 및 tracrRNA는 2개의 핵산 분자로서 별도로 제공될 수 있다. 일부 실시형태에서, 가이드 RNA는 구조를 포함하되, 가이드 서열은 표적 서열에 상보적인 서열을 포함한다. 가이드 서열은 전형적으로 20개의 뉴클레오타이드 길이이다. napDNAbp:핵산 편집 효소/도메인 융합 단백질을 특정 게놈 표적 부위로 표적화하기에 적합한 가이드 RNA의 서열은 본 개시내용에 기초하여 당업자에게 명백할 것이다. 이러한 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오타이드의 상류 또는 하류에서 50개의 뉴클레오타이드 이내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 임의의 제공된 융합 단백질을 특정 표적 서열로 표적화하기에 적합한 일부 예시적인 가이드 RNA 서열이 본 명세서에 제공된다.

[1861] sgRNA의 고유한 부분은 Cas9(예를 들어, SpyCas9) 및/또는 DNA 표적과 상호작용하는 다양한 특징을 형성할 것으로 예측된다. 6개의 보존된 모듈이 Cas9 엔도뉴클레이스 활성을 지시하는 천연 crRNA:tracrRNA 이중체 및 단일 가이드 RNA(sgRNA) 내에서 확인되었다(문헌[Briner *et al.*, Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339] 참조). 6개의 모듈은 DNA 표적화를 담당하는 스페이서, 상부 줄기, 벌지, CRISPR 반복:tracrRNA 이중체에 의해 형성된 하부 줄기, 넥서스(Nexus) 및 tracrRNA의 3' 말단으로부터의 헤어핀을 포함한다. 상부 및 하부 줄기는 주로 포스페이트 백본과의 서열-독립적 상호작용을 통해 Cas9와 상호작용한다. 일부 실시형태에서, 상부 줄기는 불필요하다. 일부 실시형태에서, 하부 줄기의 베이스에 있는 보존된 우라실 뉴클레오타이드 서열은 불필요하다. 벌지는 Cas9의 Rec1 도메인과의 특정 측쇄 상호작용에 참여한다. U44의 핵염기는 Tyr 325 및 His 328의 측쇄와 상호작용하는 반면, G43은 Tyr 329와 상호작용한다. 넥서스는 sgRNA:Cas9 상호작용의 코어를 형성하며, sgRNA와 Cas9 및 표적 DNA 사이의 교차점에 있다. A51 및 A52의 핵염기는 Phe 1105의 측쇄와 상호작용하고; U56은 Arg 457 및 Asn 459과 상호작용하며; U59의 핵염기는 Arg 74, Asn 77, Pro 475, Leu 455, Phe 446 및 Ile 448의 측쇄에 의해 정의된 소수성 포켓에 삽입되고; C60은 Leu 455, Ala 456 및 Asn 459와 상호작용하며, C61은 Arg 70의 측쇄와 상호작용하여 차례로 C15와 상호작용한다. 일부 실시형태에서, 이들 돌연변이 중 하나 이상이 sgRNA:Cas9 상호작용을 최적화하기 위해 Cas9(예를 들어, spyCas9)에 대한 sgRNA의 벌지 및/또는 넥서스에서 만들어진다.

[1862] 또한, tracrRNA 넥서스 및 헤어핀은 Cas9 페어링에 중요하며, 직교 Cas9 단백질을 추가로 활용하는데 중요한 이질적인 Cas9 단백질을 분리하는 교차 직교 장벽(cross orthogonality barrier)으로 교환될 수 있다. 일부 실시형태에서, 넥서스 및 헤어핀은 직교 Cas9 단백질을 표적화하도록 교환된다. 일부 실시형태에서, sgRNA는 보다 조밀하고 및 형태적으로 안정적인 가이드 RNA를 설계하기 위해 상부 줄기, 헤어핀 1 및/또는 하부 줄기의 서열 가요성에서 공급된다. 일부 실시형태에서, 모듈은 다양한 키메라 가이드를 갖는 단일 Cas9를 사용하거나 또는 키메라 sgRNA의 상이한 조합을 갖는 직교 시스템을 동시에 사용함으로써 다중 편집을 최적화하도록 변형된다. 가이드 기능적 모듈 및 이의 방법에 관한 세부사항은, 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Briner *et al.*, Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339]에 기술되어 있다.

[1863] 본 명세서에 개시된 염기 편집기의 도메인은 임의의 순서로 배열될 수 있다. 예를 들어, 폴리뉴클레오타이드-프로그램 가능한 뉴클레오타이드-결합 도메인(예를 들어, Cas9 또는 Cas12) 및 데아미네이스 도메인(예를 들어, 사이티딘 또는 아데노신 데아미네이스)을 포함하는 융합 단백질을 포함하는 염기 편집기의 비제한적인 예는 하

기와 같이 배열될 수 있다:

[1864] $NH_2$-[핵염기 편집 도메인]-링커1-[핵염기 편집 도메인]-COOH;

[1865] $NH_2$-[데아미네이스]-링커1-[핵염기 편집 도메인]-COOH;

[1866] $NH_2$-[데아미네이스]-링커1-[핵염기 편집 도메인]-링커2-[UGI]-COOH;

[1867] $NH_2$-[데아미네이스]-링커1-[핵염기 편집 도메인]-COOH;

[1868] $NH_2$-[아데노신 데아미네이스]-링커1-[핵염기 편집 도메인]-COOH;

[1869] $NH_2$-[핵염기 편집 도메인]-[데아미네이스]-COOH;

[1870] $NH_2$-[데아미네이스]-[핵염기 편집 도메인]-[이노신 BER 저해제]-COOH;

[1871] $NH_2$-[데아미네이스]-[이노신 BER 저해제]-[핵염기 편집 도메인]-COOH;

[1872] $NH_2$-[이노신 BER 저해제]-[데아미네이스]-[핵염기 편집 도메인]-COOH;

[1873] $NH_2$-[핵염기 편집 도메인]-[데아미네이스]-[이노신 BER 저해제]-COOH;

[1874] $NH_2$-[핵염기 편집 도메인]-[이노신 BER 저해제]-[데아미네이스]-COOH;

[1875] $NH_2$-[이노신 BER 저해제]-[핵염기 편집 도메인]-[데아미네이스]-COOH;

[1876] NH2-[핵염기 편집 도메인]-링커1-[데아미네이스]-링커2-[핵염기 편집 도메인]-COOH;

[1877] NH2-[핵염기 편집 도메인]-링커1-[데아미네이스]-[핵염기 편집 도메인]-COOH;

[1878] NH2-[핵염기 편집 도메인]-[데아미네이스]-링커2-[핵염기 편집 도메인]-COOH;

[1879] NH2-[핵염기 편집 도메인]-[데아미네이스]-[핵염기 편집 도메인]-COOH;

[1880] NH2-[핵염기 편집 도메인]-링커1-[데아미네이스]-링커2-[핵염기 편집 도메인]-[이노신 BER 저해제]-COOH;

[1881] NH2-[핵염기 편집 도메인]-링커1-[데아미네이스]-[핵염기 편집 도메인]-[이노신 BER 저해제]-COOH;

[1882] NH2-[핵염기 편집 도메인]-[데아미네이스]-링커2-[핵염기 편집 도메인]-[이노신 BER 저해제]-COOH;

[1883] NH2-[핵염기 편집 도메인]-[데아미네이스]-[핵염기 편집 도메인]-[이노신 BER 저해제]-COOH;

[1884] NH2-[이노신 BER 저해제]-[핵염기 편집 도메인]-링커1-[데아미네이스]-링커2-[핵염기 편집 도메인]-COOH;

[1885] NH2-[이노신 BER 저해제]-[핵염기 편집 도메인]-링커1-[데아미네이스]-[핵염기 편집 도메인]-COOH;

[1886] NH2-[이노신 BER 저해제]-[핵염기 편집 도메인]-[데아미네이스]-링커2-[핵염기 편집 도메인]-COOH; 또는

[1887] NH2-[이노신 BER 저해제]NH2-[핵염기 편집 도메인]-[데아미네이스]-[핵염기 편집 도메인]-COOH.

[1888] 일부 실시형태에서, 본 명세서에 제공된 염기 편집 융합 단백질은, 예를 들어, 표적 염기가 정의된 영역(예를 들어, "탈아미노화 창") 내에 배치되는 정확한 위치에 위치되어야 한다. 일부 실시형태에서, 표적은 4-염기 영역 내에 있을 수 있다. 일부 실시형태에서, 이러한 정의된 표적 영역은 PAM의 상류에서 대략 15개의 염기일 수 있다. 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); 및 Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)] 참조.

[1889] 정의된 표적 영역은 탈아미노화 창일 수 있다. 탈아미노화 창은 염기 편집기가 표적 뉴클레오타이드에 작용하고 이를 탈아미노화하는 정의된 영역일 수 있다. 일부 실시형태에서, 탈아미노화 창은 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개의 염기 영역 내에 있다. 일부 실시형태에서, 탈아미노화 창은 PAM의 상류에서 5개,

6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개 또는 25개의 염기이다.

[1890] 본 개시내용의 염기 편집기는 표적 폴리뉴클레오타이드 서열의 편집을 용이하게 하는 임의의 도메인, 특징 또는 아미노산 서열을 포함할 수 있다. 예를 들어, 일부 실시형태에서, 염기 편집기는 핵 국부화 서열(NLS)을 포함한다. 일부 실시형태에서, 염기 편집기의 NLS는 데아미네이스 도메인과 napDNAbp 도메인 사이에 국부화된다. 일부 실시형태에서, 염기 편집기의 NLS는 napDNAbp 도메인의 C-말단에 국부화된다.

[1891] 융합 단백질에 포함될 수 있는 단백질 도메인의 비제한적인 예는 다음 활성 중 하나 이상을 갖는 데아미네이스 도메인(예를 들어, 아데노신 데아미네이스 또는 사이티딘 데아미네이스), 우라실 글리코실레이스 저해제(UGI) 도메인, 에피토프 태그, 리포터 유전자 서열 및/또는 단백질 도메인: 메틸레이스 활성, 데메틸레이스 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 유전자 침묵 활성, 염색질 변형 활성, 후성유전학적 변형 활성, 히스톤 변형 활성, RNA 절단 활성 및 핵산 결합 활성을 포함한다. 추가적인 도메인은 이종성 기능적 도메인일 수 있다. 이러한 이종성 기능적 도메인은, 예를 들어, 히스톤 메틸화, 히스톤 아세틸화, 히스톤 유비퀴틴화 등을 유발하는 표적 DNA(예를 들어, 히스톤, DNA 결합 단백질 등)와 관련된 표적 폴리펩타이드의 변형과 같은 기능 활성을 부여할 수 있다. 부여된 다른 기능 및/또는 활성은 트랜스포세이즈 활성, 인테그레이스 활성, 재조합효소 활성, 라이게이스 활성, 유비퀴틴 라이게이스 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모화(SUMOylation) 활성, 탈수모화 활성 또는 위의 임의의 조합을 포함할 수 있다.

[1892] 부여된 다른 기능은 메틸트랜스퍼레이스 활성, 데메틸레이스 활성, 탈아미노화 활성, 디스뮤테이스 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그레이스 활성, 트랜스포세이즈 활성, 재조합효소 활성, 폴리머레이스 활성, 라이게이스 활성, 헬리케이스 활성, 포토라이에이스 활성 또는 글리코실레이스 활성, 아세틸트랜스퍼레이스 활성, 데아세틸레이스 활성, 카이네이스 활성, 포스파테이스 활성, 유비퀴틴 라이게이스 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모화 활성, 탈수모화 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 리모델링 활성, 프로테이스 활성, 옥시도리덕테이스 활성, 트랜스퍼레이스 활성, 하이드롤레이스 활성, 라이에이스 활성, 아이소머레이스 활성, 신테이스 활성, 신테테이스 활성 및 탈미리스토일화 활성 또는 이들의 임의의 조합을 포함할 수 있다.

[1893] 도메인은 에피토프 태그, 리포터 단백질, 다른 결합 도메인으로 검출되거나 또는 표지될 수 있다. 에피토프 태그의 비제한적인 예는 히스티딘(His) 태그, V5 태그, FLAG 태그, 인플루엔자 헤마글루티닌(HA) 태그, Myc 태그, VSV-G 태그 및 티오레독신(thioredoxin: Trx) 태그를 포함한다. 리포터 유전자의 예는 글루타티온-5-트랜스퍼레이스(GST), 서양고추냉이 과산화효소(horseradish peroxidase: HRP), 클로람페니콜 아세틸트랜스퍼레이스(chloramphenicol acetyltransferase: CAT) 베타-갈락토시데이스, 베타-글루쿠로니데이스, 루시퍼레이스, 녹색 형광 단백질(GFP), HcRed, DsRed, 청록색 형광 단백질(CFP), 황색 형광 단백질(YFP) 및 청색 형광 단백질(BFP)을 포함한 자가형광 단백질을 포함하지만, 이들로 제한되지 않는다. 추가적인 단백질 서열은 DNA 분자에 결합하거나 또는 말토스 결합 단백질(MBP), S-태그, Lex A DNA 결합 도메인(DBD) 융합, GAL4 DNA 결합 도메인 융합 및 단순 포진 바이러스(HSV) BP16 단백질 융합을 포함하지만 이들로 제한되지 않는 다른 세포 분자에 결합하는 아미노산 서열을 포함할 수 있다.

[1894] ***사이티딘 또는 아데노신 데아미네이스 및 Cas9 도메인을 포함하는 융합 단백질의 사용 방법***

[1895] 본 개시내용의 일부 양태는 본 명세서에 제공된 융합 단백질 또는 복합체의 사용 방법을 제공한다. 예를 들어, 본 개시내용의 일부 양태는 DNA 분자를 본 명세서에 제공된 임의의 융합 단백질 및 적어도 하나의 가이드 RNA와 접촉시키는 단계를 포함하는 방법을 제공하되, 가이드 RNA는 약 15개 내지 100개의 뉴클레오타이드 길이이고, 표적 서열에 상보적인 적어도 10개의 연속 뉴클레오타이드의 서열을 포함한다. 일부 실시형태에서, 표적 서열의 3' 말단은 표준 PAM 서열(NGG)에 바로 인접한다. 일부 실시형태에서, 표적 서열의 3' 말단은 표준 PAM 서열(NGG)에 바로 인접하지 않는다. 일부 실시형태에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG 또는 CAA 서열에 바로 인접한다. 일부 실시형태에서, 표적 서열의 3' 말단은 NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN 또는 5'(TTTV) 서열에 바로 인접한다.

[1896] 일부 실시형태에서, 본 발명의 융합 단백질 관심 표적을 돌연변이시키기 위해 사용된다. 특히, 본 명세서에 기재된 사이티딘 데아미네이스 또는 아데노신 데아미네이스 핵염기 편집기는 표적 서열 내에서 다중 돌연변이를 만들 수 있다. 이러한 돌연변이는 표적의 기능에 영향을 미칠 수 있다. 예를 들어, 사이티딘 데아미네이스 또는 아데노신 데아미네이스 핵염기 편집기가 조절 영역을 표적화하는데 사용되는 경우, 조절 영역의 기능이 변경되고 하류 단백질의 발현이 감소되거나 또는 제거된다.

[1897] 각 서열에서 특정 위치 또는 잔기의 넘버링은 사용된 특정 단백질 및 넘버링 방식에 의존한다는 것이 이해될 것이다. 넘버링은, 예를 들어, 성숙 단백질의 전구체 및 성숙 단백질 자체에서 다를 수 있으며, 종마다 서열의 차이가 넘버링에 영향을 미칠 수 있다. 당업자는 당업계에 잘 알려진 방법, 예를 들어, 서열 정렬 및 상동 잔기의 결정에 의해 임의의 상동 단백질 및 각각의 암호화 핵산에서 각 잔기를 확인할 수 있을 것이다.

[1898] 본 명세서에 개시된 바와 같은 Cas9 도메인 및 사이티딘 또는 아데노신 데아미네이스를 포함하는 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위로 표적화하기 위해, 전형적으로 가이드 RNA, 예를 들어, sgRNA와 함께 융합 단백질을 공동 발현하는 것이 필요하다는 것이 당업자에게 명백할 것이다. 본 명세서의 다른 곳에 더 상세히 설명된 바와 같이, 가이드 RNA는 전형적으로 Cas9 결합을 가능하게 하는 tracrRNA 프레임워크 및 Cas9:핵산 편집 효소/도메인 융합 단백질에 서열 특이성을 부여하는 가이드 서열을 포함한다. 대안적으로, 가이드 RNA 및 tracrRNA는 2개의 핵산 분자로서 별도로 제공될 수 있다. 일부 실시형태에서, 가이드 RNA는 구조를 포함하되, 가이드 서열은 표적 서열에 상보적인 서열을 포함한다. 가이드 서열은 전형적으로 20개의 뉴클레오타이드 길이이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특정 게놈 표적 부위로 표적화하기에 적합한 가이드 RNA의 서열은 본 개시내용에 기초하여 당업자에게 명백할 것이다. 이러한 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오타이드의 상류 또는 하류에서 50개의 뉴클레오타이드 이내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 표적화 임의의 제공된 융합 단백질을 특정 표적 서열로 표적화하기에 적합한 일부 예시적인 가이드 RNA 서열이 본 명세서에 제공된다.

[1899] ***염기 편집기 효율성***

[1900] 일부 실시형태에서, 본 명세서에 제공된 방법의 목적은 유전자 편집을 통해 유전자 산물의 정상적인 기능을 방해하는 것이다. 본 명세서에 제공된 핵염기 편집 단백질은 시험관내 유전자 편집-기반 인간 치료제에 대해 검증될 수 있다. 본 명세서에 제공된 핵염기 편집 단백질, 예를 들어, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인(예를 들어, Cas9) 및 핵염기 편집 도메인(예를 들어, 아데노신 데아미네이스 도메인 또는 사이티딘 데아미네이스 도메인)을 포함하는 융합 단백질은 A에서 G로 또는 C에서 T로 뉴클레오타이드를 편집하는데 사용될 수 있음을 당업자는 이해할 것이다.

[1901] CRISPR-Cas9 뉴클레이스는 표적화된 게놈 편집을 매개하기 위해 널리 사용되어 왔다. 대부분의 게놈 편집 응용에서, Cas9는 가이드 폴리뉴클레오타이드(예를 들어, 단일 가이드 RNA(sgRNA))와 복합체를 형성하고, sgRNA 서열에 의해 지정된 표적 부위에서 이중-가닥 DNA 파손(DSB)을 유도한다. 세포는 주로 비-상동 말단-연결(non-homologous end-joining: NHEJ) 복구 경로를 통해 이러한 DSB에 반응하며, 이는 유전자를 방해하는 프레임시프트 돌연변이를 일으킬 수 있는 확률론적 삽입 또는 결실(indel)을 초래한다. DSB의 측면에 있는 서열에 높은 수준의 상동성을 갖는 공여체 DNA 주형의 존재하에, 상동성 유도 복구(homology directed repair: HDR)로 알려진 대체 경로를 통해 유전자 교정이 달성될 수 있다. 불행히도, 대부분의 비-교란 조건하에서, HDR은 비효율적이고, 세포 상태 및 세포 유형에 따라 다르며, 더 많은 빈도의 indel에 의해 지배된다. 인간의 질환과 관련하여 알려진 유전적 변이의 대부분은 점 돌연변이이며, 보다 효율적이고 깔끔하게 정확한 점 돌연변이를 만들 수 있는 방법이 필요하다. 본 명세서에 제공된 바와 같은 염기 편집 시스템은 이중-가닥 DNA 파손을 생성하지 않고, 공여체 DNA 주형을 필요로 하지 않으며, 확률론적 삽입 및 결실을 과도하게 유도하지 않고 게놈 편집을 제공하는 새로운 방법을 제공한다.

[1902] 일부 실시형태에서, 본 개시내용은 상당한 수의 의도하지 않은 돌연변이, 예컨대, 의도하지 않은 점 돌연변이를 생성하지 않고 핵산(예를 들어, 대상체의 게놈 내의 핵산)에서 종결 코돈과 같은 의도된 돌연변이를 효율적으로 생성하는 염기 편집기를 제공한다. 일부 실시형태에서, 의도된 돌연변이는 의도된 돌연변이를 생성하도록 특별히 설계된 가이드 폴리뉴클레오타이드(예를 들어, gRNA)에 결합된 특정 염기 편집기(예를 들어, 아데노신 염기 편집기 또는 사이티딘 염기 편집기)에 의해 생성된 돌연변이이다. 일부 실시형태에서, 의도된 돌연변이는 질환 또는 장애, 예를 들어, T-세포 또는 NK-세포 악성종양과 관련된 표적 항원과 관련된 유전자에 있다. 일부 실시형태에서, 의도된 돌연변이는 질환 또는 장애, 예를 들어, T-세포 또는 NK-세포 악성종양과 관련된 표적 항원과 관련된 유전자에서 아데닌(A)에서 구아닌(G)으로의 점 돌연변이(예를 들어, SNP)이다. 일부 실시형태에서, 의도된 돌연변이는 유전자(예를 들어, 조절 영역 또는 요소)의 코딩 영역 또는 비-코딩 영역 내의 아데닌(A)에서 구아닌(G)으로의 점 돌연변이이다. 일부 실시형태에서, 의도된 돌연변이는 질환 또는 장애, 예를 들어, T-세포 또는 NK-세포 악성종양과 관련된 표적 항원과 관련된 유전자에서 사이토신(C)에서 티민(T)으로의 점 돌연변이(예를 들어, SNP)이다. 일부 실시형태에서, 의도된 돌연변이는 유전자의 코딩 영역 또는 비-코딩 영역(예를 들어, 조절 영역 또는 요소) 내의 사이토신(C)에서 티민(T)으로의 점 돌연변이이다. 일부 실시형태에서, 의도된 돌연변이는 종결 코돈, 예를 들어, 유전자의 코딩 영역 내에 조기 종결 코돈을 생성하는 점 돌연변이이다. 일부 실

시형태에서, 의도된 돌연변이는 종결 코돈을 제거하는 돌연변이이다.

[1903] 본 발명의 염기 편집기 유리하게는 상당한 비율의 indel을 생성하지 않으면서 단백질을 암호화하는 특정 뉴클레오타이드 염기를 변형시킨다. 본 명세서에서 사용되는 "indel"은 핵산 내의 뉴클레오타이드 염기의 삽입 또는 결실을 지칭한다. 이러한 삽입 또는 결실은 유전자의 코딩 영역 내에서 프레임 쉬프트 돌연변이를 유발할 수 있다. 일부 실시형태에서, 핵산에서 다수의 삽입 또는 결실(즉, indel)을 생성하지 않으면서 핵산 내의 특정 뉴클레오타이드를 효율적으로 변형(예를 들어, 돌연변이)시키는 염기 편집기를 생성하는 것이 바람직하다. 일부 실시형태에서, 핵산에서 다수의 삽입 또는 결실(즉, indel)을 생성하지 않으면서 핵산 내의 특정 뉴클레오타이드를 효율적으로 변형(예를 들어, 돌연변이 또는 메틸화)시키는 염기 편집기를 생성하는 것이 바람직하다. 소정의 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 의도된 변형(예를 들어, 메틸화) 대 indel의 더 큰 비율을 생성할 수 있다. 소정의 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 의도된 변형(예를 들어, 돌연변이) 대 indel의 더 큰 비율을 생성할 수 있다.

[1904] 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 1:1보다 큰 의도된 돌연변이 대 indel(즉, 의도된 점 돌연변이:의도하지 않은 점 돌연변이)의 비를 생성할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 200:1, 적어도 300:1, 적어도 400:1, 적어도 500:1, 적어도 600:1, 적어도 700:1, 적어도 800:1, 적어도 900:1 또는 적어도 1000:1 이상인 의도된 돌연변이 대 indel의 비를 생성할 수 있다. 의도된 돌연변이와 indel의 수는 임의의 적합한 방법을 사용하여 결정될 수 있다.

[1905] 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 핵산의 영역에서 indel의 형성을 제한할 수 있다. 일부 실시형태에서, 영역은 염기 편집기에 의해 표적화된 뉴클레오타이드 또는 염기 편집기에 의해 표적화된 뉴클레오타이드의 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개의 뉴클레오타이드 내의 영역에 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 핵산의 영역에서 indel의 형성을 1% 미만, 1.5% 미만, 2% 미만, 2.5% 미만, 3% 미만, 3.5% 미만, 4% 미만, 4.5% 미만, 5% 미만, 6% 미만, 7% 미만, 8% 미만, 9% 미만, 10% 미만, 12% 미만, 15% 미만 또는 20% 미만으로 제한할 수 있다. 핵산 영역에서 형성된 indel의 수는 핵산(예를 들어, 세포의 게놈 내의 핵산)이 염기 편집기에 노출된 시간의 양에 따라 달라질 수 있다. 일부 실시형태에서, indel의 수 또는 비율은 핵산(예를 들어, 세포의 게놈 내의 핵산)을 염기 편집기에 노출시킨 지 적어도 1시간, 적어도 2시간, 적어도 6시간, 적어도 12시간, 적어도 24시간, 적어도 36시간, 적어도 48시간, 적어도 3일, 적어도 4일, 적어도 5일, 적어도 7일, 적어도 10일 또는 적어도 14일 후에 결정된다.

[1906] 본 개시내용의 일부 양태는 본 명세서에 제공된 임의의 염기 편집기가 상당한 수의 의도하지 않은 돌연변이(예를 들어, 거짓 표적외 편집 또는 방관자 편집)를 생성하지 않으면서 핵산(예를 들어, 대상체의 게놈 내의 핵산)에서 의도된 돌연변이를 효율적으로 생성할 수 있다는 인식을 기반으로 한다. 일부 실시형태에서, 의도된 돌연변이는 의도된 돌연변이를 생성하도록 특별히 설계된 gRNA에 결합된 특정 염기 편집기에 의해 생성된 돌연변이이다. 일부 실시형태에서, 의도된 돌연변이는 종결 코돈, 예를 들어, 유전자의 코딩 영역 내의 조기 종결 코돈을 생성하는 돌연변이이다. 일부 실시형태에서, 의도된 돌연변이는 종결 코돈을 제거하는 돌연변이이다. 일부 실시형태에서, 의도된 돌연변이는 유전자의 스플라이싱을 변경하는 돌연변이이다. 일부 실시형태에서, 의도된 돌연변이는 유전자의 조절 서열(예를 들어, 유전자 프로모터 또는 유전자 억제자)을 변경하는 돌연변이이다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 1:1보다 큰 의도된 돌연변이 대 의도하지 않은 돌연변이(예를 들어, 의도된 돌연변이:의도하지 않은 돌연변이)의 비를 생성할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 150:1, 적어도 200:1, 적어도 250:1, 적어도 500:1 또는 적어도 1000:1 이상인 의도된 돌연변이 대 의도하지 않은 돌연변이의 비를 생성할 수 있다. 본 명세서에 기재된 염기 편집기의 특성은 임의의 융합 단백질, 또는 본 명세서에 제공된 융합 단백질을 사용하는 방법에 적용될 수 있음을 이해하여야 한다.

[1907] 염기 편집은 종종 유전적 변형, 유전자 변형 및 핵산 서열의 변형과 같은 과 같은 "변형"으로 지칭되며, 변형이 염기 편집 변형이라는 맥락에 기초하여 명확하게 이해할 수 있다. 따라서, 염기 편집 변형은, 예를 들어, 본 개시내용 전반에 걸쳐 논의된 데아미네이스 활성의 결과로서 뉴클레오타이드 염기 수준에서의 변형이며, 이는 이

어서 유전자 서열의 변화를 초래하고 유전자 산물에 영향을 미칠 수 있다. 따라서 본질적으로, 본 명세서에 기재된 유전자 편집 변형은 구조적으로 및/또는 기능적으로 유전자의 변형을 초래할 수 있되, 유전자의 발현 산물은 변형될 수 있고, 예를 들어, 유전자의 발현은 넉아웃되거나; 또는 반대로 향상되거나, 또는 일부 상황에서는 유전자 기능 또는 활성이 변형될 수 있다. 본 명세서에 개시된 방법을 사용하면, 염기 편집 효율은 염기 편집이 수행되는 유전자의 넉다운 효율로서 결정될 수 있되, 염기 편집은 유전자의 발현을 넉다운시키도록 의도된다. 넉다운 수준은, 예를 들어, 유세포 분석에 의한 단백질 발현 수준에 대한 검정과 같은 임의의 검출 검정; 정량적 RT-PCR, 노던 블롯 분석, 또는 파이로시퀀싱과 같은 임의의 기타 적합한 검정과 같은 RNA 발현을 검출하기 위한 검정에 의해 발현 수준을 결정함으로써 정량적으로 검증될 수 있고; 뉴클레오타이드 시퀀싱 반응에 의해 정성적으로 검증될 수 있다.

[1908] 일부 실시형태에서, 변형, 예를 들어, 단일 염기 편집은 유전자 표적화된 발현의 적어도 10%의 감소를 초래한다. 일부 실시형태에서, 염기 편집 효율은 유전자 표적화된 발현의 적어도 10%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 유전자 표적화된 발현의 적어도 20%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 유전자 표적화된 발현의 적어도 30%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 유전자 표적화된 발현의 적어도 40%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 유전자 표적화된 발현의 적어도 50%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 60%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 70%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 80%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 90%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 91%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 92%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 93%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 94%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 95%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 96%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 97%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 98%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자 발현의 적어도 99%의 감소를 초래할 수 있다. 일부 실시형태에서, 염기 편집 효율은 표적화된 유전자의 넉아웃(유전자 발현의 100% 넉다운)을 초래할 수 있다.

[1909] 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집 시스템은 표적 폴리뉴클레오타이드 서열에서 50% 미만, 40% 미만, 30% 미만, 20% 미만, 19% 미만, 18% 미만, 17% 미만, 16% 미만, 15% 미만, 14% 미만, 13% 미만, 12% 미만, 11% 미만, 10% 미만, 9% 미만, 8% 미만, 7% 미만, 6% 미만, 5% 미만, 4% 미만, 3% 미만, 2% 미만, 1% 미만, 0.9% 미만, 0.8% 미만, 0.7% 미만, 0.6% 미만, 0.5% 미만, 0.4% 미만, 0.3% 미만, 0.2% 미만, 0.1% 미만, 0.09% 미만, 0.08% 미만, 0.07% 미만, 0.06% 미만, 0.05% 미만, 0.04% 미만, 0.03% 미만, 0.02% 미만 또는 0.01% 미만의 indel 형성을 초래한다.

[1910] 일부 실시형태에서, 표적화된 변형, 예를 들어, 단일 염기 편집은 상이한 가이드 RNA를 이용한 염기 편집을 위해 적어도 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12 13개, 14개, 15개, 16개, 17 개,18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개 또는 50개의 상이한 내인성 서열을 표적화하기 위해 동시에 사용된다. 일부 실시형태에서, 표적화된 변형, 예를 들어, 단일 염기 편집은 상이한 가이드 RNA를 사용한 염기 편집을 위해 적어도 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12 13개, 14개, 15개, 16개, 17 개,18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 이상의 상이한 내인성 유전자 서열을 표적화하기 위해 순차적으로 사용된다.

[1911] 본 개시내용의 일부 양태는 본 명세서에 제공된 임의의 염기 편집기가 상당한 수의 의도하지 않은 돌연변이, 예컨대, 의도하지 않은 점 돌연변이(즉, 방관자 돌연변이)를 생성하지 않으면서 핵산(예를 들어, 대상체의 게놈 내의 핵산)에서 점 돌연변이와 같은 의도된 돌연변이를 효율적으로 생성할 수 있다는 인식을 기반으로 한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 적어도 0.01%의 의도된 돌연변이를 생성(즉, 적어도 0.01%의 염기 편집 효율)할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 적어도 0.01%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95% 또는 99%의

의도된 돌연변이를 생성할 수 있다.

[1912] 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 표적 폴리뉴클레오타이드 서열에서 50% 미만, 40% 미만, 30% 미만, 20% 미만, 19% 미만, 18% 미만, 17% 미만, 16% 미만, 15% 미만, 14% 미만, 13% 미만, 12% 미만, 11% 미만, 10% 미만, 9% 미만, 8% 미만, 7% 미만, 6% 미만, 5% 미만, 4% 미만, 3% 미만, 2% 미만, 1% 미만, 0.9% 미만, 0.8% 미만, 0.7% 미만, 0.6% 미만, 0.5% 미만, 0.4% 미만, 0.3% 미만, 0.2% 미만, 0.1% 미만, 0.09% 미만, 0.08% 미만, 0.07% 미만, 0.06% 미만, 0.05% 미만, 0.04% 미만, 0.03% 미만, 0.02% 미만 또는 0.01% 미만의 indel 형성을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 표적 폴리뉴클레오타이드 서열에서 0.8% 미만의 indel 형성을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 표적 폴리뉴클레오타이드 서열에서 최대 0.8%의 indel 형성을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 표적 폴리뉴클레오타이드 서열에서 0.3% 미만의 indel 형성을 초래한다. 일부 실시형태에서, 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7 염기 편집기 중 하나를 포함하는 염기 편집 시스템과 비교하여 표적 폴리뉴클레오타이드 서열에서 더 낮은 indel 형성을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7.10을 포함하는 염기 편집 시스템과 비교하여 표적 폴리뉴클레오타이드 서열에서 더 낮은 indel 형성을 초래한다.

[1913] 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7 염기 편집기 중 하나를 포함하는 염기 편집 시스템과 비교하여 indel 빈도가 감소한다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7 염기 편집기 중 하나를 포함하는 염기 편집 시스템과 비교하여 적어도 0.01%, 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 또는 적어도 95% indel 빈도가 감소한다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 염기 편집 시스템은 ABE7.10을 포함하는 염기 편집 시스템과 비교하여 적어도 0.01%, 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 또는 적어도 95% indel 빈도가 감소한다.

[1914] 본 발명은 증가된 효율성 및 특이성을 갖는 아데노신 데아미네이스 변이체(예를 들어, ABE8 변이체)를 제공한다. 특히, 본 명세서에 기재된 아데노신 데아미네이스 변이체는 폴리뉴클레오타이드 내에서 목적하는 염기를 편집할 가능성이 더 높고, 변경되도록 의도되지 않은 염기(예를 들어, "방관자")를 편집할 가능성이 적다.

[1915] 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 감소된 방관자 편집 또는 돌연변이를 감소시켰다. 일부 실시형태에서, 의도하지 않은 편집 또는 돌연변이는 방관자 돌연변이 또는 방관자 편집, 예를 들어, 표적 뉴클레오타이드 서열의 표적 창의 의도하지 않은 또는 비-표적 위치에서 표적 염기(예를 들어, A 또는 C)의 염기 편집이다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집 시스템과 비교하여 방관자 편집 또는 돌연변이를 감소시켰다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 방관자 편집 또는 돌연변이를 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집 시스템과 비교하여 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99% 감소시켰다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 방관자 편집 또는 돌연변이를 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집 시스템과 비교하여 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배 또는 적어도 3.0배 감소시켰다.

[1916] 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 거짓 편집을 감소시켰다. 일부 실시형태에서, 의도하지 않은 편집 또는 돌연변이는 거짓 돌연변이 또는 거짓

편집, 예를 들어, 게놈의 의도하지 않은 또는 비-표적 영역에서 표적 염기(예를 들어, A 또는 C)의 비-특이적 편집 또는 가이드 독립적 편집이다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집 시스템과 비교하여 거짓 편집을 감소시켰다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 거짓 편집을 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집 시스템과 비교하여 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99% 감소시켰다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 거짓 편집을 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집 시스템과 비교하여 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배 또는 적어도 3.0배 감소시켰다.

[1917] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 적어도 0.01%, 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99%의 염기 편집 효율을 갖는다. 일부 실시형태에서, 염기 편집 효율은 세포의 집단에서 편집된 핵염기의 백분율을 계산함으로써 측정될 수 있다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 세포의 집단에서 편집된 핵염기에 의해 측정되는 바와 같이 적어도 0.01%, 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99%의 염기 편집 효율을 갖는다.

[1918] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기와 비교하여 더 높은 염기 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기, 예를 들어, ABE7.10과 비교하여 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 105%, 적어도 110%, 적어도 115%, 적어도 120%, 적어도 125%, 적어도 130%, 적어도 135%, 적어도 140%, 적어도 145%, 적어도 150%, 적어도 155%, 적어도 160%, 적어도 165%, 적어도 170%, 적어도 175%, 적어도 180%, 적어도 185%, 적어도 190%, 적어도 195%, 적어도 200%, 적어도 210%, 적어도 220%, 적어도 230%, 적어도 240%, 적어도 250%, 적어도 260%, 적어도 270%, 적어도 280%, 적어도 290%, 적어도 300%, 적어도 310%, 적어도 320%, 적어도 330%, 적어도 340%, 적어도 350%, 적어도 360%, 적어도 370%, 적어도 380%, 적어도 390%, 적어도 400%, 적어도 450% 또는 적어도 500% 더 높은 염기 편집 효율을 갖는다.

[1919] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기, 예를 들어, ABE7.10과 비교하여 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배, 적어도 3.0배, 적어도 3.1배, 적어도 3.2, 적어도 3.3배, 적어도 3.4배, 적어도 3.5배, 적어도 3.6배, 적어도 3.7배, 적어도 3.8배, 적어도 3.9배, 적어도 4.0배, 적어도 4.1배, 적어도 4.2배, 적어도 4.3배, 적어도 4.4배, 적어도 4.5배, 적어도 4.6배, 적어도 4.7배, 적어도 4.8배, 적어도 4.9배 또는 적어도 5.0배 더 높은 염기 편집 효율을 갖는다.

[1920] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 적어도 0.01%, 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99%의 표적에 대한 염기 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 세포의 집단에서 편집된 표적 핵염기에 의해 측정된 바와 같이 적어도 0.01%, 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또

는 적어도 99%의 표적에 대한 염기 편집 효율을 갖는다.

[1921] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기와 비교하여 더 높은 표적에 대한 염기 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기, 예를 들어, ABE7.10과 비교하여 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 105%, 적어도 110%, 적어도 115%, 적어도 120%, 적어도 125%, 적어도 130%, 적어도 135%, 적어도 140%, 적어도 145%, 적어도 150%, 적어도 155%, 적어도 160%, 적어도 165%, 적어도 170%, 적어도 175%, 적어도 180%, 적어도 185%, 적어도 190%, 적어도 195%, 적어도 200%, 적어도 210%, 적어도 220%, 적어도 230%, 적어도 240%, 적어도 250%, 적어도 260%, 적어도 270%, 적어도 280%, 적어도 290%, 적어도 300%, 적어도 310%, 적어도 320%, 적어도 330%, 적어도 340%, 적어도 350%, 적어도 360%, 적어도 370%, 적어도 380%, 적어도 390%, 적어도 400%, 적어도 450% 또는 적어도 500% 더 높은 표적에 대한 염기 편집 효율을 갖는다.

[1922] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기, 예를 들어, ABE7.10과 비교하여 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배, 적어도 3.0배, 적어도 3.1배, 적어도 3.2배, 적어도 3.3배, 적어도 3.4배, 적어도 3.5배, 적어도 3.6배, 적어도 3.7배, 적어도 3.8배, 적어도 3.9배, 적어도 4.0배, 적어도 4.1배, 적어도 4.2배, 적어도 4.3배, 적어도 4.4배, 적어도 4.5배, 적어도 4.6배, 적어도 4.7배, 적어도 4.8배, 적어도 4.9배 또는 적어도 5.0배 더 높은 표적에 대한 염기 편집 효율을 갖는다.

[1923] 본 명세서에 기재된 ABE8 염기 편집기 변이체는 플라스미드, 벡터, LNP 복합체 또는 mRNA를 통해 숙주 세포에 전달될 수 있다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 mRNA로서 숙주 세포에 전달될 수 있다. 일부 실시형태에서, 핵산 기반 전달 시스템, 예를 들어, mRNA를 통해 전달된 ABE8 염기 편집기는 편집된 핵염기에 의해 측정되는 바와 같이 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99%의 표적에 대한 편집 효율을 갖는다. 일부 실시형태에서, mRNA 시스템에 의해 전달된 ABE8 염기 편집기는 플라스미드 또는 벡터 시스템에 의해 전달된 ABE8 염기 편집기와 비교하여 더 높은 염기 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 105%, 적어도 110%, 적어도 115%, 적어도 120%, 적어도 125%, 적어도 130%, 적어도 135%, 적어도 140%, 적어도 145%, 적어도 150%, 적어도 155%, 적어도 160%, 적어도 165%, 적어도 170%, 적어도 175%, 적어도 180%, 적어도 185%, 적어도 190%, 적어도 195%, 적어도 200%, 적어도 210%, 적어도 220%, 적어도 230%, 적어도 240%, 적어도 250%, 적어도 260%, 적어도 270%, 적어도 280%, 적어도 290%, 적어도 300% 더 높은, 적어도 310%, 적어도 320%, 적어도 330%, 적어도 340%, 적어도 350%, 적어도 360%, 적어도 370%, 적어도 380%, 적어도 390%, 적어도 400%, 적어도 450% 또는 적어도 500%의 표적에 대한 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배, 적어도 3.0배, 적어도 3.1배, 적어도 3.2배, 적어도 3.3배, 적어도 3.4배, 적어도 3.5배, 적어도 3.6배, 적어도 3.7배, 적어도 3.8배, 적어도 3.9배, 적어도 4.0배, 적어도 4.1배, 적어도 4.2배, 적어도 4.3배, 적어도 4.4배, 적어도 4.5배, 적어도 4.6배, 적어도 4.7배, 적어도 4.8배, 적어도 4.9배 또는 적어도 5.0배 더 높은 표적에 대한 편집 효율을 갖는다.

[1924] 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 표적 폴리뉴클레오타이드 서열에서 50% 미만, 40% 미만, 30% 미만, 20% 미만, 19% 미만, 18% 미만, 17%

미만, 16% 미만, 15% 미만, 14% 미만, 13% 미만, 12% 미만, 11% 미만, 10% 미만, 9% 미만, 8% 미만, 7% 미만, 6% 미만, 5% 미만, 4% 미만, 3% 미만, 2% 미만, 1% 미만, 0.9% 미만, 0.8% 미만, 0.7% 미만, 0.6% 미만, 0.5% 미만, 0.4% 미만, 0.3% 미만, 0.2% 미만, 0.1% 미만, 0.09% 미만, 0.08% 미만, 0.07% 미만, 0.06% 미만, 0.05% 미만, 0.04% 미만, 0.03% 미만, 0.02% 미만 또는 0.01% 미만의 표적외 편집을 초래한다.

[1925] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 더 낮은 가이드된 표적외 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99% 더 낮은 가이드된 표적외 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배 또는 적어도 3.0배 더 낮은 가이드된 표적외 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 가이드된 표적외 편집 효율을 적어도 약 2.2배 감소시켰다.

[1926] 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 더 낮은 가이드-독립적 표적외 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 또는 적어도 99% 더 낮은 가이드-독립적 표적외 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배, 적어도 3.0배, 적어도 5.0배, 적어도 10.0배, 적어도 20.0배, 적어도 50.0배, 적어도 70.0배, 적어도 100.0배, 적어도 120.0배, 적어도 130.0배 또는 적어도 150.0배 더 낮은 가이드-독립적 표적외 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달될 때와 비교하여 mRNA 시스템에 의해 전달될 때 가이드-독립적 표적외 편집 효율(예를 들어, 거짓 RNA 탈아미노화)를 134.0배 감소시켰다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체는 게놈 전반에 걸쳐 가이드-독립적 돌연변이 비율을 증가시키지 않는다.

[1927] 일부 실시형태에서, (예를 들어, 형질도입, 형질감염, 전기천공법 또는 임의의 다른 방법에 의한) 단일 유전자 전달 이벤트는 세포의 게놈 내에서 5개 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 6개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 7개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 전기천공법 이벤트는 세포의 게놈 내에서 8개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 9개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 10개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 20개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 30개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 40개의 서열의 염기 편집을 표적화하는데 사용될 수 있다. 일부 실시형태에서, 단일 유전자 전달 이벤트는 세포의 게놈 내에서 50개의 서열의 염기 편집을 표적화하는데 사용될 수 있다.

[1928] 일부 실시형태에서, 본 명세서에 기재된 방법, 예를 들어, 염기 편집 방법은 표적외 효과가 최소이거나 전혀 없

다.

[1929] 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 50%의 성공적으로 편집된 세포 집단(즉, 성공적으로 조작된 세포)을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 55%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 60%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 65%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 70%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 75%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 80%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 85%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 90%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 적어도 95%의 성공적으로 편집된 세포 집단을 초래한다. 일부 실시형태에서, 본 명세서에 기재된 염기 편집 방법은 약 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%의 성공적으로 편집된 세포 집단을 초래한다.

[1930] 일부 실시형태에서, 염기 편집 개입 후 살아있는 세포 회수는 염기 편집 이벤트의 시점에서 시작 세포 집단의 적어도 60%, 70%, 80%, 90%보다 크다. 일부 실시형태에서, 위에 기재된 바와 같은 살아있는 세포 회수는 약 70%이다. 일부 실시형태에서, 위에 기재된 바와 같은 살아있는 세포 회수는 약 75%이다. 일부 실시형태에서, 위에 기재된 바와 같은 살아있는 세포 회수는 약 80%이다. 일부 실시형태에서, 위에 기재된 바와 같은 살아있는 세포 회수는 약 85%이다. 일부 실시형태에서, 위에 기재된 바와 같은 살아있는 세포 회수는 염기 편집 이벤트의 시점에서 집단의 세포의 약 90% 또는 약 91%, 92%, 93%, 94% 95%, 96%, 97%, 98% 또는 99% 또는 100%이다.

[1931] 일부 실시형태에서, 조작된 세포 집단은 시험관내에서 약 2배, 약 3-배, 약 4-배, 약 5-배, 약 6-배, 약 7-배, 약 8-배, 약 9-배, 약 10-배, 약 15-배, 약 20-배, 약 25-배, 약 30-배, 약 35-배, 약 40-배, 약 45-배, 약 50-배 또는 약 100-배 추가로 확장될 수 있다.

[1932] 의도된 돌연변이 및 indel의 수는, 예를 들어, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 PCT 특허 PCT/2017/045381(WO2018/027078) 및 PCT/US2016/058344(WO2017/070632); 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); 및 Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)]에 기술되어 있는 바와 같은 임의의 적합한 방법을 사용하여 결정될 수 있다.

[1933] 일부 실시형태에서, indel 빈도를 계산하기 위해, indel이 발생할 수 있는 창의 양쪽 측면에 있는 2개의 10-bp 서열과 정확히 일치하는지에 대해 시퀀싱 판독물을 스캔한다. 정확히 일치하는 항목이 없으면, 판독물은 분석에서 제외된다. 이 indel 창의 길이가 참조 서열과 정확히 일치하면, 판독물은 indel을 포함하지 않는 것으로 분류된다. indel 창이 참조 서열보다 2개 이상의 염기가 더 길거나 또는 더 짧으면, 시퀀싱 판독물은 각각 삽입 또는 결실로 분류된다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 핵산의 영역에서 indel의 형성을 제한할 수 있다. 일부 실시형태에서, 영역은 염기 편집기에 의해 표적화된 뉴클레오타이드 또는 염기 편집기에 의해 표적화된 뉴클레오타이드의 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개의 뉴클레오타이드 내의 영역에 있다.

[1934] 표적 뉴클레오타이드 영역에서 형성된 indel의 수는 핵산(예를 들어, 세포의 게놈 내의 핵산)이 염기 편집기에 노출되는 시간의 양에 의존할 수 있다. 일부 실시형태에서, indel의 수 또는 비율은 표적 뉴클레오타이드 서열(예를 들어, 세포의 게놈 내의 핵산)을 염기 편집기에 노출시킨 지 적어도 1시간, 적어도 2시간, 적어도 6시간, 적어도 12시간, 적어도 24시간, 적어도 36시간, 적어도 48시간, 적어도 3일, 적어도 4일, 적어도 5일, 적어도 7일, 적어도 10일 또는 적어도 14일 후에 결정된다. 본 명세서에 기재된 바와 같은 염기 편집기의 특성은 임의의 융합 단백질, 또는 본 명세서에 제공된 융합 단백질을 사용하는 방법에 적용될 수 있음을 이해하여야 한다.

[1935] 염기 편집기 효율성에 대한 세부사항은 PCT 특허 PCT/2017/045381(WO 2018/027078) 및 PCT/US2016/058344(WO 2017/070632)에 기술되어 있으며, 이들 각각은 그 전문이 참조에 의해 본 명세서에 원용된다. 또한, 그 전체 내용이 참조에 의해 본 명세서에 원용되어 있는 문헌[Komor, A.C., *et al.*, "Programmable editing of a target

base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); 및 Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)] 참조. 일부 실시형태에서, 본 명세서에 제공된 방법을 사용하여 하나 이상의 유전자에서 복수의 핵염기쌍을 편집하면 적어도 하나의 의도된 돌연변이가 형성된다. 일부 실시형태에서, 상기 적어도 하나의 의도된 돌연변이의 상기 형성은 유전자의 정상적인 기능의 붕괴를 초래한다. 일부 실시형태에서, 상기 적어도 하나의 의도된 돌연변이의 상기 형성은 유전자에 의해 암호화되는 단백질의 발현을 감소시키거나 또는 제거한다. 다중 편집은 본 명세서에 제공된 임의의 방법 또는 방법의 조합을 사용하여 달성될 수 있음을 이해하여야 한다.

[1936] ***다중 편집***

[1937] 일부 실시형태에서, 본 명세서에 제공된 염기 편집 시스템은 하나 이상의 유전자에서 복수의 핵염기쌍을 다중 편집할 수 있다. 일부 실시형태에서, 복수의 핵염기쌍은 동일한 유전자에 위치한다. 일부 실시형태에서, 복수의 핵염기쌍은 하나 이상의 유전자에 위치하되, 적어도 하나의 유전자는 상이한 유전자좌에 위치한다. 일부 실시형태에서, 다중 편집은 하나 이상의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 하나 이상의 염기 편집 시스템을 포함할 수 있다. 일부 실시형태에서, 다중 편집은 단일 가이드 폴리뉴클레오타이드를 갖는 하나 이상의 염기 편집 시스템을 포함할 수 있다. 일부 실시형태에서, 다중 편집은 복수의 가이드 폴리뉴클레오타이드를 갖는 하나 이상의 염기 편집 시스템을 포함할 수 있다. 일부 실시형태에서, 다중 편집은 단일 염기 편집 시스템을 갖는 하나 이상의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드 및 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드의 혼합물을 포함할 수 있다. 본 명세서에 기재된 바와 같은 임의의 염기 편집기를 사용하는 다중 편집의 특성은 본 명세서에 제공된 임의의 염기 편집기를 사용하는 방법의 임의의 조합에 적용될 수 있음을 이해하여야 한다. 또한, 본 명세서에 기재된 바와 같은 임의의 염기 편집기를 사용하는 다중 편집은 복수의 핵염기쌍의 순차적 편집을 포함할 수 있음을 이해하여야 한다.

[1938] 일부 실시형태에서, 복수의 핵염기쌍은 하나 이상의 유전자에 있다. 일부 실시형태에서, 복수의 핵염기쌍은 동일한 유전자에 있다. 일부 실시형태에서, 하나 이상의 유전자 중 적어도 하나의 유전자는 상이한 유전자좌에 위치한다.

[1939] 일부 실시형태에서, 편집은 적어도 하나의 단백질 코딩 영역에서의 복수의 핵염기쌍의 편집이다. 일부 실시형태에서, 편집은 적어도 하나의 단백질 비-코딩 영역에서의 복수의 핵염기쌍의 편집이다. 일부 실시형태에서, 편집은 적어도 하나의 단백질 코딩 영역 및 적어도 하나의 단백질 비-코딩 영역에서의 복수의 핵염기쌍의 편집이다.

[1940] 일부 실시형태에서, 편집은 하나 이상의 가이드 폴리뉴클레오타이드와 함께 수행된다. 일부 실시형태에서, 염기 편집 시스템은 하나 이상의 염기 편집 시스템을 포함할 수 있다. 일부 실시형태에서, 염기 편집 시스템은 단일 가이드 폴리뉴클레오타이드와 함께 하나 이상의 염기 편집 시스템을 포함할 수 있다. 일부 실시형태에서, 염기 편집 시스템은 복수의 가이드 폴리뉴클레오타이드와 함께 하나 이상의 염기 편집 시스템을 포함할 수 있다. 일부 실시형태에서, 편집은 단일 염기 편집 시스템을 갖는 하나 이상의 가이드 폴리뉴클레오타이드와 함께 수행된다. 일부 실시형태에서, 편집은 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드와 함께 수행된다. 일부 실시형태에서, 편집은 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드와 함께 수행된다. 일부 실시형태에서, 편집은 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드 및 표적 폴리뉴클레오타이드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드의 혼합물과 함께 수행된다. 본 명세서에 기재된 바와 같은 임의의 염기 편집기를 사용하는 다중 편집의 특성은 본 명세서에 제공된 임의의 염기 편집기를 사용하는 방법의 임의의 조합에 적용될 수 있음을 이해하여야 한다. 또한,

편집은 복수의 핵염기쌍의 순차적 편집을 포함할 수 있음을 이해하여야 한다.

[1941] 일부 실시형태에서, 하나 이상의 유전자에서 복수의 핵염기쌍의 다중 편집이 가능한 염기 편집 시스템은 ABE9 염기 편집기 중 하나를 포함한다. 일부 실시형태에서, 하나 이상의 유전자에서 복수의 핵염기쌍의 다중 편집이 가능한 염기 편집 시스템은 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함한다. 일부 실시형태에서, 하나 이상의 유전자에서 복수의 핵염기쌍의 다중 편집이 가능한 염기 편집 시스템은 ABE7 염기 편집기 중 하나를 포함한다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 다중 편집이 가능한 염기 편집 시스템은 ABE7 염기 편집기 중 하나를 포함하는 다중 편집이 가능한 염기 편집 시스템과 비교하여 더 높은 다중 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 다중 편집이 가능한 염기 편집 시스템은 ABE7 염기 편집기 중 하나를 포함하는 다중 편집이 가능한 염기 편집 시스템과 비교하여 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 105%, 적어도 110%, 적어도 115%, 적어도 120%, 적어도 125%, 적어도 130%, 적어도 135%, 적어도 140%, 적어도 145%, 적어도 150%, 적어도 155%, 적어도 160%, 적어도 165%, 적어도 170%, 적어도 175%, 적어도 180%, 적어도 185%, 적어도 190%, 적어도 195%, 적어도 200%, 적어도 210%, 적어도 220%, 적어도 230%, 적어도 240%, 적어도 250%, 적어도 260%, 적어도 270%, 적어도 280%, 적어도 290%, 적어도 300% 더 높은, 적어도 310%, 적어도 320%, 적어도 330%, 적어도 340%, 적어도 350%, 적어도 360%, 적어도 370%, 적어도 380%, 적어도 390%, 적어도 400%, 적어도 450% 또는 적어도 500% 더 높은 다중 편집 효율을 갖는다. 일부 실시형태에서, 본 명세서에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 다중 편집이 가능한 염기 편집 시스템은 ABE7 염기 편집기 중 하나를 포함하는 다중 편집이 가능한 염기 편집 시스템과 비교하여 적어도 1.1배, 적어도 1.2배, 적어도 1.3배, 적어도 1.4배, 적어도 1.5배, 적어도 1.6배, 적어도 1.7배, 적어도 1.8배, 적어도 1.9배, 적어도 2.0배, 적어도 2.1배, 적어도 2.2배, 적어도 2.3배, 적어도 2.4배, 적어도 2.5배, 적어도 2.6배, 적어도 2.7배, 적어도 2.8배, 적어도 2.9배, 적어도 3.0배, 적어도 3.1배, 적어도 3.2배, 적어도 3.3배, 적어도 3.4배, 적어도 3.5배, 적어도 4.0배, 적어도 4.5배, 적어도 5.0배, 적어도 5.5배 또는 적어도 6.0배 더 높은 다중 편집 효율을 갖는다.

[1942] **전달 시스템**

[1943] 유전자(예를 들어, *CD2*)에서 하나 이상의 뉴클레오타이드를 표적화하기 위한 핵염기 편집기의 적합성은 본 명세서에 기재된 바와 같이 평가된다. 일 실시형태에서, 단일 관심 세포는 리포터(예를 들어, GFP)를 암호화하는 소량의 벡터를 이용하여 본 명세서에 기재된 염기 편집 시스템을 암호화하는 핵산 분자 또는 분자로 형질감염, 형질도입 또는 그렇지 않으면 변형된다. 이들 세포는 면역 세포(예를 들어, T-세포 또는 NK-세포) 또는 불멸화된 인간 세포주, 예컨대, 293T, K562 또는 U20S를 포함하는 당업계에 공지된 임의의 세포일 수 있다. 대안적으로, 일차 세포(예를 들어, 인간)가 사용될 수 있다. 세포는 또한 조직 생검, 수술, 혈액, 혈장, 혈청 또는 다른 생물학적 유체와 같은 대상체 또는 개체로부터 얻을 수 있다. 이러한 세포는 최종 세포 표적과 관련될 수 있다.

[1944] 전달은 바이러스 벡터를 사용하여 수행될 수 있다. 일 실시형태에서, 형질감염은 지질 형질감염(예컨대, Lipofectamine 또는 Fugene)을 사용하거나 또는 전기천공법에 의해 수행될 수 있다. 형질감염 후, 일관되고 높은 수준의 형질감염을 확인하기 위해 리포터(예를 들어, GFP)의 발현이 형광 현미경 검사 또는 유세포 분석에 의해 결정될 수 있다. 이러한 예비 형질감염은 가장 큰 활성을 제공하는 편집기 조합을 결정하기 위해 상이한 핵염기 편집기를 포함할 수 있다. 시스템은 하나 이상의 상이한 벡터를 포함할 수 있다. 일 실시형태에서, 염기 편집기는 목적하는 세포 유형, 우선적으로 진핵생물 세포, 바람직하게는 포유동물 세포 또는 인간 세포의 발현을 위해 최적화된 코돈이다.

[1945] 핵염기 편집기의 활성은 본 명세서에 기재된 바와 같이, 즉, 표적 서열의 변경을 검출하기 위해 세포의 게놈을 시퀀싱함으로써 평가된다. 생어 시퀀싱의 경우, 정제된 PCR 앰플리콘은 플라스미드 백본에 클로닝되고, 형질전환되고, 미니프렙되고, 단일 프라이머로 시퀀싱된다. 시퀀싱은 또한 차세대 시퀀싱(NGS) 기법을 사용하여 수행될 수 있다. 차세대 시퀀싱을 사용하는 경우, 앰플리콘은 의도된 절단 부위가 비대칭으로 배치된 300bp 내지 500bp일 수 있다. PCR 후, 차세대 시퀀싱 어댑터 및 바코드(예를 들어, Illumina 다중 어댑터 및 인덱스)가, 예를 들어, 고속대량 시퀀싱(예를 들어, Illumina MiSeq)에 사용하기 위해 앰플리콘의 말단에 추가될 수 있다. 초기 테스트에서 가장 높은 수준의 표적 특이적 변경을 유도하는 융합 단백질이 추가 평가를 위해 선택될 수 있다.

[1946] 특정 실시형태에서, 핵염기 편집기는 관심 폴리뉴클레오타이드를 표적화하는데 사용된다. 일 실시형태에서, 본 발명의 핵염기 편집기는 세포의 게놈 내에서 하나 이상의 관심 핵산 서열을 표적화하는데 사용되는 하나 이상의 가이드 RNA와 함께 세포(예를 들어, 면역 세포(예를 들어, T-세포 또는 NK-세포))에 전달되어 표적 유전자(들)(예를 들어, *CD2*)를 변경한다. 일부 실시형태에서, 염기 편집기는 하나 이상의 관심 유전자(예를 들어, CD2, TRAC, B2M, CIITA, TRBC1, TRBC2, PD-1, CD52)의 서열에 하나 이상의 편집을 도입하기 위해 하나 이상의 가이드 RNA에 의해 표적화된다. 일부 실시형태에서, 하나 이상의 관심 유전자의 서열에 대한 하나 이상의 편집은 숙주 세포(예를 들어, 면역 세포(예를 들어, T-세포 또는 NK-세포))에서 유전자에 의해 암호화되는 단백질의 발현을 감소시키거나 또는 제거한다. 일부 실시형태에서, 하나 이상의 관심 유전자(예를 들어, *CD2*)에 의해 암호화되는 하나 이상의 단백질의 발현은 숙주 세포(예를 들어, 면역 세포(예를 들어, T-세포 또는 NK-세포))에서 완전히 넉아웃되거나 또는 제거된다.

[1947] 일부 실시형태에서, 숙주 세포는 포유동물 세포이다. 일부 실시형태에서, 숙주 세포는 인간 세포이다. 일부 실시형태에서, 숙주 세포는 면역 세포이다. 일부 실시형태에서, 면역 세포는 T 세포이다. 일부 실시형태에서, 면역 세포는 NK 세포이다. 일부 실시형태에서, 하나 이상의 편집이 *CD2*, *CD3*, *CD5*, *CD7*, *CD52*, *B2M*, *CIITA*, *TRBC1*, *TRBC2*, *TRAC* 및 *PD-1* 또는 이들의 조합으로부터 선택되는 하나 이상의 유전자에 도입된다. 일부 실시형태에서, 하나 이상의 편집이 *CD2* 유전자에 도입된다. 일부 실시형태에서, 하나 이상의 편집이 *CD5* 유전자에 도입된다. 일부 실시형태에서, 하나 이상의 편집이 *CD7* 유전자에 도입된다. 일부 실시형태에서, 하나 이상의 편집이 *CD2*, *CD52*, *TRAC* 및 *PD-1* 유전자에 도입된다. 일부 실시형태에서, 하나 이상의 편집이 *CD5*, *CD52*, *TRAC* 및 *PD-1* 유전자에 도입된다. 일부 실시형태에서, 하나 이상의 편집이 *CD7*, *CD3*, *CD52* 및 *PD-1* 유전자에 도입된다.

[1948] ***핵염기 편집기 및 gRNA의 핵산-기반 전달***

[1949] 본 개시내용에 따른 사이티딘 또는 아데노신 데아미네이스 핵염기 편집기를 암호화하는 핵산은 대상체에게 투여되거나, 또는 당업계에 공지된 방법에 의해 또는 본 명세서에 기재된 바와 같이 시험관내 또는 생체내 세포 내로 전달될 수 있다. 예를 들어, 사이티딘 또는 아데노신 데아미네이스 핵염기 편집기는, 예를 들어, 벡터(예를 들어, 바이러스 또는 비-바이러스 벡터), 비-벡터 기반 방법(예를 들어, 네이키드 DNA, DNA 복합체, 지질 나노입자를 사용하여) 또는 이들의 조합에 의해 전달될 수 있다.

[1950] 사이티딘 또는 아데노신 데아미네이스 핵염기 편집기를 암호화하는 핵산은, 예를 들어, 형질감염 또는 전기천공법에 의해 네이키드 DNA 또는 RNA로서 세포(예를 들어, 면역 세포)에 직접 전달될 수 있거나, 또는 표적 세포에 의한 흡수를 촉진하는 분자(예를 들어, N-아세틸갈락토사민)에 접합될 수 있다. 본 명세서에 기재된 벡터와 같은 핵산 벡터도 또한 사용될 수 있다. 특정 실시형태에서, 폴리뉴클레오타이드, 예를 들어, 염기 편집기 또는 이의 기능적 성분을 암호화하는 mRNA는 본 명세서에 기재된 바와 같은 다중 가이드 RNA와의 조합으로 공동 전기천공될 수 있다.

[1951] 핵산 벡터는 본 명세서에 기재된 융합 단백질의 도메인을 암호화하는 하나 이상의 서열을 포함할 수 있다. 벡터는 또한 단백질을 코딩하는 서열과 관련된(예를 들어, 이에 삽입되거나 또는 융합된) 신호 펩타이드를 암호화하는 서열(예를 들어, 핵 국부화, 인 국부화 또는 미토콘드리아 국부화를 위해)을 포함할 수 있다. 일례로서, 핵산 벡터는 하나 이상의 핵 국부화 서열(예를 들어, SV40으로부터의 핵 국부화 서열) 및 하나 이상의 데아미네이스를 포함하는 Cas9 코딩 서열을 포함할 수 있다.

[1952] 핵산 벡터는 또한 임의의 적합한 수의 조절/제어 요소, 예를 들어, 프로모터, 인핸서, 인트론, 폴리아데닐화 신호, 코작 컨센서스 서열 또는 내부 리보솜 진입 부위(internal ribosome entry site: IRES)를 포함할 수 있다. 이들 요소는 당업계에 잘 알려져 있다.

[1953] 본 개시내용에 따른 핵산 벡터는 재조합 바이러스 벡터를 포함한다. 예시적인 바이러스 벡터는 위의 본 명세서에 기재되어 있다. 당업계에 공지된 다른 바이러스 벡터가 또한 사용될 수 있다. 또한, 바이러스 입자가 핵산 및/또는 펩타이드 형태의 게놈 편집 시스템 성분을 전달하는데 사용될 수 있다. 예를 들어, "빈(empty)" 바이러스 입자는 임의의 적합한 카고(cargo)를 포함하도록 조립될 수 있다. 바이러스 벡터 및 바이러스 입자는 또한 표적 조직 특이성을 변경하기 위해 표적화 리간드를 혼입하도록 조작될 수 있다.

[1954] 바이러스 벡터에 더하여, 비-바이러스 벡터가 본 개시내용에 따른 게놈 편집 시스템을 암호화하는 핵산을 전달하기 위해 사용될 수 있다. 비-바이러스 핵산 벡터의 중요한 범주 중 하나는 유기 또는 무기일 수 있는 나노입자이다. 나노입자는 당업계에 잘 알려져 있다. 임의의 적합한 나노입자 설계가 게놈 편집 시스템 성분 또는 이러한 성분을 암호화하는 핵산을 전달하는데 사용될 수 있다. 예를 들어, 유기(예를 들어, 지질 및/또는 중합체)

나노입자는 본 개시내용의 소정의 실시형태에서 전달 비히클로서 사용하기에 적합할 수 있다. 나노입자 제형에 사용하기 위한 예시적인 지질 및/또는 유전자 전달이 표 18(아래)에 나타나 있다.

표 18

| 유전자 전이를 위해 사용된 지질 | | |
|---|---|---|
| 지질 | 약어 | 특징 |
| 1,2-다이올레오일-sn-글리세로-3-포스파티딜콜린 | DOPC | 헬퍼 |
| 1,2-다이올레오일-sn-글리세로-3-포스파티딜에탄올아민 | DOPE | 헬퍼 |
| 콜레스테롤 | | 헬퍼 |
| N-[1-(2,3-다이올레일옥시)프로필]N,N,N-트라이메틸암모늄 클로라이드 | DOTMA | 양이온성 |
| 1,2-다이올레오일옥시-3-트라이메틸암모늄-프로페인 | DOTAP | 양이온성 |
| 다이옥타데실아미도글리실스페르민 | DOGS | 양이온성 |
| N-(3-아미노프로필)-N,N-다이메틸-2,3-비스(도데실옥시)-1-프로판아미늄 브로마이드 | GAP-DLRIE | 양이온성 |
| 세틸트라이메틸암모늄 브로마이드 | CTAB | 양이온성 |
| 6-라우록시헥실 오르니티네이트 | LHON | 양이온성 |
| 1-(2,3-다이올레오일옥시프로필)-2,4,6-트라이메틸피리디늄 | 2Oc | 양이온성 |
| 2,3-다이올레일옥시-N-[2(스페르민카복사미도-에틸]-N,N-다이메틸-1-프로판아미늄 트라이플루오로아세테이트 | DOSPA | 양이온성 |
| 1,2-다이올레일-3-트라이메틸암모늄-프로페인 | DOPA | 양이온성 |
| N-(2-하이드록시에틸)-N,N-다이메틸-2,3-비스(테트라데실옥시)-1-프로판아미늄 브로마이드 | MDRIE | 양이온성 |
| 다이미리스토옥시프로필 다이메틸 하이드록시에틸 암모늄 브로마이드 | DMRI | 양이온성 |
| 3β-[N-(N',N'-다이메틸아미노에테인)-카바모일]콜레스테롤 | DC-Chol | 양이온성 |
| 비스-구아니듐-트렌-콜레스테롤 | BGTC | 양이온성 |
| 1,3-다이오데옥시-2-(6-카복시-스페르밀)-프로필아마이드 | DOSPER | 양이온성 |

[1955]

| 유전자 전이를 위해 사용된 지질 | | |
|---|---|---|
| 지질 | 약어 | 특징 |
| 다이메틸옥타데실암모늄 브로마이드 | DDAB | 양이온성 |
| 다이옥타데실아미도글리실스페르미딘 | DSL | 양이온성 |
| rac-[(2,3-다이옥타데실옥시프로필)(2-하이드록시에틸)]-다이메틸암모늄 클로라이드 | CLIP-1 | 양이온성 |
| rac-[2(2,3-다이헥사데실옥시프로필-옥시메틸옥시)에틸]트라이메틸암모늄브로마이드 | CLIP-6 | 양이온성 |
| 에틸다이미리스토일포스파티딜콜린 | EDMPC | 양이온성 |
| 1,2-다이스테아릴옥시-N,N-다이메틸-3-아미노프로페인 | DSDMA | 양이온성 |
| 1,2-다이미리스토일-트라이메틸암모늄 프로페인 | DMTAP | 양이온성 |
| O,O'-다이미리스틸-N-라이실 아스파테이트 | DMKE | 양이온성 |
| 1,2-다이스테아로일-sn-글리세로-3-에틸포스포콜린 | DSEPC | 양이온성 |
| N-팔미토일 D-에리트로-스핑고실 카바모일-스페르민 | CCS | 양이온성 |
| N-t-뷰틸-N0-테트라데실-3-테트라데실아미노프로피온아미딘 | 다이 C14-아미딘 | 양이온성 |
| 옥타데센오일옥시[에틸-2-헵타데센일-3 하이드록시에틸] 이미다졸리늄 클로라이드 | DOTIM | 양이온성 |
| N1-콜레스테릴옥시카보닐-3,7-다이아자노네인-1,9-다이아민 | CDAN | 양이온성 |
| 2-(3-[비스(3-아미노-프로필)-아미노]프로필아미노)-N-다이테트라데실카바모일메-에틸-아세트아마이드 | RPR209120 | 양이온성 |
| 1,2-다이리놀레일옥시-3-다이메틸아미노프로페인 | DLinDMA | 양이온성 |
| 2,2-다이리놀레일-4-다이메틸아미노에틸-[1,3]-다이옥솔레인 | DLin-KC2-DMA | 양이온성 |
| 다이리놀레일-메틸-4-다이메틸아미노뷰티레이트 | DLin-MC3-DMA | 양이온성 |

[1956]

[1957] 표 19는 유전자 전달 및/또는 나노입자 제형에서 사용하기 위한 예시적인 중합체를 열거한다.

표 19

| 유전자 전이에 사용되는 중합체 | |
|---|---|
| 중합체 | 약어 |
| 폴리(에틸렌)글리콜 | PEG |
| 폴리에틸렌이민 | PEI |
| 다이티오비스 (석신이미딜프로피오네이트) | DSP |
| 다이메틸-3,3'-다이티오비스프로피온이미데이트 | DTBP |
| 폴리(에틸렌 이민)바이카바메이트 | PEIC |
| 폴리(L-라이신) | PLL |
| 히스티딘 변형 PLL | |
| 폴리(N-바이닐피롤리돈) | PVP |
| 폴리(프로필렌이민) | PPI |
| 폴리(아미도아민) | PAMAM |
| 폴리(아미도에틸렌이민) | SS-PAEI |
| 트라이에틸렌테트라민 | TETA |
| 폴리(β-아미노에스터) | |
| 폴리(4-하이드록시-L-프롤린 에스터) | PHP |
| 폴리(알릴아민) | |
| 폴리(α-[4-아미노뷰틸]-L-글리콜산) | PAGA |
| 폴리(D,L-락트산-코-글리콜산) | PLGA |
| 폴리(N-에틸-4-바이닐피리디늄 브로마이드) | |
| 폴리(포스파겐)s | PPZ |
| 폴리(포스포에스터)s | PPE |
| 폴리(포스포라미데이트)s | PPA |
| 폴리(N-2-하이드록시프로필메타크릴아마이드) | pHPMA |
| 폴리(2-(다이메틸아미노)에틸 메타크릴레이트) | pDMAEMA |
| 폴리(2-아미노에틸 프로필렌 포스페이트) | PPE-EA |
| 키토산 | |
| 갈락토실화 키토산 | |

[1958]

| 유전자 전이에 사용되는 중합체 | |
|---|---|
| 중합체 | 약어 |
| N-도데실화 키토산 | |
| 히스톤 | |
| 콜라겐 | |
| 덱스트란-스페르민 | D-SPM |

[1959]

[1960] 표 20은 본 명세서에 기재된 융합 단백질을 암호화하는 폴리뉴클레오타이드에 대한 전달 방법을 요약한다.

표 20

| 전달 | 벡터/방식 | 비-분열 세포로의 전달 | 발현 기간 | 게놈 통합 | 전달된 분자의 유형 |
|---|---|---|---|---|---|
| 물리적 | (예를 들어, 전기천공법, 입자총법, 칼슘 포스페이트 형질감염 | 전달됨 | 일시적 | 통합되지 않음 | 핵산 및 단백질 |
| 바이러스 | 레트로바이러스 | 전달되지 않음 | 안정적 | 통합됨 | RNA |
| | 렌티바이러스 | 전달됨 | 안정적 | 통합됨/변형 없음 | RNA |
| | 아데노바이러스 | 전달됨 | 일시적 | 통합되지 않음 | DNA |
| | 아데노-관련 바이러스(AAV) | 전달됨 | 안정적 | 통합되지 않음 | DNA |
| | 백시니아 바이러스 | 전달됨 | 매우 일시적 | 통합되지 않음 | DNA |
| | 단순 포진 바이러스 | 전달됨 | 안정적 | 통합되지 않음 | DNA |
| 비-바이러스 | 양이온성 리포솜 | 전달됨 | 일시적 | 무엇을 전달하느냐에 따름 | 핵산 및 단백질 |
| | 중합체성 나노입자 | 전달됨 | 일시적 | 무엇을 전달하느냐에 따름 | 핵산 및 단백질 |
| 생물학적 비-바이러스 전달 비히클 | 약화된 세균 | 전달됨 | 일시적 | 통합되지 않음 | 핵산 |
| | 조작된 박테리오파지 | 전달됨 | 일시적 | 통합되지 않음 | 핵산 |
| | 포유동물 바이러스-유사 입자 | 전달됨 | 일시적 | 통합되지 않음 | 핵산 |
| | 생물학적 리포솜: 파열 적혈구 및 엑소솜 | 전달됨 | 일시적 | 통합되지 않음 | 핵산 |

[1961]

[1962]

[1963] 또 다른 양태에서, 게놈 편집 시스템 성분 또는 이러한 성분을 암호화하는 핵산, 예를 들어, 핵산 결합 단백질,

예컨대, 예를 들어, Cas9 또는 이의 변이체, 및 관심 게놈 핵산 서열을 표적화하는 gRNA의 전달은 리보핵단백질(RNP)을 세포에 전달함으로써 달성될 수 있다. RNP는 표적화 gRNA와의 복합체로 핵산 결합 단백질, 예를 들어, Cas9를 포함한다. RNP는 예를 들어, 문헌[Zuris, J.A. *et al.*, 2015, *Nat. Biotechnology*, 33(1):73-80]에 의해 보고된 바와 같은 전기천공법, 뉴클레오펙션 또는 양이온성 지질-매개성 방법과 같은 공지의 방법을 사용하여 세포에 전달될 수 있다. RNP는 CRISPR 염기 편집 시스템, 특히 일차 세포와 같이 형질감염이 어려운 세포에 사용하기에 유리하다. 또한, RNP는 또한 특히 CRISPR 플라스미드에 사용될 수 있는 진핵생물 프로모터, 예를 들어, CMV 또는 EF1A가 잘 발현되지 않을 때, 세포에서 단백질 발현과 함께 발생할 수 있는 어려움을 완화시킬 수 있다. 유리하게는, RNP의 사용은 외래 DNA를 세포로 전달할 필요가 없다. 또한, 핵산 결합 단백질 및 gRNA 복합체를 포함하는 RNP는 시간이 지남에 따라 분해되기 때문에, RNP의 사용은 표적외 효과를 제한할 가능성이 있다. 플라스미드 기반 기법과 유사한 방식으로, RNP는 결합 단백질(예를 들어, Cas9 변이체)을 전달하고 상동성 유도 복구(HDR)를 지시하는데 사용될 수 있다.

[1964] 핵산 분자 발현을 코딩하는 염기 편집기를 구동하는데 사용되는 프로모터는 AAV ITR을 포함할 수 있다. 이는 벡터에서 공간을 차지할 수 있는 추가적인 프로모터 요소의 필요성을 제거하는데 유리할 수 있다. 확보된 추가적인 공간은 가이드 핵산 또는 선별 마커와 같은 추가적인 요소의 발현을 구동하는데 사용될 수 있다. ITR 활성은 상대적으로 약하므로, 선택된 뉴클레이스의 과발현으로 인한 잠재적 독성을 감소시키는데 사용될 수 있다.

[1965] 임의의 적합한 프로모터는 염기 편집기 및, 적절한 경우, 가이드 핵산의 발현을 구동하는데 사용될 수 있다. 편재된 발현을 위해, 사용될 수 있는 프로모터는 CMV, CAG, CBh, PGK, SV40, 페리틴 중쇄 또는 경쇄 등을 포함한다. 뇌 또는 기타 CNS 세포 발현을 위해, 적합한 프로모터는 모든 뉴런의 경우 시냅신I, 흥분성 뉴런의 경우 CaMKII알파, 가바작동성(GABAergic) 뉴런의 경우 GAD67 또는 GAD65 또는 VGAT 등을 포함할 수 있다. 간세포 발현을 위해, 적합한 프로모터 알부민 프로모터를 포함한다. 폐 세포 발현을 위해, 적합한 프로모터는 SP-B를 포함할 수 있다. 내피 세포를 위해, 적합한 프로모터는 ICAM을 포함할 수 있다. 조혈 세포를 위해, 적합한 프로모터는 IFN베타 또는 CD45를 포함할 수 있다. 골아세포를 위해, 적합한 프로모터는 OG-2를 포함할 수 있다.

[1966] 일부 실시형태에서, 본 개시내용의 염기 편집기는 별도의 프로모터가 동일한 핵산 분자 내에서 염기 편집기 및 양립 가능한 가이드 핵산의 발현을 구동할 수 있도록 충분히 작은 크기이다. 예를 들어, 벡터 또는 바이러스 벡터는 염기 편집기를 암호화하는 핵산에 작동 가능하게 연결된 제1 프로모터 및 가이드 핵산에 작동 가능하게 연결된 제2 프로모터를 포함할 수 있다.

[1967] 가이드 핵산의 발현을 구동하는데 사용되는 프로모터는 Pol III 프로모터, 예컨대, U6 또는 H1을 포함할 수 있다. Pol II 프로모터 및 인트론 카세트를 사용하여 gRNA 아데노 관련 바이러스(AAV)를 발현한다.

[1968] 특정 실시형태에서, 본 발명의 융합 단백질은 바이러스 벡터(예를 들어, 아데노-관련 바이러스(AAV), AAV3, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh8, AAV10 및 이의 변이체) 또는 임의의 바이러스 벡터의 적합한 캡시드 단백질에 존재하는 폴리뉴클레오타이드에 의해 암호화된다. 따라서, 일부 양태에서, 본 개시내용은 융합 단백질의 바이러스 전달에 관한 것이다. 바이러스 벡터의 예는 레트로바이러스 벡터(예를 들어, 말로니 뮤린 백혈병 바이러스, MML-V), 아데노바이러스 벡터(예를 들어, AD100), 렌티바이러스 벡터(HIV 및 FIV-기반 벡터), 헤르페스바이러스 벡터(예를 들어, HSV-2)를 포함한다.

[1969] 일부 양태에서, 면역 세포에서 특정 유전자를 편집하기 위한 본 명세서에 기재된 방법은 CAR-T 세포를 유전적으로 변형시키는데 사용될 수 있다. 이러한 CAR-T 세포 및 이러한 CAR-T 세포를 생산하기 위한 방법은 국제 특허 PCT/US2016/060736, PCT/US2016/060734, PCT/US2016/034873, PCT/US2015/040660, PCT/EP2016/055332, PCT/IB2015/058650, PCT/EP2015/067441, PCT/EP2014/078876, PCT/EP2014/059662, PCT/IB2014/061409, PCT/US2016/019192, PCT/US2015/059106, PCT/US2016/052260, PCT/US2015/020606, PCT/US2015/055764, PCT/CN2014/094393, PCT/US2017/059989, PCT/US2017/027606 및 PCT/US2015/064269에 기술되어 있으며, 각각의 내용은 그 전문이 참조에 의해 본 명세서에 원용된다.

[1970] ***바이러스 벡터***

[1971] 따라서, 본 명세서에 기재된 염기 편집기는 바이러스 벡터로 전달될 수 있다. 일부 실시형태에서, 본 명세서에 개시된 염기 편집기는 바이러스 벡터에 포함된 핵산상에 암호화될 수 있다. 일부 실시형태에서, 하나 이상의 염기 편집 시스템의 성분은 하나 이상의 바이러스 벡터상에 암호화될 수 있다. 예를 들어, 염기 편집기 및 가이드 핵산은 단일 바이러스 벡터상에 암호화될 수 있다. 다른 실시형태에서, 염기 편집기 및 가이드 핵산은 상이한 바이러스 벡터상에 암호화된다. 어느 경우든, 염기 편집기 및 가이드 핵산은 각각 프로모터 및 종결인자에 작동

가능하게 연결될 수 있다. 바이러스 벡터상에 암호화된 성분의 조합은 선택된 바이러스 벡터의 카고 크기 제약에 의해 결정될 수 있다.

[1972] 염기 편집기의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 배양물 또는 숙주에서 특정 세포에 바이러스를 표적화하고 핵 또는 숙주 세포 게놈에 바이러스 페이로드를 이동시키기 위한 고도로 진화된 과정을 이용한다. 바이러스 벡터는 배양물의 세포, 환자(생체내)에 직접 투여될 수 있거나, 또는 시험관내에서 세포를 처리하는데 사용될 수 있으며, 변형된 세포는 선택적으로 환자(생체외)에게 투여될 수 있다. 종래의 바이러스 기반 시스템은 유전자 전달을 위한 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-관련 및 단순 포진 바이러스 벡터를 포함할 수 있다. 숙주 게놈의 통합은 레트로바이러스, 렌티바이러스 및 아데노-관련 바이러스 유전자 전달 방법으로 가능하며, 종종 삽입된 이식유전자의 장기 발현을 초래한다. 추가적으로, 다양한 세포 유형 및 표적 조직에서 높은 형질도입 효율이 관찰되었다.

[1973] 바이러스 벡터는 렌티바이러스(예를 들어, HIV 및 FIV-기반 벡터), 아데노바이러스(예를 들어, AD100), 레트로바이러스(예를 들어, 말로니 뮤린 백혈병 바이러스, MML-V), 헤르페스바이러스 벡터(예를 들어, HSV-2) 및 아데노-관련 바이러스(AAV) 또는 특히, 예를 들어, 미국 특허 제8,454,972호(아데노바이러스에 대한 제형, 용량), 미국 특허 제8,404,658호(AAV에 대한 제형, 용량) 및 미국 특허 제5,846,946호(DNA 플라스미드에 대한 제형, 용량)으로부터의 제형 및 용량과 렌티바이러스, AAV 및 아데노바이러스와 관련된 임상 시험에 관한 임상 시험 및 간행물로부터의 제형 및 용량을 사용하는 다른 플라스미드 또는 바이러스 벡터 유형을 포함할 수 있다. 예를 들어, AAV의 경우, 투여 경로, 제형 및 용량은 미국 특허 제8,454,972호 및 AAV와 관련된 임상 시험에서와 같을 수 있다. 아데노바이러스의 경우, 투여 경로, 제형 및 용량은 미국 특허 제8,404,658호 및 아데노바이러스와 관련된 임상시험에서와 같을 수 있다. 플라스미드 전달의 경우, 투여 경로, 제형 및 용량은 미국 특허 제5,846,946호 및 플라스미드와 관련된 임상 연구에서와 같을 수 있다. 용량은 평균 70kg의 개인(예를 들어, 성인 인간 남성)을 기준으로 하거나 또는 이에 대해 추정될 수 있으며, 다양한 체중과 종의 환자, 대상체, 포유동물에 대해 조정될 수 있다. 투여 빈도는 환자 또는 대상체의 연령, 성별, 일반적인 건강, 기타 병태 및 고려될 특정 병태 또는 증상을 포함하는 일반적인 요인에 따라 개업의 또는 수의사(예를 들어, 의사, 수의사)의 범위 내이다. 바이러스 벡터는 관심 조직에 주입될 수 있다. 세포-유형 특이적 염기 편집의 경우, 염기 편집기 및 선택적 가이드 핵산의 발현은 세포-유형 특이적 프로모터에 의해 구동될 수 있다.

[1974] 레트로바이러스의 지향성(tropism)은 외래 외피 단백질을 혼입하여 표적 세포의 잠재적 표적 집단을 확장시킴으로써 변경될 수 있다. 렌티바이러스 벡터는 비-분열 세포를 형질도입 또는 감염시킬 수 있고 전형적으로 높은 바이러스 역가를 생성할 수 있는 레트로바이러스 벡터이다. 따라서, 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에 따라 달라진다. 레트로바이러스 벡터는 최대 6kb 내지 10kb의 외래 서열에 대한 패키징 용량을 갖는 시스-작용성 긴 말단 반복부로 구성된다. 최소 시스-작용성 LTR은 벡터의 복제 및 패키징에 충분하며, 이는 이어서 치료 유전자를 표적 세포에 통합시켜 영구적인 이식유전자 발현을 제공하는데 사용된다. 널리 사용되는 레트로바이러스 벡터는 뮤린 백혈병 바이러스(murine leukemia virus: MuLV), 긴팔원숭이 백혈병 바이러스(gibbon ape leukemia virus: GaLV), 시미안 면역 결핍 바이러스(Simian Immuno deficiency virus: SIV), 인간 면역 결핍 바이러스(human immuno deficiency virus: HIV) 및 이들의 조합을 기반으로 하는 것들을 포함한다(예를 들어, 문헌[Buchscher *et al.*, J. Virol. 66:2731-2739 (1992); Johann *et al.*, J. Virol. 66:1635-1640 (1992); Sommnerfelt *et al.*, Virol. 176:58-59 (1990); Wilson *et al.*, J. Virol. 63:2374-2378 (1989); Miller *et al.*, J. Virol. 65:2220-2224 (1991)]; PCT/US94/05700 참조).

[1975] 레트로바이러스 벡터, 특히 렌티바이러스 벡터는 표적 세포로의 효율적인 통합을 위해 주어진 길이보다 더 작은 폴리뉴클레오타이드 서열을 필요로 할 수 있다. 예를 들어, 길이가 9kb보다 긴 레트로바이러스 벡터는 더 작은 크기의 레트로바이러스와 비교하여 바이러스 역가가 낮을 수 있다. 일부 양태에서, 본 개시내용의 염기 편집기는 레트로바이러스 벡터를 통해 표적 세포로의 효율적인 패키징 및 전달을 가능하게 하기에 충분한 크기이다. 일부 실시형태에서, 염기 편집기는 가이드 핵산 및/또는 표적화 가능한 뉴클레이스 시스템의 다른 성분과 함께 발현되는 경우에도 효율적인 패킹 및 전달을 가능하게 하는 크기이다.

[1976] 패키징 세포는 전형적으로 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성하는데 사용된다. 이러한 세포는 아데노바이러스를 패키징하는 293 세포 및 레트로바이러스를 패키징하는 psi.2 세포 또는 PA317 세포를 포함한다. 유전자 요법에 사용되는 바이러스 벡터는 일반적으로 핵산 벡터를 바이러스 입자로 패키징하는 세포주를 생성함으로써 생성된다. 벡터는 전형적으로 패키징 및 숙주로의 후속 통합에 필요한 최소한의 바이러스 서열을 포함하고, 다른 바이러스 서열은 발현될 폴리뉴클레오타이드(들)에 대한 발현 카세트로 대체된다. 누락된 바이러스 기능은 전형적으로 패키징 세포주에 의해 트랜스로(in trans) 제공된다. 예를 들어, 유전자 요법에 사용되는

아데노-관련 바이러스("AAV") 벡터는 전형적으로 패키징 및 숙주 게놈으로의 통합에 필요한 AAV 게놈으로부터의 ITR 서열만을 보유한다. 바이러스 DNA는 다른 AAV 유전자, 즉 rep 및 cap를 암호화하는 헬퍼 플라스미드를 포함하지만 ITR 서열이 결여되어 있는 세포에 패키징될 수 있다. 세포주는 또한 헬퍼로서 아데노바이러스에 감염될 수 있다. 헬퍼 바이러스는 AAV 벡터의 복제 및 헬퍼 플라스미드로부터 AAV 유전자의 발현을 촉진할 수 있다. 헬퍼 플라스미드는 일부 경우에 ITR 서열의 결여로 인해 충분한 양으로 패키징 되지 않는다. 아데노바이러스에 의한 오염은, 예를 들어, 아데노바이러스가 AAV보다 더 민감한 열처리에 의해 감소될 수 있다.

[1977] 일시적 발현이 선호되는 적용에서, 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 많은 세포 유형에서 매우 높은 형질도입 효율이 가능하며, 세포 분열이 필요하지 않다. 이러한 벡터로, 높은 역가 및 발현 수준이 얻어졌다. 이러한 벡터는 비교적 간단한 시스템에서 대량으로 생산될 수 있다. 아데노-관련 바이러스("AAV") 벡터는 또한, 예를 들어, 핵산 및 펩타이드의 시험관내 생산에서 그리고 생체내 및 생체외 유전자 요법 절차를 위해 표적 핵산으로 세포를 형질도입하는데 사용될 수 있다(예를 들어, 문헌[West *et al.*, Virology 160:38-47 (1987)]; 미국 특허 제4,797,368호; WO 93/24641; 문헌[Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994)] 참조). 재조합 AAV 벡터의 작제물은 미국 특허 제5,173,414호; 문헌[Tratschin *et al.*, Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, *et al.*, Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); 및 Samulski *et al.*, J. Virol. 63:03822-3828 (1989)]을 포함하는 다수의 간행물에 기술되어 있다.

[1978] AAV는 파보바이러스 패밀리에 속하는 작은 단일-가닥 DNA 의존성 바이러스이다. 4.7kb 야생형(wt) AAV 게놈은 각각 4개의 복제 단백질과 3개의 캡시드 단백질을 암호화하는 2개의 유전자로 구성되며, 145-bp의 역 말단 반복부(inverted terminal repeat: ITR)가 양쪽에 인접해 있다. 비리온은 Vp1, Vp2 및 Vp3의 세 가지 캡시드 단백질로 구성되어 있으며, 동일한 오픈 리딩 프레임에서 1:1:10의 비로 생성되지만, 차별적 스플라이싱(Vp1) 및 대체 번역 시작 부위(각각 Vp2 및 Vp3)로부터 생성된다. Vp3은 비리온에서 가장 풍부한 서브유닛이며, 바이러스의 지향성을 정의하는 세포 표면에서 수용체 인식에 참여한다. 바이러스 감염에서 기능하는 포스포리페이스 도메인은 Vp1의 고유한 N 말단에서 확인되었다.

[1979] wt AAV와 유사하게, 재조합 AAV(rAAV)는 시스-작용성 145-bp ITR을 사용하여 벡터 이식유전자 카세트의 측면에 배치하여, 외래 DNA의 패키징을 위해 최대 4.5kb를 제공한다. 감염 후, rAAV는 본 발명의 융합 단백질을 발현할 수 있고, 원형의 수미식 연쇄체(head-to-tail concatemer)에서 에피솜으로 존재함으로써 숙주 게놈으로의 통합 없이 지속될 수 있다. 시험관내 및 생체내에서 이 시스템을 사용한 rAAV의 성공의 많은 예가 있지만, 제한된 패키징 용량은 유전자의 코딩 서열의 길이가 wt AAV 게놈보다 크거나 같을 때 AAV-매개성 유전자 전달의 사용을 제한하였다.

[1980] 바이러스 벡터는 적용에 기초하여 선택될 수 있다. 예를 들어, 생체내 유전자 전달의 경우, AAV는 다른 바이러스 벡터보다 유리할 수 있다. 일부 실시형태에서, AAV는 면역 반응을 활성화할 수 있는 세포 입자의 초-원심분리가 필요하지 않은 정제 방법으로 인해 낮은 독성을 허용한다. 일부 실시형태에서, AAV는 숙주 게놈에 통합되지 않기 때문에 삽입 돌연변이 유발 가능성이 낮다. 아데노바이러스는 이들이 유도하는 강한 면역원성 반응 때문에 일반적으로 백신으로 사용된다. 바이러스 벡터의 패키징 용량은 벡터에 패키징될 수 있는 염기 편집기의 크기를 제한할 수 있다.

[1981] AAV는 2개의 145개 염기 역 말단 반복부(ITR)를 포함하는 약 4.5Kb 또는 4.75Kb의 패키징 용량을 갖는다. 이는 개시된 염기 편집기뿐만 아니라 프로모터 및 전사 종결인자가 단일 바이러스 벡터에 적합할 수 있음을 의미한다. 4.5Kb 또는 4.75Kb보다 큰 작제물은 바이러스 생산을 크게 감소시킬 수 있다. 예를 들어, SpCas9는 상당히 크고, 유전자 자체가 4.1Kb를 초과하여 AAV에 패킹하기 어렵다. 따라서, 본 개시내용의 실시형태는 종래의 염기 편집기보다 길이가 짧은 개시된 염기 편집기를 이용하는 것을 포함한다. 일부 예에서, 염기 편집기는 4kb 미만이다. 개시된 염기 편집기는 4.5kb, 4.4kb, 4.3kb, 4.2kb, 4.1kb, 4kb, 3.9kb, 3.8kb, 3.7kb, 3.6kb, 3.5kb, 3.4kb, 3.3kb, 3.2kb, 3.1kb, 3kb, 2.9kb, 2.8kb, 2.7kb, 2.6kb, 2.5kb, 2 kb 또는 1.5kb 미만일 수 있다. 일부 실시형태에서, 개시된 염기 편집기는 4.5kb 이하의 길이이다.

[1982] AAV는 AAV1, AAV2, AAV5 또는 이들의 임의의 조합일 수 있다. 표적화될 세포와 관련하여 AAV의 유형을 선택할 수 있다; 예를 들어, 뇌 또는 신경 세포를 표적화하기 위해 AAV 혈청형 1, 2, 5 또는 하이브리드 캡시드 AAV1, AAV2, AAV5 또는 이들의 임의의 조합을 선택할 수 있고; 표적화 심장 조직을 표적화하기 위해 AAV4를 선택할 수 있다. AAV8은 간으로의 전달에 유용하다. 이들 세포에 관한 소정의 AAV 혈청형의 도표는 문헌[Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)]에서 찾을 수 있다.

[1983] 일부 실시형태에서, 렌티바이러스 벡터는 키메라 항원 수용체로 변형된 면역 세포(예를 들어, T-세포 또는 NK-세포)를 형질도입하는데 사용된다. 렌티바이러스는 유사분열 및 유사분열 후 세포 돌 다에서 유전자를 감염시키고 이를 발현하는 능력을 갖는 복잡한 레트로바이러스이다. 가장 일반적으로 알려진 렌티바이러스는 다른 바이러스의 외피 당단백질을 사용하여 광범위한 세포 유형을 표적으로 하는 인간 면역결핍바이러스(HIV)이다.

[1984] 렌티바이러스는 다음과 같이 제조될 수 있다. pCasES10(렌티바이러스 전달 플라스미드 백본을 포함함)을 클로닝한 후, 항생제 없이 10% 소 태아 혈청이 포함된 DMEM에서 형질감염시키기 전날 낮은 계대(p=5)의 HEK293FT를 T-75 플라스크에 50%의 컨플루언스로 시딩하였다. 20시간 후, 배지를 OptiMEM(무혈청) 배지로 변경하고, 4시간 후에 형질감염을 수행하였다. 세포를 10㎍의 렌티바이러스 전달 플라스미드(pCasES10) 및 다음 패키징 플라스미드로 형질감염시켰다: 5㎍의 pMD2.G(VSV-g 슈도형) 및 7.5㎍의 psPAX2(gag/pol/rev/tat). 형질감염은 양이온성 지질 전달제(50㎕의 Lipofectamine 2000 및 100㎕의 Plus 시약)를 사용하여 4㎖의 OptiMEM에서 수행할 수 있다. 6시간 후, 배지를 10% 소 태아 혈청이 포함된 무항생제 DMEM으로 변경하였다. 이러한 방법은 배양 중에 혈청을 사용하지만, 무혈청 방법이 선호된다.

[1985] 렌티바이러스는 다음과 같이 정제될 수 있다. 바이러스 상청액을 48시간 후 수확하였다. 상청액은 먼저 찌꺼기를 제거하고, 0.45㎛의 저단백 결합 필터(protein binding filter: PVDF)를 통해 여과하였다. 그런 다음, 이를 초원심분리기에서 24,000rpm에서 2시간 동안 회전시켰다. 바이러스 펠릿을 4℃에서 밤새 50㎕의 DMEM에 재현탁시켰다. 그런 다음, 이를 분취하고, -80℃에서 즉시 동결시켰다.

[1986] 또 다른 실시형태에서, 말 감염성 빈혈 바이러스(equine infectious anemia virus: EIAV)를 기반으로 하는 최소 비영장류 렌티바이러스 벡터가 또한 상정된다. 또 다른 실시형태에서, 망막하 주사를 통해 전달되는 혈관신생억제 단백질 엔도스타틴 및 안지오스타틴을 발현하는 말 감염성 빈혈 바이러스-기반 렌티바이러스 유전자 요법 벡터 RetinoStat®가 상정된다. 또 다른 실시형태에서, 자기-불활성화 렌티바이러스 벡터의 사용이 상정된다.

[1987] 시스템의 임의의 RNA, 예를 들어, 가이드 RNA 또는 염기 편집기-암호화 mRNA는 RNA의 형태로 전달될 수 있다. 염기 편집기-암호화 mRNA는 시험관내 전사를 사용하여 생성될 수 있다. 예를 들어, 뉴클레이스 mRNA는 다음 요소를 포함하는 PCR 카세트를 사용하여 합성될 수 있다: T7 프로모터, 선택적 코작 서열(GCCACC), 뉴클레이스 서열 및 3' UTR, 예컨대, 베타 글로빈-폴리A 꼬리로부터의 3' UTR. 카세트는 T7 폴리머레이스에 의한 전사에 사용될 수 있다. 가이드 폴리뉴클레오타이드(예를 들어, gRNA)는 또한 T7 프로모터, 이어서 서열 "GG" 및 가이드 폴리뉴클레오타이드 서열을 포함하는 카세트로부터의 시험관내 전사를 사용하여 전사될 수 있다.

[1988] 발현을 향상시키고 가능한 독성을 감소시키기 위해, 염기 편집기-코딩 서열 및/또는 가이드 핵산은, 예를 들어, 유사-U 또는 5-메틸-C를 사용하여 하나 이상의 변형된 뉴클레오사이드를 포함하도록 변형될 수 있다.

[1989] AAV 벡터의 작은 패키징 용량은 이러한 크기를 초과하는 많은 유전자의 전달 및/또는 큰 생리학적 조절 요소의 사용을 어렵게 한다. 이러한 문제는, 예를 들어, 전달될 단백질(들)을 2개 이상의 단편으로 나눔으로써 해결될 수 있되, N-말단 단편은 스플릿 인테인-N에 융합되고, C-말단 단편은 스플릿 인테인-C에 융합된다. 그런 다음, 이 단편은 2개 이상의 AAV 벡터로 패키징 된다. 본 명세서에서 사용되는 "인테인"은 측면 N-말단 및 C-말단 엑스테인(예를 들어, 연결될 단편)을 결찰하는 자기-스플라이싱 단백질 인트론(예를 들어, 펩타이드)을 지칭한다. 이종성 단백질 단편을 연결하기 위한 소정의 인테인의 사용은, 예를 들어, 문헌[Wood *et al.*, J. Biol. Chem. 289(21); 14512-9 (2014)]에 기술되어 있다. 예를 들어, 별도의 단백질 단편에 융합될 때, 인테인 IntN 및 IntC는 서로를 인식하고, 스스로를 스플라이스하며, 동시에 이들이 융합된 단백질 단편의 측면 N-말단 및 C-말단 엑스테인을 결찰하여 2개의 단백질 단편으로부터 전장 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 명백할 것이다.

[1990] 본 발명의 융합 단백질의 단편은 길이가 다양할 수 있다. 일부 실시형태에서, 단백질 단편은 2개의 아미노산 내지 약 1000개의 아미노산 길이 범위이다. 일부 실시형태에서, 단백질 단편은 약 5개의 아미노산 내지 약 500개의 아미노산 길이 범위이다. 일부 실시형태에서, 단백질 단편은 약 20개의 아미노산 내지 약 200개의 아미노산 길이 범위이다. 일부 실시형태에서, 단백질 단편은 약 10개의 아미노산 내지 약 100개의 아미노산 길이 범위이다. 다른 길이의 적합한 단백질 단편은 당업자에게 명백할 것이다.

[1991] 일 실시형태에서, 이중 AAV 벡터는 큰 이식유전자 발현 카세트를 2개의 개별 절반(5' 및 3' 말단 또는 헤드 및 꼬리)으로 분할함으로써 생성되되, 카세트의 각 절반은 단일 AAV 벡터(5kb 미만)에 패키징 된다. 그런 다음, 전장 이식유전자 발현 카세트의 재조립이 다음 이중 AAV 벡터에 의한 동일한 세포의 동시-감염시에 달성된다: (1) 5' 및 3' 게놈 간의 상동 재조합(HR)(이중 AAV 중첩 벡터); (2) 5' 및 3' 게놈의 ITR-매개성 수미식 연쇄화

(concatemerization)(이중 AAV 트랜스-스플라이싱 벡터); 또는 (3) 이들 2개의 메커니즘의 조합(이중 AAV 하이브리드 벡터). 생체내 이중 AAV 벡터의 사용은 전장 단백질의 발현을 초래한다. 이중 AAV 벡터 플랫폼의 사용은 4.7kb 초과의 크기의 이식유전자에 대한 효율적이고 및 실행 가능한 유전자 전달 전략을 나타낸다.

[1992] ***인테인***

[1993] 인테인(개재 단백질)은 단백질 스플라이싱으로 알려진 과정을 수행하는 다양한 가지각색의 유기체에서 발견되는 자가-처리 도메인이다. 단백질 스플라이싱은 펩타이드 결합의 절단 및 형성으로 구성된 다단계 생화학 반응이다. 단백질 스플라이싱의 내인성 기질은 인테인-함유 유기체에서 발견되는 단백질이지만, 인테인은 사실상 임의의 폴리펩타이드 백본을 화학적으로 조작하는데 사용될 수 있다.

[1994] 단백질 스플라이싱에서, 인테인은 2개의 펩타이드 결합을 절단함으로써 전구체 폴리펩타이드로부터 스스로를 절단하여 새로운 펩타이드 결합의 형성을 통해 측면 엑스테인(외부 단백질) 서열을 결찰한다. 이러한 재배열은 번역 후(또는 가능한 동시-번역으로) 발생한다. 인테인-매개성 단백질 스플라이싱은 자발적으로 발생하여 인테인 도메인의 접힘만을 필요로 한다.

[1995] 약 5%의 인테인이 스플릿 인테인이며, 이는 각각 하나의 엑스테인에 융합된 N-인테인 및 C-인테인인 2개의 별도의 폴리펩타이드로 전사 및 번역된다. 번역시, 인테인 단편은 트랜스로 단백질 스플라이싱을 수행하기 위해 자발적이고 비공유적으로 표준 인테인 구조로 조립된다. 단백질 스플라이싱의 메커니즘은 인테인-엑스테인 접합부에서 2개의 펩타이드 결합이 절단되고 N-엑스테인과 C-엑스테인 사이에 새로운 펩타이드 결합을 형성되는 일련의 아실-전달 반응을 수반한다. 이러한 과정은 연결 N-엑스테인과 인테인의 N-말단을 연결하는 펩타이드 결합의 활성화에 의해 개시된다. 사실상 모든 인테인은 N-말단에 C-말단 N-엑스테인 잔기의 카보닐 탄소를 공격하는 시스테인 또는 세린을 갖는다. 이러한 N에서 O/S로의 아실-이동은 일반적으로 발견되는 아스파테이트와 함께 보존된 트레오닌 및 히스티딘(TXXH 모티프로 지칭됨)에 의해 촉진되어 선형 (티오)에스터 중간체를 형성한다. 다음으로, 이러한 중간체는 시스테인, 세린 또는 트레오닌인 첫 번째 C-엑스테인 잔기(+1)의 친핵성 공격에 의해 트랜스-(티오)에스터화된다. 생성된 분지형 (티오)에스터 중간체는 인테인의 고도로 보존된 C-말단 아스파라긴의 고리화라는 독특한 변형을 통해 분해된다. 이러한 과정은 히스티딘(고도로 보존된 HNF 모티프에서 발견됨)과 끝에서 두 번째의 히스티딘에 의해 촉진되며, 아스파테이트를 포함할 수도 있다. 이러한 석신이미드 형성 반응은 반응성 복합체로부터 인테인을 제거하고, 비-펩티드 연결을 통해 부착된 엑스테인을 남긴다. 이러한 구조는 인테인-독립적인 방식으로 안정적인 펩타이드 결합으로 신속하게 재배열된다.

[1996] 일부 실시형태에서, 뉴클레이스(예를 들어, Cas9)의 일부 또는 단편이 인테인에 융합된다. 뉴클레이스는 인테인의 N-말단 또는 C-말단에 융합될 수 있다. 일부 실시형태에서, 융합 단백질의 일부 또는 단편은 인테인에 융합되고, AAV 캡시드 단백질에 융합된다. 인테인, 뉴클레이스 및 캡시드 단백질은 임의의 배열(예를 들어, 뉴클레이스-인테인-캡시드, 인테인-뉴클레이스-캡시드, 캡시드-인테인-뉴클레이스 등)로 함께 융합될 수 있다. 일부 실시형태에서, 염기 편집기(예를 들어, ABE, CBE)의 N-말단 단편은 스플릿 인테인-N에 융합되고, C-말단 단편은 스플릿 인테인-C에 융합된다. 이들 단편은 이어서 2개 이상의 AAV 벡터에 패키징 된다. 일부 실시형태에서, 인테인의 N-말단은 융합 단백질의 C-말단에 융합되고, 인테인의 C-말단은 AAV 캡시드 단백질의 N-말단에 융합된다.

[1997] 일 실시형태에서, 인테인은 AAV 캡시드 단백질에 이식된 사이티딘 또는 아데노신 데아미네이스 염기 편집기 단백질의 단편 또는 일부를 연결하는데 사용된다. 이종성 단백질 단편을 연결하기 위한 소정의 인테인의 사용은, 예를 들어, 문헌[Wood *et al.*, J. Biol. Chem. 289(21); 14512-9 (2014)]에 기술되어 있다. 예를 들어, 별도의 단백질 단편에 융합될 때, 인테인 IntN 및 IntC는 서로를 인식하고, 스스로를 스플라이싱하며, 동시에 이들이 융합된 단백질 단편의 측면 N-말단 및 C-말단 엑스테인을 결찰하여 2개의 단백질 단편으로부터 전장 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 명백할 것이다.

[1998] 일부 실시형태에서, ABE는 SpCas9의 선택된 영역 내의 Ala, Ser, Thr 또는 Cys 잔기에서 N-말단 및 C-말단 단편으로 스플릿 되었다. 이들 영역은 Cas9 결정 구조 분석에 의해 확인된 루프 영역에 해당한다. 아미노산 위치 S303, T310, T313, S355, A456, S460, A463, T466, S469, T472, T474, C574, S577, A589 및 S590에서 각 단편의 N-말단은 인테인-N에 융합되고, 각 단편의 C-말단은 인테인 C에 융합되며, 이는 아래 서열에서 대문자로 표시되어 있다.

```
   1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntdr hsikknliga llfdsgetae
  61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
 121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
 181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
 241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
 301 llSdilrvnT eiTkaplsas mikrydehhq dltllkalvr qqlpekykei ffdqSkngya
 361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
 421 ailrrqedfy pflkdnreki ekiltfripy yvgplArgnS rfAwmTrkSe eTiTpwnfee
 481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kieCfdSvei sgvedrfnAS lgtyhdllki
 601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg
 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
 721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
 841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd
```

[1999]

[2000] **핵염기 편집기의 사용**

[2001] 신생물 세포와 관련된 표적 항원은 또한 건강한 면역 세포에서 발현될 수 있다. 따라서, 변형된 면역 세포(예를 들어, CAR-T)를 사용하여, 신생물 세포 상의 항원을 표적으로 하는 세포는 또한 달리 동족 살해 또는 자기-살해로 알려져 있는 해당 항원을 발현하는 건강한 면역 세포의 살해를 초래할 수 있다. 동족 살해 저항성을 제공하기 위해 변형된 면역 세포에서 표적 항원(예를 들어, CD2)의 발현을 감소시키거나 또는 제거할 수 있는 염기 편집기와 조합한 변형된 면역 세포의 사용은 치료 및 기초 연구에 적용되는 유전자 교정을 사용하기 위한 새로운 전략을 제공한다.

[2002] 본 개시내용은 동족 살해 저항성의 변형된 면역 세포를 생산하는 방법을 제공한다. 일부 실시형태에서, 핵염기 편집기(예를 들어, 아데노신 데아미네이스 염기 편집기 또는 사이티딘 데아미네이스 염기 편집기) 및 하나 이상의 가이드 RNA를 사용하여 건강한 대상체 또는 공여체로부터 얻어진 면역 세포(예를 들어, T-세포 또는 NK-세포)를 편집하여 면역 세포(예를 들어, T-세포 또는 NK-세포)에서 표적 항원(예를 들어, CD2)의 발현을 감소시키거나 또는 제거하는 단계를 포함하는 방법이 제공된다. 그런 다음, 면역 세포는 표적 항원(예를 들어, CD2)에 대해 지시된 키메라 항원 수용체로 형질도입된다. 생성된 변형된 면역 세포(예를 들어, CAR-T)는 동족 살해 저항성이며, 건강한 면역 세포를 표적으로 하지 않고 신생물 세포를 표적화할 수 있다. 일부 실시형태에서, 동족 살해 저항성의 변형된 면역 세포는 **실시예 2**에 제공된 바와 같이 생성된다. 일부 실시형태에서, 동족 살해 저항성 CD2 CAR-T 세포를 생성하는 방법이 제공된다.

[2003] 본 개시내용은 또한 본 명세서에 제공된 바와 같은 변형된 면역 세포를 사용하여 질환(예를 들어, T-세포 또는 NK-세포 악성종양)이 있는 것으로 진단된 대상체의 치료 방법을 제공한다. 일부 실시형태에서, 유효량의 본 명세서에 제공된 바와 같이 생성된 변형된 면역 세포를 질환(예를 들어, T-세포 또는 NK-세포 악성종양)을 갖거나 또는 갖는 경향이 있는 대상체에게 투여하는 단계를 포함하는 방법이 제공된다. 본 개시내용은 데아미네이스 매개성 유전자 편집에 의해 교정될 수 있는 표적 항원과 관련된 T-세포 또는 NK-세포 악성종양의 치료 방법을 제공한다. 본 명세서에 제공된 바와 같은 전략 및 변형된 면역 세포로 치료될 수 있는 적합한 질환은 본 개시내용에 기초하여 당업자에게 명백할 것이다.

[2004] 표적 뉴클레오타이드 서열에서 핵염기를 편집하기 위해 염기 편집기 또는 염기 편집 시스템을 사용하는 방법이 본 명세서에 제공된다. 일부 실시형태에서, 염기 편집 시스템(예를 들어, 데아미네이스, Cas9 도메인 및 하나 이상의 가이드 RNA를 포함)의 활성은 스플라이싱 부위의 파괴를 초래하거나 또는 조기 종결 코돈을 표적 뉴클레오타이드 서열에 도입시킨다.

[2005] 일부 실시형태에서, 본 명세서에 제공된 데아미네이스(예를 들어, 아데노신 데아미네이스 또는 사이티딘 데아미

네이스)는 DNA의 데옥시아데노신 잔기를 탈아미노화할 수 있다. 본 개시내용의 다른 양태는 데아미네이스(예를 들어, 아데노신 데아미네이스 또는 사이티딘 데아미네이스) 및 특정 뉴클레오타이드 서열에 결합할 수 있는 도메인(예를 들어, Cas9 또는 Cpf1 단백질)을 포함하는 융합 단백질을 제공한다. 예를 들어, 아데노신은 전형적으로 사이토신 잔기와 염기쌍을 이루는 이노신 잔기로 전환될 수 있다. 이러한 융합 단백질은 특히 핵산 서열의 표적화된 편집에 유용하다. 이러한 융합 단백질은, 예를 들어, 돌연변이체 세포 또는 동물의 생성을 위해 시험관내에서 DNA의 표적화된 편집을 위해; 표적화된 돌연변이의 도입을 위해, 예를 들어, 생체외 세포, 예를 들어, 동일한 또는 또 다른 대상체에게 후속적으로 재도입되는 대상체로부터 얻은 세포의 유전적 결함의 교정을 위해; 그리고 생체내 표적화된 돌연변이의 도입, 예를 들어, 유전적 결함의 교정을 위해 사용될 수 있거나, 또는 G에서 A로 또는 T에서 C로 돌연변이에서 질환-관련 유전자의 불활성화 돌연변이의 도입은 본 명세서에 제공된 핵염기 편집기를 사용하여 치료될 수 있다. 본 개시내용은 데아미네이스 및 핵염기 편집기를 이용하는 변형된 면역세포, 융합 단백질, 핵산, 벡터, 조성물, 방법, 키트, 시스템 등을 제공한다.

[2006] ***핵산을 편집하는 방법***

[2007] 본 개시내용의 일부 양태는 핵산을 편집하는 방법을 제공한다. 일부 실시형태에서, 방법은 핵산의 핵염기(예를 들어, 이중-가닥 DNA 서열의 염기쌍)를 편집하는 방법이다. 일부 실시형태에서, 방법은 a) 핵산의 표적 영역(예를 들어, 이중-가닥 DNA 서열)을 염기 편집기(예를 들어, 사이티딘 또는 아데노신 데아미네이스에 융합된 Cas9 도메인) 및 가이드 핵산(예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계로서, 표적 영역은 표적화된 핵염기쌍을 포함하는, 상기 접촉시키는 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 표적 영역의 단일 가닥에 있는 상기 표적 핵염기쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계 및 d) 상기 표적 영역의 1개 이하의 가닥을 절단하는 단계를 포함하되, 제1 핵염기에 상보적인 제3 핵염기는 제2 핵염기에 상보적인 제4 핵염기로 대체된다. 일부 실시형태에서, 방법은 핵산에서 20% 미만의 indel 형성을 초래한다. 일부 실시형태에서, 단계 b가 생략됨을 이해하여야 한다. 일부 실시형태에서, 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만 또는 0.1% 미만의 indel 형성을 초래한다. 일부 실시형태에서, 방법은 제2 핵염기를 제4 핵염기에 상보적인 제5 핵염기로 대체하여 의도된 편집된 염기쌍(예를 들어, C·G에서 T·A로)을 생성하는 단계를 더 포함한다. 일부 실시형태에서, 의도된 염기쌍의 적어도 5%가 편집된다. 일부 실시형태에서, 의도된 염기쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45% 또는 50%가 편집된다.

[2008] 일부 실시형태에서, 표적 뉴클레오타이드에서 의도된 생성물 대 의도하지 않은 생성물의 비는 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1 또는 200:1 이상이다. 일부 실시형태에서, 의도된 돌연변이 대 indel 형성의 비는 1:1, 10:1, 50:1, 100:1, 500:1 또는 1000:1 이상이다. 일부 실시형태에서, 절단된 단일 가닥(틈이 있는 가닥)은 가이드 핵산에 혼성화된다. 일부 실시형태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥 반대편에 있다. 일부 실시형태에서, 염기 편집기는 Cas9 도메인을 포함한다. 일부 실시형태에서, 염기 편집기는 편집되지 않은 가닥을 보호하거나 또는 이에 결합한다. 일부 실시형태에서, 염기 편집기는 닉케이스 활성을 포함한다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 상류에 있다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 상류에서 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드이다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 하류에 있다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 하류에서 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드이다. 일부 실시형태에서, 방법은 표준(예를 들어, NGG) PAM 부위를 필요로 하지 않는다. 일부 실시형태에서, 핵염기 편집기는 링커를 포함한다. 일부 실시형태에서, 링커는 1개 내지 25개의 아미노산 길이이다. 일부 실시형태에서, 링커는 5개 내지 20개의 아미노산 길이이다. 일부 실시형태에서, 링커는 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 아미노산 길이이다. 일 실시형태에서, 링커는 32개의 아미노산 길이이다. 또 다른 실시형태에서, "긴 링커"는 적어도 약 60개의 아미노산 길이이다. 다른 실시형태에서, 링커는 약 3개 내지 100개의 아미노산 길이이다. 일부 실시형태에서, 표적 영역은 표적 창을 포함하되, 표적 창은 표적 핵염기쌍을 포함한다. 일부 실시형태에서, 표적 창은 1개 내지 10개의 뉴클레오타이드를 포함한다. 일부 실시형태에서, 표적 창은 1개 내지 9개, 1개 내지 8개, 1개 내지 7개, 1개 내지 6개, 1개 내지 5개, 1개 내지 4개, 1개 내지 3개, 1개 내지 2개 또는 1개의 뉴클레오타이드 길이이다. 일부 실시형태에서, 표적 창은 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드 길이이다. 일부 실시형태에서, 의도된 편집된 염기쌍은 표적 창 내에 있다. 일부 실시형태에서, 표적 창은 의도된 편집된 염기쌍을 포함한다. 일부 실시형태에서, 방법은 본 명세서에 제공된 임의의 염기 편집기를 사용하여 수행된다. 일부 실시형태에서, 표적 창은 메틸화 창

이다.

[2009] 일부 실시형태에서, 본 개시내용은 뉴클레오타이드를 편집하는 방법을 제공한다. 일부 실시형태에서, 본 개시내용은 이중-가닥 DNA 서열의 핵염기쌍을 편집하는 방법을 제공한다. 일부 실시형태에서, 방법은 a) 이중-가닥 DNA 서열의 표적 영역을 염기 편집기 및 가이드 핵산(예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계로서, 표적 영역은 표적 핵염기쌍을 포함하는, 상기 접촉시키는 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 표적 영역의 단일 가닥에서 상기 표적 핵염기쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, d) 상기 표적 영역의 1개 이하의 가닥을 절단하는 단계를 포함하되, 제1 핵염기에 상보적인 제3 핵염기는 제2 핵염기에 상보적인 제4 핵염기로 대체되고, 제2 핵염기는 제4 핵염기에 상보적인 제5 핵염기로 대체되어 의도된 편집된 염기쌍을 생성하고, 의도된 편집된 염기쌍을 생성하는 효율은 적어도 5%이다. 일부 실시형태에서, 단계 b는 생략됨을 이해하여야 한다. 일부 실시형태에서, 의도된 염기쌍의 적어도 5%가 편집된다. 일부 실시형태에서, 의도된 염기쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45% 또는 50%가 편집된다. 일부 실시형태에서, 본 명세서에 기재된 방법에 의한 염기 편집은 임의의 특정 유전자 부위에서 적어도 10%의 염기 전환 효율을 가질 수 있다. 일부 실시형태에서, 본 명세서에 기재된 방법에 의한 염기 편집은 임의의 특정 유전자 부위에서 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50% 적어도 55% 또는 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 또는 적어도 95%, 96%, 97%, 98% 또는 적어도 99%의 염기 전환 효율을 가질 수 있다. 일부 실시형태에서, 본 명세서에 기재된 방법에 의한 염기 편집은 임의의 특정 유전자 부위에서 적어도 70%의 염기 전환 효율을 가질 수 있다. 일부 실시형태에서, 본 명세서에 기재된 방법에 의한 염기 편집은 임의의 특정 유전자 부위에서 적어도 80%의 염기 전환 효율을 가질 수 있다. 일부 실시형태에서, 본 명세서에 기재된 방법에 의한 염기 편집은 임의의 특정 유전자 부위에서 적어도 90%의 염기 전환 효율을 가질 수 있다.

[2010] 일부 실시형태에서, 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만 또는 0.1% 미만의 indel 형성을 유발한다. 일부 실시형태에서, 표적 뉴클레오타이드에서 의도된 생성물 대 의도하지 않은 생성물의 비는 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1 또는 200:1 이상이다. 일부 실시형태에서, 의도된 돌연변이 대 indel 형성의 비는 1:1, 10:1, 50:1, 100:1, 500:1 또는 1000:1 이상이다. 일부 실시형태에서, 절단된 단일 가닥은 가이드 핵산에 혼성화된다. 일부 실시형태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥의 반대편에 있다. 일부 실시형태에서, 핵염기 편집기는 닉케이스 활성을 포함한다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 상류에 있다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 상류에서 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드이다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 하류에 있다. 일부 실시형태에서, 의도된 편집된 염기쌍은 PAM 부위의 하류에서 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드이다. 일부 실시형태에서, 방법은 표준(예를 들어, NGG) PAM 부위를 필요로 하지 않는다. 일부 실시형태에서, 핵염기 편집기는 링커를 포함한다. 일부 실시형태에서, 링커는 1개 내지 25개의 아미노산 길이이다. 일부 실시형태에서, 링커는 5개 내지 20개의 아미노산 길이이다. 일부 실시형태에서, 링커는 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 아미노산 길이이다. 예를 들어, 일부 실시형태에서, 표적 영역은 표적 창을 포함하되, 표적 창은 표적 핵염기쌍을 포함한다. 일부 실시형태에서, 표적 창은 1개 내지 10개의 뉴클레오타이드를 포함한다. 일부 실시형태에서, 표적 창은 1개 내지 9개, 1개 내지 8개, 1개 내지 7개, 1개 내지 6개, 1개 내지 5개, 1개 내지 4개, 1개 내지 3개, 1개 내지 2개 또는 1개의 뉴클레오타이드 길이이다. 일부 실시형태에서, 표적 창은 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개의 뉴클레오타이드 길이이다. 일부 실시형태에서, 의도된 편집된 염기쌍은 표적 창 내에서 발생한다. 일부 실시형태에서, 표적 창은 의도된 편집된 염기쌍을 포함한다. 일부 실시형태에서, 핵염기 편집기는 본 명세서에 제공된 염기 편집기 중 어느 하나이다.

[2011] ***약제학적 조성물***

[2012] 일부 양태에서, 본 발명은 본 명세서에 기재된 임의의 유전적으로 변형된 면역 세포, 염기 편집기, 융합 단백질 또는 융합 단백질-가이드 폴리뉴클레오타이드 복합체를 포함하는 약제학적 조성물을 제공한다. 보다 구체적으로, 유전적으로 변형된 면역 세포 또는 키메라 항원 수용체(CAR)를 발현하는 이러한 면역 세포의 집단을 포함하는 약제학적 조성물이 본 명세서에 제공되되, 상기 변형된 면역 세포 또는 이의 집단은 동족 살해 저항성을 제공하거나, 변형된 면역 세포의 기능을 향상시키거나 또는 변형된 면역 세포의 면역억제 또는 저해를 감소시키기 위해 적어도 하나의 편집된 유전자를 갖고, 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다.

일부 실시형태에서, 적어도 하나의 편집된 유전자는 CD3, CD5, CD7, CD33, CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M, PDCD1/PD1, CBLB, TGFBR2, ZAP70, NFATc1, TET2 또는 이들의 조합이다. 일부 실시형태에서, CD3, CD5, CD7, CD33 또는 CD123이 편집된다. 일부 실시형태에서, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PDCD1/PD1으로부터 선택되는 하나 이상의 유전자와 조합하여 CD3, CD5, CD7, CD33 또는 CD123이 편집된다.

[2013] 일부 실시형태에서, 약제학적 조성물은 CD5가 결여되어 있거나 또는 이의 수준이 감소되고 CD5 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD7이 결여되어 있거나 또는 이의 수준이 감소되고 CD7 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD33이 결여되어 있거나 또는 이의 수준이 감소되고 CD33 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD3이 결여되어 있거나 또는 이의 수준이 감소되고 CD3 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD123이 결여되어 있거나 또는 이의 수준이 감소되고 CD123 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다.

[2014] 일부 실시형태에서, 약제학적 조성물은 CD5, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PDCD1/PD1로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이들의 수준이 감소되고 CD5 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD7, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이들의 수준이 감소되고 CD7 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD33, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이들의 수준이 감소되고 CD33 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD3, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이들의 수준이 감소되고 CD3 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 CD123, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이들의 수준이 감소되고 CD123 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다.

[2015] 본 발명의 약제학적 조성물은 공지된 기법에 따라 제조될 수 있다. 예를 들어, 문헌[Remington, The Science And Practice of Pharmacy (21st ed. 2005)] 참조. 일반적으로, 면역 세포 또는 이의 집단은 투여 또는 보관 전에 적합한 담체와 혼합되고, 일부 실시형태에서, 약제학적 조성물은 약제학적으로 허용 가능한 담체를 더 포함한다. 적합한 약제학적으로 허용 가능한 담체는 일반적으로 약제학적 조성물을 대상체에게 투여하는 것을 보조하거나, 약제학적 조성물을 전달 가능한 제제로 가공하는 것을 보조하거나 또는 투여 전에 약제학적 조성물을 보관하는 것을 보조하는 비활성 물질을 포함한다. 약제학적으로 허용 가능한 담체는 제형의 형태, 일관성, 점도, pH, 약동학, 용해도를 안정화, 최적화 또는 그렇지 않으면 변경할 수 있는 작용제를 포함할 수 있다. 이러한 작용제는 완충제, 습윤제, 유화제, 희석제, 캡슐화제 및 피부 침투 강화제를 포함한다. 예를 들어, 담체는 식염수, 완충 식염수, 덱스트로스, 아르기닌, 수크로스, 물, 글리세롤, 에탄올, 소르비톨, 덱스트란, 소듐 카복시메틸 셀룰로스 및 이들의 조합을 포함할 수 있지만, 이들로 제한되지 않는다.

[2016] 약제학적으로 허용 가능한 담체로 작용할 수 있는 물질의 일부 비제한적인 예는 (1) 당, 예컨대, 락토스, 글루코스 및 수크로스; (2) 전분, 예컨대, 옥수수 전분 및 감자 전분; (3) 셀룰로스 및 이의 유도체, 예컨대, 소듐 카복시메틸 셀룰로스, 메틸셀룰로스, 에틸 셀룰로스, 미세결정질 셀룰로스 및 셀룰로스 아세테이트; (4) 분상 트래거캔스; (5) 맥아; (6) 젤라틴; (7) 윤활제, 예컨대, 마그네슘 스테아레이트, 소듐 라우릴 설페이트 및 활석; (8) 부형제, 예컨대, 코코아 버터 및 좌약 왁스; (9) 오일, 예컨대, 땅콩유, 면실유, 홍화유, 참깨유, 올리브유, 옥수수유 및 대두유; (10) 글리콜, 예컨대, 프로필렌 글리콜; (11) 폴리올, 예컨대, 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜(PEG); (12) 에스터, 예컨대, 에틸 올레이트 및 에틸 라우레이트; (13) 아가; (14) 완충제, 예컨대, 마그네슘 하이드록사이드 및 알루미늄 하이드록사이드; (15) 알긴산; (16) 무발열원수; (17) 등장성 식염수; (18) 링거액; (19) 에틸 알코올; (20) pH 완충액; (21) 폴리에스터, 폴리카보네이트 및/또는 폴

리안하이드라이드; (22) 증량제, 예컨대, 폴리펩타이드 및 아미노산; (23) 혈청 알코올(serum alcohol), 예컨대, 에탄올; 및 (23) 약제학적 제형에 사용되는 기타 무독성의 양립 가능한 물질을 포함한다. 습윤제, 착색제, 방출제, 코팅제, 감미제, 향미제, 방향제, 방부제 및 항산화제가 또한 제형에 존재할 수 있다.

[2017] 약제학적 조성물은 제형의 pH를 약 5.0 내지 약 8.0의 범위와 같은 생리학적 pH를 반영하는 미리 결정된 수준으로 유지하기 위해 1종 이상의 pH 완충 화합물을 포함할 수 있다. 수성 액체 제형에 사용되는 pH 완충 화합물은 아미노산 또는 아미노산의 혼합물, 예컨대, 히스티딘 또는 히스티딘 및 글리신과 같은 아미노산의 혼합물일 수 있다. 대안적으로, pH 완충 화합물은 바람직하게는 제형의 pH를 약 5.0 내지 약 8.0의 범위와 같은 미리 결정된 수준으로 유지하고 칼슘 이온을 킬레이트화하지 않는 작용제이다. 이러한 pH 완충 화합물의 예시적인 예는 이미다졸 및 아세테이트 이온을 포함하지만, 이들로 제한되지 않는다. pH 완충 화합물은 제형의 pH를 미리 결정된 수준으로 유지하기에 적합한 임의의 양으로 존재할 수 있다.

[2018] 약제학적 조성물은 또한 1종 이상의 삼투 조절제, 즉, 수여 개체의 혈류 및 혈액 세포에 허용 가능한 수준으로 제형의 삼투 특성(예를 들어, 긴장성, 삼투질 농도 및/또는 삼투압)을 조절하는 화합물을 포함할 수 있다. 삼투 조절제는 칼슘 이온을 킬레이트화하지 않는 작용제일 수 있다. 삼투 조절제는 제형의 삼투 특성을 조절하는 당업자에게 공지되거나 또는 이용 가능한 임의의 화합물일 수 있다. 당업자는 본 발명의 제형에 사용하기 위해 주어진 삼투 조절제의 적합성을 경험적으로 결정할 수 있다. 적합한 유형의 삼투 조절제의 예시적인 예는 염, 예컨대, 소듐 클로라이드 및 소듐 아세테이트; 당, 예컨대, 수크로스, 덱스트로스 및 만니톨; 아미노산, 예컨대, 글리신; 및 1종 이상의 이러한 작용제 및/또는 작용제 유형의 혼합물을 포함하지만, 이들로 제한되지 않는다. 삼투 조절제(들)는 제형의 삼투 특성을 조절하기에 충분한 임의의 농도로 존재할 수 있다.

[2019] 변형된 면역 세포 또는 이의 집단 및 담체 이외에, 본 발명의 약제학적 조성물은 질환의 치료에 유용한 적어도 1종의 추가적인 치료제를 포함할 수 있다. 예를 들어, 본 명세서에 기재된 약제학적 조성물의 일부 실시형태는 화학치료제를 추가로 포함한다. 일부 실시형태에서, 약제학적 조성물은 사이토카인 펩타이드 또는 사이토카인 펩타이드를 암호화하는 핵산 서열을 추가로 포함한다. 일부 실시형태에서, 변형된 면역 세포 또는 이의 집단을 포함하는 약제학적 조성물은 추가적인 치료제와 별도로 투여될 수 있다.

[2020] 일부 실시형태에서, 적어도 1종의 추가적인 치료제는 하나 이상의 추가적인 변형된 면역 세포, 또는 이의 변형된 면역 세포의 하나 이상의 집단이다. 일부 실시형태에서, 하나 이상의 추가적인 변형된 면역 효과기 세포는 편집된 유전자의 발현을 넉아웃 또는 넉다운시키기 위한 적어도 하나의 편집된 유전자를 포함한다. 일부 실시형태에서, 적어도 하나의 편집된 유전자는 CD3, CD5, CD7, CD33, CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M, PD1 또는 이들의 조합이다. 일부 실시형태에서, CD3, CD5, CD7, CD33 또는 CD123이 편집된다. 일부 실시형태에서, CD3, CD5, CD7, CD33 또는 CD123은 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및/또는 PD1으로부터 선택되는 하나 이상의 유전자와 조합하여 편집된다.

[2021] 일부 실시형태에서, 약제학적 조성물은 이의 변형된 면역 세포의 2개 이상의 집단을 포함한다. 일부 실시형태에서, 약제학적 조성물은 1) CD5, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD5 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단; 및 2) CD7, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD7 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다.

[2022] 일부 실시형태에서, 약제학적 조성물은 1) CD33, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD33 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단; 및 2) CD123, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD123 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단을 포함한다.

[2023] 본 발명의 약제학적 조성물은 자가 또는 동종이계 면역 세포 면역요법에 반응성인 임의의 질환 또는 병태를 치료하는데 사용될 수 있다. 예를 들어, 약제학적 조성물은 일부 실시형태에서 신생물의 치료에 유용하다. 일부 실시형태에서, 신생물은 고형 종양이다. 다른 실시형태에서, 신생물은 액체 종양이다. 일부 실시형태에서, 신생물은 혈액암이다. 일부 실시형태에서, 혈액암은 B 세포암이고, 일부 실시형태에서, B 세포암은 다발성 골수종이다. 일부 실시형태에서, B 세포암은 재발성/불응성 다발성 골수종의 재발이다. 일부 실시형태에서, 혈액암은 백

혈병, 골수종 및/또는 림프종이다. 일부 경우에, 백혈병은 급성 백혈병이다. 급성 백혈병은, 예를 들어, 급성 골수성 백혈병(AML)을 포함한다. 급성 백혈병 또한, 예를 들어, 급성 림프성 백혈병 또는 급성 림프구성 백혈병(ALL)을 포함하고; ALL은 B-계통 ALL; T-계통 ALL; 및 T-세포 급성 림프구성 백혈병(T-ALL)을 포함한다.

[2024] 신생물의 비제한적인 예는 T-세포 급성 림프아구성 백혈병(T-ALL), 용상 진균증(MF), 세자리 증후군(SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 $ALK^+$, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 $CD30^+$ 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종을 포함한다. 일부 실시형태에서, 신생물은 T-세포 급성 림프아구성 백혈병(T-ALL)이다. 일부 실시형태에서, 신생물은 급성 골수성 백혈병(AML)이다.

[2025] 일부 실시형태에서, 신생물(예를 들어, 백혈병)이 발생하거나 발생하는 경향이 있는 대상체에는 유효량의 1) CD5, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및 PDCD1/PD1로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD5 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단; 2) CD7, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD7 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단; 3) CD33, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD33 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단; 및 4) CD123, 및 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M으로부터 선택되는 적어도 하나의 편집된 유전자가 결여되어 있거나 또는 이의 수준이 감소되고 CD123 키메라 항원 수용체를 발현하는 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단 중 하나 이상이 투여된다.

[2026] 본 발명의 유전적으로 변형된 면역 세포의 치료적 용도에 관한 한 가지 고려사항은 최적의 또는 만족스러운 효과를 달성하는데 필요한 세포의 양이다. 투여되는 세포의 양은 치료되는 대상체에 따라 다를 수 있다. 일 실시형태에서, $10^4$개 내지 $10^{10}$개, $10^5$개 내지 $10^9$개 또는 $10^6$개 내지 $10^8$개의 본 발명의 유전적으로 변형된 면역반응성 세포가 인간 대상체에게 투여된다. 일부 실시형태에서, 적어도 약 $1\times10^8$개, $2\times10^8$개, $3\times10^8$개, $4\times10^8$ 및 $5\times10^8$개의 본 발명의 유전적으로 변형된 면역 세포가 인간 대상체에게 투여된다. 정확한 유효 용량을 결정하는 것은 크기, 연령, 성별, 체중 및 병태를 포함하여 각 개별 대상체에 대한 인자를 기반으로 할 수 있다. 투여량은 본 개시내용 및 당해 분야의 지식으로부터 당업자에 의해 용이하게 확인될 수 있다.

[2027] 당업자는 본 발명의 방법에서 투여되는 조성물에서 세포의 수 및 선택적 첨가제, 비히클 및/또는 담체의 양을 쉽게 결정할 수 있다. 전형적으로, 첨가제(활성 면역 세포(들)에 더하여)는 인산 완충 식염수 중 0.001% 내지 50%(중량) 용액의 양으로 존재하고, 활성 성분은 약 0.0001 내지 약 5 wt%, 바람직하게는 약 0.0001 내지 약 1 wt%, 훨씬 더 바람직하게는 약 0.0001 내지 약 0.05 wt% 또는 약 0.001 내지 약 20 wt%, 바람직하게는 약 0.01 내지 약 10 wt%, 훨씬 더 바람직하게는 약 0.05 내지 약 5 wt%와 같이 마이크로그램에서 밀리그램의 순서로 존재한다. 물론, 동물 또는 인간에게 투여되는 임의의 조성물 및 임의의 특정 투여 방법에 대해, 따라서 적합한 동물 모델(예를 들어, 마우스와 같은 설치류)에서 치사량(LD) 및 LD50을 결정하는 것에 의해서와 같이 독성; 및 적합한 반응을 이끌어내는 조성물(들)의 투여량, 그 안에 있는 성분의 농도 및 조성물(들)을 투여하는 시기를 결정하는 것이 바람직하다. 이러한 결정은 당업자의 지식, 본 개시내용 및 본 명세서에 인용된 문서로부터 과도한 실험을 요구하지 않는다. 그리고, 순차적 투여의 시기는 과도한 실험 없이 확인될 수 있다.

[2028] 일 실시형태에서, 본 명세서에 기재된 방법 및 조성물은 CAR을 발현하는 조작된 T 세포를 생성하는데 사용될 수 있고, 조작된 T 세포가 표적에 대한 특이적 면역 반응을 일으킬 수 있도록 하나 이상의 염기 편집된 변형을 가질 수 있다. CAR은 항원 표적에 대해 특이적으로 지시될 수 있고, 항원은 숙주의 세포에 의해 제시될 수 있다. 일부 실시형태에서, 면역 반응은 세포독성을 포함한다. 일부 실시형태에서, 조작된 T 세포는 이의 표적에 대해 향상된 세포독성 반응을 갖는다. 일부 실시형태에서, 조작된 T 세포는 조작되지 않은 T 세포와 비교하여 이의 표적에 대해 향상된 세포독성 반응을 유도한다. 일부 실시형태에서, 조작된 T 세포는 조작되지 않은 세포와 비교하여 적어도 1.1-배, 1.5-배, 2-배, 3-배, 4-배, 5-배, 6-배, 7-배, 8-배, 9-배, 10-배, 11-배, 12-배, 13-배, 14-배, 15-배, 16-배, 17-배, 18-배, 19-배, 20-배 이상 향상된 세포독성 반응을 나타낸다. 일부 실시형태에서, 조작된 T 세포는 조작되지 않은 세포보다 적어도 10%, 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%,

적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 100%, 적어도 200%, 적어도 500% 또는 적어도 1000% 이상 더 많은 표적 세포를 죽일 수 있다. 일부 실시형태에서, T 세포는 더 높은 기억 반응을 유도할 수 있다. 일부 실시형태에서, T 세포는 조작되지 않은 세포보다 더 낮은 수준의 염증성 사이토카인을 유도할 수 있으며, 즉, 조작된 세포는 사이토카인 폭풍 반응을 일으키지 않는다.

[2029] 일부 실시형태에서, 조작된 T 세포는 동종이계 숙주에 투여되되, 조작된 T 세포는 숙주에 의한 거부 반응이 없다. 일부 실시형태에서, 동종이계 T 세포는 숙주에 의한 무시해도 될 정도의 또는 최소한의 거부를 유도한다. 일부 실시형태에서, 조작된 T 세포는 동족 살해 저항성을 갖는다.

[2030] 일부 실시형태에서, 약제학적 조성물은 대상체에 전달하기 위해 제형화된다. 본 명세서에 기재된 약제학적 조성물을 투여하기에 적합한 경로는 제한 없이 국소, 피하, 경피, 피내, 병변내, 동맥내, 복강내, 방광내, 경점막, 잇몸, 치아내, 달팽이관내, 고실경유, 기관내, 경막외, 척추강내, 근육내, 정맥내, 혈관내, 골수강내, 안구주위, 종양내, 대뇌내 및 뇌실내 투여를 포함한다.

[2031] 일부 실시형태에서, 본 명세서에 기재된 약제학적 조성물은 병든 부위(예를 들어, 종양 부위)에 국부적으로 투여된다. 일부 실시형태에서, 본 명세서에 기재된 약제학적 조성물은 주사, 카테터, 좌약 또는 시알라스틱 막 또는 섬유와 같은 막을 포함하는 다공성, 비-다공성 또는 젤라틴성 물질인 임플란트에 의해 대상체에게 투여된다.

[2032] 다른 실시형태에서, 본 명세서에 기재된 약제학적 조성물은 제어 방출 시스템으로 전달된다. 일 실시형태에서, 펌프가 사용될 수 있다. (예를 들어, 문헌[Langer, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald *et al.*, 1980, Surgery 88:507; Saudek *et al.*, 1989, N. Engl. J. Med. 321:574] 참조). 또 다른 실시형태에서, 중합체 물질이 사용될 수 있다(예를 들어, 문헌[Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61. See also Levy *et al.*, 1985, Science 228: 190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71: 105] 참조). 다른 제어 방출 시스템은, 예를 들어, 상기 Langer에서 논의되어 있다.

[2033] 일부 실시형태에서, 약제학적 조성물은 대상체, 예를 들어, 인간에 대한 정맥내 또는 피하 투여에 적합화된 조성물로서 일상적인 절차에 따라 제형화된다. 일부 실시형태에서, 주사 투여용 약제학적 조성물은 주사 부위의 통증을 완화하기 위한 가용화제 및 리그노카인과 같은 국소 마취제로서 사용하는 멸균 등장성 용액이다. 일반적으로, 성분은 별도로 또는 단위 투여 형태, 예를 들어, 활성제의 양을 나타내는 앰풀 또는 사셰(sachette)와 같은 밀폐된 용기에 건조 동결건조된 분말 또는 물이 없는 농축물로 함께 혼합되어 공급된다. 의약품이 주입에 의해 투여되는 경우, 멸균 의약품 등급의 물 또는 식염수가 담긴 주입 병이 없어도 된다. 약제학적 조성물이 주사에 의해 투여되는 경우, 주사용 멸균수 또는 식염수의 앰풀이 제공되어 투여 전에 성분이 혼합될 수 있도록 제공될 수 있다.

[2034] 전신 투여를 위한 약제학적 조성물은 액체, 예를 들어, 멸균 식염수, 젖산 링거액 또는 행크 용액일 수 있다. 또한, 약제학적 조성물은 고체 형태일 수 있고, 사용 직전에 재용해되거나 또는 현탁될 수 있다. 동결건조된 형태가 또한 상정된다. 약제학적 조성물은 비경구 투여에도 적합한 리포솜 또는 미세결정과 같은 지질 입자 또는 소포 내에 포함될 수 있다. 입자는 조성물이 그 안에 포함되어 있는 한 단층 또는 다층과 같은 임의의 적합한 구조일 수 있다. 화합물은 융합성 지질 다이올레오일포스파티딜에탄올아민(DOPE), 낮은 수준(5몰% 내지 10몰%)의 양이온성 지질을 포함하는 "안정화된 플라스미드-지질 입자"(SPLP)에 의해 포획되고 폴리에틸렌글리콜(PEG) 코팅에 의해 안정화될 수 있다(문헌[Zhang Y. P. *et al.*, Gene Ther. 1999, 6: 1438-47]). N-[1-(2,3-다이올레오일옥시)프로필]-N,N,N-트라이메틸-암모늄메틸설페이트 또는 "DOTAP"와 같은 양으로 하전된 지질이 이러한 입자 및 소포에 특히 바람직하다. 이러한 지질 입자의 제제는 잘 알려져 있다. 예를 들어, 각각 참조에 의해 본 명세서에 원용되어 있는 미국 특허 제4,880,635호; 제4,906,477호; 제4,911,928호; 제4,917,951호; 제4,920,016호; 및 제4,921,757호 참조.

[2035] 예를 들어, 본 명세서에 기재된 약제학적 조성물은 단위 용량으로 투여되거나 또는 패키징될 수 있다. 본 개시내용의 약제학적 조성물과 관련하여 사용될 때 용어 "단위 용량"은 대상체에 대한 단일 투여량으로 적합한 물리적으로 별개의 단위를 지칭하며, 각 단위는 필요한 희석제와 관련하여 목적하는 치료적 효과를 생성하도록 계산된 미리 결정된 양의 활성 물질; 즉, 담체 또는 비히클을 포함한다.

[2036] 또한, 약제학적 조성물은 (a) 동결건조된 형태의 본 발명의 화합물을 포함하는 용기 및 (b) 약제학적으로 허용

가능한 희석제(예를 들어, 본 발명의 동결건조된 화합물의 재구성 또는 희석에 사용되는 멸균제)를 포함하는 제2 용기를 포함하는 약제학적 키트로 제공될 수 있다. 이러한 용기(들)와 선택적으로 관련되는 것은 의약품 또는 생물학적 제품의 제조, 사용 또는 판매를 규제하는 정부 기관에 의해 규정된 형식의 통지일 수 있으며, 이 통지는 인간 투여를 위한 제조, 사용 또는 판매 기관의 승인을 반영한다.

[2037] 또 다른 양태에서, 위에 기재된 질환의 치료에 유용한 물질을 포함하는 제조 물품이 포함된다. 일부 실시형태에서, 제조 물품은 용기 및 표지를 포함한다. 적합한 용기는, 예를 들어, 병, 바이알, 주사기 및 테스트 튜브를 포함한다. 용기는 유리 또는 플라스틱과 같은 다양한 물질로 형성될 수 있다. 일부 실시형태에서, 용기는 본 명세서에 기재된 질환을 치료하는데 효과적인 조성물을 담고 있으며, 멸균 접근 포트를 가질 수 있다. 예를 들어, 용기는 정맥내 용액 백 또는 피하 주사 바늘로 뚫을 수 있는 마개가 있는 바이알일 수 있다. 조성물의 활성제는 본 발명의 화합물이다. 일부 실시형태에서, 용기의 라벨 또는 용기와 관련된 라벨은 해당 조성물이 선택한 질환을 치료하는데 사용됨을 나타낸다. 제조 물품은 인산 완충 식염수, 링거액 또는 덱스트로스 용액과 같은 약제학적으로 허용 가능한 완충액을 포함하는 제2 용기를 추가로 포함할 수 있다. 이는 다른 완충액, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 패키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[2038] 일부 실시형태에서, 본 명세서에 기재된 임의의 융합 단백질, gRNA 및/또는 복합체가 약제학적 조성물의 일부로서 제공된다. 일부 실시형태에서, 약제학적 조성물은 본 명세서에 제공된 임의의 융합 단백질을 포함한다. 일부 실시형태에서, 약제학적 조성물은 본 명세서에 제공된 임의의 복합체를 포함한다. 일부 실시형태에서, 약제학적 조성물은 gRNA 및 양이온성 지질과 복합체를 형성하는 RNA-가이드된 뉴클레이스(예를 들어, Cas9)를 포함하는 리보핵단백질 복합체를 포함한다. 일부 실시형태에서, 약제학적 조성물은 gRNA, 핵산 프로그램 가능 DNA 결합 단백질, 양이온성 지질 및 약제학적으로 허용 가능한 부형제를 포함한다. 약제학적 조성물은 선택적으로 하나 이상의 추가적인 치료적으로 활성인 물질을 포함할 수 있다.

[2039] 일부 실시형태에서, 본 명세서에 제공된 조성물은 대상체 내에서 표적화된 게놈 변형을 수행하기 위하여 대상체, 예를 들어, 인간 대상체에게 투여된다. 일부 실시형태에서, 세포는 대상체로부터 얻어지고, 본 명세서에 제공된 임의의 약제학적 조성물과 접촉된다. 일부 실시형태에서, 대상체로부터 제거되고 생체외에서 약제학적 조성물과 접촉된 세포는 선택적으로 목적하는 게놈 변형이 세포에서 영향을 미치거나 검출된 후에 대상체에게 재도입된다. 뉴클레이스를 포함하는 약제학적 조성물을 전달하는 방법은 공지되어 있으며, 예를 들어, 모든 개시내용이 참조에 의해 본 명세서에 원용되어 있는 미국 특허 제6,453,242호; 제6,503,717호; 제6,534,261호; 제6,599,692호; 제6,607,882호; 제6,689,558호; 제6,824,978호; 제6,933,113호; 제6,979,539호; 제7,013,219호; 및 제7,163,824호에 기술되어 있다. 본 명세서에 제공된 약제학적 조성물의 설명은 주로 인간에게 투여하기에 적합한 약제학적 조성물에 관한 것이지만, 당업자는 이러한 조성물이 일반적으로 모든 종류의 동물 또는 유기체에, 예를 들어, 수의학적 용도를 위해 투여하기에 적합하다는 것을 이해할 것이다.

[2040] 다양한 동물에 대한 투여에 적합한 조성물을 제공하기 위해 인간에 대한 투여에 적합한 약제학적 조성물의 변형은 잘 이해되며, 일반적으로 숙련된 수의 약리학자는 만약에 있다면 단지 일상적인 실험을 통해 이러한 변형을 설계하고/하거나 수행할 수 있다. 약제학적 조성물의 투여가 고려되는 대상체는 인간 및/또는 다른 영장류; 포유동물, 사육 동물, 반려동물, 및 소, 돼지, 말, 양, 고양이, 개, 마우스 및/또는 래트와 같은 상업적으로 관련된 포유동물; 및/또는 닭, 오리, 거위 및/또는 칠면조와 같은 상업적으로 관련된 새를 포함하는 새를 포함하지만, 이들로 제한되지 않는다.

[2041] 본 명세서에 기재된 약제학적 조성물의 제형은 약리학 분야에서 공지되거나 또는 이후 개발될 임의의 방법에 의해 제조될 수 있다. 일반적으로, 이러한 제조 방법은 활성 성분(들)을 부형제 및/또는 하나 이상의 다른 보조 성분과 결합시킨 다음, 필요하거나 그리고/또는 바람직한 경우, 제품을 목적하는 단일-용량 또는 다중-용량 단위로 성형 및/또는 패키징하는 단계를 포함한다. 약제학적 제형은 약제학적으로 허용 가능한 부형제를 추가적으로 포함할 수 있으며, 이는 본 명세서에서 사용되는 바와 같이 목적하는 특정 투여 형태에 적합한 임의의 및 모든 용매, 분산 매질, 희석제 또는 다른 액체 비히클, 분산 또는 현탁 보조제, 표면 활성제, 등장제, 증점제 또는 유화제, 방부제, 고체 결합제, 윤활제 등을 포함한다. 문헌[Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006); 참조에 의해 본 명세서에 원용됨]은 약제학적 조성물을 제형화하는데 사용되는 다양한 부형제 및 이의 제조를 위한 공지된 기법을 개시하고 있다. 또한, 뉴클레이스를 포함하는 약제학적 조성물을 생성하기 위한 추가적인 적합한 방법, 시약, 부형제 및 용매에 대한, 그 전문이 참조에 의해 본 명세서에 원용되어 있는 PCT 출원

PCT/US2010/055131(2010년 11월 2일자로 출원된 공개 번호 WO2011/053982 A8)을 참조한다.

[2042] 임의의 바람직하지 않은 생물학적 효과를 생성하거나 또는 그렇지 않으면 약제학적 조성물의 임의의 다른 성분(들)과 유해한 방식으로 상호작용하는 것과 같이 물질 또는 이의 유도체와 양립할 수 없는 경우를 제외하고, 이의 사용이 본 개시내용의 범주 내에서 상정된다.

[2043] 위에 기재된 바와 같은 조성물은 유효량으로 투여될 수 있다. 유효량은 투여 방식, 치료되는 특정 병태 및 목적하는 결과에 따라 다를 것이다. 이는 또한 병태의 단계, 대상체의 연령 및 신체 상태, 동시 요법의 특성(있는 경우) 및 의사에게 잘 알려진 유사한 인자에 따라 달라질 수 있다. 치료적 적용의 경우, 이는 의학적으로 바람직한 결과를 달성하기에 충분한 양이다.

[2044] 일부 실시형태에서, 본 개시내용에 따른 조성물은 임의의 다양한 질환, 장애 및/또는 병태의 치료에 사용될 수 있다.

[2045] ***치료 방법***

[2046] 본 발명의 일부 양태는 치료를 필요로 하는 대상체를 치료하는 방법을 제공하며, 방법은 치료적 유효량의 본 명세서에 기재된 바와 같은 약제학적 조성물을 이를 필요로 하는 대상체에게 투여하는 단계를 포함한다. 보다 구체적으로, 치료 방법은 키메라 수용체(CAR)를 발현하고 적어도 하나의 편집된 유전자를 갖는 변형된 면역 세포의 집단을 포함하는 하나 이상의 약제학적 조성물을 이를 필요로 하는 대상체에게 투여하는 단계를 포함하되, 적어도 하나의 편집된 유전자는 동족 살해 저항성을 제공하거나, 기능을 향상시키거나 또는 변형된 면역 세포의 면역억제 또는 저해를 감소시키고, 적어도 하나의 편집된 유전자의 발현은 넉아웃되거나 또는 넉다운된다. 일부 실시형태에서, 치료 방법은 자가 면역 세포 요법이다. 다른 실시형태에서, 치료 방법은 동종이계 면역 세포 요법이다.

[2047] 소정의 실시형태에서, 면역 세포의 특이성은 본 명세서에서 상정된 키메라 항원 수용체를 발현하도록 면역 세포를 유전적으로 변형시킴으로써 대상체에서 병든 또는 변경된 세포의 표면 상에 발현되는 마커로 재지시된다. 일부 실시형태에서, 치료 방법은 본 명세서에 기재된 바와 같은 면역 세포를 대상체에게 투여하는 단계를 포함하되, 면역 세포는 신생물 세포에서 발현된 마커로 특이성을 재 지시하도록 유전적으로 변형되었다. 일부 실시형태에서, 신생물은 B 세포암; 예를 들어, B 세포암, 예컨대, 림프종, 백혈병 또는 골수종, 예를 들어, 다발성 골수종이다. 일부 실시형태에서, 신생물은 고형 종양이다. 다른 실시형태에서, 신생물은 액체 종양이다. 일부 실시형태에서, 신생물은 혈액암이다. 일부 실시형태에서, 혈액암은 백혈병, 골수종 및/또는 림프종이다. 일부 경우에, 백혈병은 전백혈병을 포함한다. 일부 경우에, 백혈병은 급성 백혈병이다. 급성 백혈병은, 예를 들어, 급성 골수성 백혈병(AML)을 포함한다. 급성 백혈병 또한, 예를 들어, 급성 림프성 백혈병 또는 급성 림프구성 백혈병(ALL)을 포함하고; ALL은 B-계통 ALL; T-계통 ALL; 및 T-세포 급성 림프구성 백혈병(T-ALL)을 포함한다.

[2048] 신생물의 비제한적인 예는 T-세포 급성 림프아구성 백혈병(T-ALL), 용상 진균증(MF), 세자리 증후군(SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 $ALK^+$, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 $CD30^+$ 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종을 포함한다. 일부 실시형태에서, 신생물은 T-세포 급성 림프아구성 백혈병(T-ALL)이다. 일부 실시형태에서, 신생물은 급성 골수성 백혈병(AML)이다. 따라서, 본 개시내용의 일부 실시형태는 대상체에서 신생물을 치료하는 방법을 제공한다.

[2049] 일부 실시형태에서, 치료 방법은 기능적 T 세포 수용체 알파 불변(TRAC), 분화 클러스터 33(CD33), 분화 클러스터 3(CD3), 분화 클러스터 5(CD5), 분화 클러스터 7(CD7), 분화 클러스터 52(CD52), 분화 클러스터 123(CD123), 예정 세포사 1(PD-1), 베타-2 마이크로글로불린(B2M), Fas 세포 표면 사멸 수용체(FAS), 림프구-활성화 유전자 3(LAG-3), 클래스 II, 주요 조직적합성 복합체, 전사촉진인자(CIITA), T 세포 수용체 베타 불변 1(TRBC1), T 세포 수용체 베타 불변 2(TRBC2) 또는 이들의 조합이 결여되어 있거나 또는 이들의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)를 신생물(예를 들어, 백혈병)이 발생하거나 발생하는 경향이 있는 대상체에게 투여하는 단계를 포함한다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD123이 결여되어 있거나 또는 이의 수준이 감소되고 CD123 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하

는 경향이 있는 대상체에는 CD5가 결여되어 있거나 또는 이의 수준이 감소되고 CD5 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD7이 결여되어 있거나 또는 이의 수준이 감소되고 CD7 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD33이 결여되어 있거나 또는 이의 수준이 감소되고 CD33 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD3이 결여되어 있거나 또는 이의 수준이 감소되고 CD3 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다.

[2050] 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 기능적 TRAC가 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 기능적 B2M의 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 기능적 FAS가 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 기능적 LAG-3이 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 기능적 TRBC1 및/또는 TRBC2가 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 기능적 CIITA가 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD52가 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 PD-1이 결여되어 있거나 또는 이의 수준이 감소된 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다.

[2051] 일부 실시형태에서, 신생물(예를 들어, 백혈병)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD123이 결여되어 있거나 또는 이의 수준이 감소되고, TRAC, LAG-3, FAS, CD52, B2M, CIITA, TRBC1, TRBC2 및/또는 PD-1 또는 이들의 조합이 결여되어 있거나 또는 이들의 수준이 감소되고, CD123 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 또는 이의 집단이 투여된다. 일부 실시형태에서, 신생물(예를 들어, 백혈병)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD5가 결여되어 있거나 또는 이의 수준이 감소되고, TRAC, LAG-3, FAS, CD52, B2M, CIITA, TRBC1, TRBC2 및/또는 PD-1 또는 이들의 조합이 결여되어 있거나 또는 이들의 수준이 감소되고, CD5 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, 백혈병)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD7이 결여되어 있거나 또는 이의 수준이 감소되고, TRAC, LAG-3, FAS, CD52, B2M, CIITA, TRBC1, TRBC2 및/또는 PD-1 또는 이들의 조합이 결여되어 있거나 또는 이들의 수준이 감소되고, CD7 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, 백혈병)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD33이 결여되어 있거나 또는 이의 수준이 감소되고, TRAC, LAG-3, FAS, CD52, B2M, CIITA, TRBC1, TRBC2 및/또는 PD-1 또는 이들의 조합이 결여되어 있거나 또는 이들의 수준이 감소되고, CD33 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다. 일부 실시형태에서, 신생물(예를 들어, 백혈병)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD3이 결여되어 있거나 또는 이의 수준이 감소되고, TRAC, LAG-3, FAS, CD52, B2M, CIITA, TRBC1, TRBC2 및/또는 PD-1 또는 이들의 조합이 결여되어 있거나 또는 이들의 수준이 감소되고, CD3 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 투여된다.

[2052] 일부 실시형태에서, 대상체에서 신생물을 치료하는 방법은 본 명세서에 기재된 바와 같은 면역 세포 및 하나 이상의 추가적인 치료제를 대상체에게 투여하는 단계를 포함한다. 예를 들어, 본 발명의 면역 세포는 하나 이상의

추가적인 변형된 면역 세포 또는 이의 집단과 병용-투여될 수 있다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD123이 결여되어 있거나 또는 이의 수준이 감소되고 CD123 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 및 CD33이 결여되어 있거나 또는 이의 수준이 감소되고 CD33 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 병용 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD5가 결여되어 있거나 또는 이의 수준이 감소되고 CD5 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 및 CD7이 결여되어 있거나 또는 이의 수준이 감소되고 CD7 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 병용 투여된다. 일부 실시형태에서, 임의의 면역 세포는 CD3, CD5, CD7, CD33, CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및/또는 PD1 또는 이들의 조합 중 하나 이상의 돌연변이를 추가로 포함한다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD3이 결여되어 있거나 또는 이의 수준이 감소되고 CD3 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 및 CD7이 결여되어 있거나 또는 이의 수준이 감소되고 CD7 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 병용 투여된다. 일부 실시형태에서, 신생물(예를 들어, T-세포 또는 NK-세포 악성종양)이 발생하거나 또는 발생하는 경향이 있는 대상체에는 CD3이 결여되어 있거나 또는 이의 수준이 감소되고 CD3 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포) 및 CD123이 결여되어 있거나 또는 이의 수준이 감소되고 CD123 키메라 항원 수용체를 발현하는 유효량의 변형된 면역 효과기 세포(예를 들어, CAR-T 세포)가 병용 투여된다.

[2053] 일부 실시형태에서, 본 발명의 면역 세포는 사이토카인과 병용 투여될 수 있다. 일부 실시형태에서, 사이토카인은 IL-2, IFN-α, IFN-γ 또는 이들의 조합이다. 일부 실시형태에서, 면역 세포는 화학치료제와 병용 투여된다. 화학치료제는 사이클로포스파마이드(cyclophosphamide), 독소루비신(doxorubicin), 빈크리스틴(vincristine), 프레드니손(prednisone) 또는 리툭시맙(rituximab) 또는 이들의 조합일 수 있다. 다른 화학치료제는 오비누투주맙(obinutuzumab), 벤다무스틴(bendamustine), 클로람부실(chlorambucil), 사이클로포스파마이드, 이브루티닙(ibrutinib), 메토트렉세이트(methotrexate), 시타라빈(cytarabine), 덱사메타손(dexamethasone), 시스플라틴(cisplatin), 보르테조밉(bortezomib), 플루다라빈, 이델라리십(idelalisib), 아칼라브루티닙(acalabrutinib), 레날리도마이드(lenalidomide), 베네토클락스(venetoclax), 사이클로포스파마이드, 아이포스파마이드(ifosfamide), 에토포시드(etoposide), 펜토스타틴(pentostatin), 멜팔란(melphalan), 카르필조밉(carfilzomib), 익사조밉(ixazomib), 파노비노스타트(panobinostat), 다라투무맙(daratumumab), 엘로투주맙(elotuzumab), 탈리도마이드(thalidomide), 레날리도마이드 또는 포말리도마이드(pomalidomide) 또는 이들의 조합을 포함한다. "병용 투여된"은 치료 과정 동안 2종 이상의 치료제 또는 약제학적 조성물을 투여하는 것을 지칭한다. 이러한 병용-투여는 동시 투여 또는 순차적 투여일 수 있다. 나중에 투여되는 치료제 또는 약제학적 조성물의 순차적 투여는 제1 약제학적 조성물 또는 치료제의 투여 후 치료 과정 동안 임의의 시간에 발생할 수 있다.

[2054] 본 발명의 일부 실시형태에서, 투여된 면역 세포는 생체내에서 증식하고, 장기간 동안 대상체에서 지속될 수 있다. 일부 실시형태에서, 본 발명의 면역 세포는 기억 면역 세포로 성숙할 수 있고, 대상체 내에서 순환 상태로 남아 있어 키메라 항원 수용체에 의해 인식되는 마커를 발현하는 병든 또는 변경된 세포의 재발에 능동적으로 반응할 수 있는 세포의 집단을 생성할 수 있다.

[2055] 면역 세포 핵염기 편집기 폴리펩타이드를 발현하거나 또는 이를 도입하고, 적어도 하나의 폴리펩타이드를 암호화하는 핵산 분자를 표적화할 수 있는 2개 이상의 가이드 RNA와 접촉시키는 방법에 의해 생성된 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻어진 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되되, 제1 폴리펩타이드는 Cbl 프로토-종양 유전자 B(CBLB)이고, 제2 폴리펩타이드는 T 세포 수용체 알파 불변(TRAC), 베타-2 마이크로글로불린(B2M) 및 예정 세포사 단백질 1(PD1) 폴리펩타이드로 이루어진 군으로부터 선택되며, 핵염기 편집기 폴리펩타이드는 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 및 사이티딘 데아미네이스를 포함한다. 이러한 방법의 한 예에서, 3 개의 가이드 RNA가 각각 B2M, TRAC, PDCD1 또는 CBLB 폴리뉴클레오타이드를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 이러한 방법의 또 다른 예에서, 4개의 가이드 RNA가 각각 B2M, TRAC, PDCD1 및 CBLB 폴리뉴클레오타이드 중 하나를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 한 예에서, 2개의 가이드 RNA가 각각 B2M 또는 TRAC 폴리뉴클레오타이드를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 3개의 가이드 RNA가 각각 B2M, TRAC, PDCD1 또는 CBLB 폴리뉴클레오타이드를 표

적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 4개의 가이드 RNA가 B2M, TRAC, PDCD1 및 CBLB 폴리뉴클레오타이드 중 하나를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 3개의 가이드 RNA가 각각 B2M, TRAC 및 PDCD1 폴리뉴클레오타이드를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 가이드 RNA는 TRAC 엑손 4 스플라이스 수용체 부위, B2M 엑손 1 스플라이스 공여체 부위 또는 PDCD1 엑손 1 스플라이스 공여체 부위를 표적으로 한다. 또 다른 실시형태에서, 변형된 면역 세포를 생산하는 방법은 세포독성 T-림프구 관련 단백질 4(CTLA4) 폴리뉴클레오타이드를 표적으로 하는 가이드 RNA를 세포에서 발현시키는 단계를 더 포함한다.

[2056] 면역 세포 핵염기 편집기 폴리펩타이드를 발현하거나 또는 이를 도입하고, 세포를 Cbl 프로토-종양 유전자 B(CBLB), T 세포 수용체 알파 불변(TRAC), 베타-2 마이크로글로불린(B2M) 및 예정 세포사 단백질 1(PD1) 폴리펩타이드로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드를 암호화하는 핵산 분자를 표적화할 수 있는 2개 이상의 가이드 RNA와 접촉시키는 방법에 의해 생성된 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되되, 핵염기 편집기 폴리펩타이드는 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 및 아데노신 데아미네이스를 포함한다. 한 예에서, 2개의 가이드 RNA가 각각 B2M 또는 TRAC 폴리뉴클레오타이드를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 3개의 가이드 RNA가 각각 B2M, TRAC, PDCD1 또는 CBLB 폴리뉴클레오타이드를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 4개의 가이드 RNA가 각각 B2M, TRAC, PDCD1 및 CBLB 폴리뉴클레오타이드 중 하나를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 3개의 가이드 RNA가 각각 B2M, TRAC 및 PDCD1 폴리뉴클레오타이드를 표적으로 하는 세포에서 발현되거나 또는 이와 접촉한다. 또 다른 예에서, 가이드 RNA는 TRAC 엑손 4 스플라이스 수용체 부위, B2M 엑손 1 스플라이스 공여체 부위 또는 PDCD1 엑손 1 스플라이스 공여체 부위를 표적으로 한다. 또 다른 실시형태에서, 변형된 면역 세포를 생산하는 방법은 세포독성 T-림프구 관련 단백질 4(CTLA4) 폴리뉴클레오타이드를 표적으로 하는 가이드 RNA를 세포에서 발현시키는 단계를 더 포함한다.

[2057] 면역 세포 핵염기 편집기 폴리펩타이드를 발현하거나 또는 이를 도입하고, 세포를 V-Set 면역조절 수용체(VISTA), T 세포 면역글로불린 뮤신 3(Tim-3), Ig 및 ITIM 도메인을 갖는 T 세포 면역수용체(TIGIT), 형질전환 성장 인자 베타 수용체 II(TGFbRII), 조절 인자 X 관련 안키린 함유 단백질(Regulatory Factor X Associated Ankyrin Containing Protein: RFXANK), PVR 관련 면역글로불린 도메인 함유(PVRIG), 림프구-활성화 유전자 3(Lag3), 세포독성 T-림프구 관련 단백질 4(CTLA-4), 키티네이스 3 유사 1(Chitinase 3 Like 1: Chi3l1), 분화 클러스터 96(CD96), B 및 T 림프구 관련(BTLA), Tet 메틸사이토신 다이옥시게네이스 2(TET2), 스프로우티 RTK 신호전달 길항제 1(Sprouty RTK Signaling Antagonist 1: Spry1), 스프로우티 RTK 신호전달 길항제 2(Spry2), 클래스 II 주요 조직적합성 복합체 전사촉진인자(CIITA), 분화 클러스터 3(CD3), 분화 클러스터 7(CD7), 분화 클러스터 33(CD33), 분화 클러스터 52(CD52), 분화 클러스터 123(CD123), T 세포 수용체 베타 불변 1(TRBC1), T 세포 수용체 베타 불변 2(TRBC2), 사이토카인 유도성 SH2 함유 단백질(CISH), 아세틸-CoA 아세틸트랜스퍼레이스 1(ACAT1), 사이토크롬 P450 패밀리 11 서브패밀리 구성원 1(Cyp11a1), GATA 결합 단백질 3(GATA3), 핵 수용체 서브패밀리 4 그룹 구성원 1(NR4A1), 핵 수용체 서브패밀리 4 그룹 구성원 2(NR4A2), 핵 수용체 서브패밀리 4 그룹 구성원 3(NR4A3), 메틸화-제어 J 단백질(MCJ), Fas 세포 표면 사멸 수용체(FAS) 및 셀렉틴 P 리간드/P-셀렉틴 당단백질 리간드-1(SELPG/PSGL1) 폴리펩타이드로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드를 암호화하는 핵산 분자를 표적화할 수 있는 2개 이상의 가이드 RNA와 접촉시키는 방법에 의해 생성된 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되되, 핵염기 편집기 폴리펩타이드는 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 및 아데노신 또는 사이티딘 데아미네이스를 포함한다. 일 실시형태에서, 세포를 생산하는 방법은 세포를 T 세포 수용체 알파 불변(TRAC), 베타-2 마이크로글로불린(B2M) 및 예정 세포사 단백질 1(PD1) 폴리펩타이드로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드를 암호화하는 핵산 분자를 표적화할 수 있는 2개 이상의 가이드 RNA와 접촉시키는 단계를 더 포함한다. 또 다른 실시형태에서, 변형된 면역 세포를 생산하는 방법은 세포독성 T-림프구 관련 단백질 4(CTLA4) 폴리뉴클레오타이드를 표적으로 하는 가이드 RNA를 세포에서 발현시키는 단계를 더 포함한다.

[2058] CBLB, PD1, CTLA4, TGFBR2, TIGIT 또는 부조직적합성(minor histocompatibility) 폴리펩타이드로 이루어진 군으로부터 선택되는 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 표적으로 하는 하나 이상의 가이드 RNA를 신생물을 갖는 대상체로부터 단리된 면역 세포에서 발현시키는 단계를 포함하는 방법에 의해 생성된 항-신생물 활성을 갖는 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T

세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포) 및 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 및 사이티딘 데아미네이스 또는 아데노신 데아미네이스를 포함하는 핵염기 편집기 폴리펩타이드로 본 명세서에 기재된 바와 같이 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공된다. 한 예에서, 가이드 RNA는 각각 CBLB 및 PDCD1 폴리뉴클레오타이드를 표적으로 한다. 한 예에서, 가이드 RNA는 각각 PDCD1 및 CTLA4 폴리뉴클레오타이드를 표적으로 한다. 한 예에서, 가이드 RNA는 각각 CBLB, PDCD1 및 CTLA4 폴리뉴클레오타이드를 표적으로 한다. 한 예에서, 가이드 RNA는 각각 CBLB, PD1 및 CTLA4 폴리뉴클레오타이드를 표적으로 한다.

[2059] 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, 가이드 RNA는 표적 폴리뉴클레오타이드에서 스플라이스 수용체 부위 또는 스플라이스 공여체 부위를 표적으로 한다. 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, 핵염기 편집기 폴리펩타이드는 표적 폴리뉴클레오타이드에서 종결 코돈을 생성한다. 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, 핵염기 편집기 폴리펩타이드는 PDCD1 엑손 2에서 종결 코돈을 생성한다. 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, 핵염기 편집기 폴리펩타이드는, 예를 들어, 이분 NLS, N-말단 NLS, C-말단 NLS 또는 하나 이상의 우라실 글리코실레이스 저해제와 같은 하나 이상의 핵 국부화 신호(NLS)를 발현하는 것을 추가로 포함한다.

[2060] 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, napDNAbp는 Cas9, CasX, CasY, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12j/Casϕ 및 BvCas12b 또는 이들의 활성 단편으로 이루어진 군으로부터 선택된다. 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, napDNAbp 도메인은 뉴클레이스 활성이 없는 Cas9(dCas9), Cas9 닉케이스(nCas9) 또는 뉴클레이스 활성이 있는 Cas9를 포함한다. 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, 사이티딘 데아미네이스는 페트로마이존 마리누스 사이토신 데아미네이스 1(pmCDA1) 또는 활성화-유도성 사이티딘 데아미네이스(AICDA)이고, 아데노신 데아미네이스는 TadA이다.

[2061] 위에 기재된 바와 같이 변형된 면역 세포를 생산하는 임의의 이러한 방법의 한 예에서, 방법은 TIGIT 폴리뉴클레오타이드, TGFBR2 폴리뉴클레오타이드, ZAP70 폴리뉴클레오타이드, NFATc1 폴리뉴클레오타이드 및 TET2 폴리뉴클레오타이드를 표적으로 하는 하나 이상의 가이드 RNA를 면역 세포에서 발현시키는 단계를 더 포함한다.

[2062] 일부 예에서, 대상체에 투여될 변형된 면역 세포는 검출 가능한 전위를 포함하지 않는다는 것이 이해될 것이다.

[2063] 감소된 면역원성 및 증가된 항-신생물 활성을 갖는 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되며, 변형된 면역 세포는 TRAC, B2M, CBLB 또는 PD1 폴리뉴클레오타이드 또는 이들의 조합에서의 돌연변이를 포함한다. 한 예에서, 변형된 면역 세포는 B2M 및 TRAC 폴리뉴클레오타이드에서의 돌연변이를 포함한다. 또 다른 예에서, 변형된 면역 세포는 B2M, TRAC 및 PDCD1 폴리뉴클레오타이드에서의 돌연변이를 포함한다. 또 다른 예에서, 변형된 면역 세포는 B2M, TRAC 및 CBLB 폴리뉴클레오타이드에서의 돌연변이를 포함한다. 또 다른 예에서, 변형된 면역 세포는 B2M, TRAC, CBLB 및 PD1 폴리뉴클레오타이드에서의 돌연변이를 포함한다. 또 다른 예에서, 변형된 면역 세포는 TIGIT, TGFBR2, ZAP70, NFATc1 또는 TET2를 암호화하는 하나 이상의 폴리뉴클레오타이드에서의 돌연변이를 포함한다. 또 다른 예에서, 변형된 면역 세포는 V-Set 면역조절 수용체(VISTA), T 세포 면역글로불린 뮤신 3(Tim-3), Ig 및 ITIM 도메인을 갖는 T 세포 면역수용체 (TIGIT), 형질전환 성장 인자 베타 수용체 II(TGFbRII), 조절 인자 X 관련 안키린 함유 단백질(RFXANK), PVR 관련 면역글로불린 도메인 함유(PVRIG), 림프구-활성화 유전자 3(Lag3), 세포독성 T-림프구 관련 단백질 4(CTLA-4), 키티네이스 3 유사 1(Chi3l1), 분화 클러스터 96(CD96), B 및 T 림프구 관련(BTLA), Tet 메틸사이토신 다이옥시게네이스 2(TET2), 스프로우티 RTK 신호전달 길항제 1(Spry1), 스프로우티 RTK 신호전달 길항제 2(Spry2), 클래스 II 주요 조직적합성 복합체 전사촉진인자(CIITA), 분화 클러스터 3(CD3), 분화 클러스터 7(CD7), 분화 클러스터 33(CD33), 분화 클러스터 52(CD52), 분화 클러스터 123(CD123), T 세포 수용체 베타 불변 1(TRBC1), T 세포 수용체 베타 불변 2(TRBC2), 사이토카인 유도성 SH2 함유 단백질(CISH), 아세틸-CoA 아세틸트랜스퍼레이스 1(ACAT1), 사이토크롬 P450 패밀리 11 서브패밀리 구성원 1(Cyp11a1), GATA 결합 단백질 3(GATA3), 핵 수용체 서브패밀리 4 그룹 구성원 1(NR4A1), 핵 수용체 서브패밀리 4 그룹 구성원 2(NR4A2), 핵 수용체 서브패밀리 4 그룹 구성원 3(NR4A3), 메틸화-제어 J 단백질(MCJ), Fas 세포 표면 사멸 수용체(FAS) 또는 셀렉틴 P 리간드/P-셀렉틴 당단백질 리간드-1(SELPG/PSGL1)을 암호화하는 하나 이상의 폴리뉴클레오타이드에서의 돌연변이를 포함한다. 또 다른 예에서, 변형된 면역 세포는, 예를 들어, 신생물과 관련된 마커(예를 들어, B 세포 성숙 항원(BCMA))에 대한 친화성을 갖는 세포외 도메인을 포함하는 키메라 항원 수용체를 발현한다.

[2064] 감소된 면역원성 및 증가된 항-신생물 활성을 갖는 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되며, 변형된 면역 세포는 CD33, PD-1, CD52, TRAC 폴리뉴클레오타이드 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD33, PD-1, CD52, TRAC 또는 이들의 조합 중 2개 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD33, PD-1, CD52, TRAC 또는 이들의 조합 중 3개 이상에서의 돌연변이를 포함한다. 한 예에서, 변형된 면역 세포는 CD33, PD-1, CD52 및 TRAC에서의 돌연변이를 포함한다.

[2065] 감소된 면역원성 및 증가된 항-신생물 활성을 갖는 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되며, 변형된 면역 세포는 CD3, PD-1, CD52, TRAC, CD7 폴리뉴클레오타이드 또는 이들의 조합 중 하나 이상에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD3, PD-1, CD52, TRAC, CD7 또는 이들의 조합 중 2개 이상에서의 돌연변이를 포함한다. 한 예에서, 변형된 면역 세포는 CD3 및 CD7에서의 돌연변이를 포함한다. 다양한 실시형태에서, 변형된 면역 세포는 CD3, PD-1, CD52, TRAC, CD7 또는 이들의 조합 중 3개 이상에서의 돌연변이를 포함한다. 한 예에서, 변형된 면역 세포는 CD3, PD-1, CD52, TRAC 및 CD7에서의 돌연변이를 포함한다.

[2066] 핵염기 편집기 폴리펩타이드; 베타-2 마이크로글로불린(B2M) 및 예정 세포사 단백질 1(PD1) 폴리펩타이드; Cas12 폴리펩타이드; TRAC를 암호화하는 핵산 분자를 표적화할 수 있는 가이드 RNA; 및 키메라 항원 수용체(CAR) 또는 T 세포 수용체를 암호화하는 DNA 공여체 주형으로 이루어진 군으로부터 선택되는 적어도 하나의 폴리펩타이드를 암호화하는 핵산 분자를 표적화할 수 있는 2개 이상의 가이드 RNA를 면역 세포에서 발현시키거나 또는 도입하는 방법에 의해 생성된 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되되, 핵염기 편집기 폴리펩타이드는 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 및 사이티딘 데아미네이스를 포함한다.

[2067] 감소된 면역원성 및 증가된 항-신생물 활성을 갖는 변형된 면역 세포(예를 들어, 건강한 대상체로부터 얻은 면역 세포 또는, 예를 들어, 세포독성 T 세포, 조절 T 세포 또는 T 헬퍼 세포와 같은 T 세포)로 신생물 또는 액체암을 치료하는 방법이 본 명세서에 제공되며, 변형된 면역 세포는 B2M 또는 PD1 폴리뉴클레오타이드 중 하나 이상에서의 돌연변이, TRAC 유전자좌에서 키메라 항원 수용체(CAR)를 암호화하는 핵산의 삽입 및 면역 세포의 세포 표면 상에서 CAR을 발현하는 것을 포함한다.

[2068] 본 발명의 일부 실시형태에서, 치료적 유효량의 본 명세서에 기재된 바와 같은 약제학적 조성물을 액체암의 치료를 필요로 하는 대상체에게 투여하는 단계를 포함하는, 액체암의 치료를 필요로 하는 대상체에서 액체암을 치료하는 방법이 본 명세서에 제공된다. 이러한 방법에 의해 치료되는 액체암은 림프종, 백혈병, 골수종 또는 이들의 조합을 포함하지만, 이들로 제한되지 않는다.

[2069] 한 예에서, 치료적 유효량의 본 명세서에 기재된 바와 같은 약제학적 조성물을 액체암의 치료를 필요로 하는 대상체에게 투여하는 단계를 포함하는, 액체암의 치료를 필요로 하는 대상체에서 액체암을 치료하는 방법이 본 명세서에 제공되되, 액체암은 림프종을 포함한다. 본 명세서에 기재된 방법을 사용하여 치료되는 액체 림프종은 악성 형질 세포 신생물, 호지킨 림프종; 결절성 림프구 우세 호지킨 림프종; 진성 적혈구증가증, 호지킨 림프종; 소림프구성 림프종(small lymphocytic lymphoma: SLL); 미만성 거대 B-세포 림프종; 면역아세포성 대세포 림프종; 전구체 B-림프아구성 림프종; 맨틀 세포 림프종; 변연부 림프종; 용상 진균증; 역형성 대세포 림프종; 세자리 증후군; 전구체 T-림프아구성 림프종; 및 미만성 거대 B-세포 림프종과 고전적 호지킨 림프종 사이의 중간 특징을 갖는 분류되지 않은 B-세포 림프종; B-세포 비-호지킨 림프종(NHL); 여포성 림프종; 버킷 림프종; 버킷-유사 림프종; 변연부 림프종; 맨틀 세포 림프종; 미만성 거대 B-세포 림프종(diffuse large B-cell lymphoma: DLBCL), 원발성 종격동 B-세포 림프종; 대세포 림프종; 전구체 B-림프아구성 림프종; 미만성 거대 B-세포 림프종과 버킷 림프종 사이의 중간 특징을 갖는 분류되지 않은 B-세포 림프종; 변연부 림프종; 점막-관련 림프 조직 림프종; 소세포 림프구성 림프종; 맨틀 세포 림프종; 원발성 종격동(흉선) 거대 B-세포 림프종; 림프형질세포성 림프종(예컨대, 발덴스트롬 마크로글로불린혈증); 결절 변연부 B 세포 림프종; 비장 변연부 림프종; 혈관내 거대 B-세포 림프종; 원발성 삼출성 림프종; 림프종양 육아종증; T 세포/조직구-풍부 거대 B-세포 림프종; 원발성 중추신경계 림프종; 원발성 피부 미만성 거대 B-세포 림프종(다리 유형); 노인의 EBV 양성 미만성 거대 B-세포 림프종; 염증과 관련된 미만성 거대 B-세포 림프종; 혈관내 거대 B-세포 림프종; ALK-양성 거대 B-세포 림프종; 형질아세포성 림프종; HHV8-관련 다심성 캐슬만병에서 발생하는 거대 B-세포 림프종; 종격동 거대

B 세포 림프종; 혈관내 거대 B 세포 림프종; 원발성 삼출성 림프종; 비강형, 장병증형 T 세포 림프종; 간비장 T 세포 림프종; 아세포 NK 세포 림프종; 원발성 피부 역형성 대세포 림프종; 림프종성 구진증; 혈관면역아세포성 T 세포 림프종; 말초 T 세포 림프종; 상세불명의 역형성 대세포 림프종; 고전적 호지킨 림프종(결절성 경화증, 혼합 세포성, 림프구-풍부, 림프구 고갈 또는 고갈되지 않음); 결절성 림프구-우세 호지킨 림프종; 성인 T 세포 림프종; B-세포 림프종-DLBCL, B-세포 림프종-DLBCL-배아 중심-유사; B-세포 림프종-DLBCL-활성화된 B-세포-유사; 비-호지킨 림프종; 원발성 중추신경계(CNS) 림프종; T-세포 림프종; 말초 T 세포 림프종, 피부 T-세포 림프종(cutaneous T-cell lymphoma: CTCL); 역형성 대세포 림프종(anaplastic large cell lymphoma: ALCL); AIDS-관련 림프종; 장질환-유형 T 세포 림프종; 간비장 T 세포 림프종; 아세포 NK 세포 림프종; 원발성 피부 역형성 대세포 림프종; 림프종성 구진증, 혈관면역아세포성 T 세포 림프종; 말초 T 세포 림프종; 상세불명의 역형성 대세포 림프종; 고전적 호지킨 림프종(결절성 경화증, 혼합 세포성, 림프구-풍부, 림프구 고갈 또는 고갈되지 않음); 및 결절성 림프구-우세 호지킨 림프종; 다른 림프종; 및 이러한 림프종의 조합을 포함하지만, 이들로 제한되지 않는다.

[2070] 한 예에서, 치료적 유효량의 본 명세서에 기재된 바와 같은 약제학적 조성물을 액체암의 치료를 필요로 하는 대상체에게 투여하는 단계를 포함하는, 액체암의 치료를 필요로 하는 대상체에서 액체암을 치료하는 방법이 본 명세서에 제공되되, 액체암은 백혈병을 포함한다. 일부 경우에, 백혈병은 전백혈병을 포함한다. 전백혈병은, 예를 들어, 골수이형성 증후군(myelodysplastic syndrome: MDS)을 포함한다. 일부 경우에, 백혈병은 급성 백혈병이다. 급성 백혈병은, 예를 들어, 급성 골수성 백혈병(AML)을 포함한다. 급성 백혈병 또한, 예를 들어, 급성 림프성 백혈병 또는 급성 림프구성 백혈병(ALL)을 포함하고; ALL은 B-계통 ALL; T-계통 ALL; 및 T-세포 급성 림프구성 백혈병(T-ALL)을 포함한다. 본 명세서에 기재된 방법으로 치료되는 다른 백혈병은 털세포 백혈병(hairy cell leukemia: HCL); 만성 골수성 백혈병(chronic myeloid leukemia: CIVIL); 만성 골수단핵구 백혈병(chronic myelomonocytic leukemia: CMML); 만성 림프구성 백혈병(chronic lymphocytic leukemia: CLL); B-세포 전림프구성 백혈병, 골수성 백혈병; 버킷 백혈병; T 세포 전림프구성 백혈병; T 세포 대형 과립 림프구성 백혈병; 공격성 천연 살해(NK) 세포 백혈병; 성인 T 세포 백혈병(adult T cell leukemia: ATL); 급성 전골수성 백혈병(acute promyeloid leukemia: APML); 전림프구성 백혈병(prolymphocytic leukemia: PLL); 적아세포성 백혈병; 급성 거핵아세포성 백혈병; 대형 과립 림프구성 백혈병(large granular lymphocytic leukemia: LGF); 만성 골수성 백혈병(CML); 다른 백혈병; 및 이러한 백혈병의 조합을 포함하지만, 이들로 제한되지 않는다.

[2071] 한 예에서, 치료적 유효량의 본 명세서에 기재된 바와 같은 약제학적 조성물을 액체암의 치료를 필요로 하는 대상체에게 투여하는 단계를 포함하는, 액체암의 치료를 필요로 하는 대상체에서 액체암을 치료하는 방법이 본 명세서에 제공되되, 액체암은 골수종을 포함한다. 본 명세서에 기재된 방법을 사용하여 치료되는 골수종은 다발성 골수종; 무증상 다발성 골수종; 형질 세포 골수종; 비-분비 골수종, IgD 골수종; 골경화 골수종; 기타 골수종; 및 이러한 골수종의 조합을 포함하지만, 이들로 제한되지 않는다.

[2072] 한 예에서, 치료적 유효량의 본 명세서에 기재된 바와 같은 약제학적 조성물을 액체암의 치료를 필요로 하는 대상체에게 투여하는 단계를 포함하는, 액체암의 치료를 필요로 하는 대상체에서 액체암을 치료하는 방법이 본 명세서에 제공되되, 액체암은 칼러병; 골수종증; 형질세포종; 뼈의 고립성 형질세포종; 골수외 형질세포종용상 진균증/세자리 증후군; 원발성 피부 CD30-양성 T 세포 림프증식성 장애; 고 호산구증가증; 만성 호산구증가증; 및 다른 조혈 세포 관련 암을 포함한다.

[2073] 일 실시형태에서, 대상체에는 적어도 $0.1\times10^5$개의 세포, 적어도 $0.5\times10^5$개의 세포, 적어도 $1\times10^5$개의 세포, 적어도 $5\times10^5$개의 세포, 적어도 $1\times10^6$개의 세포, 적어도 $0.5\times10^7$개의 세포, 적어도 $1\times10^7$개의 세포, 적어도 $0.5\times10^8$개의 세포, 적어도 $1\times10^8$개의 세포, 적어도 $0.5\times10^9$개의 세포, 적어도 $1\times10^9$개의 세포, 적어도 $2\times10^9$개의 세포, 적어도 $3\times10^9$개의 세포, 적어도 $4\times10^9$개의 세포, 적어도 $5\times10^9$개의 세포 또는 적어도 $1\times10^{10}$개의 세포가 투여된다. 특정 실시형태에서, 약 $1\times10^7$개의 세포 내지 약 $1\times10^9$개의 세포, 약 $2\times10^7$개의 세포 내지 약 $0.9\times10^9$개의 세포, 약 $3\times10^7$개의 세포 내지 약 $0.8\times10^9$개의 세포, 약 $4\times10^7$개의 세포 내지 약 $0.7\times10^9$개의 세포, 약 $5\times10^7$개의 세포 내지 약 $0.6\times10^9$개의 세포 또는 약 $5\times10^7$개의 세포 내지 약 $0.5\times10^9$개의 세포가 대상체에게 투여된다.

[2074] 일 실시형태에서, 대상체에는 적어도 $0.1\times10^4$개의 세포/kg 체중, 적어도 $0.5\times10^4$개의 세포/kg 체중, 적어도 1

$\times 10^4$개의 세포/kg 체중, 적어도 $5\times 10^4$개의 세포/kg 체중, 적어도 $1\times 10^5$개의 세포/kg 체중, 적어도 $0.5\times 10^6$개의 세포/kg 체중, 적어도 $1\times 10^6$개의 세포/kg 체중, 적어도 $0.5\times 10^7$개의 세포/kg 체중, 적어도 $1\times 10^7$개의 세포/kg 체중, 적어도 $0.5\times 10^8$개의 세포/kg 체중, 적어도 $1\times 10^8$개의 세포/kg 체중, 적어도 $2\times 10^8$개의 세포/kg 체중, 적어도 $3\times 10^8$개의 세포/kg 체중, 적어도 $4\times 10^8$개의 세포/kg 체중, 적어도 $5\times 10^8$개의 세포/kg 체중 또는 적어도 $1\times 10^9$개의 세포/kg 체중이 투여된다. 특정 실시형태에서, 약 $1\times 10^6$개의 세포/kg 체중 내지 약 $1\times 10^8$개의 세포/kg 체중, 약 $2\times 10^6$ T개의 세포/kg 체중 내지 약 $0.9\times 10^8$개의 세포/kg 체중, 약 $3\times 10^6$개의 세포/kg 체중 내지 약 $0.8\times 10^8$개의 세포/kg 체중, 약 $4\times 10^6$개의 세포/kg 체중 내지 약 $0.7\times 10^8$개의 세포/kg 체중, 약 $5\times 10^6$개의 세포/kg 체중 내지 약 $0.6\times 10^8$개의 세포/kg 체중 또는 약 $5\times 10^6$개의 세포/kg 체중 내지 약 $0.5\times 10^8$개의 세포/kg 체중이 대상체에게 투여된다.

[2075] 당업자는 특정 실시형태에서 상정되는 약제학적 조성물의 다중 투여가 목적하는 요법에 영향을 미치기 위해 필요할 수 있음을 인식할 것이다. 예를 들어, 조성물은 1주, 2주, 3주, 1개월, 2개월, 3개월, 4개월, 5개월, 6개월, 1년, 2년, 5년, 10년 이상의 기간에 걸쳐 1회, 2회, 3회, 4회, 5회, 6회, 7회, 8회, 9회 또는 10회 이상 대상체에게 투여될 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 대상체에게 투여하는 단계를 포함할 수 있고, 방법은 면역 반응을 증가시키거나 또는 감소시킨다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1일 1회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1일 2회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1일 3회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1주에 1회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1주에 2회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1주에 3회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1개월에 1회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1개월에 2회 이상 투여하는 단계를 포함할 수 있다. 임의의 이러한 방법에서, 방법은 유효량의 변형된 면역 세포를 1개월에 3회 이상 투여하는 단계를 포함할 수 있다.

[2076] 소정의 실시형태에서, 활성화된 약제학적 조성물을 대상체에게 투여한 다음 후속적으로 혈액을 다시 채혈하고(또는 성분채집을 수행함), 이로부터 T 세포를 활성화하고, 이러한 활성화되고 확장된 T 세포를 환자에게 재주입하는 것이 바람직할 수 있다. 이러한 과정은 몇 주마다 여러 번 수행될 수 있다. 소정의 실시형태에서, T 세포는 10cc 내지 400cc의 채혈로부터 활성화될 수 있다. 소정의 실시형태에서, T 세포는 20cc, 30cc, 40cc, 50cc, 60cc, 70cc, 80cc, 90cc, 100cc, 150cc, 200cc, 250cc, 300cc, 350cc 또는 400cc 이상의 채혈로부터 활성화된다.

[2077] 본 명세서에서 상정된 약제학적 조성물의 투여는 주입, 수혈 또는 비경구를 포함하지만 이들로 제한되지 않는 종래의 기법을 사용하여 수행될 수 있다. 일부 실시형태에서, 비경구 투여는 혈관내로, 정맥내로, 근육내로, 동맥내로, 척추강내로, 종양내로, 피내로, 복강내로, 기관내로, 피하로, 표피하로, 관절내로, 피막하로, 지주막하로 및 흉골내로 주입 또는 주사하는 것을 포함한다.

[2078] 개별 대상체에서의 반응은 완전 반응, 부분 반응, 안정 병변 및 진행성 병변으로 특징지어질 수 있다. 일부 실시형태에서, 반응은 완전 반응(complete response: CR)이다. 완전 반응은 모든 순환 종양 세포(circulating tumor cell: CTC) 또는 단핵 혈액 세포(mononuclear blood cell: MNBC)의 소멸로 정의될 수 있으며, 즉, 임의의 병리학적 림프절(표적이든 또는 비-표적이든)은 단축이 10㎜ 미만으로 감소되어야 한다. 일부 예에서(예를 들어, AML), 완전 반응은 다음과 같이 정의될 수 있다: 골수 아세포가 5% 미만; 아우어 막대(Auer rod)가 있는 아세포의 부재; 골수외 질환의 부재; 절대 호중구수 $1.0\times 10^9/\ell$ (1000/㎕) 초과; 혈소판수 $100\times 10^9/\ell$ (100 000/㎕) 초과; 및 적혈구 수혈의 독립성. 소정의 실시형태에서, 반응은 불완전한 회복을 동반한 CR(CR with Incomplete Recovery: CRi)이다. 일부 예에서(예를 들어, AML), 불완전한 회복을 동반한 CR은 잔류 호중성백혈구 감소증($1.0\times 10^9/\ell$ [1000/㎕] 미만) 또는 혈소판감소증($100\times 10^9/\ell$ [100 000/㎕] 미만)을 제외한 모든 CR 기준을 포함하는 것으로 정의될 수 있다. 소정의 실시형태에서, 반응은 형태학적 무백혈병 상태(morphologic leukemia free state)이다. 일부 예에서(예를 들어, AML), 형태학적 무백혈병 상태는 골수 아세포가 5% 미만; 아우어 막대가 있는 아세포의 부재; 골수외 질환의 부재; 및 혈액학적 회복이 필요하지 않음을 포함하는 것으로

정의될 수 있다. 소정의 실시형태에서, 반응은 부분 반응(partial response: PR)이다. 부분 반응은 기준치 합계 직경을 기준으로 순환 종양 세포(CTC) 또는 단핵 혈액 세포(MNBC)의 직경의 합의 적어도 30%의 감소를 의미하는 것으로 정의될 수 있다. 일부 예에서(예를 들어, AML), 부분 반응은 CR의 모든 혈액학 기준을 포함하는 것으로 정의될 수 있다; 골수 아세포 백분율이 5% 내지 25%로 감소; 및 치료전 골수 아세포 백분율이 적어도 50% 감소. 소정의 실시형태에서, 반응은 형태학적 무백혈병 상태이다. 일부 예에서(예를 들어, AML), 형태학적 무백혈병 상태는 골수 아세포가 5% 미만; 아우어 막대가 있는 아세포의 부재; 골수외 질환의 부재; 및 혈액학적 회복이 필요하지 않음을 포함하는 것으로 정의될 수 있다. 소정의 실시형태에서, 반응은 재발이다.

[2079] 일부 예에서(예를 들어, AML), 재발은 골수 아세포가 5% 미만; 아우어 막대가 있는 아세포의 부재; 골수외 질환의 부재; 및 혈액학적 회복이 필요하지 않음을 포함하는 것으로 정의될 수 있다. 일부 실시형태에서, 반응은 진행성 병변(progressive disease: PD)이다. 진행성 병변은 연구에서 가장 작은 수(이는 가장 작은 경우 기준치 수가 포함됨)를 기준으로 순환 종양 세포(CTC) 또는 단핵 혈액 세포(MNBC)의 수의 적어도 20%의 증가 및 순환 종양 세포(CTC) 또는 단핵 혈액 세포(MNBC)의 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개, 적어도 11개, 적어도 12개, 적어도 13개, 적어도 14개, 적어도 15개, 적어도 16개, 적어도 17개, 적어도 18개, 적어도 19개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개 또는 적어도 100개 이상의 절대적 증가로 정의될 수 있다.

[2080] 하나 이상의 새로운 병변의 출현은 또한 진행으로 간주될 수 있다. 일부 실시형태에서, 질환은 안정 병변(stable disease: SD)이다. 안정 병변은 연구 중에 가장 적은 수의 CTC 및/또는 MNBC를 기준으로 PR의 자격을 갖기에 충분한 액체암 세포수의 감소도 PD의 자격을 갖기에 충분한 증가로도 특징지을 수 없다. 소정의 실시형태에서, 반응은 병리학적 완전 반응이다. 예를 들어, 수술 또는 생검시에 제거된 조직의 검사 후 병리학자에 의해 결정된 바와 같은 병리학적 완전 반응은 일반적으로 수술 표본에서 침습성 및/또는 비-침습성 액체 암세포의 조직학적 증거가 없음을 지칭한다.

[2081] 일부 실시형태에서, 약제학적 조성물은 주 2회 또는 3회 투여된다. 또 다른 실시형태에서, 약제학적 조성물은 주 2회 투여된다. 또 다른 실시형태에서, 약제학적 조성물은 2주 또는 3주마다 1회 투여된다. 또 다른 실시형태에서, 약제학적 조성물은 1주 또는 2주마다 1회 투여된다. 또 다른 실시형태에서, 약제학적 조성물은 21-일 주기의 1일, 4일, 8일 및 11일에 투여된다. 또 다른 실시형태에서, 약제학적 조성물은 28-일 주기의 1일, 8일 및 15일에 투여된다.

[2082] 일부 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 0.5mg 내지 30mg이다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 0.5mg 내지 20mg이다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 0.5mg 내지 10mg이다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 0.04mg, 약 0.08mg, 약 0.16mg, 약 0.32mg, 약 0.64mg, 약 1.25mg, 약 1.28mg, 약 1.92mg, 약 2.5mg, 약 3.56mg, 약 3.75mg, 약 5.0mg, 약 7.12mg, 약 7.5mg, 약 10mg, 약 14.24mg, 약 15mg, 약 20mg 또는 약 30mg이다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 1.92mg, 약 3.75mg, 약 7.5mg, 약 15.0mg 또는 약 30.0mg이고, 화합물은 주 2회 투여된다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 1.28mg, 약 2.56mg, 약 5.0mg, 약 10mg 또는 약 20mg이고, 화합물은 주 2회 투여된다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 1.92mg, 약 3.75mg, 약 7.5mg, 약 15.0mg 또는 약 30.0mg이고, 화합물은 주 1회 투여된다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 1.28mg, 약 2.56mg, 약 5.0mg, 약 10mg 또는 약 20mg이고, 화합물은 주 1회 투여된다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 1.92mg, 약 3.75mg, 약 7.5mg, 약 15.0mg 또는 약 30.0mg이고, 화합물은 1일 1회 7일 동안 3회, 5회 또는 7회 투여된다. 또 다른 실시형태에서, 화합물은 1일 1회, 7일 동안 7회 정맥내로 투여된다. 또 다른 실시형태에서, 투여되는 화합물의 양은 인간 대상체의 체중 킬로그램당 약 1.28mg, 약 2.56mg, 약 5.0mg, 약 10mg 또는 약 20mg이고, 화합물은 1일 1회 7일 동안 3회, 5회 또는 7회 투여된다. 또 다른 실시형태에서, 화합물은 1일 1회, 7일 동안 7회 정맥내로 투여된다.

[2083] 일부 실시형태에서, 화합물은 0.25시간, 0.5시간, 1시간, 2시간, 3시간, 4시간, 5시간, 6시간, 7시간, 8시간, 9시간, 10시간, 11시간 또는 12시간의 기간에 걸쳐 투여된다. 또 다른 실시형태에서, 화합물은 0.25시간 내지 2시간의 기간에 걸쳐 투여된다. 또 다른 실시형태에서, 화합물은 1시간의 기간에 걸쳐 점진적으로 투여된다. 또 다른 실시형태에서, 화합물은 2시간의 기간에 걸쳐 점진적으로 투여된다.

[2084] 일부 실시형태에서, 화합물의 생체내 순환 반감기는 약 1시간, 2시간, 3시간, 4시간, 5시간, 6시간, 7시간, 8시

간, 9시간, 10시간 또는 12시간이다. 또 다른 실시형태에서, 화합물의 생체내 순환 반감기는 약 4시간이다. 또 다른 실시형태에서, 화합물의 생체내 순환 반감기는 약 6시간이다. 일부 실시형태에서, 화합물의 생물학적 조직 반감기는 약 1시간, 2시간, 3시간, 4시간, 5시간, 6시간, 7시간, 8시간, 9시간, 10시간 또는 12시간이다. 또 다른 실시형태에서, 화합물의 생물학적 조직 반감기는 약 10시간이다.

[2085] 일부 실시형태에서, 치료는 치료 시작 후 1개월의 기간 내에 액체 암세포의 수를 약 95%, 약 90%, 약 85%, 약 80%, 약 75%, 약 70%, 약 65%, 약 60%, 약 55%, 약 50%, 약 45%, 약 40%, 약 35%, 약 30%, 약 25%, 약 20%, 약 15%, 약 10% 또는 약 5% 감소시킨다. 또 다른 실시형태에서, 치료는 치료 시작 후 1개월의 기간 내에 액체 암세포의 수를 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90% 또는 적어도 약 95% 감소시킨다. 또 다른 실시형태에서, 치료는 치료 시작 후 1년의 기간 내에 액체 암세포의 수를 약 95%, 약 90%, 약 85%, 약 80%, 약 75%, 약 70%, 약 65%, 약 60%, 약 55%, 약 50%, 약 45%, 약 40%, 약 35%, 약 30%, 약 25%, 약 20%, 약 15%, 약 10% 또는 약 5% 감소시킨다. 또 다른 실시형태에서, 치료는 치료 시작 후 1년의 기간 내에 액체 암세포의 수를 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90% 또는 적어도 약 95% 감소시킨다. 또 다른 실시형태에서, 치료는 치료 시작 후 6개월의 기간 내에 액체 암세포의 수를 약 95%, 약 90%, 약 85%, 약 80%, 약 75%, 약 70%, 약 65%, 약 60%, 약 55%, 약 50%, 약 45%, 약 40%, 약 35%, 약 30%, 약 25%, 약 20%, 약 15%, 약 10% 또는 약 5% 감소시킨다. 또 다른 실시형태에서, 치료는 치료 시작 후 6개월의 기간 내에 액체 암세포의 수를 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90% 또는 적어도 약 95% 감소시킨다. 또 다른 실시형태에서, 치료는 치료 시작 후 3개월의 기간 내에 액체 암세포의 수를 약 95%, 약 90%, 약 85%, 약 80%, 약 75%, 약 70%, 약 65%, 약 60%, 약 55%, 약 50%, 약 45%, 약 40%, 약 35%, 약 30%, 약 25%, 약 20%, 약 15%, 약 10% 또는 약 5% 감소시킨다. 또 다른 실시형태에서, 치료는 치료 시작 후 3개월의 기간 내에 액체 암세포의 수를 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90% 또는 적어도 약 95% 감소시킨다.

[2086] 일부 실시형태에서, 치료는 인간 대상체가 화합물로 치료되지 않은 경우 인간 대상체의 예상 생존 시간과 비교하여, 인간 대상체의 생존 시간을 증가시킨다. 또 다른 실시형태에서, 인간 대상체의 생존 시간의 증가는 적어도 30일이다. 또 다른 실시형태에서, 인간 대상체의 생존 시간의 증가는 적어도 3개월이다. 또 다른 실시형태에서, 인간 대상체의 생존 시간의 증가는 적어도 6개월이다. 또 다른 실시형태에서, 인간 대상체의 생존 시간의 증가는 적어도 1년이다.

[2087] 일부 실시형태에서, 인간 대상체는 액체암의 하나 이상의 다른 치료에 대해 불응성(refractory) 및/또는 불내성(intolerant)이다. 또 다른 실시형태에서, 인간 대상체는 적어도 하나의 성공적이지 않은 액체암의 선행 치료 및/또는 요법을 받은 적이 있다.

[2088] 일부 실시형태에서, 기재된 방법으로 치료되는 인간 대상체는 어린이(예를 들어, 0세 내지 18세)이다. 다른 실시형태에서, 기재된 방법으로 치료되는 인간 대상체는 성인(예를 들어, 18세 이상)이다.

[2089] ***키트***

[2090] 본 발명은 대상체에서 신생물의 치료를 위한 키트를 제공한다. 일부 실시형태에서, 키트는 고형 종양의 치료를 위한 것이다. 다른 실시형태에서, 키트는 액체 종양의 치료를 위한 것이다. 일부 실시형태에서, 키트는 혈액암의 치료를 위한 것이다. 일부 실시형태에서, 혈액암은 B 세포암이다. 일부 실시형태에서, 혈액암은 백혈병, 골수종 및/또는 림프종이다. 일부 실시형태에서, 키트는 T-세포 급성 림프아구성 백혈병(T-ALL), 용상 진균증(MF), 세자리 증후군(SS), 말초 T/NK-세포 림프종, 역형성 대세포 림프종 ALK+, 원발성 피부 T-세포 림프종, T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 CD30 + 림프증식성 장애, 외부 결절성 NK/T-세포 림프종, 성인 T-세포 백혈병/림프종, T-세포 전림프구성 백혈병, 피하 지방층염-유사 T-세포 림프종, 원발성 피부 감마-델타 T-세포 림프종, 공격성 NK-세포 백혈병 및 장질환-관련 T-세포 림프종으로 이루어진 군으로부터 선택되는 신생물의 치료를 위한 것이다. 일부 실시형태에서, 키트는 T-세포 급성 림프아구성 백혈병(T-ALL)의 치료를 위한 것이다. 일부 실시형태에서, 키트는 급성 골수성 백혈병(AML)의 치료를 위한 것이다. 일부 실시형태에서, 키트는 인간 대상체의 치료를 위한 것이다.

[2091] 일부 실시형태에서, 키트는 본 명세서에 제공된 바와 같은 임의의 키메라 항원 수용체를 포함한다. 일부 실시형태에서, 키트는 본 명세서에 제공된 바와 같은 임의의 키메라 항원 수용체를 암호화하는 핵산을 포함한다. 일부 실시형태에서, 키트는 임의의 본 명세서에 제공된 바와 같은 변형된 면역 세포를 포함한다. 일부 실시형태에서, 키트는 CD3 CAR, CD5 CAR, CD7 CAR, CD33 CAR 및/또는 CD123 CAR을 발현하는 변형된 면역 세포를 포함한다. 일

부 실시형태에서, 임의의 면역 세포는 CD3, CD5, CD7, CD33, CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및/또는 PD1 또는 이들의 조합 중 하나 이상에서의 돌연변이를 더 포함한다.

[2092] 일부 실시형태에서, 키트는 본 명세서에 제공된 임의의 변형된 면역 세포의 집단을 포함한다. 일부 실시형태에서, 키트는 본 명세서에 제공된 바와 같은 임의의 약제학적 조성물을 포함한다. 일부 실시형태에서, 키트는 CD123 변형된 면역 세포의 집단, 또는 CD123 변형된 면역 세포 또는 변형된 면역 세포의 집단을 포함하는 약제학적 조성물을 포함한다. 일부 실시형태에서, 키트는 CD5 변형된 면역 세포의 집단, 또는 CD5 변형된 면역 세포 또는 변형된 면역 세포의 집단을 포함하는 약제학적 조성물을 포함한다. 일부 실시형태에서, 키트는 CD7 변형된 면역 세포의 집단, 또는 CD7 변형된 면역 세포 또는 변형된 면역 세포의 집단을 포함하는 약제학적 조성물을 포함한다. 일부 실시형태에서, 키트는 CD33 변형된 면역 세포의 집단, 또는 CD33 변형된 면역 세포 또는 변형된 면역 세포의 집단을 포함하는 약제학적 조성물을 포함한다. 일부 실시형태에서, 키트는 CD3 변형된 면역 세포의 집단, 또는 CD3 변형된 면역 세포 또는 변형된 면역 세포의 집단을 포함하는 약제학적 조성물을 포함한다.

[2093] 일부 실시형태에서, 키트는 염기 편집기 폴리펩타이드 또는 염기 편집기 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 추가로 포함하되, 염기 편집기 폴리펩타이드는 핵산 프로그램 가능 DNA 결합 단백질(napDNAbp) 및 데아미네이스를 포함한다. 일부 실시형태에서, napDNAbp는 Cas9 또는 Cas12이다. 일부 실시형태에서, 염기 편집기를 암호화하는 폴리뉴클레오타이드는 mRNA 서열이다. 일부 실시형태에서, 데아미네이스는 사이티딘 데아미네이스 또는 아데노신 데아미네이스이다.

[2094] 본 발명은 사이티딘 또는 아데노신 데아미네이스 핵염기 편집기를 암호화하는 뉴클레오타이드 서열, 적어도 하나의 가이드 RNA를 포함하는 핵산 작제물을 포함하는 키트를 제공하며, 각 가이드 RNA는 TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M, PD1, CBLB 및/또는 CTLA4를 암호화하는 유전자의 핵산 서열에 대해 적어도 85% 상보적인 핵산 서열을 갖는다. 일부 실시형태에서, 사이티딘 또는 아데노신 데아미네이스를 암호화하는 뉴클레오타이드 서열은 사이티딘 또는 아데노신 데아미네이스 핵염기 편집기의 발현을 구동하는 이종성 프로모터를 포함한다.

[2095] 본 개시내용의 일부 양태는 핵염기 편집기를 암호화하는 뉴클레오타이드 및 가이드 RNA를 포함하는 핵산 작제물을 포함하는 키트를 제공한다. 일부 실시형태에서, 핵산 작제물은 핵염기 편집기의 발현을 구동하는 이종성 프로모터를 포함한다. 일부 실시형태에서, 본 개시내용은 (a) 본 명세서에 제공된 바와 같은 사이티딘 또는 아데노신 데아미네이스에 융합된 Cas9 도메인을 암호화하는 뉴클레오타이드 서열; 및 (b) (a) 서열의 발현을 구동하는 이종성 프로모터를 포함하는 핵산 작제물을 포함하는 키트를 제공한다. 일부 실시형태에서, Cas9 도메인은 사이티딘 데아미네이스에 융합된다. 일부 실시형태에서, Cas9 도메인은 아데노신 데아미네이스에 융합된다.

[2096] 일부 실시형태에서, 키트는 사이티딘 데아미네이스 핵염기 편집기 및 가이드 RNA를 포함한다. 일부 실시형태에서, 키트는 아데노신 데아미네이스 핵염기 편집기 및 가이드 RNA를 포함한다. 일부 실시형태에서, 키트는 하나 이상의 가이드 핵산 서열을 추가로 포함한다. 일부 실시형태에서, 하나 이상의 가이드 핵산 서열은 *CD3, CD5, CD7, CD33* 또는 *CD123*을 표적으로 한다. 일부 실시형태에서, 하나 이상의 가이드 핵산 서열은 CD3, CD5, CD7, CD33, CD123, TRAC, LAG-3, FAS, CIITA, TRBC1, TRBC2, CD52, B2M 및/또는 PD1 중 하나를 각각 표적으로 한다. 일부 실시형태에서, 가이드 RNA는 CD5를 암호화하는 유전자의 핵산 서열에 대해 적어도 85% 상보적인 핵산 서열을 갖는다. 일부 실시형태에서, 키트는 핵염기 편집기 및 CD5 가이드 RNA를 포함한다. 일부 실시형태에서, 가이드 RNA는 CD7을 암호화하는 유전자의 핵산 서열에 대해 적어도 85% 상보적인 핵산 서열을 갖는다. 일부 실시형태에서, 키트는 핵염기 편집기 및 CD7 가이드 RNA를 포함한다. 일부 실시형태에서, 가이드 RNA는 CD33을 암호화하는 유전자의 핵산 서열에 대해 적어도 85% 상보적인 핵산 서열을 갖는다. 일부 실시형태에서, 키트는 핵염기 편집기 및 CD33 가이드 RNA를 포함한다. 일부 실시형태에서, 가이드 RNA는 CD3을 암호화하는 유전자의 핵산 서열에 대해 적어도 85% 상보적인 핵산 서열을 갖는다. 일부 실시형태에서, 키트는 핵염기 편집기 및 CD3 가이드 RNA를 포함한다. 일부 실시형태에서, 가이드 RNA는 CD123을 암호화하는 유전자의 핵산 서열에 대해 적어도 85% 상보적인 핵산 서열을 갖는다. 일부 실시형태에서, 키트는 사이티딘 데아미네이스 핵염기 편집기 및 CD123 가이드 RNA를 포함한다. 일부 실시형태에서, 키트는 하나 이상의 추가적인 가이드 RNA를 추가로 포함할 수 있으며, 각 가이드 RNA는 TRAC, LAG-3, FAS, PD1, B2M, CIITA, TRBC1, TRBC2 및/또는 CD52를 암호화하는 유전자의 핵산 서열에 대해 적어도 85% 상보적인 핵산 서열을 갖는다.

[2097] 신생물 치료 키트는 신생물의 치료에서 변형된 면역 세포를 사용하기 위한 서면 설명서를 추가로 포함할 수 있다. 다른 실시형태에서, 설명서는 다음 중 적어도 하나를 포함한다: 주의사항; 경고; 임상 연구; 및/또는 참조. 설명서는 용기(존재하는 경우)에 직접, 또는 용기에 부착된 라벨에 또는 용기 내부 또는 용기와 함께 제공되는

별도의 시트, 팜플렛, 카드 또는 폴더에 인쇄될 수 있다. 추가의 실시형태에서, 키트는 적합한 제어 변수(operational parameter)에 대한 라벨 또는 별도의 삽입물(패키지 삽입물) 형태의 설명서를 포함할 수 있다. 또 다른 실시형태에서, 키트는 검출, 보정 또는 정규화를 위한 표준(들)으로 사용되는 적절한 양성 및 음성 대조군 또는 대조군 샘플을 갖는 하나 이상의 용기를 포함할 수 있다. 키트는 (멸균) 인산 완충 식염수, 링거액 또는 덱스트로스 용액과 같은 약제학적으로 허용 가능한 완충액을 포함하는 제2 용기를 추가로 포함할 수 있다. 이는 다른 완충액, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 패키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[2098] 본 발명의 실시는 달리 지시되지 않는 한 당업자의 범위 내에 있는 분자 생물학(재조합 기법 포함), 미생물학, 세포 생물학, 생화학 및 면역학의 종래의 기법을 사용한다. 이러한 기법은 문헌, 예컨대, 문헌["Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991)]에 충분히 설명되어 있다. 이들 기법은 본 발명의 폴리뉴클레오타이드 및 폴리펩타이드의 생산에 적용 가능하며, 그 자체로 본 발명을 만들고 실시하는 데 고려될 수 있다. 특정 실시형태에 특히 유용한 기법은 다음 부문에서 논의될 것이다.

[2099] 하기 실시예는 본 발명의 검정, 스크리닝 및 치료적 방법을 만들고 사용하는 방법에 대한 완전한 개시 및 설명을 당업자에게 제공하기 위해 제시되며, 본 발명자들이 발명으로 간주하는 것의 범위를 제한하는 것으로 의도되지 않는다.

[2100] **실시예**

[2101] **실시예 1: 면역 세포에서 발현되는 유전자에서의 스플라이싱 부위의 파괴 및 종결 코돈의 도입**

[2102] 핵염기 편집기 BE4를 사용하여 면역 세포에서 발현된 유전자의 스플라이싱 부위를 파괴하고 이의 서브세트에 종결 코돈을 삽입하였다. 플라스미드 작제물 pCMV_BE4max는 사이티딘 데아미네이스 활성을 갖는 APOBEC-1 사이티딘 데아미네이스 도메인, D10A 돌연변이를 포함하며 닉케이스 활성을 갖는 Cas9 도메인 및 2개의 우라실 DNA 글리코실레이스 저해제(UGI) 도메인을 포함하는 BE4를 암호화한다. UGI는 면역 세포에서 발현되는 소정의 유전자의 스플라이싱 부위를 편집하기 위해 강력하게 차단하는 바실러스 서브틸리스 박테리오파지 PBS1로부터 유래되는 83개의 아미노산 잔기 단백질이다. BE4는 N-말단 및 C-말단 핵 국부화 신호(NLS)를 추가로 포함한다.

[2103] >pCMV_BE4max

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA
GTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC
ATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCC
AAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC
TATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGA
CTCACTATAGGGAGAGCCGCCACCATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAA
AGAAGAAGCGGAAAGTCTCCTCAGAGACTGGGCCTGTCGCCGTCGATCCAACCCTGCGCCGCCG
GATTGAACCTCACGAGTTTGAAGTGTTCTTTGACCCCCGGGAGCTGAGAAAGGAGACATGCCTG
CTGTACGAGATCAACTGGGGAGGCAGGCACTCCATCTGGAGGCACACCTCTCAGAACACAAATA
AGCACGTGGAGGTGAACTTCATCGAGAAGTTTACCACAGAGCGGTACTTCTGCCCCAATACCAG
ATGTAGCATCACATGGTTTCTGAGCTGGTCCCCTTGCGGAGAGTGTAGCAGGGCCATCACCGAG
TTCCTGTCCAGATATCCACACGTGACACTGTTTATCTACATCGCCAGGCTGTATCACCACGCAG
ACCCAAGGAATAGGCAGGGCCTGCGCGATCTGATCAGCTCCGGCGTGACCATCCAGATCATGAC
AGAGCAGGAGTCCGGCTACTGCTGGCGGAACTTCGTGAATTATTCTCCTAGCAACGAGGCCCAC
TGGCCTAGGTACCCACACCTGTGGGTGCGCCTGTACGTGCTGGAGCTGTATTGCATCATCCTGG
GCCTGCCCCCTTGTCTGAATATCCTGCGGAGAAAGCAGCCCCAGCTGACCTTCTTTACAATCGC
CCTGCAGTCTTGTCACTATCAGAGGCTGCCACCCCACATCCTGTGGGCCACAGGCCTGAAGTCT
GGAGGATCTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAG
AGAGCAGTGGCGGCAGCAGCGGCGGCAGCGACAAGAAGTACAGCATCGGCCTGGCCATCGGCAC
CAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTG
CTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCG
GCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAA
CCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTC

[2104]

CACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCG
GCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAA
ACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGC
TGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGG
CGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATC
GCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCC
TGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGA
CACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTT
CTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGA
TCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGAC
CCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAG
AGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCA
TCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGA
CCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAG
CTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGA
TCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAG
ATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTG
GACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCA
ACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGAC
CAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAG
GCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACT
ACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGC
CTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAG
GAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGA
TCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCG
GCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG
TCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGC
TGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGG
CGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCG
TGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAAT
GAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAA
AACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACG
TGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAG

[2105]

CTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAG
AGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGA
ACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
CGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCAC
GTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGG
AAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTA
CAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGA
ACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGT
ACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTT
CTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCA
GACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGA
AAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGC
TGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGC
TACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAA
ACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCC
CTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAG
GATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGC
AGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGC
CTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC
CTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT
ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGA
GACACGGATCGACCTGTCTCAGCTGGGAGGTGACAGCGGCGGGAGCGGCGGGAGCGGGGGGAGC
ACTAATCTGAGCGACATCATTGAGAAGGAGACTGGGAAACAGCTGGTCATTCAGGAGTCCATCC
TGATGCTGCCTGAGGAGGTGGAGGAAGTGATCGGCAACAAGCCAGAGTCTGACATCCTGGTGCA
CACCGCCTACGACGAGTCCACAGATGAGAATGTGATGCTGCTGACCTCTGACGCCCCCGAGTAT
AAGCCTTGGGCCCTGGTCATCCAGGATTCTAACGGCGAGAATAAGATCAAGATGCTGAGCGGAG
GATCCGGAGGATCTGGAGGCAGCACCAACCTGTCTGACATCATCGAGAAGGAGACAGGCAAGCA
GCTGGTCATCCAGGAGAGCATCCTGATGCTGCCCGAAGAAGTCGAAGAAGTGATCGGAAACAAG
CCTGAGAGCGATATCCTGGTCCATACCGCCTACGACGAGAGTACCGACGAAAATGTGATGCTGC
TGACATCCGACGCCCCAGAGTATAAGCCCTGGGCTCTGGTCATCCAGGATTCCAACGGAGAGAA
CAAAATCAAAATGCTGTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAG

[2106] AAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGC

CTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGA
CAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGG
GGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGT
GAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTA
GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCG
GGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTA
TTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGC
GGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAG
AACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT
TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTC
ATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCA
CGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCG
GTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATG
TAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATT
TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC
AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAA
AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAACTC
ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAA
AAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCC
CGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGC
GCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAG
AGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACAT
GATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAA
GTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCA
TCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTT
AAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTG
AGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCA

[2107]

GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACG
GAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACAT
TTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTC
GACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTG
TTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACA
ATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATA
TACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCAT
AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCA
ACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTT
CCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT

[2108]

C

[2109] 면역 세포에서 단백질 발현을 넉다운 또는 넉아웃시키는 BE4의 효과를 확인하기 위해, 제1 면역 세포의 집단을

BE4를 암호화하는 mRNA와 특정 표적 부위에 따라 TRAC 엑손 2, TRAC 엑손 4 또는 B2M 엑손 1의 공여체 또는 수용체 스플라이싱 부위의 G 염기에 상보적인 C 염기를 표적으로 하는 sgRNA로 공동 형질감염시켰다. 시험관내 전사(트라이링크 바이오테크놀로지스(TriLinK Biotechnologies))에 의해 mRNA를 생성하였다. 간략하게는, 4㎍의 BE4 mRNA 및 2㎍의 합성 gRNA를 1M CD3+ T 세포(Nucleofector™ 플랫폼, 론자 바이오사이언스(Lonza Bioscience))에 전기천공하였다. 그런 다음, 세포를 3일 동안 배양하여 염기-편집하기에 충분한 시간을 제공하였다. 비교를 위해, 제2 면역 세포의 집단을 Cas9 뉴클레이스를 암호화하는 mRNA와 B2M 엑손 1의 공여체 스플라이싱 부위의 G 염기를 표적으로 하는 sgRNA로 공동 형질감염시켰다. BE4 편집과 Cas9 편집 사이에 눈의 띄는 차이는 관찰되지 않았고, 각 편집된 유전자에 대한 넉다운은 90%보다 큰 반면, 전기천공되지 않은 대조군 세포는 유의한 넉다운이 없었다(도 2).

[2110] 세포가 단일 가닥 닉을 촉매하는 BE4를 암호화하는 mRNA 또는 이중-가닥 파손을 촉매하는 Cas9 뉴클레이스로 형질감염되는 경우, 표적화된 유전자의 관찰된 넉다운의 원인이 되는 유전적 변형이 다를 것이라고 가정하였다. 이러한 가설을 테스트하기 위해, 면역 세포를 2㎍의 BE4/1㎍의 sgRNA(중간) 또는 4㎍의 BE4/2㎍의 sgRNA(고)의 BE4 염기 편집기를 암호화하는 mRNA와 B2M 엑손 1의 공여체 스플라이싱 부위의 G 염기를 표적으로 하는 sgRNA로 공동 형질감염시켰다. 3일, 5일 및 7일 동안 인큐베이션한 후, DNA를 수집하고, 시퀀싱하였다. 도 3을 참조하면, 대부분의 염기 편집은 예상되는 방식으로 스플라이싱 부위만의 파괴를 나타내었다(즉, 안티센스 가닥에서 C에서 T로의 전이가 혼입되어 센스 가닥에서 G에서 A로의 전이가 발생함). 이러한 결과는 Cas9 뉴클레이스로 형질감염된 세포로부터 얻어진 결과와 대조되며, 이는 Cas9 형질감염된 세포에서 대부분의 편집이 indel임을 나타낸다(도 3).

[2111] 스플라이싱 부위의 파괴 및 종결 코돈의 도입은 표적 유전자의 발현을 넉다운시키는데 효과적일 수 있다. TRAC 엑손 4의 스플라이스 수용체와 B2M 엑손 1 및 PDCD1 엑손 1의 스플라이스 공여체의 BE4-매개성 편집은 전장 단백질의 발현을 감소시켰다(도 4 및 도 5). 스플라이싱 부위에서 관찰된 BE4-매개성 변화는 C에서 T로의 전이였지만, indel 및 C에서 G로의 전환도 관찰되었다. 엑손의 연속적인 사이티딘 잔기가 표적화되고 티미딘 잔기로 편집되는 PDCD1 유전자의 엑손 2에 오커(ochre) 종결 코돈을 삽입하면 TRAC 및 B2M 유전자에 대해 관찰된 것보다 덜 감소하더라도 유전자 발현의 유의한 넉다운이 발생하였다(도 4). 이러한 결과는 면역 세포에서 발현된 유전자의 BE4-매개성 단일 또는 연속적인 사이티딘 염기 편집이 유전자 발현의 효율적인 감소를 초래한다는 것을 추가로 시사한다.

[2112] **실시예 2: 스플라이싱 부위 파괴 및 종결 코돈 삽입의 인 실리코(In silico) 분석**

[2113] 설계된 gRNA가 오프-사이트(off-site) 표적에 결합할 것인지 결정하기 위해, CAS-OFFinder를 사용하여 gRNA의 핵산 서열을 분석하였다. 도 6을 참조하면, "X" 벌지 유형은 gRNA가 게놈 DNA와 정렬되고, 임의의 차이가 불일치임을 나타낸다. 불일치의 수가 1개에서 4개로 증가함에 따라, 잠재적인 오프-사이트 결합이 증가한다. 예를 들어, TRAC 엑손 4 스플라이스 수용체에 대한 결과는 3개의 불일치가 있는 경우 26개의 오프사이트 결합 가능성이 있는 반면, 4개의 불일치가 있는 경우 164개의 가능성이 있음을 보여준다.

[2114] gRNA가 20개의 염기쌍을 갖지만 게놈 DNA의 19개의 염기쌍과 정렬되는 gRNA가 벌지를 갖는 경우, 벌지가 발생한다. 다시, 도 6을 참조하면, TRAC 엑손 4 스플라이스 수용체 gRNA가 한 염기쌍의 벌지를 가질 때, 불일치가 증가함에 따라 오프사이트 결합 가능성의 수는 증가하고; 그러나, 가능성의 수는 벌지가 없을 때(즉, 벌지 크기가 0일 때)보다 유의하게 더 낮다.

[2115] **실시예 3: 면역 세포에서 다중 염기 편집**

[2116] BE4가 다중-넉다운 세포를 생성하기 위해 다중 유전자의 염기 편집을 매개할 수 있는지 여부를 결정하기 위해, 면역 세포를 B2M, TRAC, PD1 또는 이들의 조합의 특정 부위를 표적으로 하는 sgRNA와 함께 BE4 염기 편집기를 암호화하는 mRNA로 공동 형질감염시켰다. 도 7을 참조하면, BE4 시스템은 단일, 이중 및 삼중 유전자 편집에서 단백질 생산이 감소된 세포의 백분율을 확인하기 위해 유세포 분석에 의해 측정된 바와 같이 효과적인 넉다운을 유도하였다. 세포를 B2M 및 CD3 발현에 대해 게이팅하고, CD3 발현은 TRAC 발현에 대한 대용물(대용물) 역할을 하였다. PD1 염색은 비효율적이기 때문에, 이 단백질을 발현하는 세포의 직접적인 측정은 수행하지 않았다. 단일, 이중 및 삼중 유전자 편집이 있는 세포 집단 간에 차이가 관찰되지 않았으며, B2M, TRAC 및 PD1의 넉다운 발현(삼중 유전자 편집)을 위해 변형된 면역 세포는 변형되지 않은 대조군 면역 세포와 검출 가능하게 구별된다(도 8).

[2117] 감소된 단백질 발현을 담당하는 유전자에 대한 변형은 도 9에 요약되어 있다. 구체적으로 그리고 실시예 1에 기

재된 단일 유전자 변형에서 발현의 감소를 초래하는 메커니즘과 유사하게, C에서 T로의 전이는 변형된 B2M 단일 변형된 유전자 세포 집단과 B2M+PD1, B2M+TRAC 및 B2M+TRAC+PD1 다중 변형된 유전자 세포 집단에서 관찰된 많은 수의 편집을 구성한다. indel 및 전환은 편집된 유전자에서 관찰된 유전적 변화의 적은 소수를 구성한다.

[2118] 따라서, 염기 편집에 의한 3개의 유전자 좌위의 동시 변형은 단일 방향 표적화 시퀀싱(UDiTaS; 문헌[Giannoukos *et al.*, BMC Genomics. 2018 Mar 21;19(1):212. doi: 10.1186/s12864-018-4561-9])에 의해 평가되는 바와 같이 검출 가능한 전위 이벤트 없이 고효율 유전자 넉아웃을 생성하였다. 추가적으로, 전위는 BE4-편집된 유전자에서 검출되지 않았다. 액적 디지털 폴리머레이스 사슬 반응(ddPCR) 전략(도 10)을 사용하여 B2M, TRAC 및 PD1 BE4-편집된 유전자 사이의 전위를 검출하였다. B2M+TRAC+PD1 편집을 생성하기 위해 BE4 또는 Cas9로 변형된 세포로부터 추출된 DNA를 QX200 액적 디지털 기기(바이오-래드(Bio-Rad))를 사용하여 차세대 시퀀싱(NGS)으로 분석하여 BE4 및 Cas9 편집의 정확한 서열을 결정하였다. 도 11의 왼쪽 패널에 나타낸 바와 같이, 대부분의 세포에서 B2M, TRAC 및 PD1 유전자가 변형되었다. ddPCR 분석은 전위가 BE4-편집된 세포에는 존재하지 않았지만, Cas9-편집된 세포의 대략 1.7%에서 관찰되었음을 보여주었다(도 11, 오른쪽 패널). 표 21은 BE4-편집된 세포에서 전위가 관찰되지 않았음을 추가로 예시한다.

*표 21*

[2119]

| 염기 편집 | 전위 | 대조 앰플리콘 액적 | 실험 앰플리콘 액적 |
|---|---|---|---|
| Cas9 뉴클레이스 | *B2M-TRAC* | 61,206 | 585 |
| | *B2M-PDCD1* | 55,970 | 291 |
| | *PDCD1-TRAC* | 59,600 | 112 |
| BE4 | *B2M-TRAC* | 90,717 | 0 |
| | *B2M-PDCD1* | 89,028 | 0 |
| | *PDCD1-TRAC* | 83,501 | 0 |

[2120] **실시예 4: Cbl 프로토-종양 유전자 B(CBLB)의 BE4-매개성 편집**

[2121] Cbl-b는 TCR 복합체 신호전달을 음성적으로 조절하는 T 세포 수용체(TCR) 신호전달 단백질이다(도 12). T 세포는 Cbl-b 신호전달이 저해될 때 더 낮은 활성화 역치를 갖기 때문에, 이 유전자를 넉아웃 또는 넉다운시키면 T 세포 또는 CAR을 발현하는 T 세포의 효율성을 크게 개선시킬 수 있다. Cbl-b 유전자가 사이티딘 탈아미노화 매개성 변형에 민감한지 결정하기 위해, 세포를 BE4를 암호화하는 mRNA와 엑손 8 및 16의 스플라이싱 부위 수용체, 엑손 8, 10, 11 및 12의 스플라이싱 부위 공여체를 표적으로 하거나 또는 엑손 1, 4 및 8에 종결 코돈의 삽입을 촉진할 수 있는 sgRNA로 공동 형질감염시켰다. 생성된 세포를 유세포 분석으로 분석하였다.

[2122] 도 13을 참조하면, 엑손 12의 스플라이싱 부위 공여체 및 엑손 8의 스플라이싱 부위 수용체의 파괴는 Cbl-b 발현의 가장 큰 감소를 초래하였다(각각 67.2% 및 57.4%). 그리고, 엑손 8의 스플라이싱 부위 수용체 및 엑손 12의 스플라이싱 부위 공여체 sgRNA로 형질감염된 세포 중, 60%를 약간 넘는 세포가 성공적으로 편집되었다(도 13, 막대 그래프).

[2123] **실시예 5: 면역 세포에서 Cas12b 뉴클레이스 특성화**

[2124] Cas12b/c2c1 부위는 이중 가닥 핵산 분자의 두 가닥 모두를 특이적으로 표적화하고 절단한다. 표적 핵산 분자에서 indel을 매개하는 효소의 성향을 결정함으로써 2개의 상이한 Cas12b/c2c1 단백질인 BhCas12b 및 BvCas12b를 특성화하였다. Cas12b/c2c1 단백질을 암호화하는 mRNA를 GRIN2B 유전자의 유전자좌와 및 DNMT1 유전자의 유전자좌에 특이적인 가이드 RNA와 함께 T 세포에 전기천공하였다. 세포를 3일 내지 5일 동안 배양한 후 세포 DNA를 단리하였다. 차세대 시퀀싱에 의해 Indel 비율을 결정하였다. 도 14를 참조하면, BhCas12b 단백질로 처리된 세포로부터 단리된 DNA는 BvCas12b 단백질로 처리된 세포로부터 단리된 DNA(대략 20%)보다 GRIN2B 유전자에서 훨씬 더 높은 백분율(대략 75%)의 indel을 가졌다. DNMT1 유전자의 indel은 또한 BvCas12b로 처리된 세포로부터 단리된 DNA에서 관찰된 것(대략 0%)보다 BhCas12b로 처리된 세포로부터 단리된 DNA에서 더 높은 비율(대략 20%)로 관찰되었다.

[2125] BhCas12b(V4) 단백질을 사용하여 TRAC 유전자를 파괴하였다. GRIN2B, DNMT1 및 TRAC 유전자의 유전자좌에 특이적인 가이드 RNA와 함께 BhCas12b(V4) 단백질을 암호화하는 mRNA로 전기천공법을 통해 T 세포를 형질도입시켰다. 전기천공법 96시간 후, 형광 보조 세포 분류(FACS) 분석을 사용하여 세포를 평가하고, 세포를 CD3(TRAC에 대한 대용물)에 대해 게이팅하였다. 도 15를 참조하면, GFP를 암호화하는 플라스미드로, 또는 BhCas12b(V4) 및 GRIN2B 또는 DNMT1에 특이적인 가이드 RNA로 형질도입된 세포의 대략 95%가 CD3+였다. BhCas12b(V4) 및 TRAC 유전자의 유전자좌에 특이적인 가이드 RNA를 발현하도록 형질도입된 이러한 세포는 CD3+일 가능성이 적다(사용된 가이드 RNA에 따라 대략 2% 내지 대략 50%). 테스트된 11개의 TRAC 가이드 RNA 중 3개가 대략 100%의 BhCas12b(V4)-매개성 indel을 유도하였다.

[2126] **실시예 6: CAR-P2A-mCherry 렌티바이러스 발현 특성화**

[2127] CAR-P2A-mCherry 렌티바이러스를 사용하여 키메라 항원 수용체(CAR)를 발현하도록 세포를 형질도입시키고, 형광 보조 세포 분류(FACS)를 사용하여 CAR 발현에 대해 분석하였다. 세포를 염색하지 않고, R-피코에리트린(phycoerythrin: PE) 또는 플루오레세인 아이소티오사이아네이트(fluorescein isothiocyanate: FITC)에 접합된 BCMA 단백질과 함께 인큐베이션하였다. BCMA는 CAR의 표적 항원이기 때문에, CAR을 발현하는 세포는 염료-표지된 BCMA에 결합할 것이다. 도 16을 참조하면, 염색되지 않은 세포의 경우, FACS 분석은 형질도입된 샘플에서 mCherry의 존재만을 검출하였으며, 일부는 PE 채널로 스필오버(spillover)되었다. BCMA-PE 채널은 스필오버에서 볼 수 있는 것 이상의 매우 양성적인 신호를 나타내며, 이러한 결과는 BCMA-FITC와 함께 인큐베이션된 세포에서 확인되었다. 염료-표지된 BCMA 단백질 검출 결과는 mCherry에서 볼 수 있는 것과 거의 동일한 CAR의 발현을 시사한다. 도 17을 참조하면, 폴리(1,8-옥테인다이올 시트레이트)(POC) 렌티바이러스 벡터로 형질도입된 세포에서 FACS 분석을 통해 85%의 CAR 발현이 검출되었다.

[2128] **실시예 7: BE4는 높은 생성물 순도로 효율적이고, 지속 가능한 유전자 넉아웃을 생성한다**

[2129] BE4는 다중 유전자의 염기 편집을 매개하여 다중-넉다운 세포를 생성한다. 면역 세포를 B2M, TRAC, PD1 또는 이들의 조합의 특정 부위를 표적으로 하는 sgRNA와 함께 BE4 염기 편집기를 암호화하는 mRNA로 공동 형질감염시켰다. 시퀀싱 데이터에 의해 나타낸 바와 같이, 염기 편집은 세포를 변형시키는데 효율적이었으며, 적어도 7일까지 지속되었다(도 18). C에서 T로의 전이가 관찰된 수많은 편집을 구성하기 때문에 높은 생성물 순도가 관찰되었다. indel 및 C에서 G로의 그리고 C에서 A로의 전환은 편집된 유전자에서 관찰된 유전적 변화의 적은 소수를 구성하였다. 염기 편집은 또한 목적하는 변형을 생성할 때 spCas9 뉴클레이스만큼 효율적이었다.

[2130] BE4 시스템은 감소된 표면 발현을 갖는 세포의 백분율을 식별하는 유세포 분석에 의해 측정된 바와 같이 효과적인 넉다운을 유도하였다(도 19a). B2M 발현에 대해 게이팅된 세포는 세포 표면에서 B2M 단백질의 손실을 나타내었다. 유세포 분석에 의해 측정된 바와 같이, 염기 편집은 또한 B2M 단백질 넉아웃을 생성할 때 spCas9 뉴클레이스만큼 효율적이었다.

[2131] **실시예 8: 직교 전위 검출 검정은 삼중-편집된 T 세포에서 BE4-유도성 재배열을 검출할 수 없다**

[2132] 면역 세포를 B2M, TRAC 및 PD1의 특정 부위를 표적으로 하는 sgRNA와 함께 BE4 염기 편집기를 암호화하는 mRNA로 공동 형질감염시켰다. B2M, TRAC 및 PD1 표적 유전자 사이에 바람직하지 않은 특정 전위를 검출할 수 있는 전위 검출 검정을 사용하여 삼중-편집된 T 세포를 평가하였다(도 20). 특히, 이들 특정 전위 중 어느 것도 임의의 BE4-편집된 유전자에서 검출되지 않았다(표 22). 대조적으로, Cas9-처리된 세포는 낮았지만 검출 가능한 수준의 전위를 나타내었다. 따라서, BE4 염기 편집기를 사용한 T 세포의 다중 편집은 Cas9 뉴클레이스를 사용한 다중 편집과 대조적으로 전위를 생성하지 않았다.

*표 22*

| 유형 | 모의 (%) | BE4-처리 (%) | Cas9-처리 (%) |
|---|---|---|---|
| 표적에 대한 변형 (B2M/TRAC/PDCD1) | 0 | 89.9 / 97.9 / 89.1 | 53.0 / 77.2 / 55.2 |
| B2M-A / TRAC-A | 0 | 0 | 0.925 |
| B2M-A / TRAC-B | 0 | 0 | 0.353 |
| B2M-A / PDCD1-A | 0 | 0 | 1.647 |
| B2M-A / PDCD1-B | 0 | 0 | 0.508 |
| B2M-B / TRAC-A* | 0 | 0 | 0.505 |

$LLoD_{BE4}$ = 0.1%

**B2M-B* 는 전위가 *B2M* 유전자좌에 국부 재배열을 포함하는 경우에만 이 실험에서 측정 가능함

[2133]

[2134] **실시예 9: 다중 염기 편집은 세포 확장을 크게 손상시키지 않는다**

[2135] BE4 및 spCas9 sgRNA 둘 다를 사용하여 B2M, TRAC 및 PD1 표적에 대해 광범위한 가이드 스크리닝을 수행하였다. 단일-플렉스 테스트를 기반으로 높은 편집 효율과 확장을 위한 가이드를 선택하였다. 최종 세포 수율을 BE4 및 spCas9를 사용하여 1, 2 및 3 편집 사이에서 비교하였으며, 전기천공만 대조군으로 정규화하였다. 목적하는 편집이 있는 BE4 편집된 세포는 최대 3개의 편집이 이루어진 경우 높은 수율을 나타내었다(도 21). 대조적으로, spCas9 편집된 세포는 다중 편집의 수가 증가할 때 감소된 수율을 보였다. 따라서, 다중 염기 편집된 세포는 최대 3개의 편집이 이루어진 경우에도 높은 세포 확장을 유지하였다. 따라서, BE4는 spCas9 처리된 샘플과 비교하여 높은 세포 확장을 유지하면서 검출 가능한 게놈 재배열이 없이 다중-편집된 T 세포를 생성하였다.

[2136] **실시예 10: BE4는 spCas9와 유사한 표적에 대한 편집 효율 및 세포 표현형을 갖는 삼중-편집된 T 세포를 생성한다**

[2137] T 세포를 B2M, TRAC 및 PD1의 특정 부위를 표적으로 하는 sgRNA와 함께 BE4 염기 편집기를 암호화하는 mRNA로 공동 형질감염시켰다. 시퀀싱 데이터에 의해 나타낸 바와 같이, 염기 편집은 3개의 부위 모두에서 세포를 변형시키는데 효율적이었다(도 22). 염기 편집에 의한 유전자의 변형은 spCas9 뉴클레이스를 사용한 것과 유사하였다. 유세포 분석은 또한 B2M 및 CD3의 감소된 표면 발현을 나타내었다(도 23, 상부 패널). 전기천공법 단독 대조군 세포와 비교하여, BE4 및 Cas9 다중 편집된 세포는 세포 표면에서 B2M 및 CD3 단백질의 상당한 감소를 나타내었다(95% 미만의 $CD3^-/B2M^-$). PD1 염색이 덜 효율적이지만, PD1의 상당한 감소(대략 90%)가 전기천공법 단독 대조군 세포와 비교하여 BE4 및 Cas9 다중 편집된 세포에서 관찰되었다(도 23, 하부 패널).

[2138] **실시예 11: BE4 편집은 CAR 발현 또는 항원-의존적 세포 살해를 변경하지 않는다**

[2139] T 세포를 B2M, TRAC 및 PD1의 특정 부위를 표적으로 하는 sgRNA와 함께 BE4 염기 편집기를 암호화하는 mRNA로 공동 형질감염시켰다. BCMA를 표적으로 하는 키메라 항원 수용체(CAR)를 항-BCMA CAR을 암호화하는 렌티바이러스 벡터의 통합에 의해 도입하였다. 렌티바이러스 벡터를 받지 않은 처리되지 않은 세포와 비교하여, CAR 발현이 BE4 및 Cas9 편집된 세포에서 유세포 분석에 의해 관찰되었다(도 24). BCMA를 발현하는 세포의 핵 염색 및 세포 사멸을 나타내는 핵 염색의 손실을 검출함으로써 CAR-T 세포의 세포 살해에 대해 평가하였다. 항원 의존성 세포 살해가 BE4 및 Cas9 편집된 T 세포를 포함하여 벡터로 형질도입되고 CAR을 발현하는 세포에서 관찰되었다(도 25). 대조적으로, 벡터로 형질도입되지 않은 처리되지 않은 세포는 세포 살해 활성을 나타내지 않았다. 따라서, BE4-생성 CAR-T 세포는 뉴클레이스만 있는 대응물과 비교하여 유사한 유전자 파괴, 세포 표현형 및 항원-의존적 세포 살해를 보여주었다.

[2140] **실시예 12: Cas12b 및 BE4는 T 세포에서 고효율 다중 편집을 위해 쌍을 이룰 수 있다**

[2141] BE4만을 사용하거나 또는 BE4와 Cas12b를 사용하여 $CD3^-$, $B2M^-$ T 세포를 생성하였다. BE4만을 사용하여 생성된 T 세포의 경우, T 세포를 B2M 및 TRAC의 특정 부위를 표적으로 하는 sgRNA와 함께 BE4 염기 편집기를 암호화하는 mRNA로 공동 형질감염시켰다. BE4와 Cas12b를 사용하여 생성된 T 세포의 경우, T 세포를 BE4 염기 편집기를

암호화하는 mRNA, 및 B2M의 특정 부위를 표적으로 하는 sgRNA, BhCas12b(V4)를 암호화하는 mRNA, 및 TRAC 유전자를 파괴하는데 사용되는 TRAC 유전자의 엑손 4를 표적으로 하는 Cas12b sgRNA로 공동 형질감염시켰다. 형광 보조 세포 분류(FACS) 분석을 사용하여 생성된 T 세포를 평가하여 B2M 및 CD3 세포 표면 발현을 검출하였다. BE4만을 사용한 넉아웃은 BE4와 Cas12b를 사용한 것과 유사한 프로파일을 나타내었다. 특히, 높은 백분율의 T 세포가 $CD3^-$, $B2M^-$: 86%(BE4 단독) 및 88%(BE4+Cas12b)였고, 다른 가능한 표현형은 $CD3^-$, $B2M^+$; $CD3^+$, $B2M^+$ T 세포였으며; 그리고 $CD3^+$, $B2M^-$는 세포 집단에서 더 적게 나타났다(도 26). 대조적으로, 전기천공법 단독 대조군은 높은 백분율(97.8%)의 $CD3^+$ $B2M^+$ 세포 및 매우 낮은 백분율의 $CD3^-$, $B2M^-$ 세포를 갖는 집단을 나타내었다.

[2142] Cas12b를 사용하여 $CD3^-$, $CAR^+$ T 세포를 생성하였다. T 세포를 BhCas12b(V4)를 암호화하는 mRNA, TRAC 유전자의 엑손 4를 표적으로 하는 Cas12b sgRNA 및 항-BCMA CAR을 암호화하는 이중-가닥 DNA(dsDNA) 공여체 주형으로 공동 형질감염시켰다. 형광 보조 세포 분류(FACS) 분석을 사용하여 T 세포를 평가하여 CD3 및 BCMA 세포 표면 발현을 검출하였다. 증가하는 양의 Cas12b가 sgRNA의 존재하에 세포에 도입되었을 때, $CD3^-$ 사분면으로의 세포 집단의 이동에 의해 보여지는 바와 같이 CD3 발현이 감소되었다(도 27). 동일한 조건하에서 세포에 공여체 주형의 양을 증가시키고 도입하였을 때, 세포 집단에서 $CD3^-$, $CAR^+$ 사분면으로의 이동이 관찰되었다.

[2143] 따라서, Cas12b는 BE4와 쌍을 이루어 다중-편집된 T 세포를 생성하여, 다중 이중-가닥 파손에 의해 유발되는 게놈 재배열을 최소화할 수 있다.

[2144] **실시예 13: 동물 모델에서의 생체내 효능**

[2145] 생체내에서 본 명세서에 기재된 화합물의 항-종양 활성을 결정하기 위해, 화합물을, 예를 들어, 단독으로(IP, IV, PO, 흡입 또는 비강 경로에 의해) 또는 관련 화학요법(예를 들어, 사이클로포스파마이드, 독소루비신, 에토포시드)의 최적 이하의 용량과 조합하여 제공하였다. 일례에서, 안정적으로 루시퍼레이스를 발현하는 $5\times10^6$개의 RS4;11 세포(급성 림프아구성 백혈병이 있는 대상체의 골수로부터 확립됨)를 전신 방사선 조사를 받은 지 3시간(hrs) 후 NOD-SCID 마우스의 꼬리 정맥으로 주사하였다. 치료하지 않고 그대로 두면, 이러한 형태의 백혈병은 이 모델에서 3주 내에 죽음을 초래한다. 백혈병은, 예를 들어, 마우스에 D-루시페린(60 mg/kg)을 주사하고 마취된 동물을 영상화(예를 들어, Xenogen 생체내 이미징 시스템, 캘리퍼 라이프 사이언시스(Caliper Life Sciences), 매사추세츠주 홉킨턴 소재)함으로써 쉽게 모니터링된다. 전신 생물발광을 Living Image 소프트웨어(캘리퍼 라이프 사이언시스, 매사추세츠주 홉킨턴 소재)에 의한 광자 플럭스(광자/초)의 통합에 의해 정량화하였다. 단독으로 또는 최적 이하의 용량의 관련 화학치료제와 조합하여 본 명세서에 기재된 화합물을, 예를 들어, 7일 내지 21일 동안 0.1 mg/kg 내지 50 mg/kg 범위의 용량으로 꼬리 정맥 또는 IP 경로에 의해 백혈병 마우스(주사 후 제10일/실험 제1일, 14 내지 16의 생물발광 범위)에 투여하였다. 선택적으로, 격일로 실험 전체에 걸쳐 마우스를 이미지화하고, 실험 기간 동안 매일 생존을 모니터링하였다. 죽은 마우스는 선택적으로 실험이 끝날 때 부검하였다. 또 다른 동물 모델은 루시퍼레이스를 안정적으로 발현하는 인간 여포성 림프종으로부터 유래되는 세포인 DoHH2를 NOD-SCID 마우스에 이식한 것이다. 이러한 생체내 테스트는 선택적으로 예비 약동학, 약력학 및 독성학 데이터를 생성한다.

[2146] **실시예 14: 임상 시험**

[2147] 인간의 치료를 위한 본 명세서에 기재된 화합물의 적합성을 결정하기 위해, 임상 시험을 수행하였다. 예를 들어, 액체암으로 진단되고 치료를 필요로 하는 대상체가 치료 그룹 및 하나 이상의 대조군 그룹에서 선택 및 분리될 수 있되, 치료 그룹에는 본 명세서에 기재된 화합물이 투여되는 반면, 대조군 그룹에는 플라시보 또는 공지의 항암제가 투여된다. 따라서, 본 명세서에 기재된 화합물의 치료 안전성 및 효능은 생존 및 삶의 질과 같은 인자와 관련하여 대상체 그룹의 비교를 수행함으로써 평가될 수 있다. 이 실시예에서, 본 명세서에 기재된 화합물로 처리된 대상체 그룹은 플라시보로 처리된 대상체 대조군 그룹과 비교하여 개선된 장기 생존을 나타낼 수 있다.

[2148] ***연구 설계***

[2149] 이 연구는 WT p53 단백질을 발현하는 진행성 액체암(예를 들어, 백혈병, 골수종 또는 액체 림프종)이 있는 대상체에서 28-일 또는 21-일 주기의 두 가지의 상이한 투여 양생법을 사용하여 IV 주입에 의해 투여된 화합물의 안전성, 내약성, PK, PD 및 항-액체 암세포 효과를 평가하도록 설계된 용량-증량, 2개군(2-arm) 연구를 포함한다. 예를 들어, 본 명세서에 기재된 화합물이 재발성/불응성 급성 골수성 백혈병(AML) 및/또는 급성 림프구성 백혈

병(ALL)이 있는 대상체에 사용될 수 있다. 대상체에는 28-일 주기의 경우 연속 3주 동안 매주 1회 또는 21-일 주기의 경우 연속 2주 동안 매주 2회 본 명세서에 기재된 화합물이 투여된다. 액체 림프종이 있는 많은 대상체가 말초 혈액에 순환 종양 세포(CTC)를 가지고 있으며, 이는 유세포 분석을 사용하여 검출 및 분석될 수 있다. 따라서, 이들 세포에서 연구 약물-특이적 표적 결합의 검출이 가능하다.

[2150] 연구는 용량 증량 단계(Dose Escalation Phase: DEP) 및 확장 단계(Expansion Phase: EXP)로 구성된다. DEP는 본 명세서에 기재된 화합물의 MTD 또는 OBD를 확립하기 위한 "3+3" 용량 증량 설계이다. EXP는 임상 안전성 프로파일 및 용량 수준의 잠재적 효능을 추가로 조사하기 위해 MTD 또는 OBD에서 특정 액체암이 있는 대상체의 최대 2개의 개별 그룹을 등록한다. EXP에 대한 대상체의 선택은 DEP의 결과뿐만 아니라 추가적인 비임상 약리학 연구로부터의 데이터를 기반으로 최종 결정된다. 후자는 액체암 유형 전반에 걸쳐 그리고 액체암 유형 내에서 본 명세서에 기재된 화합물에 대한 세포주의 감수성의 비교를 용이하게 하는 여러 액체 암세포주(예를 들어, 백혈병, 액체 림프종, 골수종 등)의 조사를 포함한다.

[2151] 연구의 용량 증량 및 용량 확장 단계에서 대상체의 치료는 질환 진행의 문서화, 허용할 수 없는 독성, 또는 대상체 또는 의사가 요법을 중단하기로 결정할 때까지 계속된다.

[2152] MTD 또는 OBD가 확립된 후, 이 용량 수준 및 특히 대상체 또는 액체 암세포 유형에서 추가 경험을 얻기 위해 추가적인 대상체가 최대 2개의 개별 확장 코호트(확장 코호트당 대략 30명의 대상체)에 등록될 수 있다. 대상체 또는 액체 암세포 유형의 선택은 연구의 용량 증량 부분에서 이루어진 관찰에 기초하여 부분적으로 결정된다.

[2153] 안전성은 이상 반응(Adverse Event: AE)의 발생률, 중증도, 기간, 인과관계, 심각도 및 유형 및 대상체의 신체 검사, 바이탈 사인 및 임상 실험실 결과의 변화를 기반으로 평가된다. 연구원은 NCI CTCAE 버전 4.0을 사용하여 AE의 중증도를 평가한다.

[2154] 이 연구의 1차 목적이 안전성 및 PK를 기반으로 하기 때문에, 통계적 분석은 본질적으로 기술적이며, 완전 반응(CR) 또는 부분 반응(PR)을 달성한 대상체 또는 IWG(2014) 기준에 기초하여 안정 병변(SD)을 유지하는 대상체를 포함하여 연구된 모든 용량 및 모든 관찰된 반응을 설명한다. 본 명세서에 기재된 화합물의 적어도 1회 용량을 투여 받은 대상체는 안전성 집단을 구성하며, 이는 모든 안전성 분석에 포함된다. 본 명세서에 기재된 화합물의 적어도 하나의 주기를 완료하고 치료 후 객관적인 질환 평가를 받는 대상체는 효능-평가 가능한 대상체 집단을 구성한다.

[2155] **실시예 15: 유전자를 확인하고 침묵시키는 예시적인 방법**

[2156] 이 실시예에서, CAR-T에 대한 표적의 유전자 침묵을 위한 방법이 제공된다. 도 28a는 표적화될 유전자의 다이어그램 및 이에 대한 고려사항을 제공한다. 표적은, 예를 들어, CD4, CD5, CD7, CS-1, TRAC, CAR, FAS, LAG-3, PD-1, β2M, CD52 및 TCR 중 하나 이상을 포함한다.

[2157] 염기 편집기를 사용하여 침묵시키는 두 가지 전략이 있다. 첫 번째 전략은 표적 단백질을 평가한 다음, 글루타민, 아르기닌 또는 트립토판 중 하나 이상을 종결 코돈으로 변경하여 단백질 발현을 침묵시킨다(도 28b의 항목 1 참조). 두 번째 전략은 단백질 발현을 침묵시키기 위한 CBE를 이용한 스플라이스 파괴(도 28b의 항목 2 참조)를 포함하였다.

[2158] **실시예 16: 다수의 CAR-T 집단이 AML 또는 T-ALL에서 질환의 클론성을 다루기 위해 사용된다**

[2159] 백혈병은 종양 미세환경에서의 선택압(selection pressure)으로 인한 클론 확장 및 선택의 반복적인 과정을 통한 질환의 진화로 인해 소정의 요법에 비추어 저항성 또는 재발이 될 것으로 상정되었다. 예로서, AML 암세포의 상이한 클론은 CD33+ CD123-; CD33+CD123+ 또는 CD33-CD123+와 같은 다양한 항원을 나타낼 수 있다. CD33+ 암세포를 표적화하는 치료제는 CD33-CD123+와 같은 이러한 특정 항원이 결여되어 있는 세포를 해결하지 않을 것이다. 따라서, 본 명세서에 기재된 다수의 CAR-T 세포 집단은 질환의 클론성을 해결하는 것으로 상정된다.

[2160] 예시적인 방법에서, CD33 및 CD123 CAR-T 세포를 도 29에 예시된 바와 같이 생성하였다. CD33 및 CD123 CAR-T 세포를 CD33+CD123+ AML에 대한 마커를 갖는 AML 환자에게 투여하였다. CD33 및 CD123 CAR-T 세포는 CD33+ CD123-, CD33+CD123+ 및 CD33-CD123+인 환자에서 AML 암세포를 제거하였다.

[2161] 또 다른 예시적인 방법에서, CD3 및 CD7 CAR-T 세포를 본 명세서에 기재된 방법에 의해 도 30에 예시된 바와 같이 생성하였다. CD3 및 CD7 CAR-T 세포를 CD3+CD7+ T-ALL에 대한 마커를 갖는 T-ALL 환자에게 투여하였다. CD3 및 CD7 CAR-T 세포는 CD3+ CD7-, CD3+CD7+ 및 CD3-CD7+인 T-ALL 암세포를 제거하였다.

[2162] **실시예 17: 동족 살해를 제거하기 위해 다수의 CAR-T 집단이 변형된다**

[2163] 전통적인 T 세포 및 CAR-T 요법의 주요 한계는 활성화된 세포가 세포 제조, 보관 동안 또는 환자에게 주입될 때 자기-살해 또는 동족 살해를 겪을 수 있다는 것이다. 동족 살해를 피하기 위해, 본 명세서에 기재된 CAR-T 세포는 세포의 CAR이 특이적인 항원을 암호화하는 폴리뉴클레오타이드에 돌연변이를 포함하도록 변형된다. 예로서, CD7 CAR-T 세포는 동족 살해를 방지하기 위해 CD7 폴리뉴클레오타이드에 돌연변이를 포함하도록 변형된다. 또 다른 예로서, CD3 CAR-T 세포는 동족 살해를 방지하기 위해 CD3 폴리뉴클레오타이드에 돌연변이를 포함하도록 변형된다. 다수의 CAR-T 집단에서, 동족간 살해(즉, 두 CAR-T 집단 모두에서 CD3 및 CD7 폴리뉴클레오타이드 둘 다의 돌연변이)를 방지하기 위해 추가적인 변형이 수행된다. 추가적인 변형은 면역억제, 이식편 대 숙주, 숙주 대 이식편 등에 관련될 수 있는 유전자에 대한 돌연변이를 포함한다.

[2164] 예시적인 방법에서, AML의 치료를 위한 CD33 및 CD123 CAR-T 세포를 도 29에 예시된 바와 같이 생성하였다. 두 CD33 및 CD123 CAR-T 세포는 모두 CD33에 돌연변이를 갖도록 변형되어 세포가 자신을 표적으로 삼지 않는다. PD-1, CD52 및 TRAC의 추가적인 돌연변이는 면역억제, 숙주 대 이식편 및 이식편 대 숙주를 각각 감소시키거나 또는 방지한다.

[2165] 예시적인 방법에서, T-ALL의 치료를 위한 CD3 및 CD7 CAR-T 세포를 도 30에 예시된 바와 같이 생성하였다. 두 CD3 및 CD7 CAR-T 세포는 모두 CD3(TRAC) 및 CD7에 돌연변이를 갖도록 변형되어 세포가 자신 또는 다른 집단을 표적으로 삼지 않는다. PD-1, CD52 및 TRAC의 추가적인 돌연변이는 면역억제, 숙주 대 이식편 및 이식편 대 숙주를 각각 감소시키거나 또는 방지한다.

[2166] **실시예 18: T-ALL CAR-T 세포에 대한 4개의 동시 염기 편집은 뉴클레이스와 비교하여 수율에 영향을 미치지 않는다**

[2167] 쿼드-T-ALL 편집의 맥락에서 이론적인 수율은 다음 성분이 테스트되는 검정에서 결정된다. 대조군은 전기천공법 없음(EP 없음), EP 단독, CBE 변이체 1, CBE 변이체 2 및 Cas9이다. 결과는 도 31a에 나타나 있으며, 상응하는 생존능력 결과는 도 31b에 나타나 있다. T-ALL CAR-T에 대한 4개의 동시 염기 편집은 뉴클레이스와 비교하여 수율에 영향을 미치지 않는 것으로 결정되었다.

[2168] **실시예 19: 확장된 론자 전기천공법은 T-ALL CAR-T 세포 프로그램에 대한 고효율의 다중 T 세포 염기 편집을 유도한다.**

[2169] T 세포를 활성화한 후 3일 동안 배양하였다. $5\times10^6$개의 세포를 편집기(CBE 1, CBE 2 또는 spCas9)를 위한 10㎍의 mRNA 및 T-ALL 가이드(TRAC, CD52, CD7, PD1)를 위한 5㎍의 sgRNA로 전기천공시켰다. 적합한 가이드는 아래 표 23에 제공되어 있다.

*표 23*

**T\ALL CAR\T 에 대한 가이드 RNA**

| 가이드 RNA (gRNA) | 가이드 핵산 서열 |
|---|---|
| TRAC gRNA | ususcsGUAUCUGUAAAACCAAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu |
| CD52 gRNA | ususasCCUGUACCAUAACCAGGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu |
| CD52 gRNA | csuscsUUACCUGUACCAUAACCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu |
| CD7 gRNA | usgscsACCUCUGGGGAGGACCUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu |
| CD7 gRNA | cscsusACCUGUCACCAGGACCAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu |
| PD1 gRNA | csascsCUACCUAAGAACCAUCCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUsususu |

[2170]

[2171] 2'-O-메틸 유사체 및 3' 포스포로티오에이트 뉴클레오타이드간 연결이 처음 3개의 5' 및 3' 말단 RNA 잔기에서 사용되며, 각 말단의 화학적 변형은 뉴클레오타이드 다음에 "s"로 표시된다.

[2172] CBE 1은 래트-APOBEC-BE4를 지칭하고, CBE2는 오랑우탄 "BE4"(ppAPOBEC)를 지칭한다. EP 없음 및 EP 단독은 단지 편집기/가이드 없이 zap이 없거나 또는 zap이 있는 대조군이었다.

[2173] 이론적 수율은 다음과 같이 측정하였다: AO/DAPI 염색을 사용한 생존능력 측정으로 세포를 매일 계수하였다(또한 EP 생존능력 판독 후 시간을 들여 카운트-AML 수치로 수행된 생존능력 판독과 함께). 세포를 매일 $0.5\times10^6$개의 세포/㎖로 다시 스플릿하였다(이론적으로 스플릿되고 실제로 모든 세포를 확장하지 않았기 때문에). 그런 다음, 배양물이 역스플릿되지 않았다면 이론적 수율이 어떠했을지 결정하기 위한 계산을 하였다. 유전자 변형의 백분율을 결정하였다(도 32).

[2174] NGS를 통해 세포를 취하고, 이론적 세포 수율을 결정하였다(도 33).

[2175] 모델 종양 세포에서 형광 마커의 손실에 의해 입증된 세포 살해(도 34).

[2176] 테스트된 모든 CAR은 CD8a 힌지, CD8 막관통 및 CD28z 신호전달 도메인이 뒤따르는 중쇄에서 경쇄(H2L) 또는 경쇄에서 중쇄(L2H) 배향의 결합제 클론을 포함한다.

[2177] **실시예 20: 예시적인 CAR-T 세포 처리 흐름 및 위탁 제조**

[2178] 건강한 공여자로부터의 동결된 성분채집 류코팩(leukopak)을 얻고, 제조사의 지침에 따라 Pasmatherm을 사용하여 해동시켰다. 예를 들어, 제조사의 지침에 따라 CliniMACS Prodigy를 사용하여 CD4 및 CD8 T-세포를 단리하였다. 그런 다음, 제조사의 지침에 따라 TransAct를 사용하여 T 세포를 활성화하였다. 제조사의 지침에 따라 Lonza 4D Nucleofactor를 사용하여 mRNA/RNP 염기 편집기의 전기천공 전달을 일으켜 CD3, CD7, CD52 및/또는 PD1을 넉아웃시켰다. 당업계에 인정된 기법을 사용하여 TCR α/β의 고갈을 수행하였다. T-세포가 확장되고, 3CAR 및 7CAR의 렌티바이러스 형질도입이 발생하였다. CAR-T 세포를 자동 온도조절 동결기(control rate freezers: CRF)로 동결보존하였다.

[2179] **실시예 21: CAR-T 세포를 이용한 T-ALL의 치료**

[2180] T-세포 급성 림프아구성 백혈병(T-ALL)은 골수의 공격성 악성 신생물이다. 모든 사례의 약 12% 내지 15%가 소아에서 진단된다. 재발 후 5년 생존율은 25% 미만이다. 치료 표준은 화학요법을 사용하여 두 번째 관해(remission)를 유도한 후 동종이계 조혈 줄기 세포 이식(alloHSCT)을 하는 것이다. 그러나, 많은 환자가 화학요법에 불응성이거나 또는 높은 종양 부담을 가지며, alloHSCT의 가교로서 깊은 관해(deep remission)를 유도할 수 없다. 또한, 사전 치료를 많이 받은 환자는 종종 자가 CAR-T 치료의 대상이 아니다. 따라서, T-ALL의 치료에는 대체 치료 옵션이 필요하다.

[2181] alloHSCT를 이용한 치료를 가교하기 위해 편집된 CAR-T 세포가 환자에서 깊은 관해를 유도할 수 있는지 여부를 조사하기 위해, 다중 편집을 사용하여 동족 살해 저항성 CAR-T 세포를 생성하였다. 도 35에 나타낸 바와 같이, $CD3^-$ $CD7^+$ 또는 $CD3^+$ $CD7^+$ T-ALL 종양 세포를 표적화하기 위해 CD7 키메라 항원 수용체를 발현하는 T 세포를 CD7, TRAC, CD52 및 PD-1의 발현을 감소시키거나 또는 제거하도록 편집하였다. 도 36에 나타낸 바와 같이, -7일에 약 20명 내지 30명의 T-ALL 환자에서 Cy/Flu/캄파스로 림프제거를 수행하였다. 제0일에, 환자에게 CD7 CAR-T 세포를 주입하였다. 그런 다음, 환자를 제65일에 Cy/Flu/TBI/ATG로 사전 적응시켰다. 제70일에, 환자를 alloHSCT로 치료하였다. CD7 CAR-T 세포를 사용한 치료는 T 세포 무형성(aplasia)을 유도하고, T-ALL 환자는 가교 치료 없이 alloHSCT로 치료될 수 있을 것으로 예상된다.

[2182] **실시예 22: CD7 CAR-T 세포의 생산**

[2183] TALL017 CD7 CAR-T 세포를 도 37a에 나타낸 바와 같이 생산하기 위해, 건강한 공여자로부터 동결된 성분채집을 받았다. 그런 다음, 성분채집을 Plasmatherm(바키 게엠베하 앤 씨오. 케이지(Barkey GmbH & Co. KG))로 해동시켰다. CliniMACS Prodigy(밀텐이 바이오텍(Miltenyi Biotec))를 사용하여 해동된 성분채집으로부터 CD4 및 CD8 T-세포를 단리하였다. 제2일에, Lonza SD Nucleofector를 사용하여 전기천공법에 의해 mRNA 염기 편집기 및 가이드 RNA를 T-세포에 전달하였다. 염기 편집 후, 렌티바이러스 형질도입(7CAR8)을 통해 CD7 키메라 항원 수용체(CAR)를 T-세포에 전달하였다. 그런 다음, 2일 내지 8일간 배양에서 T 세포를 확장시켰다. 확장 후, 수확된

CAR-T 세포(>1B)를 동결보존하였다. 해동시, TALL017 처리는 비균질한 세포 집단을 생성하였다.

[2184] 형질도입되지 않은(UTD) 및 형질도입된 TALL017 세포를 해동 후 24시간에 비교하였다. CD2+ 및 CD4+ 세포에서 세포를 게이팅함으로써 유세포 분석을 사용하여 CAR 활성화를 확인하였다(도 37c). 도 37c에 나타낸 바와 같이, 항-CD7 CAR에 의해 유도된 중간-과정 동족 살해는 TALL017 동결보존 시점에 여전히 활성화된 7CAR-T를 생성하였다. 활성화 중간에 동결보존된 7CAR-T는 제대로 해동되지 않았다.

[2185] TALL038 CD7 CAR-T 세포를 도 37b에 나타낸 바와 같이 생산하기 위해, 건강한 공여자로부터 동결 성분채집을 받았다. 그런 다음, 성분채집을 Plasmatherm(바키 게엠베하 앤 씨오. 케이지)으로 해동시켰다. CliniMACS Prodigy(밀텐이 바이오텍)를 사용하여 해동된 성분채집으로부터 CD4 및 CD8 T-세포를 단리하였다. 제2일에, Lonza SD Nucleofector를 사용하여 전기천공법에 의해 mRNA 염기 편집기 및 가이드 RNA를 T-세포에 전달하였다. 염기 편집 후, 렌티바이러스 형질도입(7CAR8)을 통해 CD7 키메라 항원 수용체(CAR)를 T-세포에 전달하였다. 그런 다음, T-세포를 2일 내지 10일간 배양에서 확장시켰다. 확장 후, TCRα/β+ 세포를 고갈시켰다. 그런 다음, T-세포를 10일 내지 12일간 배양에서 추가로 확장시켰다. 그런 다음, 제12일에 수확된 CAR-T 세포(>2B)를 동결보존하였다. 해동시, TALL038 처리는 균질한, 단분산 세포 집단을 생성하였다.

[2186] **실시예 23: TALL017 및 TALL038 CD7 CAR-T 세포의 평가**

[2187] TALL017 및 TALL038 CD7 CAR-T 세포의 편집 효율을 차세대 시퀀싱(NGS)을 사용하여 비교하였다. 도 38에 나타낸 바와 같이, 두 유형의 CD7 CAR-T 세포에서 CD52, CD7, TRAC 및 PD-1의 편집은 80% 내지 99%의 효율 범위였다. 형질도입되지 않은(UTD) 세포를 대조군으로 사용하였다.

[2188] TCRα/β, CD52 및 CD7의 해동 후 24시간 단백질 발현을 TALL017 및 TALL038 CD7 CAR-T 세포에서 조사하였다(도 39a 및 도 39b). TALL038 세포는 0.4% 미만의 TCRα/β+ 세포를 보여주었다. 단백질 발현 데이터는 TALL017 및 TALL038 CD7 CAR-T 세포 둘 다에 대한 사전-동결 메트릭스(metrics) 및 해동 후 NGS 데이터와 일치하였다. 형질도입되지 않은(UTD) 세포를 대조군으로 사용하였다.

[2189] CD7 CAR-T 세포 집단에서 세포의 정체를 평가하기 위해, 도 40에 나타낸 바와 같은 게이팅 방식을 사용하였다. 세포를 다음 순서로 게이팅하였다: FSC/SSC, 단일항, 살아있는/활성이 없는, CD45(모든 백혈구), CD19-(B 세포 제외), CD56-/CD14-(NK/단핵구 제외), CD2+(T 세포) 및 CD2+의 CD4:CD8. 도 41에 나타낸 바와 같이, TALL017 및 TALL038 CD7 CAR-T 세포 집단의 정체는 CD2+/- 및 CD56+/- 세포의 혼합이었다. 검출 가능한 단세포 또는 B 세포는 없었다.

[2190] 마지막으로, CAR 발현을 평가하였다. 도 42에 나타낸 바와 같이, TALL017 및 TALL038 CD7 CAR-T 세포 집단 모두 해동 후 48시간까지 CAR+ 세포 발현을 유지하였다. 형질도입되지 않은(UTD) 세포를 대조군으로 사용하였다.

[2191] CD25 및 CD69 발현을 또한 TALL017 및 TALL038 CD7 CAR-T 세포에서 해동 후 24시간 및 48시간에 평가하였다. 도 43에 나타낸 바와 같이, 해동 후 CD25 발현은 TALL017 세포와 비교하여 TALL038 세포에서 더 낮았다. CD25는 중기 T 세포 활성화의 마커이다. 대부분의 CAR과 달리, 항-CD7 CAR은 7CART를 추가로 활성화하는 중기-배양(mid-culture) 동족 살해를 유도한다. 활성화된 상태에서 TALL017 세포를 동결하면 세포 건강에 해로울 수 있다. TALL038 세포는 TALL017에 비해 CD25 발현이 극적으로 낮으며, 이는 동결에 대한 더 높은 저항성을 시사할 수 있다.

[2192] **실시예 24: TALL038 CD7 CAR-T 세포의 시험관내 특성화**

[2193] 시험관내에서 TALL038 CD7 CAR-T 세포를 특성화하기 위해, 비드 기반 효능 방법을 개발하였다(도 44). 제0일에, CD7을 Dynabead 및 CD7 CAR-T 세포와 커플링하고, 형질도입되지 않은(UTD) 세포를 해동하였다. 제1일에, CD7 비드를 정제하고, CD7 CAR-T 세포를 공동 배양하고, 플레이팅하고, 37℃에서 인큐베이션하였다. 제2일에, 상청액을 수집하고, 효소-결합 렉틴 검정(ELLA)을 사용하여 판독값을 얻었다.

[2194] 이 검정을 사용하여, TALL038 CD7 CAR-T 세포를 IFNɣ, TNFα, IL-10 및 IL-2 방출에 대해 평가하였다. CD7 항원에 대한 반응으로, TALL038 CD7 CAR-T 세포는 IFNɣ, TNFα 및 IL-2를 방출하였다(도 45a 및 도 45b). 형질도입되지 않은(UTD) 세포를 대조군으로 사용하였다.

[2195] 시험관내에서 TALL038 CD7 CAR-T 세포를 추가로 특성화하기 위해, 표적 종양 세포의 CAR 지시된 T-세포 살해를 측정하기 위한 공동 배양 방법을 개발하였다. 도 46은 검정의 개요를 제공한다. 먼저, 종양 세포를 시딩하였다. 그런 다음, CAR-T 세포를 시딩된 종양 세포에 첨가하였다. 이어서, 세포를 IncuCyte® 스캐닝 시스템을 사용하

여 분석하였다. 마지막으로, IncuCyte® 소프트웨어를 사용하여 판독값을 얻었다.

[2196] TALL038 CD7 CAR-T 세포를 시딩된 CCRF-CEM GFP+ 종양 세포에 첨가하였다. 표 24의 T-세포 로트를 4:1 CAR:CCRF, 2:1 CAR:CCRF 또는 1:1 CAR:CCRF의 E:T 비로 테스트하였다:

*표 24*

**T\세포 로트**

[2197]

| 로트 | CAR+% |
|---|---|
| TALL017 UTD | 0% CAR+ |
| TALL017 7CAR8 | 약 50% CAR+ |
| TALL038 UTD | 0% CAR+ |
| TALL038 7CAR8 | 약 50% CAR+ |

[2198] 도 47a 및 도 47b에 나타낸 바와 같이, TALL038 CD7 CAR-T 세포는 TALL017 CAR T 세포와 비교하여 재처리하였을 때 증가된 CCRF를 나타내었다.

[2199] **실시예 25: TALL038 CD7 CAR-T 세포의 생체내 특성화**

[2200] 이전 연구는 CCRF 모델이 CD7-지시된 CAR-T 요법에 대한 반응을 나타내었음을 보여주었다(예를 들어, 문헌 [Gomes-Silva D., *et al.*, CD7-edited T cells expressing a CD7-specific CAR for the therapy of T-cell malignancies. Blood. 2017;130:285-296] 참조).

[2201] 생체내 TALL038 CD7 CAR-T 세포를 특성화하기 위해, T-ALL의 CCRF 모델에서 생체내 성장에 대한 TALL038 약물 제품의 효능을 평가하기 위한 연구를 개발하였다. 표 25는 연구 프로토콜을 요약한다.

*표 25*

**생체내 연구에서 CCRF 모델**

[2202]

| 그룹 | 종양 | 마우스당 CCRF\GFP+ 세포 | 그룹당 마우스 수 | 처리 | 마우스당 (T 세포) 세포* | 경로 | 조직 수집 |
|---|---|---|---|---|---|---|---|
| 1 | CCRF GFP+ | 1e5 | 10 | DPBS | 없음 | IV | 혈액 |
| 2 | CCRF GFP+ | 1e5 | 10 | UTD CAR | 1.00E+07 | IV | 혈액 |
| 3 | CCRF GFP+ | 1e5 | 10 | 7CAR8 | 1.00E+07 | IV | 혈액 |
| 4 | CCRF GFP+ | 1e5 | 10 | 7CAR8 | 3.00E+06 | IV | 혈액 |
| 5 | CCRF GFP+ | 1e5 | 10 | 7CAR8 | 1.00E+06 | IV | 혈액 |
| 6 | CCRF GFP+ | 1e5 | 10 | 무작위화 중 이상값으로 제외됨 | | | |

[2203] 10마리의 6주 내지 7주령 암컷 NSG 마우스(잭슨 랩스(Jackson Labs): 스톡 005557)에 CCRF-GFP+ 세포(1e5)를 정맥내(IV) 이식하였다. 제11일에, 마우스에 총 250㎕의 TALL038 1e7세포/마우스를 IV를 통해 이식하였다. 종양 부담을 모니터링하기 위해, IVIS 이미징 측정을 2주 내지 3주 동안 주사 후 3회/주로 수행하였다. 전혈을 2주 내지 3주 동안 1회/주로 수집하였다. 이식 후 3회/주의 체중 및 임상 관찰을 하였다.

[2204] 도 48은 CCRF 이식 후 제10일/-1일에서 CD7 CAR-T 처리까지의 생물발광 광 데이터(평균, SEM)를 보여준다. 이 시점에서, 마우스는 CD7 CAR-T 치료가 필요한 측정 가능한 종양 부담을 나타내지 않았다. 도 49a 및 도 49b는 CCRF 이식 후 제19일/7CAR8 처리 후 제8일의 생물발광 광 데이터(평균, SEM)를 보여준다. 이 시점에서, 용량-의존적 종양 성장 저해가 CD7 CAR-T 처리에 대한 반응으로 관찰되었다. 마우스는 CD7 CAR-T 처리에 대한 반응으로 체중이 감소하지 않았다. 도 50a 및 도 50b는 CCRF 이식 후 제38일/CD7 CAR-T 처리 후 제27일의 생물발광 광 데이터(평균, SEM)를 보여준다. 각 그룹의 첫 번째 마우스가 인도적 종료시점(humane endpoint)(플럭스 > 1e10)에 도달한 날짜 이후의 평균 및 표준 오차는 더 이상 플롯팅하지 않았다. 이 시점에서, 지속된 용량-의존적 저해가 재발 전 각 그룹에서 관찰되었다. 도 50b에서 로그 척도를 사용하여, CD7 CAR-T 그룹의 1e7에서의 재발을 가시화하였다. CCRF 이식 후 제38일/CD7 CAR-T 처리 후 제27일의 생물발광 광 데이터는 도 51a, 도 51b, 도 52a 및 도 52b의 개별 마우스에 대해 나타나 있다. 위의 실험에 대해, UTD 및 비히클을 대조군으로 사용하였다.

[2205] **실시예 26: T-ALL 치료 및 림프종 단독요법을 위한 CD5-표적화 CAR-T**

[2206] 이전 연구는 여러 암 항원을 표적화함으로써 개선된 CAR-T 생성물을 생성할 필요성을 인식하였다(예를 들어, 문헌[Xinjie Xu, *et al.*, Mechanisms of Relapse After CD19 CAR T-Cell Therapy for Acute Lymphoblastic Leukemia and Its Prevention and Treatment Strategies. *Front Immunol*. 2019; 10: 2664, Nov. 12, 2019; E. Mejstríková, *et al.*, CD19-negative relapse of pediatric B-cell precursor acute lymphoblastic leukemia following blinatumomab treatment. *Blood Cancer Journal*, 7: 659 (2017); Hanren Dai, *et al.*, Bispecific CAR-T cells targeting both CD19 and CD22 for therapy of adults with relapsed or refractory B cell acute lymphoblastic leukemia. *J. Hematology & Oncology*. 13:30 (2020); Robbie G. Majzner and Crystal L. Mackall, Tumor Antigen Escape from CAR T-cell Therapy. *Cancer Discov*. 2018 Oct; 8(10):1219-1226] 참조).

[2207] T-ALL, MF/SS, 원발성 피부 T-세포 림프종(NOS), T-세포 대형 과립 림프구성 백혈병, 혈관면역아세포성 T/NK-세포 림프종, 간비장 T-세포 림프종, 원발성 피부 CD30 + LPD, 성인 T-세포 백혈병/림프종 및 T-세포 전림프구성 백혈병을 포함하지만 이들로 제한되지 않는 T-세포 및 NK-세포 악성종양에서 CD5 발현이 확인되었다. CD5와 CD7을 조합하면 T-ALL, MF/SS, 원발성 피부 T-세포 림프종(NOS), 혈관면역아세포성 T/NK-세포 림프종 및 성인 T-세포 백혈병/림프종과 같은 소정의 암에 대한 액체암 요법의 더 큰 가능성을 열 수 있다.

[2208] T 세포에서 CD5의 발현을 감소시키거나 또는 제거하기 위해, 아래 표 26에 제공된 바와 같이 42개의 사이티딘 염기 편집기 단일 가이드 RNA(sgRNA)를 생성하였다(애질런트(Agilent)).

*표 26*

**CD5 가이드 RNA**

| 가이드 명칭 | 가이드 그룹 | 가이드 서열 |
|---|---|---|
| RCM sgRNA 102 | Ex1 SD (Pos 6) | ACUCACCCAGCAUCCCCAGC |
| RCM sgRNA 103 | Ex2 SA (Pos 6) | AGCGACUGCAGAAAGAAGAG |
| RCM sgRNA 104 | Ex2 STOP (pos 5/6) | CAUACCAGCUGAGCCGUCCG |
| RCM sgRNA 105 | Ex3 SA (Pos 8) | UGGAAAUCUGGGGGUCAGAA |
| RCM sgRNA 106 | Ex3 SD (Pos 9) | GUUACCCACCUAAGCAGGUC |
| RCM sgRNA 107 | Ex3 STOP (pos 5) | CUGCCAGCGGCUGAACUGUG |
| RCM sgRNA 108 | Ex3 STOP (Pos 5/6) | CCUCCCACUGCUUGGAGCUC |
| RCM sgRNA 109 | Ex3 STOP (pos 6) | UCUGCCAGCGGCUGAACUGU |
| RCM sgRNA 110 | Ex3 STOP (pos 7) | GUCUGCCAGCGGCUGAACUG |
| RCM sgRNA 111 | Ex3 STOP (Pos 8) | GAAGUGCCAGGGCCAGCUGG |
| RCM sgRNA 112 | Ex3 STOP (Pos 8/9) | CCAUGUGCCAUCCGUCCUUG |
| RCM sgRNA 113 | Ex3 STOP (Pos 9) | UUUGCAGCCAGAGCUGGGGC |
| RCM sgRNA 114 | Ex4 SA (Pos 5) | GGUUCUGCAAUGAGACACUC |
| RCM sgRNA 115 | Ex5 SA (Pos 5) | GAGCUAGGAGAGGAGAGAGC |
| RCM sgRNA 116 | Ex5 SD (Pos 8) | UCACUUACCUGAGCAAAGGA |
| RCM sgRNA 117 | Ex5 SD (Pos 9) | CUCACUUACCUGAGCAAAGG |
| RCM sgRNA 118 | Ex5 STOP (Pos 4) | CUGCAGCUGGUGGCACAGUC |
| RCM sgRNA 119 | Ex5 STOP (Pos 5) | CGGCCAGCACUGUGCCGGCG |
| RCM sgRNA 120 | Ex5 STOP (Pos 6/7) | AUCUUCCAUUGGAUUGGCAA |
| RCM sgRNA 121 | Ex5 STOP (Pos 7/8) | GAUCUUCCAUUGGAUUGGCA |
| RCM sgRNA 122 | Ex6 SA (Pos 4) | AACCUGAGAGGGGAAGCAAU |
| RCM sgRNA 123 | Ex6 SA (Pos 5) | AAACCUGAGAGGGGAAGCAA |
| RCM sgRNA 124 | Ex6 STOP (Pos 4) | GUGCAGAGCCGUCUGGUGGG |
| RCM sgRNA 125 | Ex6 STOP (Pos 4/5) | CUCCCACCGCAGCGAGCUCC |
| RCM sgRNA 126 | Ex6 STOP (Pos 4/5) | CUCCCAAAGUUCGUGGCACU |
| RCM sgRNA 127 | Ex6 STOP (Pos 5) | AGGUGCAGAGCCGUCUGGUG |
| RCM sgRNA 128 | Ex6 STOP (Pos 5) | GGUGCAGAGCCGUCUGGUGG |
| RCM sgRNA 129 | Ex6 STOP (Pos 5/6) | UCUCCCAAAGUUCGUGGCAC |
| RCM sgRNA 130 | Ex6 STOP (Pos 6) | UGUCCCAGUGCCACGAACUU |
| RCM sgRNA 131 | Ex6 STOP (Pos 7) | AAGGUGCAGAGCCGUCUGGU |
| RCM sgRNA 132 | Ex6 STOP (Pos 7) | UCCUAUCGAGUGCUGGACGC |
| RCM sgRNA 133 | Ex6 STOP (Pos 8) | CAAGGUGCAGAGCCGUCUGG |
| RCM sgRNA 134 | Ex6 STOP (Pos 8) | GGUGCGCCAGGGGGCUCAGU |
| RCM sgRNA 135 | Ex6 STOP (Pos 8/9) | GGGCUGCCCACUGAGCCCCC |
| RCM sgRNA 136 | Ex6 STOP (Pos 9) | AGGUGCGCCAGGGGGCUCAG |
| RCM sgRNA 137 | Ex7 STOP (Pos 4) | GGCCAGGAUCCAAACCCCGC |
| RCM sgRNA 138 | Ex8 STOP (Pos 4) | CGCCAGUGGAUUGGCCCAAC |
| RCM sgRNA 139 | Ex8 STOP (Pos 7) | AAGAAGCAGCGCCAGUGGAU |
| RCM sgRNA 140 | Ex9 SA (Pos 8) | AAAGACACUGGGCAGAUGGU |
| RCM sgRNA 141 | Ex9 SD (Pos 6) | GCUUACCUGGAUAAGCUGAC |
| RCM sgRNA 142 | Ex10 SA (Pos 9) | UUCCAGAGCUGGGGAAAGAA |

[2209]

| | | |
|---|---|---|
| RCM sgRNA 143 | Ex10 STOP (Pos 9) | AUGGGGCUCAGAGGCUGUAA |

[2210]

[2211] 유세포 분석, 차세대 시퀀싱(NGS), RNA 및 웨스턴 블롯 분석을 사용하여 전기천공법 72시간 후 최소 n=3의 공여체 및 최소 n=5의 항체 클론에 대해 가이드를 스크리닝함으로써 CD5 편집을 확인하였다(도 53). 전기천공법 단독(EP)을 음성 대조군으로 사용하였다. CD5 가이드 RNA와 함께 사이티딘 염기 편집기를 사용하면 T 세포에서 CD5 발현이 효율적으로 감소되거나 또는 제거되었다.

[2212] BE4와 조합하여 사용될 때 NGS를 통한 염기 편집 효율에 대한 추가 테스트를 위해 2개의 후보 가이드인 sgRNA 103(스플라이스 파괴) 및 sgRNA 104(종결 코돈)를 선택하였다. 두 sgRNA 103 및 sgRNA 104는 모두 약 85%의 편집 효율을 보였다. 아래 표 27은 sgRNA 103 및 sgRNA 104를 사용하여 수행된 인 실리코 OT 분석의 결과를 제공

한다.

*표 27*

인실리코 OT 분석

| 벌지_유형 | | | 없음 | 없음 | 없음 | 없음 | 없음 | 없음 | DNA | DNA | RNA | RNA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 벌지_길이 | | | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| num_mm | | | 0 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 1 | 2 |
| 표적_부위_명칭 | pam | ot_범주 | | | | | | | | | | |
| sgRCM103 | NAA | 엑손 | 0 | 1 | 5 | 46 | 323 | 1902 | 0 | 3 | 0 | 2 |
| sgRCM103 | NAA | 없음 | 0 | 3 | 52 | 647 | 5918 | 40065 | 0 | 8 | 0 | 80 |
| sgRCM103 | NAA | TS 엑손 | 0 | 0 | 0 | 3 | 27 | 94 | 0 | 0 | 0 | 0 |
| sgRCM103 | NAG | 엑손 | 0 | 2 | 4 | 41 | 340 | 1883 | 0 | 0 | 0 | 2 |
| sgRCM103 | NAG | 없음 | 0 | 2 | 34 | 444 | 3633 | 24615 | 0 | 5 | 1 | 37 |
| sgRCM103 | NAG | TS 엑손 | 0 | 0 | 1 | 2 | 19 | 104 | 0 | 0 | 0 | 2 |
| sgRCM103 | NGA | 엑손 | 0 | 0 | 0 | 25 | 200 | 1214 | 0 | 2 | 0 | 4 |
| sgRCM103 | NGA | 없음 | 0 | 0 | 23 | 381 | 3357 | 22500 | 2 | 5 | 0 | 29 |
| sgRCM103 | NGA | TS 엑손 | 0 | 0 | 0 | 0 | 18 | 60 | 0 | 1 | 0 | 0 |
| sgRCM103 | NGG | 엑손 | 1 | 1 | 2 | 29 | 202 | 1202 | 0 | 1 | 0 | 2 |
| sgRCM103 | NGG | 없음 | 0 | 0 | 29 | 294 | 2710 | 18612 | 1 | 4 | 0 | 28 |
| sgRCM103 | NGG | TS 엑손 | 0 | 0 | 0 | 0 | 10 | 65 | 0 | 0 | 0 | 0 |
| sgRCM104 | NAA | 엑손 | 0 | 0 | 0 | 4 | 26 | 256 | 0 | 0 | 0 | 1 |
| sgRCM104 | NAA | 없음 | 0 | 0 | 2 | 26 | 386 | 3483 | 0 | 0 | 0 | 3 |
| sgRCM104 | NAA | TS 엑손 | 0 | 0 | 0 | 0 | 1 | 7 | 0 | 0 | 0 | 0 |
| sgRCM104 | NAG | 엑손 | 0 | 0 | 0 | 5 | 44 | 502 | 0 | 0 | 0 | 0 |
| sgRCM104 | NAG | 없음 | 0 | 0 | 2 | 36 | 461 | 5459 | 0 | 0 | 0 | 4 |
| sgRCM104 | NAG | TS 엑손 | 0 | 0 | 0 | 0 | 0 | 27 | 0 | 0 | 0 | 0 |
| sgRCM104 | NGA | 엑손 | 0 | 0 | 0 | 2 | 38 | 364 | 0 | 0 | 0 | 1 |
| sgRCM104 | NGA | 없음 | 0 | 0 | 0 | 51 | 1101 | 9541 | 0 | 0 | 1 | 0 |
| sgRCM104 | NGA | TS 엑손 | 0 | 0 | 0 | 0 | 1 | 27 | 0 | 0 | 0 | 0 |
| sgRCM104 | NGG | 엑손 | 1 | 0 | 0 | 2 | 70 | 506 | 0 | 0 | 0 | 0 |
| sgRCM104 | NGG | 없음 | 0 | 0 | 1 | 28 | 461 | 4267 | 0 | 2 | 0 | 1 |
| sgRCM104 | NGG | TS 엑손 | 0 | 0 | 0 | 0 | 2 | 34 | 0 | 0 | 0 | 0 |

[2213]

[2214] **실시예 27: CD5-표적화 CAR-T의 생산**

[2215] CD5-표적화 CAR-T 세포를 생산하기 위해, CD5 CAR을 암호화하는 렌티바이러스 벡터(LVV)를 조작하였다(도 55). 환자의 T 세포를 CD5 CAR LVV로 형질도입하였다. 환자에게 재주입하기 전에 CAR-T 세포의 시험관내 7일 내지 14일의 확장을 수행하였다. 도 55는 예시적인 CD5 CAR 작제물 설계를 보여준다. 예시적인 CD5 CAR 작제물 설계는 항-CD5 scFv, 힌지/스페이서, 막관통 도메인, 신호 2 공동-자극 도메인 및 CD3ζ 공동-자극 도메인을 포함한다. 힌지/스페이서 영역은 CH3, CD8α 또는 CD28을 포함할 수 있다. 막관통 도메인은 CD8α 또는 CD28을 포함할 수 있다. 공동-자극 도메인은 41bb 또는 CD28을 포함할 수 있다.

[2216] 30개의 LVV를 생성하였다. 아래 표 28은 다양한 LVV CD5 CAR 작제물을 제공한다.

표 28

**CD5 CAR LVV 작제물**

| ID | 서열 |
|---|---|
| pCAR_BTx063 | MND\5CAR\CH3\CD28TM\CD28\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx064 | MND\5CAR\CH3\CD28TM\CD28\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx065 | MND\5CAR\CH3\CD28TM\CD28\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx066 | MND\5CAR\CD8aH\CD28TM\CD28\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx067 | MND\5CAR\CD8aH\CD28TM\CD28\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx068 | MND\5CAR\CD8aH\CD28TM\CD28\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx069 | MND\5CAR\CD28H\CD28TM\CD28\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx070 | MND\5CAR\CD28H\CD28TM\CD28\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx071 | MND\5CAR\CD28H\CD28TM\CD28\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx072 | MND\5CAR\CH3\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx073 | MND\5CAR\CH3\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx074 | MND\5CAR\CH3\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx075 | MND\5CAR\CD8aH\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx076 | MND\5CAR\CD8aH\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx077 | MND\5CAR\CD8aH\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx078 | PGK\5CAR\CH3\CD28TM\CD28\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx079 | PGK\5CAR\CH3\CD28TM\CD28\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx080 | PGK\5CAR\CH3\CD28TM\CD28\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx081 | PGK\5CAR\CD8aH\CD28TM\CD28\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx082 | PGK\5CAR\CD8aH\CD28TM\CD28\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx083 | PGK\5CAR\CD8aH\CD28TM\CD28\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx084 | PGK\5CAR\CD28H\CD28TM\CD28\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx085 | PGK\5CAR\CD28H\CD28TM\CD28\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx086 | PGK\5CAR\CD28H\CD28TM\CD28\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx087 | PGK\5CAR\CH3\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx088 | PGK\5CAR\CH3\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx089 | PGK\5CAR\CH3\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 3 |
| pCAR_BTx090 | PGK\5CAR\CD8aH\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 1 |
| pCAR_BTx091 | PGK\5CAR\CD8aH\CD8aTM\41BB\CD3Z\WPRE 코돈 opt 2 |
| pCAR_BTx092 | PGK\5CAR\CD28H\CD28TM\CD28\CD3Z\WPRE 코돈 opt 1 |

[2217]

[2218] MND 프로모터 또는 PGK 프로모터를 갖는 표 27의 각각의 CD5 CAR LVV 작제물을 스크리닝하고 sgRNA 103 및 sgRNA 104를 사용하여 1차 인간 T 세포에서 CAR 발현을 조사하였다. 형질도입되지 않은(UTD) 세포를 대조군으로 사용하였다. 도 56에 나타낸 바와 같이, CD5 CAR LVV 작제물은 0% 내지 82% 범위의 CAR 발현을 초래하였다. 24개의 작제물이 10일의 배양 기간이 끝날 때 T 세포의 표면에서 검출 가능한 CAR 발현을 보여주었다.

[2219] 힌지/스페이서 CAR 작제물 서열이 CAR 발현에 영향을 미치는지 여부를 결정하기 위해, 상이한 힌지/스페이서 서열(굵은 글씨체)을 포함하는 표 29의 CAR 작제물을 sgRNA 103 및 sgRNA 104와 조합하여 조사하였다. 형질도입되지 않은(UTD) 세포를 대조군으로 사용하였다. 도 57에 나타낸 바와 같이, 힌지/스페이서 서열은 CAR 발현에 영향을 미치지 않았다.

*표 29*

**CD5 CAR 작제물**

| ID | 서열 |
|---|---|
| UTD | 형질전환되지 않음(렌티바이러스 없음) |
| pCAR_BTx063 | MND\5CAR**CH3**\CD28TM\CD28\CD3Z\WPRE |
| pCAR_BTx067 | MND\5CAR**CD8aH**\CD28TM\CD28\CD3Z\WPRE |
| pCAR_BTx071 | MND\5CAR**CD28H**\CD28TM\CD28\CD3Z\WPRE |

[2220]

[2221] 시험관내 CD5 LVV CAR-T 세포를 추가로 특성화하기 위해, 라이브 이미징 세포 살해 검정을 수행하였다. CD5 LVV CAR-T 세포(LV66, LV67, LV68 및 LV69)를 CCRF 종양 세포에 첨가하였다. 도 57A 내지 도 57D에 나타낸 바와 같이, 편집된 및 편집되지 않은 작제물은 CD5+ CCRF 세포에 대해 세포독성이었다.

[2222] CD5 CAR 작제물(LV63-LV69)을 T 세포 단독 또는 CCRF 세포에서 sgRNA 103 및 sgRNA 104를 사용하여 IFNγ 생산에 대해 시험관내에서 평가하였다(도 59a 및 도 59b). 모든 CD5 CAR 작제물이 CD5+ CCRF-CEM 백혈병 세포주의 존재하에 IFNγ를 방출하였다.

[2223] **실시예 28: 다중 변형된 면역 효과기 세포를 사용한 병용 요법**

[2224] 다중 유전자 변형 CAR-T 세포의 사용이 액체암과 같은 T-세포 또는 NK-세포 악성종양이 있는 환자의 치료를 향상시킬 수 있을지 여부를 조사하기 위하여, 환자에게 이들의 면역표현형을 기준으로 하나 이상의 변형된 CAR-T 세포를 투여하였다. 환자 샘플로부터의 면역 세포를 유세포 분석 및/또는 서열 분석을 사용하여 CD5, CD7, CD33 및 CD123의 존재 또는 부재에 대해 면역표현형을 지정하였다. mRNA 염기 편집기 및 가이드 RNA를 전기천공법에 의해 면역 세포에 전달하고, TRAC, LAG-3, FAS, CD52, B2M, CIITA, TRBC1, TRBC2 및 PD-1 중 하나 이상과 조합하여 CD5, CD7, CD33 또는 CD123을 편집하여 면역 세포에서 이들 유전자의 발현을 감소시키거나 또는 제거하였다. 염기 편집 후, CD5 CAR, CD7 CAR, CD33 CAR 또는 CD123 CAR은 렌티바이러스 형질도입을 통해 변형된 면역 세포에 전달되어 각각 CD5, CD7, CD33 및 CD123 CAR-T 세포를 생성하였다.

[2225] $CD5^+$ $CD7^-$로 면역표현형이 지정된 환자는 CD5 CAR-T 세포 또는 이의 집단으로 치료하였다. $CD5^-$ $CD7^+$로 면역표현형이 지정된 환자는 CD7 CAR-T 세포 또는 이의 집단으로 치료하였다. $CD5^+$ $CD7^+$로 면역표현형이 지정된 환자는 CD5 CAR-T 세포 또는 이의 집단 및/또는 CD7 CAR-T 세포 또는 이의 집단으로 치료하였다. CD33- CD123+로 면역표현형이 지정된 환자는 CD123 CAR-T 세포 또는 이의 집단으로 치료하였다. CD33+ CD123-로 면역표현형이 지정된 환자는 CD33 CAR-T 세포 또는 이의 집단으로 치료하였다. CD33+ CD123+로 면역표현형이 지정된 환자는 CD33 CAR-T 세포 또는 이의 집단 및/또는 CD123 CAR-T 세포 또는 이의 집단으로 치료하였다. $CD5^+$ $CD7^+$ $CD33^+$ CD123+로 면역표현형이 지정된 환자는 CD33 CAR-T 세포 또는 이의 집단, CD5 CAR-T 세포 또는 이의 집단, CD7 CAR-T 세포 또는 이의 집단 및/또는 CD123 CAR-T 세포 또는 이의 집단의 임의의 조합으로 치료하였다.

[2226] **기타 실시형태**

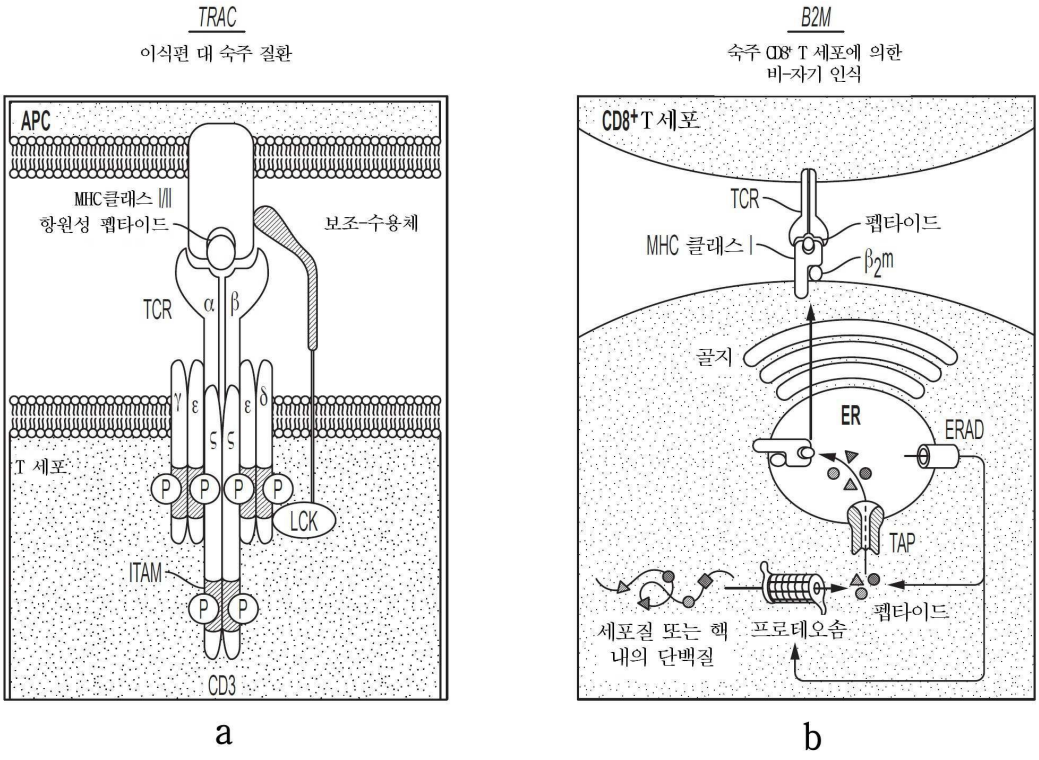
[2227] 전술한 설명으로부터, 다양한 용법 및 조건에 채택하기 위해 본 명세서에 기재된 발명에 변형 및 수정이 이루어질 수 있음이 명백할 것이다. 이러한 실시형태는 또한 다음 청구범위의 범주 내에 있다.

[2228] 본 명세서에서 변수의 임의의 정의에서 요소 목록의 인용은 열거된 요소의 임의의 단일 요소 또는 조합(또는 하위조합)으로서 해당 변수의 정의를 포함한다. 본 명세서에서 실시형태의 인용은 해당 실시형태를 임의의 단일 실시형태로서 또는 임의의 다른 실시형태 또는 이의 일부와 조합하여 포함한다.

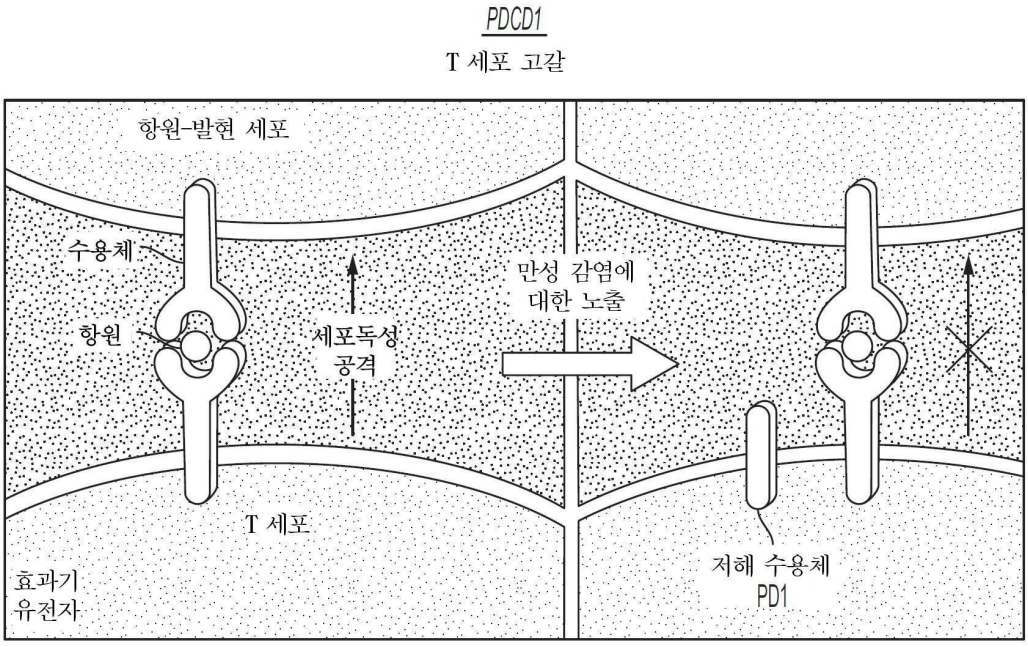
[2229] 본 명세서에 언급된 모든 간행물, 특허 및 특허 출원은 각각의 개별 간행물, 특허 또는 특허 출원이 참조에 의해 원용되는 것으로 구체적이고 개별적으로 표시된 바와 같이 동일한 정도로 참조에 의해 본 명세서에 원용된다. 달리 표시되지 않는 한, 본 명세서에 언급된 간행물, 특허 및 특허 출원은 그 전문이 참조에 의해 본 명세서에 원용된다.

## 도면

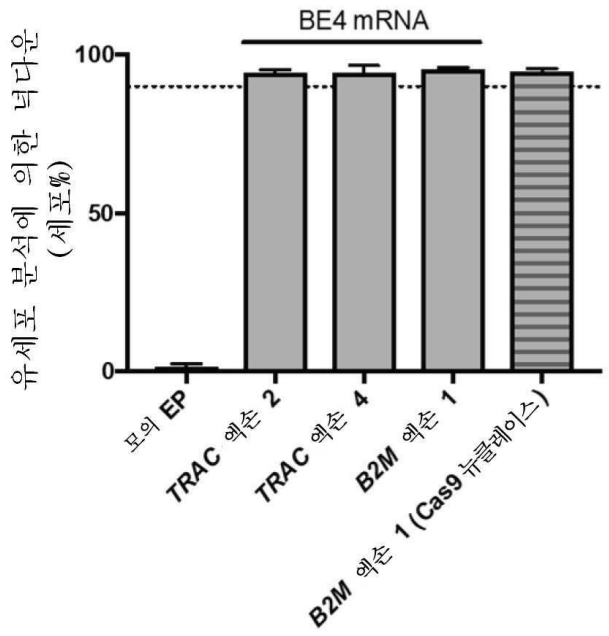
### 도면1ab



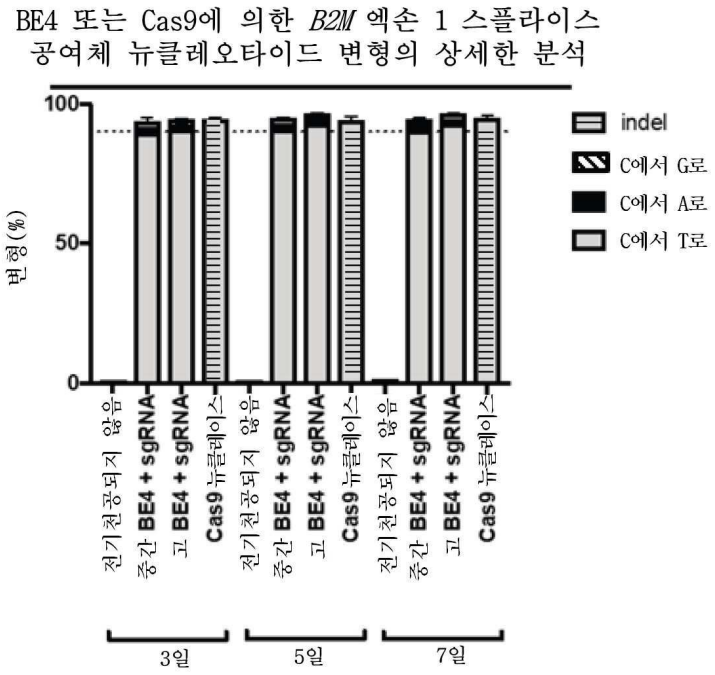

### 도면1c



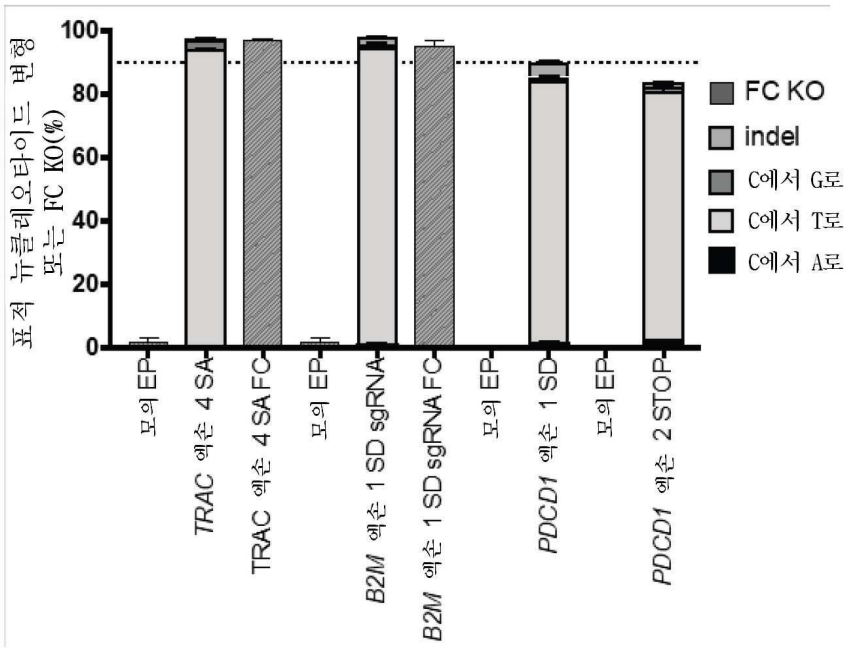
도면2


BE4 mRNA
유세포 분석에 의한 넉다운 (세포%)
100
50
0
모의 EP
TRAC 엑손 2
TRAC 엑손 4
B2M 엑손 1
B2M 엑손 1 (Cas9 뉴클레이스)

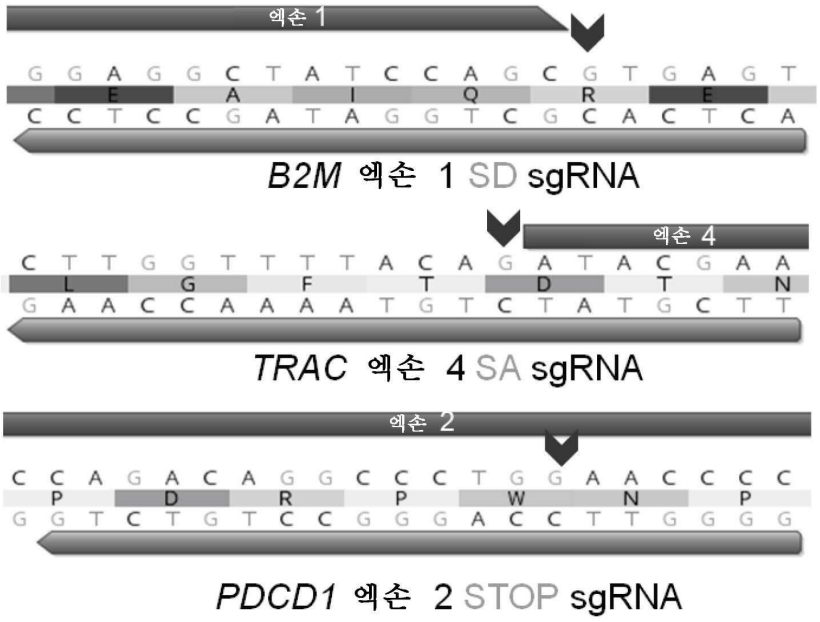
도면3


BE4 또는 Cas9에 의한 B2M 엑손 1 스플라이스 공여체 뉴클레오타이드 변형의 상세한 분석
변형(%)
100
50
0
indel
C에서 G로
C에서 A로
C에서 T로
전기천공되지 않음
중간 BE4 + sgRNA
고 BE4 + sgRNA
Cas9 뉴클레이스
3일
5일
7일

도면4


표적 뉴클레오타이드 변형 또는 FC KO(%)
100
80
60
40
20
0
FC KO
indel
C에서 G로
C에서 T로
C에서 A로
모의 EP
TRAC 엑손 4 SA
TRAC 엑손 4 SA FC
모의 EP
B2M 엑손 1 SD sgRNA
B2M 엑손 1 SD sgRNA FC
모의 EP
PDCD1 엑손 1 SD
모의 EP
PDCD1 엑손 2 STOP

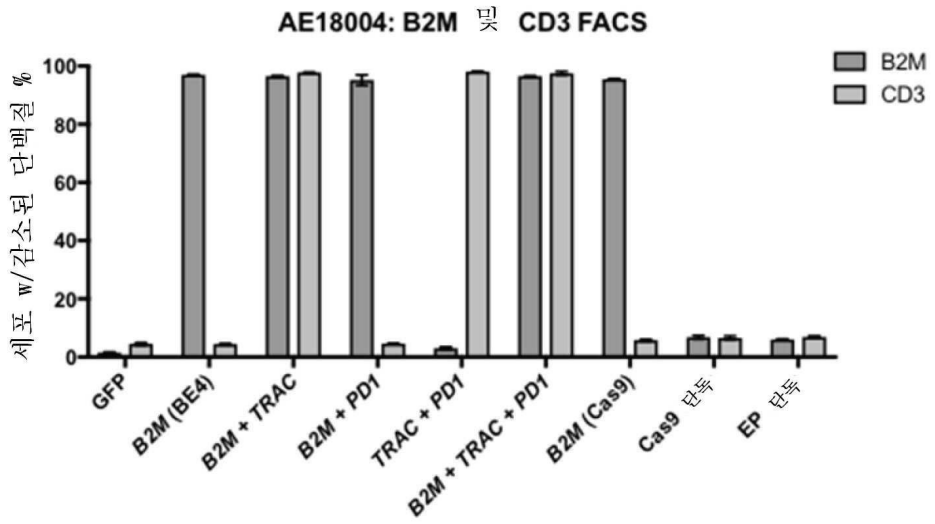
도면5


엑손 1
B2M 엑손 1 SD sgRNA
엑손 4
TRAC 엑손 4 SA sgRNA
엑손 2
PDCD1 엑손 2 STOP sgRNA

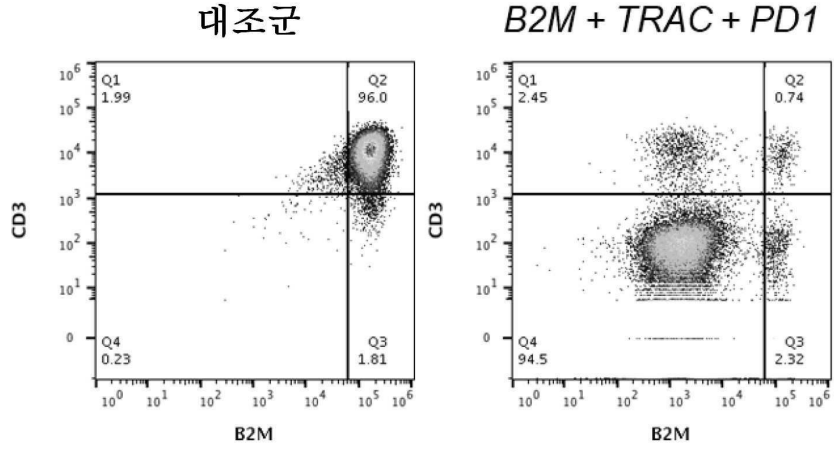
도면6

| 벌지 유형 | 벌지 크기 | 불일치 | TRAC 엑손 2 SD | TRAC 엑손 4 SA | B2M 엑손 1 SD | PD1 엑손 1 SD | PD1 엑손 2 STOP | RNF2 |
|---|---|---|---|---|---|---|---|---|
| X | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| X | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| X | 0 | 2 | 4 | 0 | 0 | 0 | 2 | 0 |
| X | 0 | 3 | 46 | 26 | 2 | 3 | 41 | 6 |
| X | 0 | 4 | 387 | 164 | 52 | 80 | 336 | 116 |
| RNA | 1 | 0 | 0 | 1 | 0 | 0 | 3 | 2 |
| RNA | 1 | 1 | 14 | 1 | 2 | 3 | 11 | 7 |
| RNA | 1 | 2 | 185 | 55 | 17 | 38 | 163 | 30 |
| | | | | | | | | |
| | | | 넉아웃: | 96% | 95% | | 80% | |
| | | | Indel: | 0.6% | 2.5% | | 1.5% | |

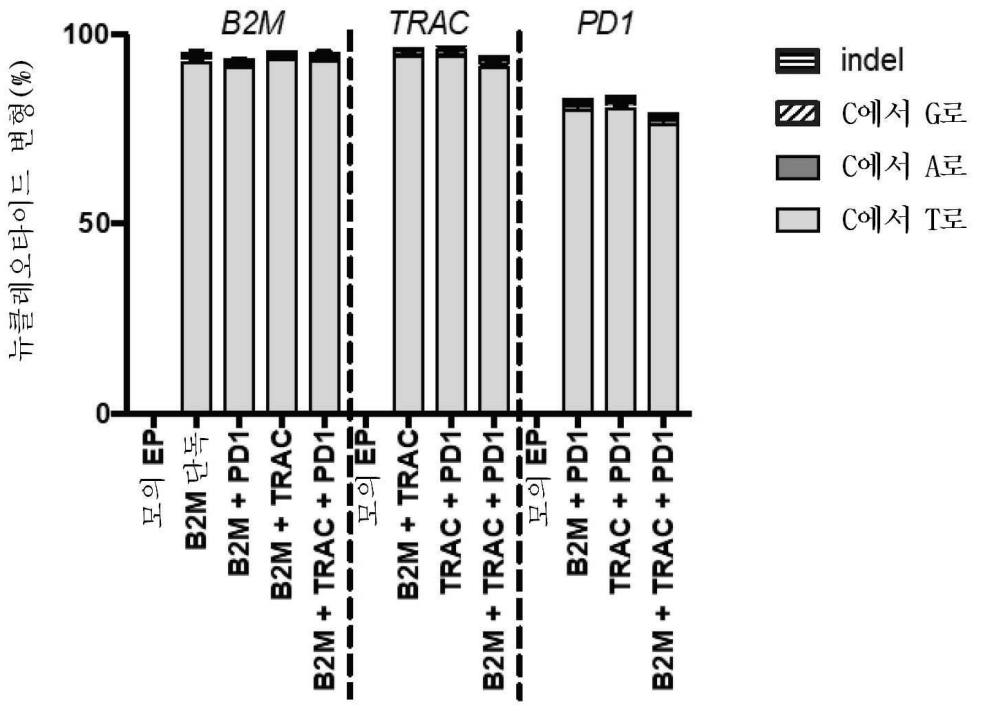
도면7



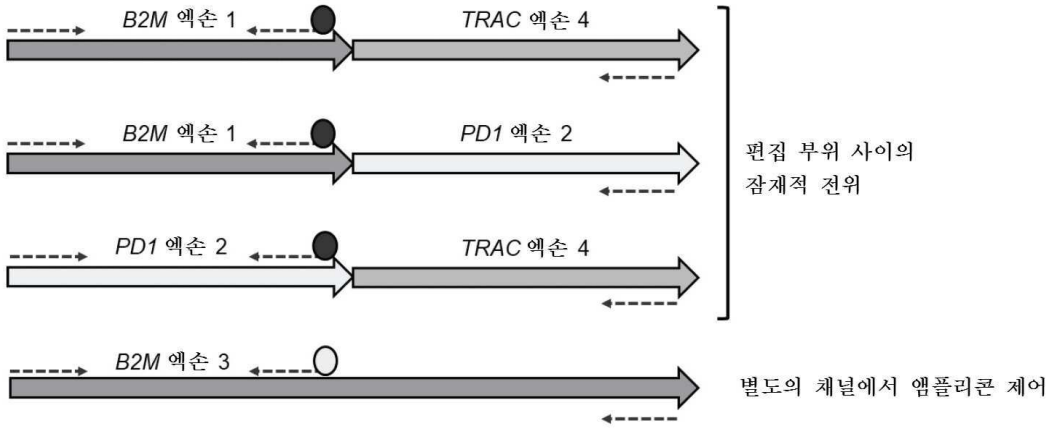

도면8



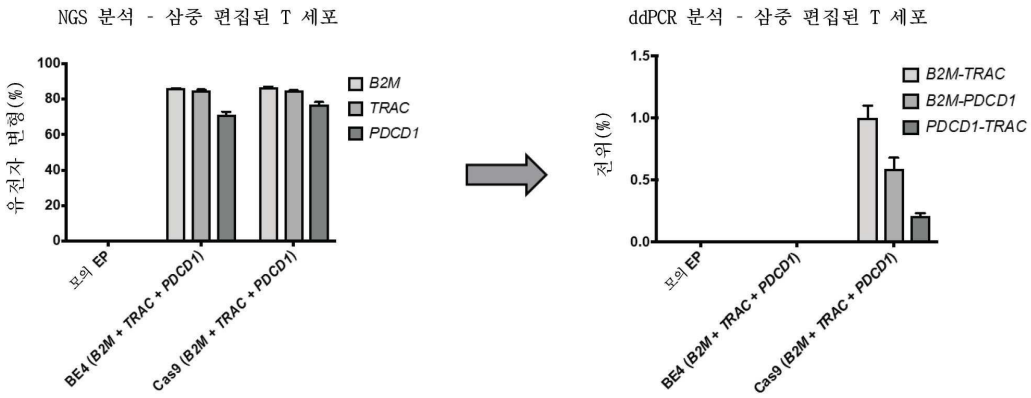
도면9


B2M
TRAC
PD1
100
50
0
뉴클레오타이드 변형(%)
indel
C에서 G로
C에서 A로
C에서 T로
모의 EP
B2M 단독
B2M + PD1
B2M + TRAC
B2M + TRAC + PD1
모의 EP
B2M + TRAC
TRAC + PD1
B2M + TRAC + PD1
모의 EP
B2M + PD1
TRAC + PD1
B2M + TRAC + PD1

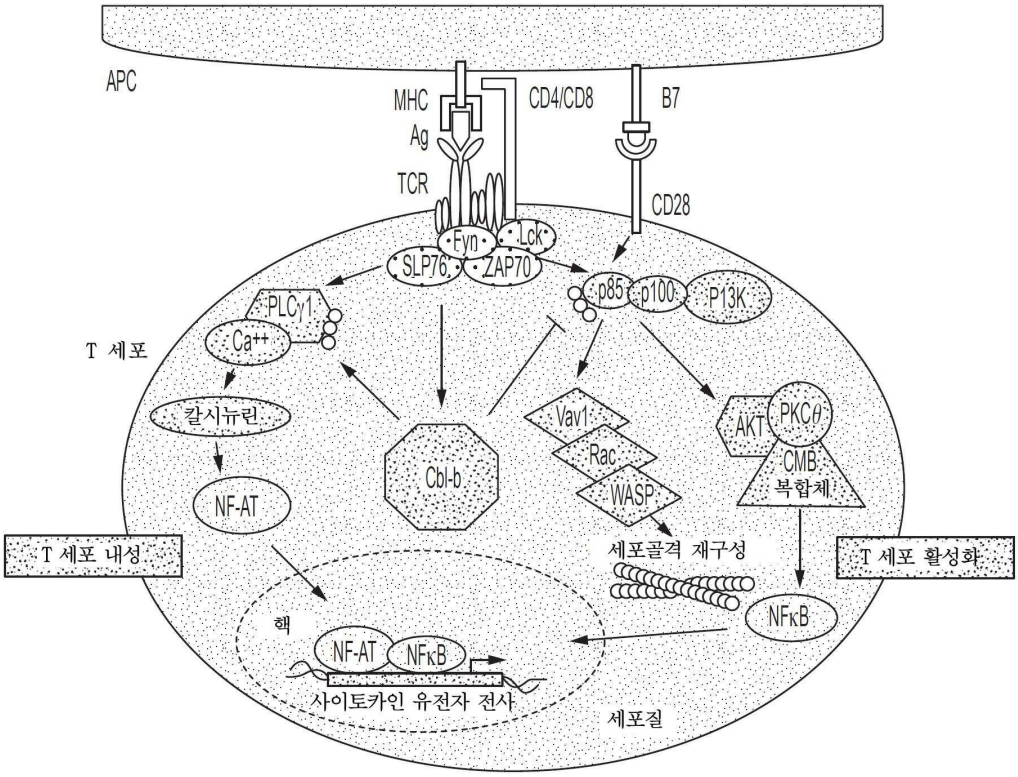
도면10


B2M 엑손 1
TRAC 엑손 4
B2M 엑손 1
PD1 엑손 2
PD1 엑손 2
TRAC 엑손 4
편집 부위 사이의
잠재적 전위
B2M 엑손 3
별도의 채널에서 앰플리콘 제어

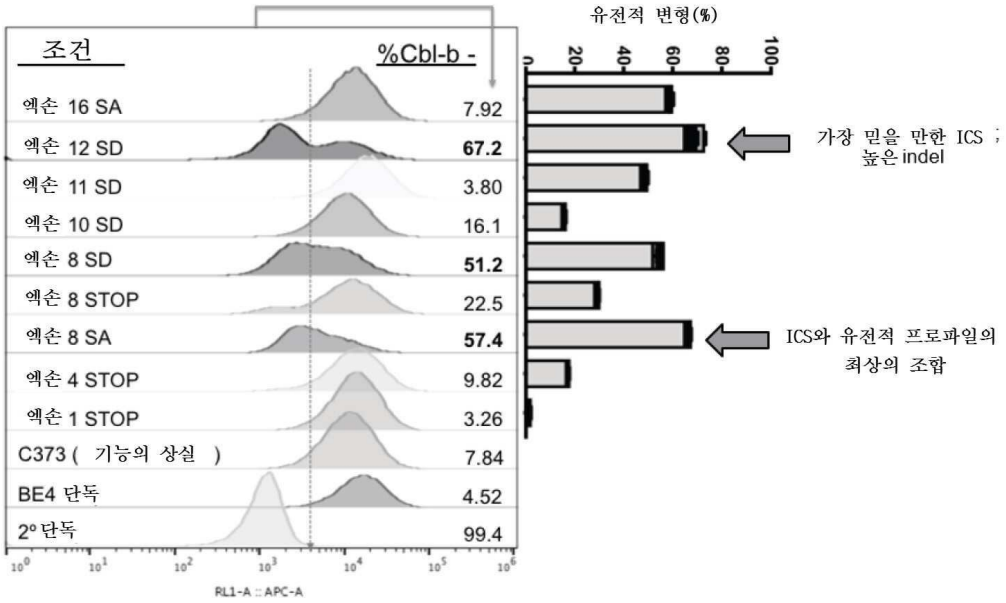
도면11


NGS 분석 - 삼중 편집된 T 세포
유전자 변형(%)
100
80
60
40
20
0
B2M
TRAC
PDCD1
모의 EP
BE4 (B2M + TRAC + PDCD1)
Cas9 (B2M + TRAC + PDCD1)
ddPCR 분석 - 삼중 편집된 T 세포
전위(%)
1.5
1.0
0.5
0.0
B2M-TRAC
B2M-PDCD1
PDCD1-TRAC
모의 EP
BE4 (B2M + TRAC + PDCD1)
Cas9 (B2M + TRAC + PDCD1)

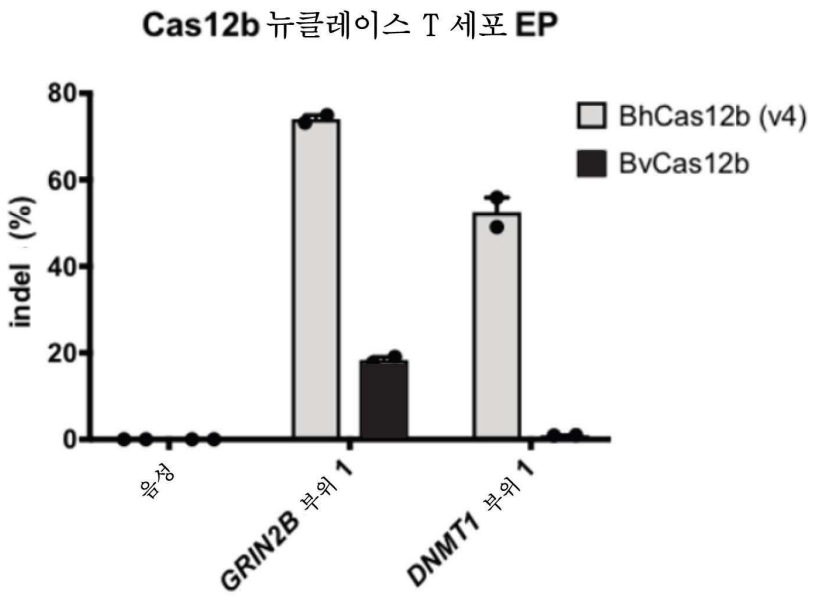
도면12


APC
MHC
Ag
TCR
CD4/CD8
B7
CD28
Fyn
Lck
SLP76
ZAP70
p85
p100
P13K
PLCγ1
Ca++
T 세포
칼시뉴린
NF-AT
Cbl-b
Vav1
Rac
WASP
AKT
PKCθ
CMB
복합체
T 세포 내성
T 세포 활성화
세포골격 재구성
NFκB
핵
NF-AT
NFκB
사이토카인 유전자 전사
세포질

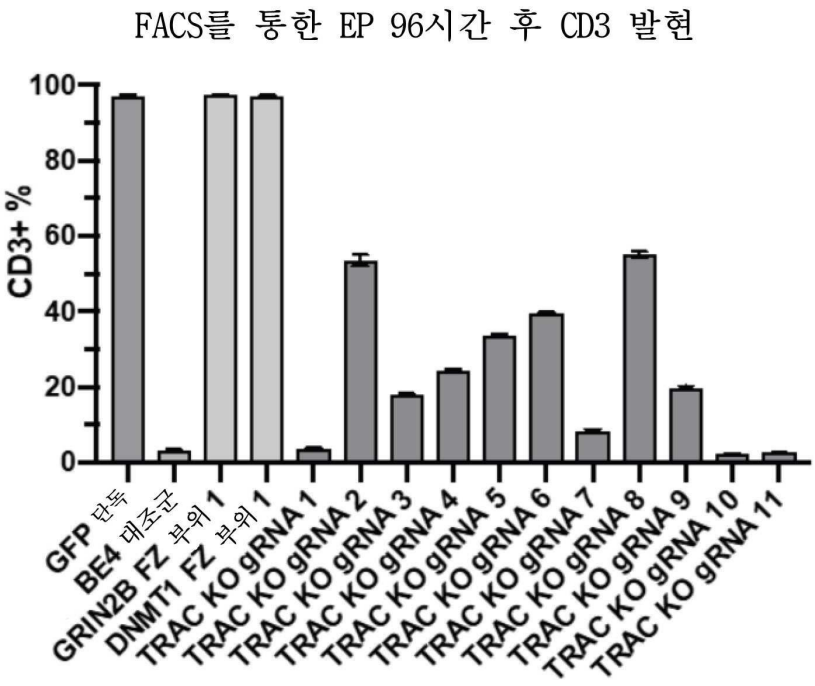
도면13


조건
%Cbl-b -
엑손 16 SA
7.92
엑손 12 SD
67.2
엑손 11 SD
3.80
엑손 10 SD
16.1
엑손 8 SD
51.2
엑손 8 STOP
22.5
엑손 8 SA
57.4
엑손 4 STOP
9.82
엑손 1 STOP
3.26
C373 ( 기능의 상실 )
7.84
BE4 단독
4.52
2° 단독
99.4
RL1-A :: APC-A
유전적 변형(%)
0
20
40
60
80
100
가장 믿을 만한 ICS ;
높은 indel
ICS와 유전적 프로파일의
최상의 조합

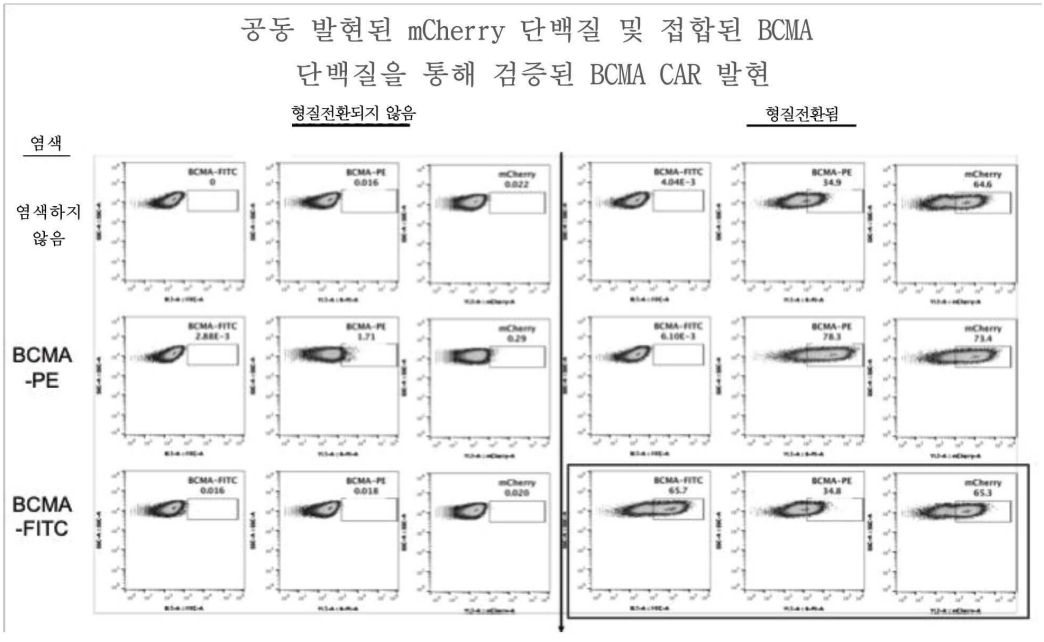
도면14


Cas12b 뉴클레이스 T 세포 EP
BhCas12b (v4)
BvCas12b
indel (%)
80
60
40
20
0
음성
GRIN2B 부위 1
DNMT1 부위 1

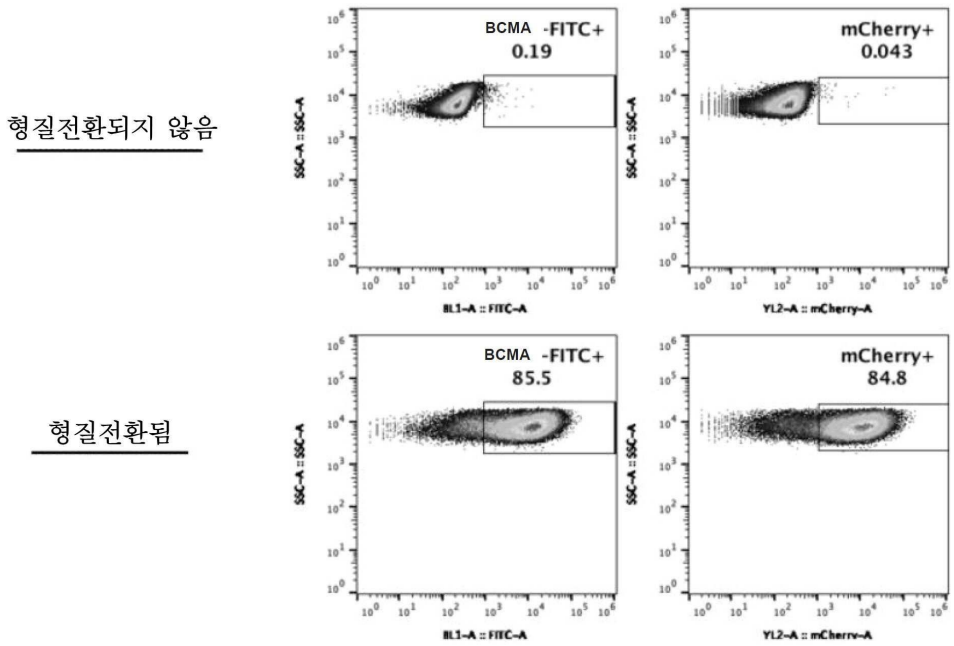
도면15


FACS를 통한 EP 96시간 후 CD3 발현
CD3+ %
100
80
60
40
20
0
GFP 단독
BE4 대조군
GRIN2B FZ 부위 1
DNMT1 FZ 부위 1
TRAC KO gRNA 1
TRAC KO gRNA 2
TRAC KO gRNA 3
TRAC KO gRNA 4
TRAC KO gRNA 5
TRAC KO gRNA 6
TRAC KO gRNA 7
TRAC KO gRNA 8
TRAC KO gRNA 9
TRAC KO gRNA 10
TRAC KO gRNA 11

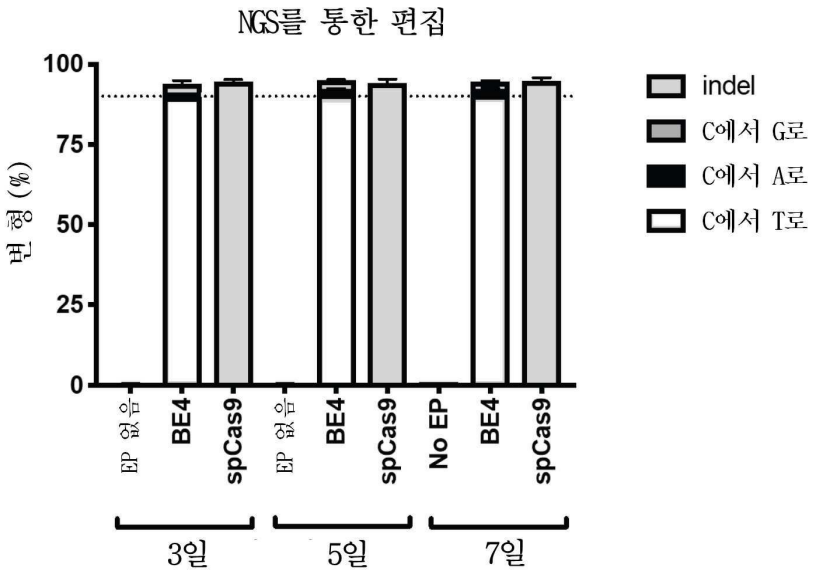
도면16


공동 발현된 mCherry 단백질 및 접합된 BCMA
단백질을 통해 검증된 BCMA CAR 발현
형질전환되지 않음
형질전환됨
염색
염색하지 않음
BCMA -PE
BCMA -FITC

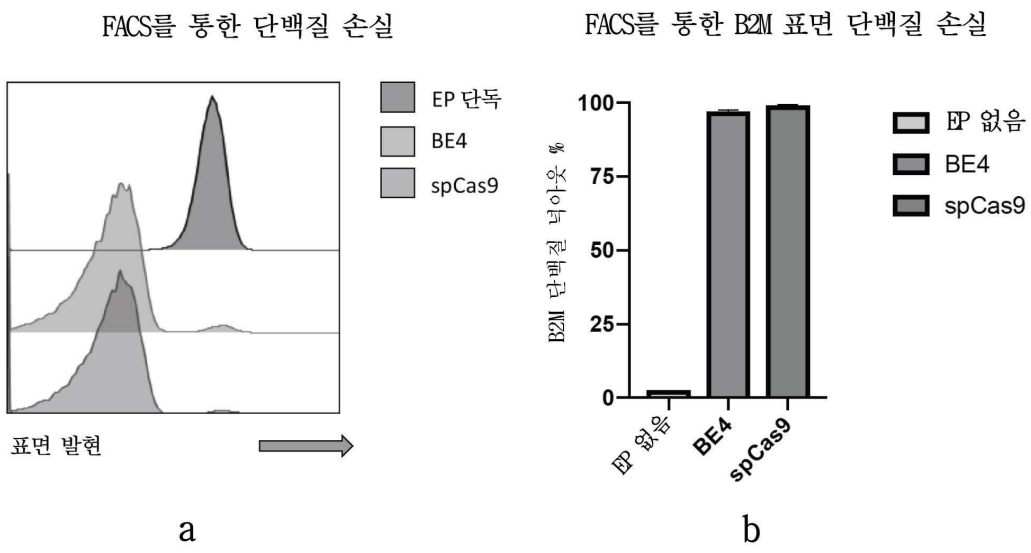
도면17


형질전환되지 않음
BCMA -FITC+
0.19
mCherry+
0.043
형질전환됨
BCMA -FITC+
85.5
mCherry+
84.8

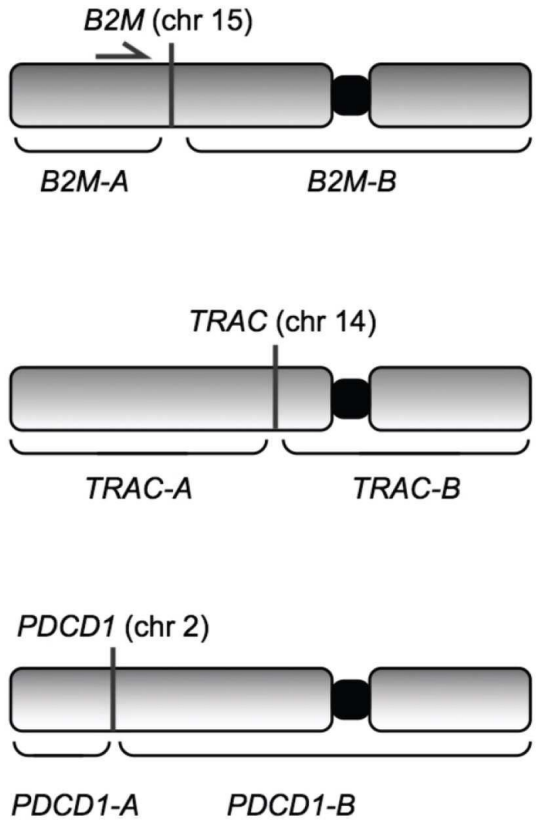
도면18


NGS를 통한 편집
편집 (%)
100
75
50
25
0
EP 없음
BE4
spCas9
EP 없음
BE4
spCas9
No EP
BE4
spCas9
3일
5일
7일
indel
C에서 G로
C에서 A로
C에서 T로

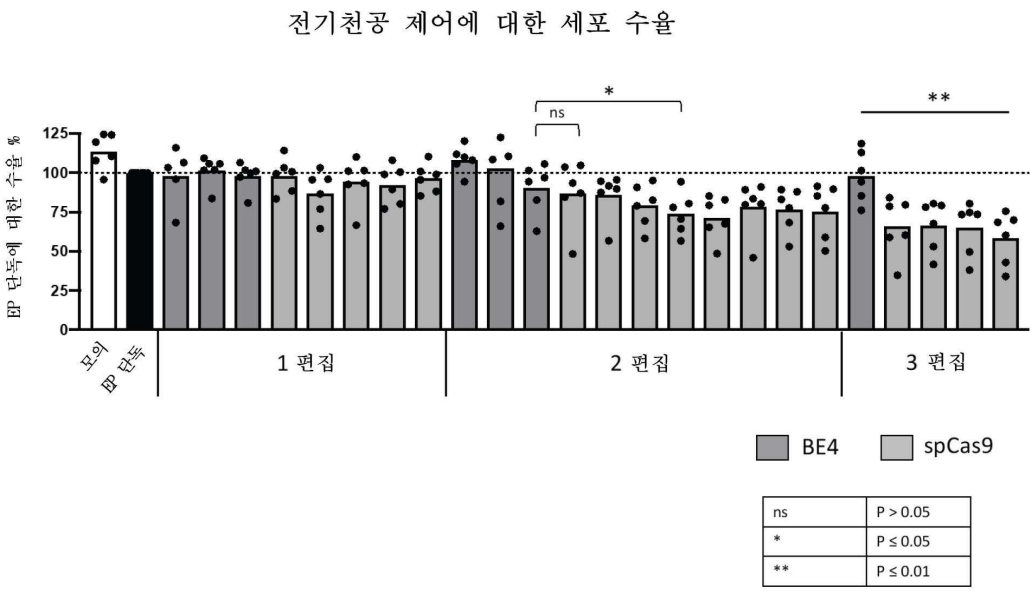
도면19


FACS를 통한 단백질 손실
EP 단독
BE4
spCas9
표면 발현
a
FACS를 통한 B2M 표면 단백질 손실
B2M 단백질 녀아웃 %
100
75
50
25
0
EP 없음
BE4
spCas9
EP 없음
BE4
spCas9
b

도면20

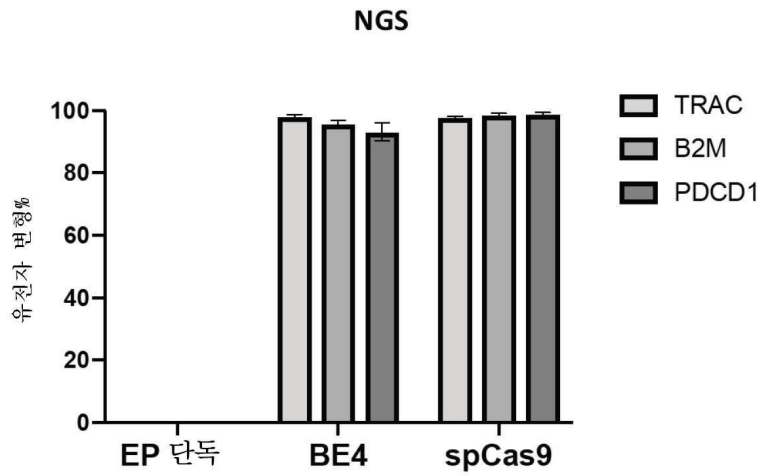

B2M (chr 15)
B2M-A
B2M-B
TRAC (chr 14)
TRAC-A
TRAC-B
PDCD1 (chr 2)
PDCD1-A
PDCD1-B

도면21

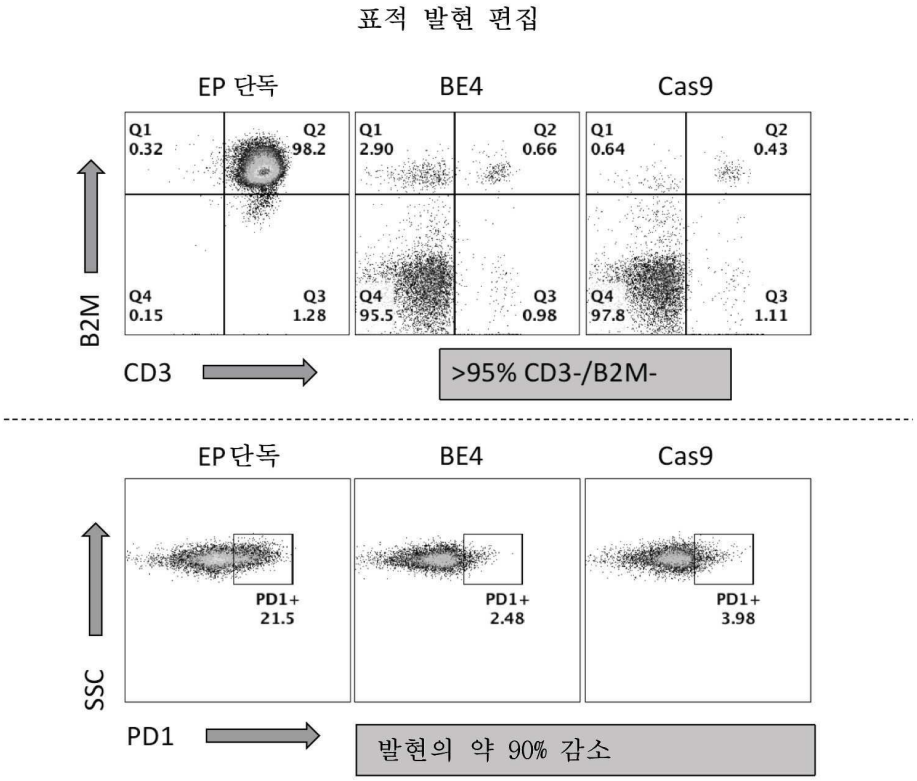



| | |
|---|---|
| ns | P > 0.05 |
| * | P ≤ 0.05 |
| ** | P ≤ 0.01 |

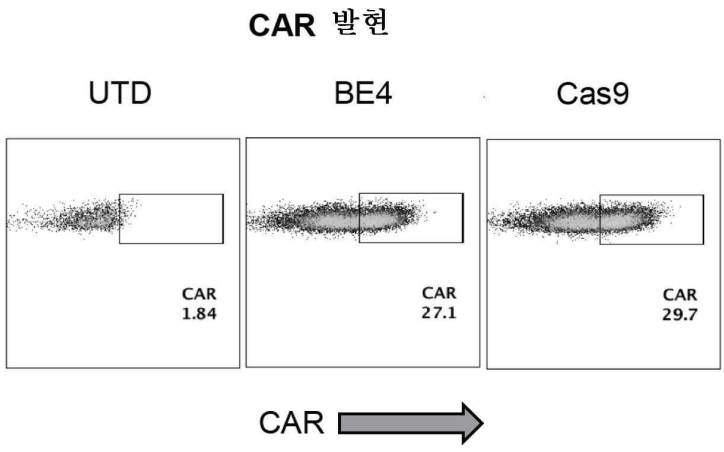
도면22



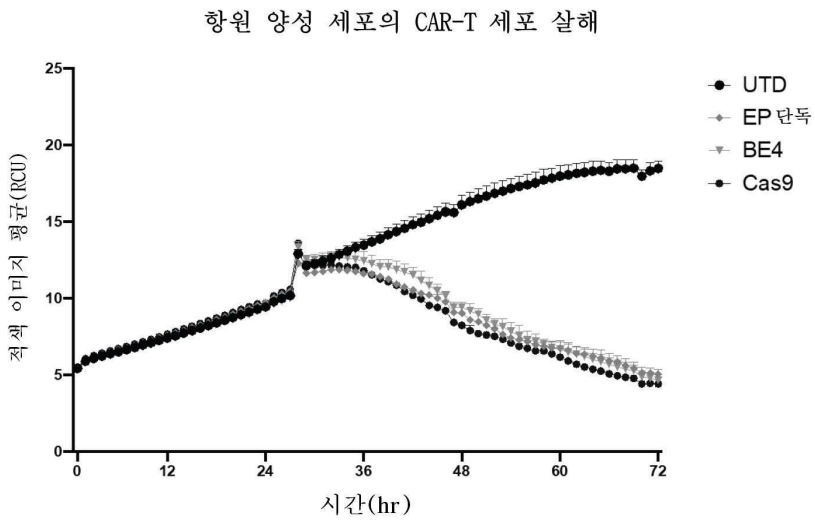
도면23


표적 발현 편집
EP 단독
BE4
Cas9
Q1 0.32
Q2 98.2
Q4 0.15
Q3 1.28
Q1 2.90
Q2 0.66
Q4 95.5
Q3 0.98
Q1 0.64
Q2 0.43
Q4 97.8
Q3 1.11
B2M
CD3
>95% CD3-/B2M-
EP단독
BE4
Cas9
PD1+ 21.5
PD1+ 2.48
PD1+ 3.98
SSC
PD1
발현의 약 90% 감소

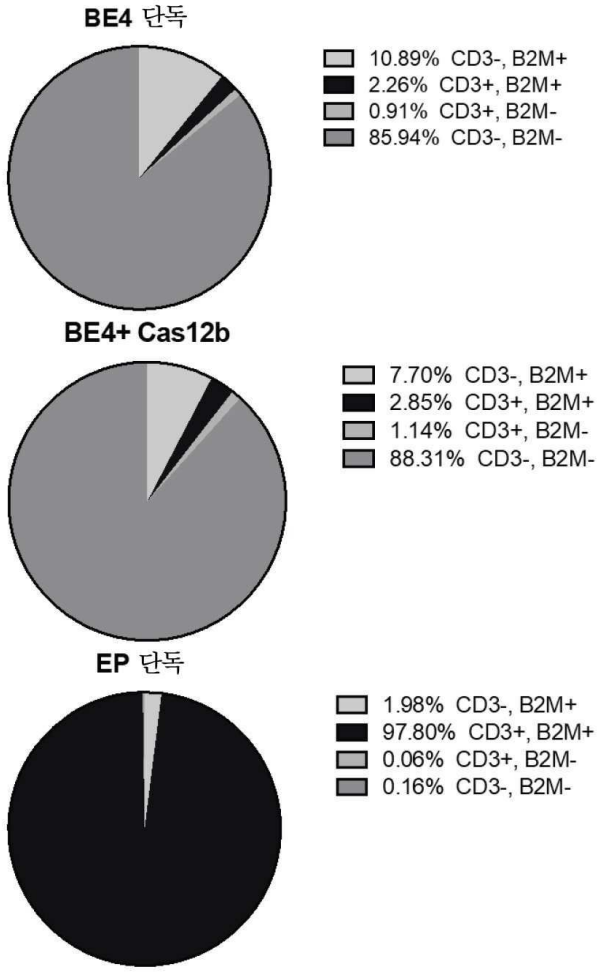
도면24


CAR 발현
UTD
BE4
Cas9
CAR 1.84
CAR 27.1
CAR 29.7
CAR

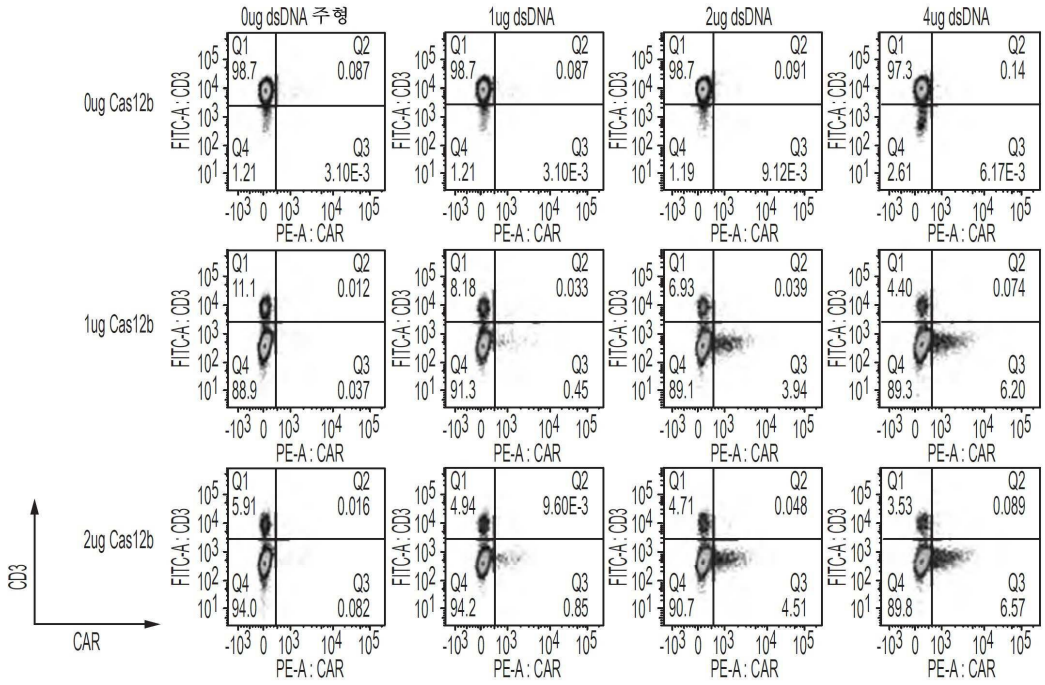
도면25


항원 양성 세포의 CAR-T 세포 살해
적색 이미지 평균(RCU)
25
20
15
10
5
0
0
12
24
36
48
60
72
시간(hr)
UTD
EP 단독
BE4
Cas9

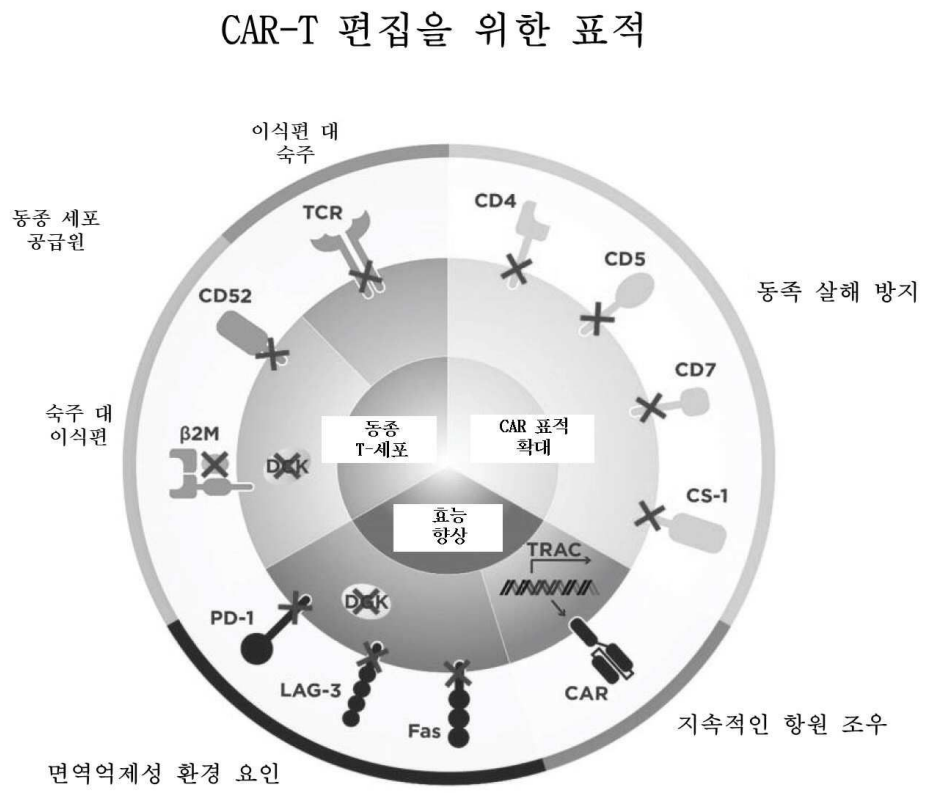
도면26


BE4 단독
10.89% CD3-, B2M+
2.26% CD3+, B2M+
0.91% CD3+, B2M-
85.94% CD3-, B2M-
BE4+ Cas12b
7.70% CD3-, B2M+
2.85% CD3+, B2M+
1.14% CD3+, B2M-
88.31% CD3-, B2M-
EP 단독
1.98% CD3-, B2M+
97.80% CD3+, B2M+
0.06% CD3+, B2M-
0.16% CD3-, B2M-

도면27


0ug dsDNA 주형
1ug dsDNA
2ug dsDNA
4ug dsDNA
0ug Cas12b
1ug Cas12b
2ug Cas12b
CD3
CAR
FITC-A : CD3
PE-A : CAR
Q1 98.7
Q2 0.087
Q4 1.21
Q3 3.10E-3
Q1 98.7
Q2 0.087
Q4 1.21
Q3 3.10E-3
Q1 98.7
Q2 0.091
Q4 1.19
Q3 9.12E-3
Q1 97.3
Q2 0.14
Q4 2.61
Q3 6.17E-3
Q1 11.1
Q2 0.012
Q4 88.9
Q3 0.037
Q1 8.18
Q2 0.033
Q4 91.3
Q3 0.45
Q1 6.93
Q2 0.039
Q4 89.1
Q3 3.94
Q1 4.40
Q2 0.074
Q4 89.3
Q3 6.20
Q1 5.91
Q2 0.016
Q4 94.0
Q3 0.082
Q1 4.94
Q2 9.60E-3
Q4 94.2
Q3 0.85
Q1 4.71
Q2 0.048
Q4 90.7
Q3 4.51
Q1 3.53
Q2 0.089
Q4 89.8
Q3 6.57

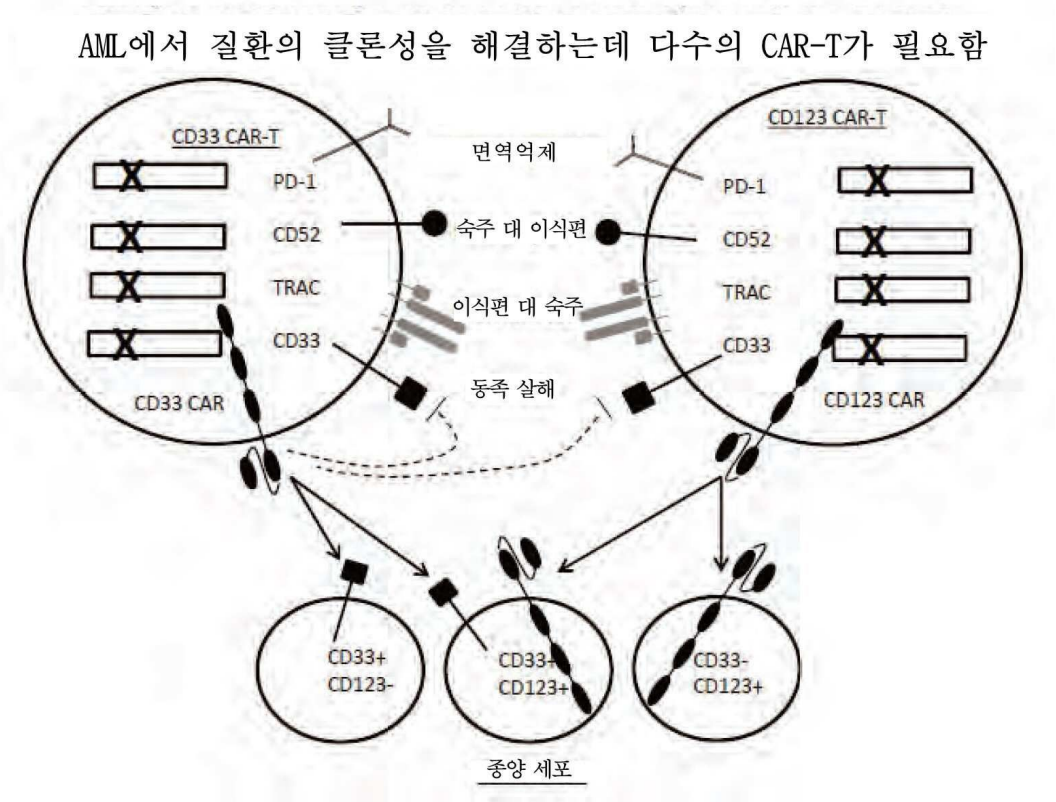
도면28a


CAR-T 편집을 위한 표적
이식편 대 숙주
동종 세포 공급원
숙주 대 이식편
면역억제성 환경 요인
동족 살해 방지
지속적인 항원 조우
TCR
CD4
CD5
CD52
CD7
CS-1
β2M
PD-1
LAG-3
Fas
CAR
TRAC
동종 T-세포
CAR 표적 확대
효능 향상

도면28b

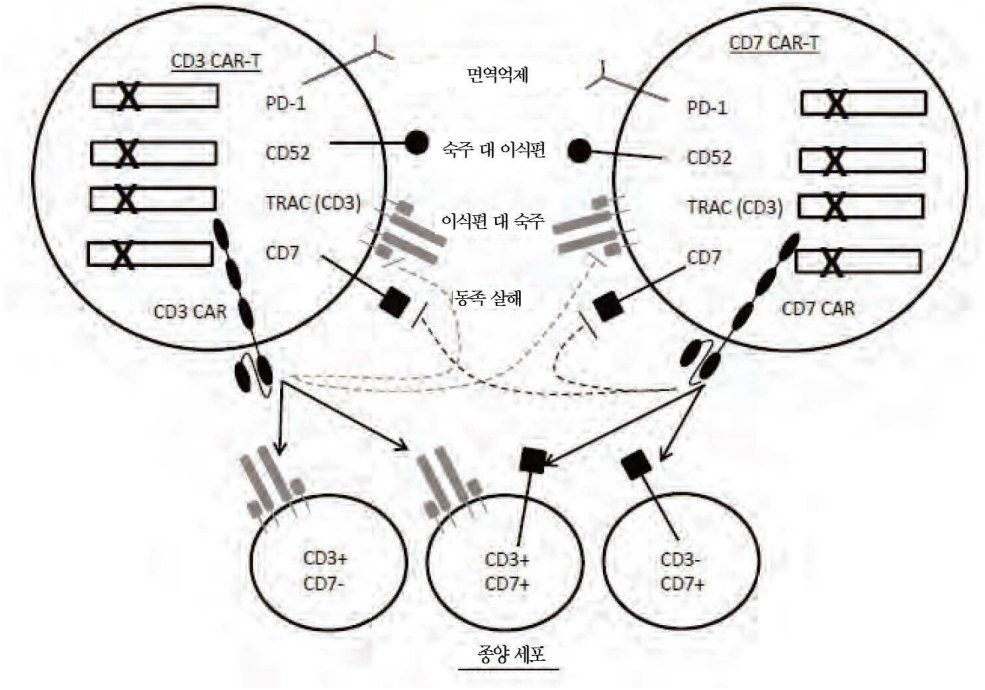

염기 편집기로 침묵시키기 위한 두 가지 전략
1. CBE로 종결 코돈을 생성함
글루타민 CAG → TAG 종결 코돈
CAA → TAA
아르기닌 CGA → TGA
트립토판 TGG → TGA
TGG → TAG
TGG → TAA
침묵 단백질 발현
2. CBE로 스플라이싱 파괴
스플라이스 수여체
인트론
엑손
스플라이스 공여체
엑손
인트론
침묵 단백질 발현

도면29


AML에서 질환의 클론성을 해결하는데 다수의 CAR-T가 필요함
CD33 CAR-T
PD-1
CD52
TRAC
CD33
CD33 CAR
면역억제
숙주 대 이식편
이식편 대 숙주
동족 살해
CD123 CAR-T
PD-1
CD52
TRAC
CD33
CD123 CAR
CD33+ CD123-
CD33+ CD123+
CD33- CD123+
종양 세포

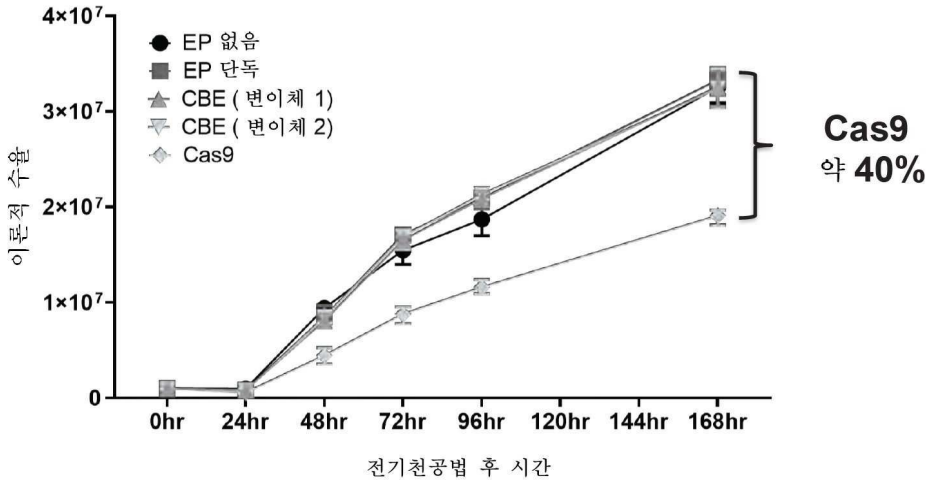
도면30

T-ALL CAR-T 조합에서 동족 살해를 제거하는데 다수의 CAR-T가 필요함

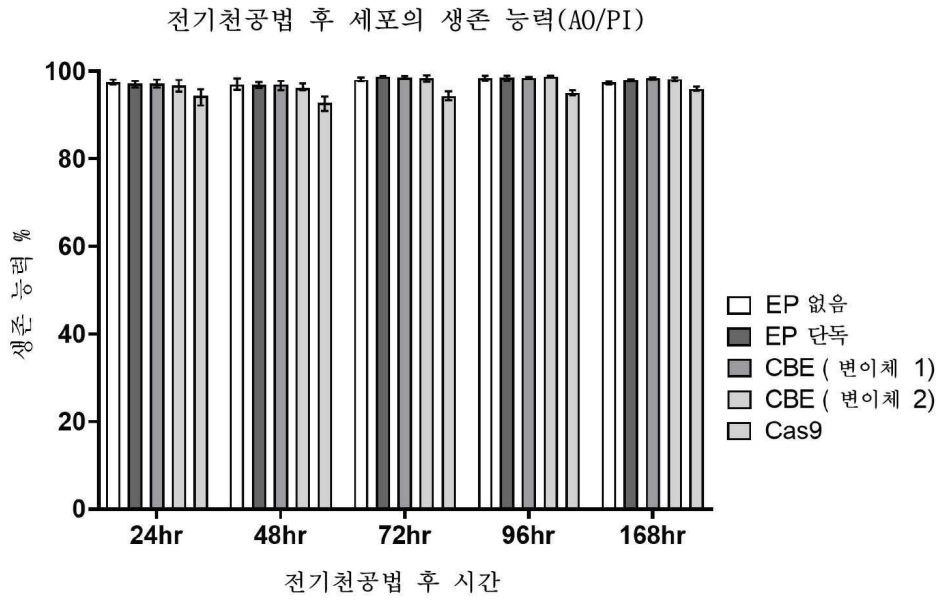



도면31a

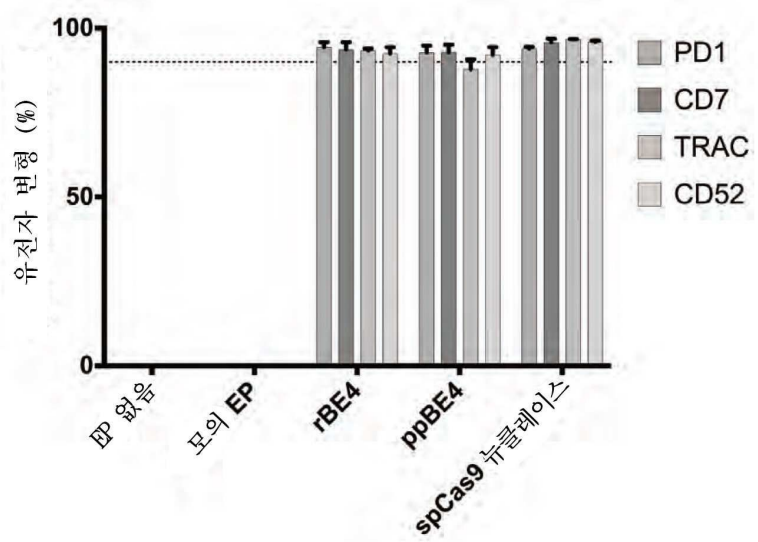
쿼드 T-ALL 편집의 맥락에서 이론적인 수율



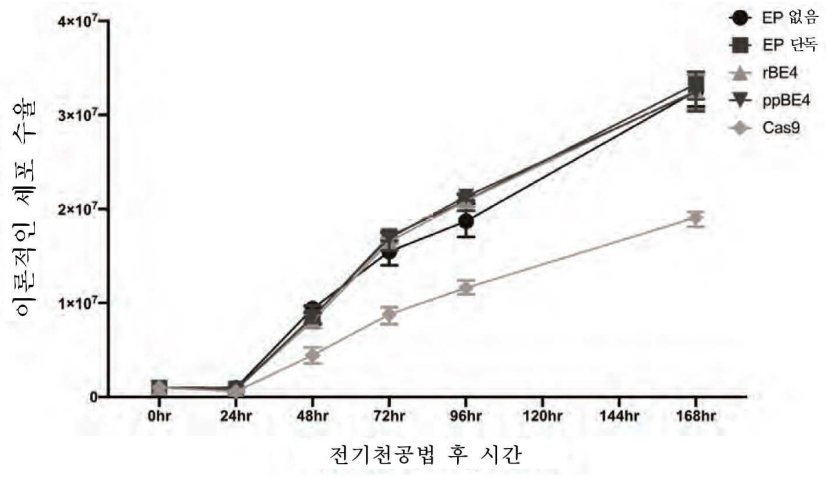
도면31b


전기천공법 후 세포의 생존 능력(AO/PI)
생존 능력 %
100
80
60
40
20
0
24hr
48hr
72hr
96hr
168hr
전기천공법 후 시간
EP 없음
EP 단독
CBE ( 변이체 1)
CBE ( 변이체 2)
Cas9

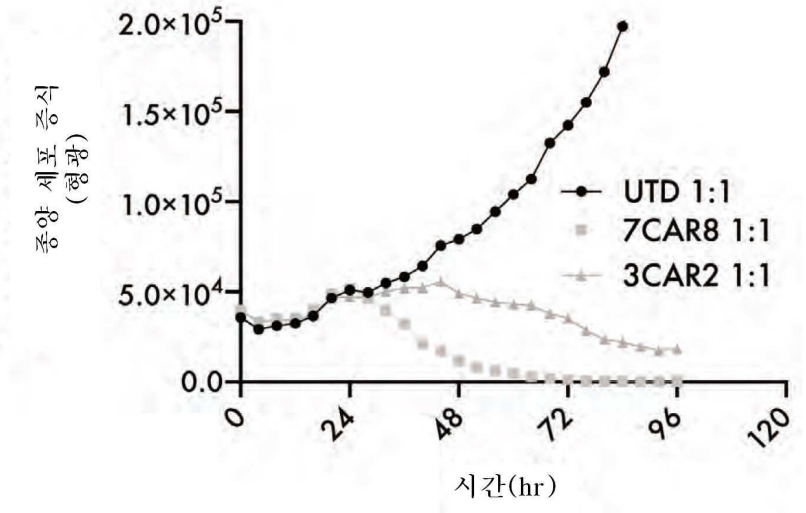
도면32


유전자 변형 (%)
100
50
0
EP 없음
모의 EP
rBE4
ppBE4
spCas9 뉴클레이즈
PD1
CD7
TRAC
CD52

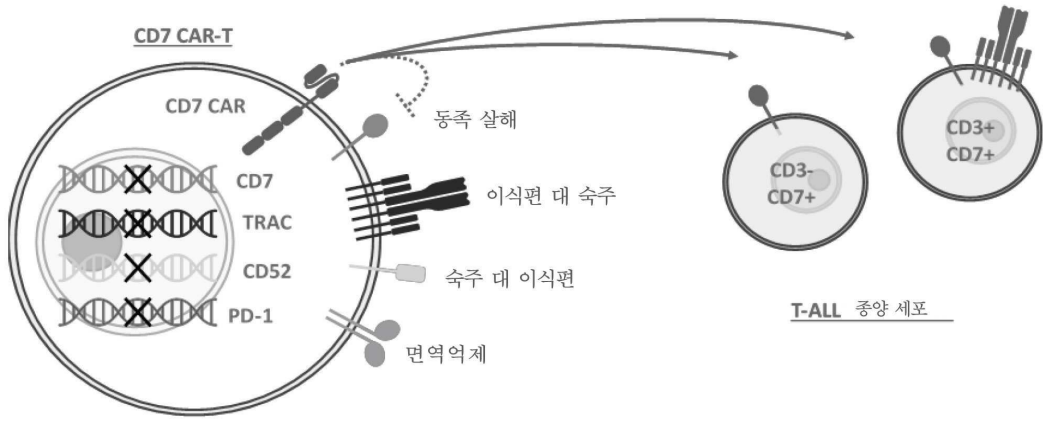
도면33



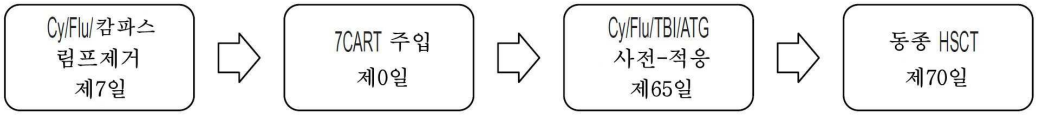

도면34



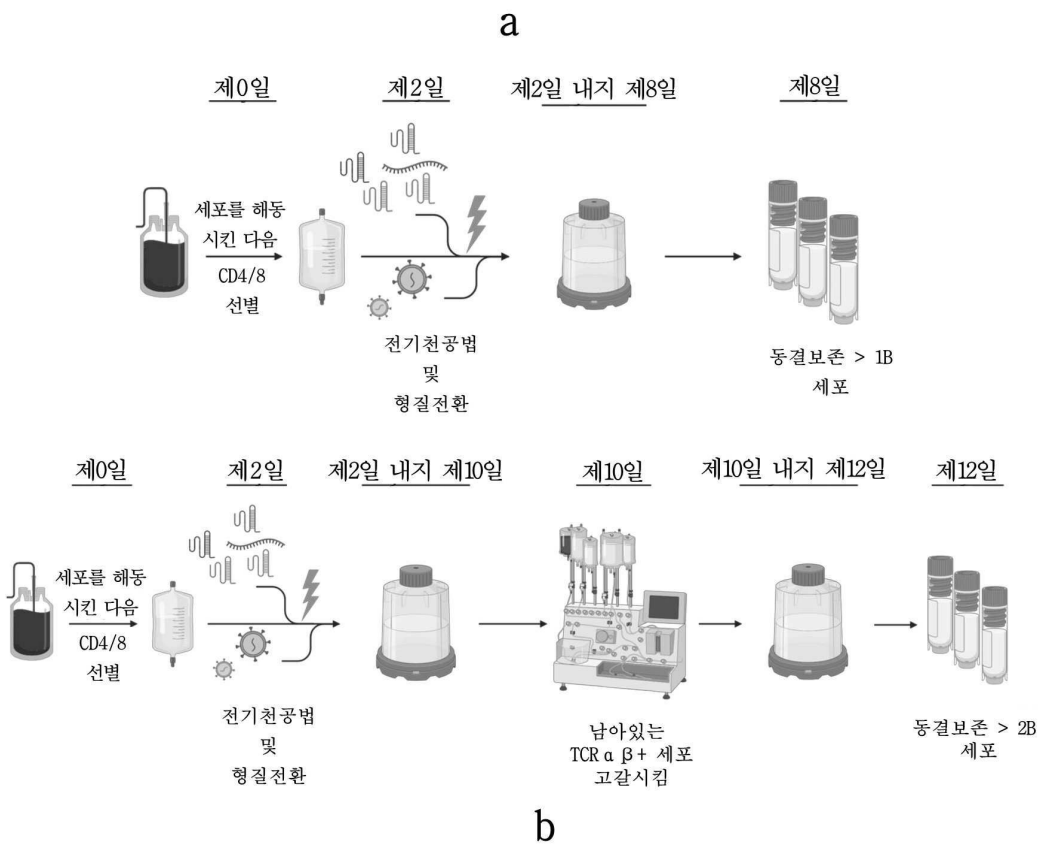

도면35



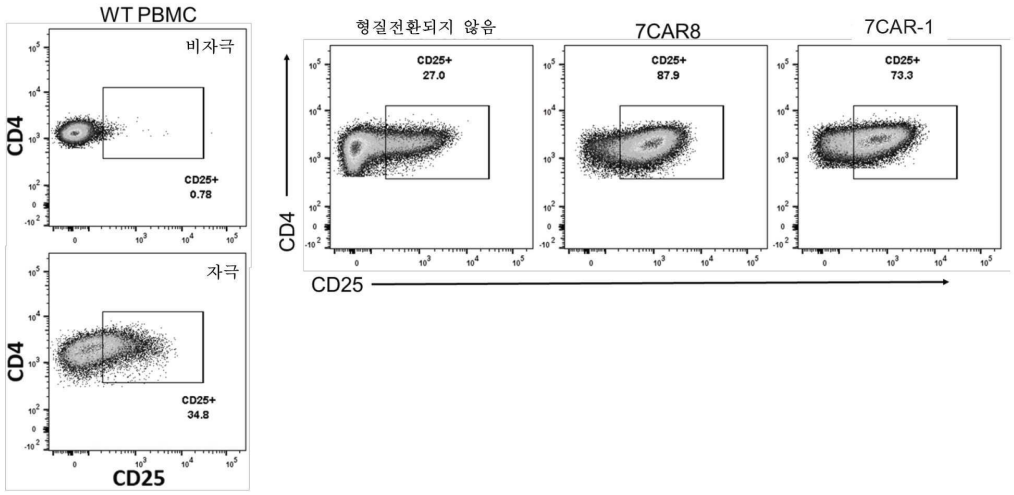
도면36



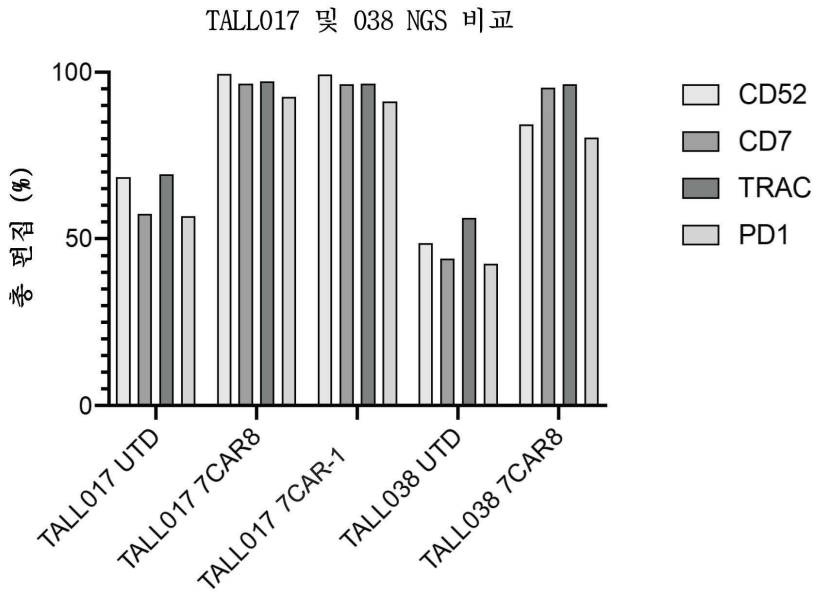

도면37ab



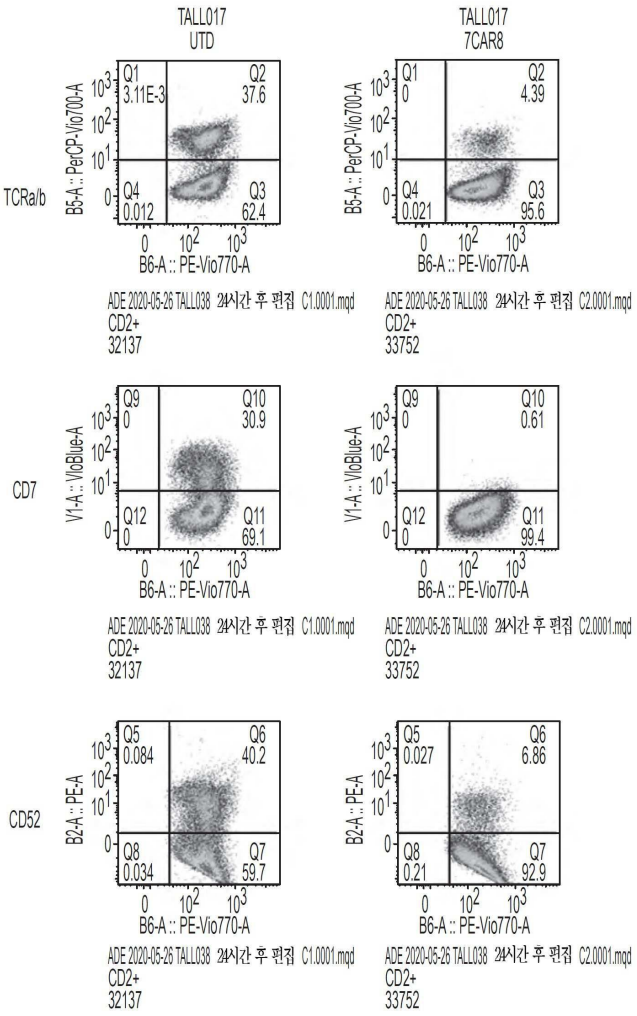

도면37c



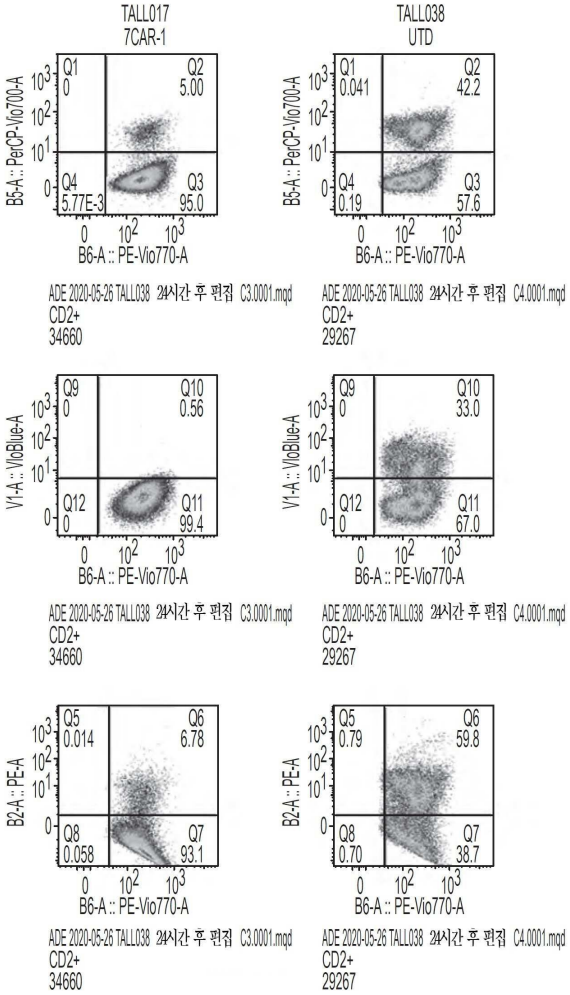
도면38


TALL017 및 038 NGS 비교
100
50
0
총 편집 (%)
CD52
CD7
TRAC
PD1
TALL017 UTD
TALL017 7CAR8
TALL017 7CAR-1
TALL038 UTD
TALL038 7CAR8

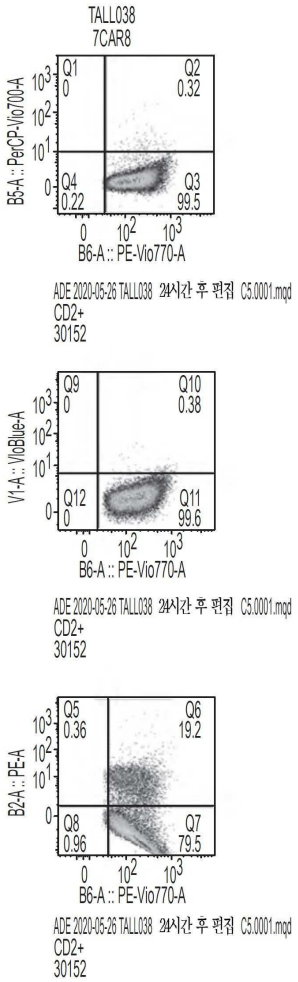
도면39aa


TALL017
UTD
TALL017
7CAR8
TCRa/b
CD7
CD52
B5-A :: PerCP-Vio700-A
V1-A :: VioBlue-A
B2-A :: PE-A
B6-A :: PE-Vio770-A
Q1 3.11E-3
Q2 37.6
Q4 0.012
Q3 62.4
Q1 0
Q2 4.39
Q4 0.021
Q3 95.6
Q9 0
Q10 30.9
Q12 0
Q11 69.1
Q9 0
Q10 0.61
Q12 0
Q11 99.4
Q5 0.084
Q6 40.2
Q8 0.034
Q7 59.7
Q5 0.027
Q6 6.86
Q8 0.21
Q7 92.9
ADE 2020-05-26 TALL038 24시간 후 편집 C1.0001.mqd
CD2+
32137
ADE 2020-05-26 TALL038 24시간 후 편집 C2.0001.mqd
CD2+
33752

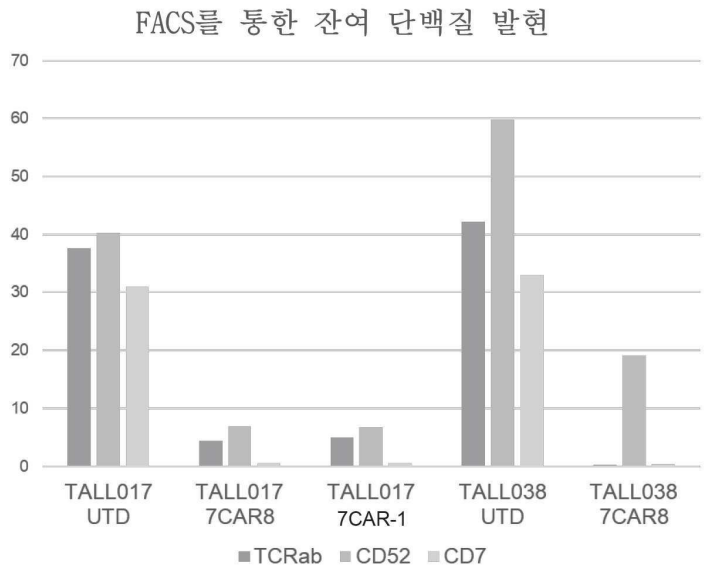
도면39ab


TALL017
7CAR-1
TALL038
UTD
B5-A :: PerCP-Vio700-A
B6-A :: PE-Vio770-A
V1-A :: VioBlue-A
B2-A :: PE-A
Q1 0
Q2 5.00
Q4 5.77E-3
Q3 95.0
Q1 0.041
Q2 42.2
Q4 0.19
Q3 57.6
Q9 0
Q10 0.56
Q12 0
Q11 99.4
Q9 0
Q10 33.0
Q12 0
Q11 67.0
Q5 0.014
Q6 6.78
Q8 0.058
Q7 93.1
Q5 0.79
Q6 59.8
Q8 0.70
Q7 38.7
ADE 2020-05-26 TALL038 24시간 후 편집 C3.0001.mqd
CD2+
34660
ADE 2020-05-26 TALL038 24시간 후 편집 C4.0001.mqd
CD2+
29267

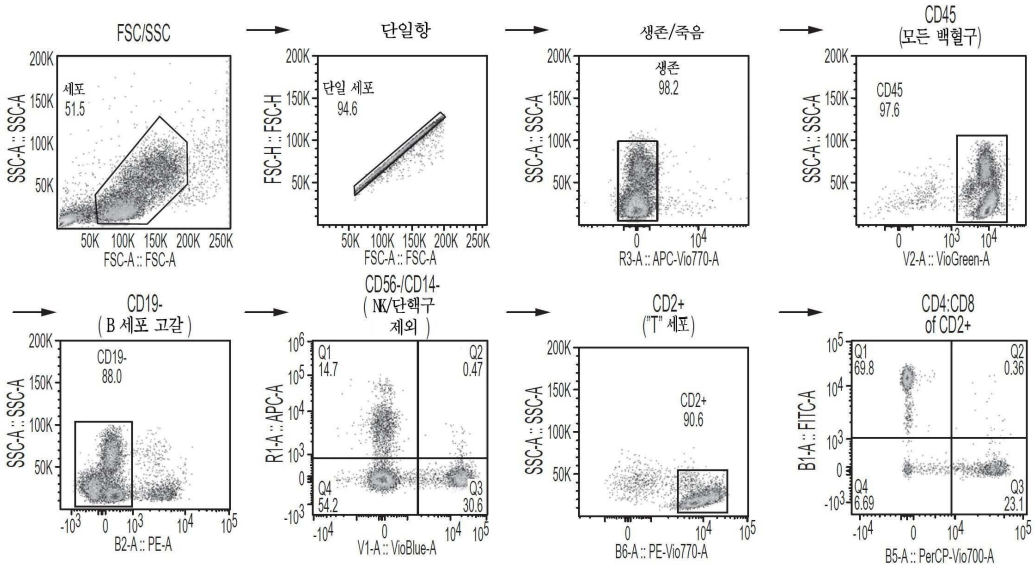
도면39ac



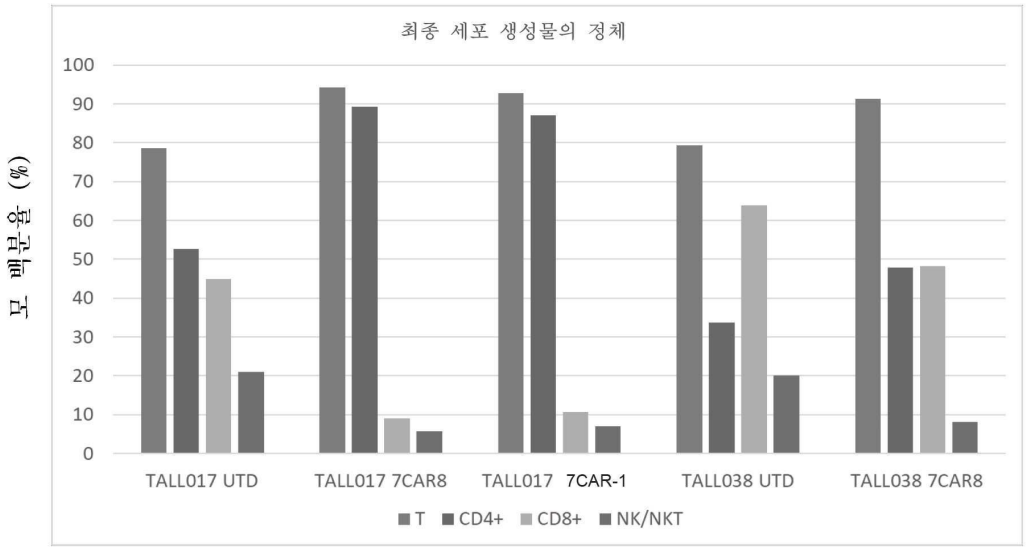

도면39b



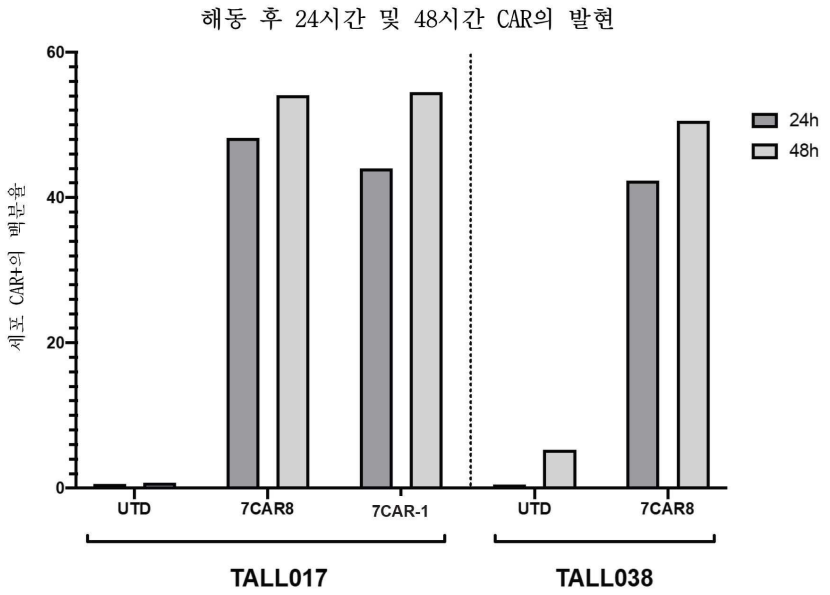
도면40


FSC/SSC
세포
51.5
SSC-A :: SSC-A
FSC-A :: FSC-A
단일항
단일 세포
94.6
FSC-H :: FSC-W
FSC-A :: FSC-A
생존/죽음
생존
98.2
R3-A :: APC-Vio770-A
CD45
(모든 백혈구)
CD45
97.6
V2-A :: VioGreen-A
CD19-
( B 세포 고갈 )
CD19-
88.0
B2-A :: PE-A
CD56-/CD14-
( NK/단핵구
제외 )
Q1
14.7
Q2
0.47
Q4
54.2
Q3
30.6
R1-A :: APC-A
V1-A :: VioBlue-A
CD2+
("T" 세포)
CD2+
90.6
B6-A :: PE-Vio770-A
CD4:CD8
of CD2+
Q1
69.8
Q2
0.36
Q4
6.69
Q3
23.1
B1-A :: FITC-A
B5-A :: PerCP-Vio700-A

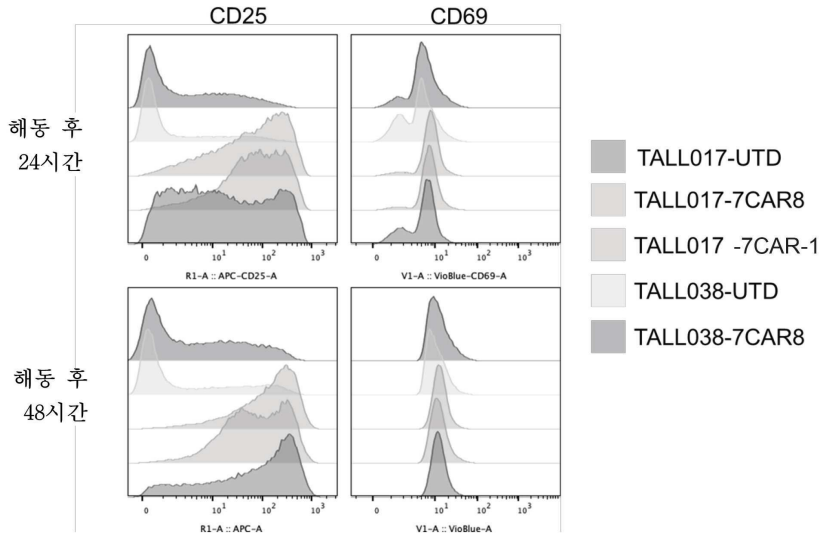
도면41


최종 세포 생성물의 정체
모 백분율 (%)
100
90
80
70
60
50
40
30
20
10
0
TALL017 UTD
TALL017 7CAR8
TALL017 7CAR-1
TALL038 UTD
TALL038 7CAR8
T
CD4+
CD8+
NK/NKT

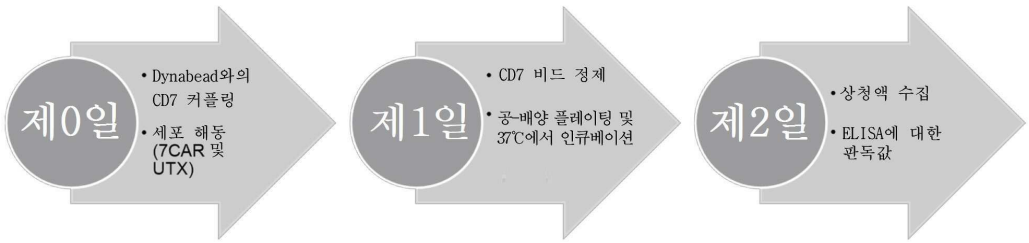
도면42



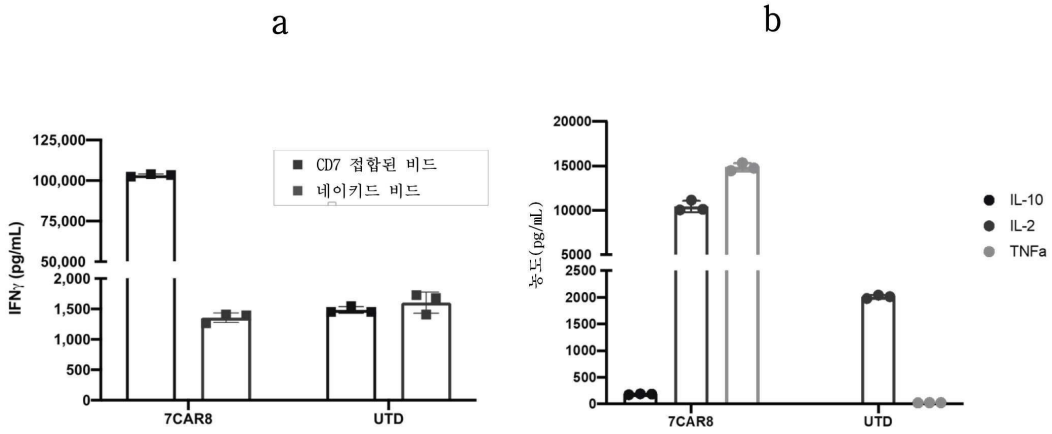

도면43



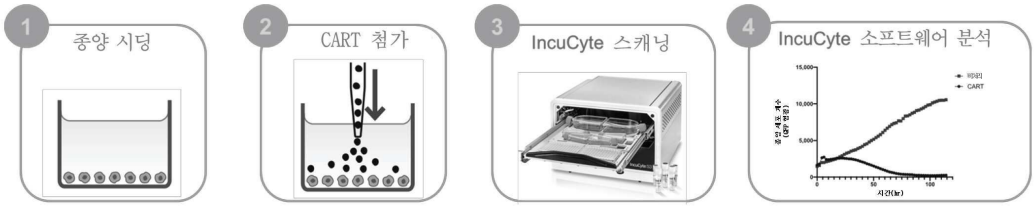

도면44



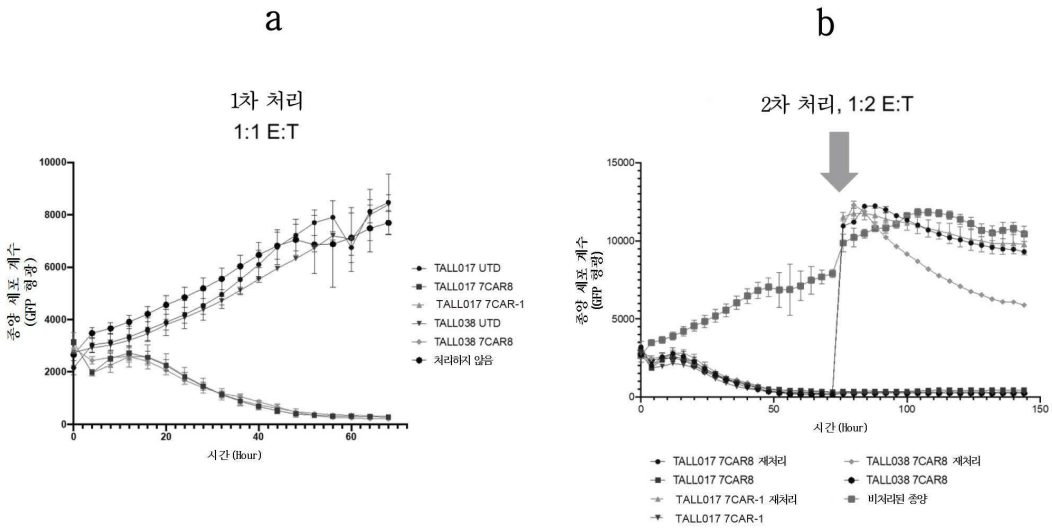
도면45



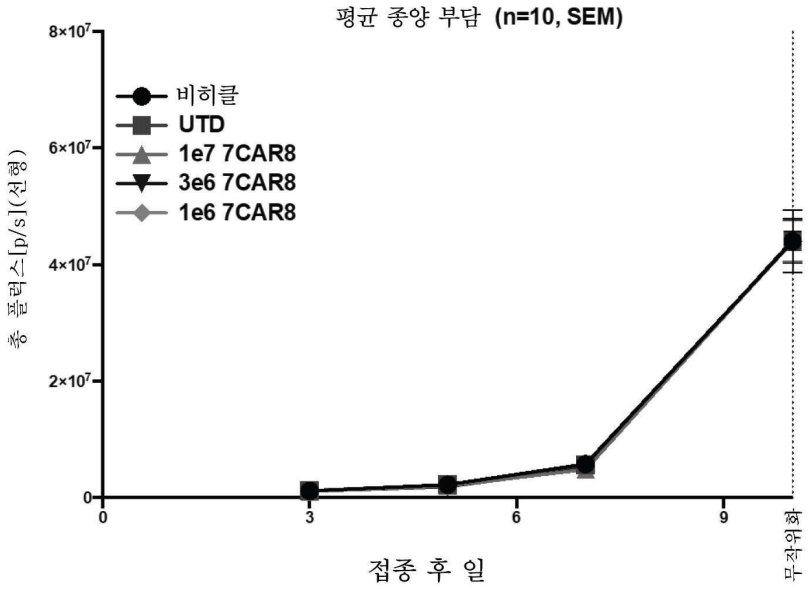

도면46



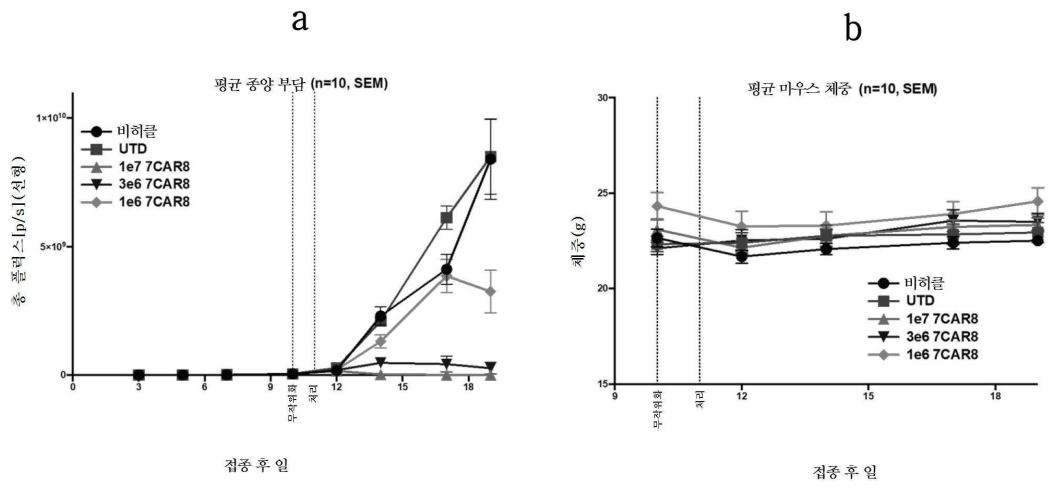

도면47



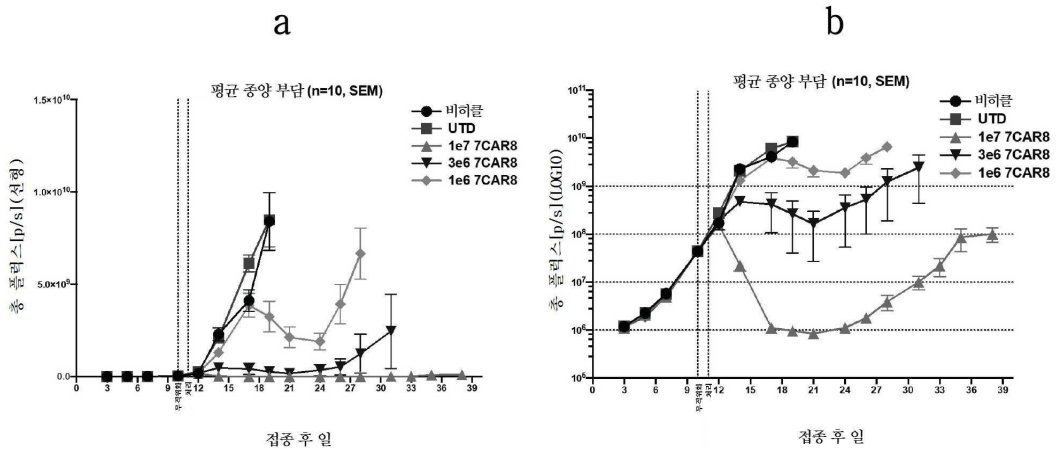
도면48


평균 종양 부담 (n=10, SEM)
비히클
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
총 플럭스[p/s](선형)
8×10^7
6×10^7
4×10^7
2×10^7
0
0
3
6
9
무작위화
접종 후 일

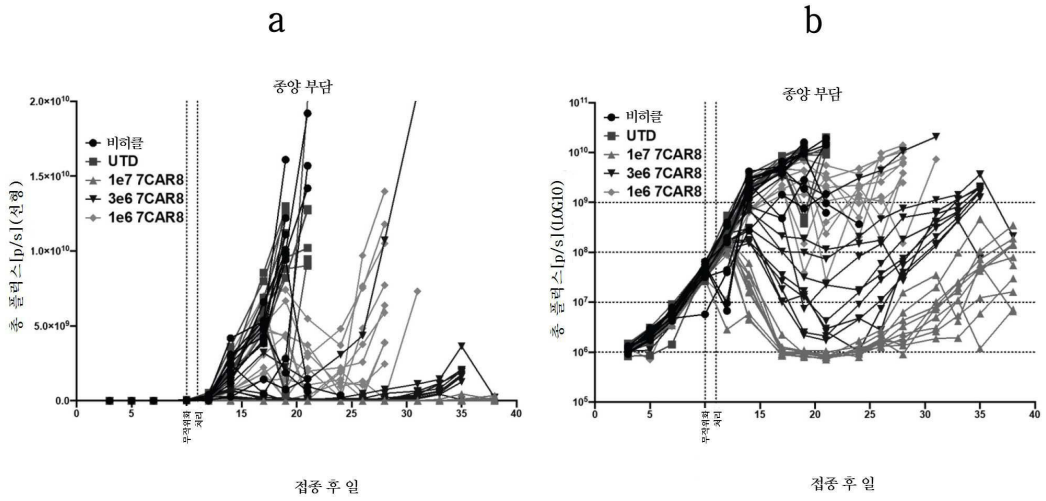
도면49


a
평균 종양 부담 (n=10, SEM)
비히클
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
총 플럭스[p/s](선형)
1×10^10
5×10^9
0
0
3
6
9
12
15
18
무작위화
처리
접종 후 일
b
평균 마우스 체중 (n=10, SEM)
체중(g)
30
25
20
15
9
12
15
18
무작위화
처리
비히클
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
접종 후 일

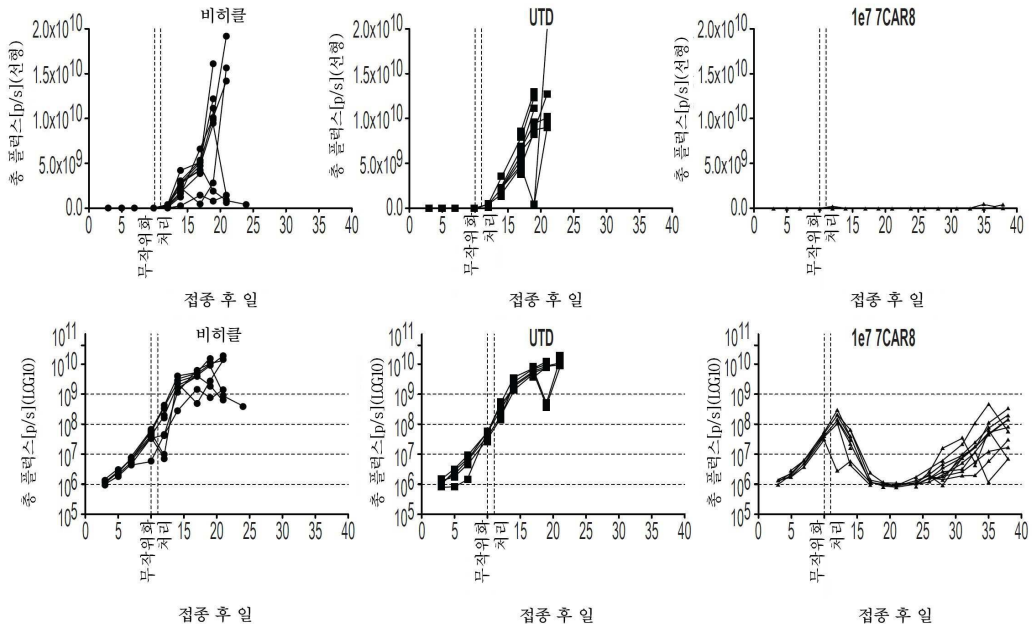
도면50


a
b
평균 종양 부담 (n=10, SEM)
비히클
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
총 플럭스[p/s](선형)
총 플럭스[p/s](LOG10)
무작위화
처리
접종 후 일

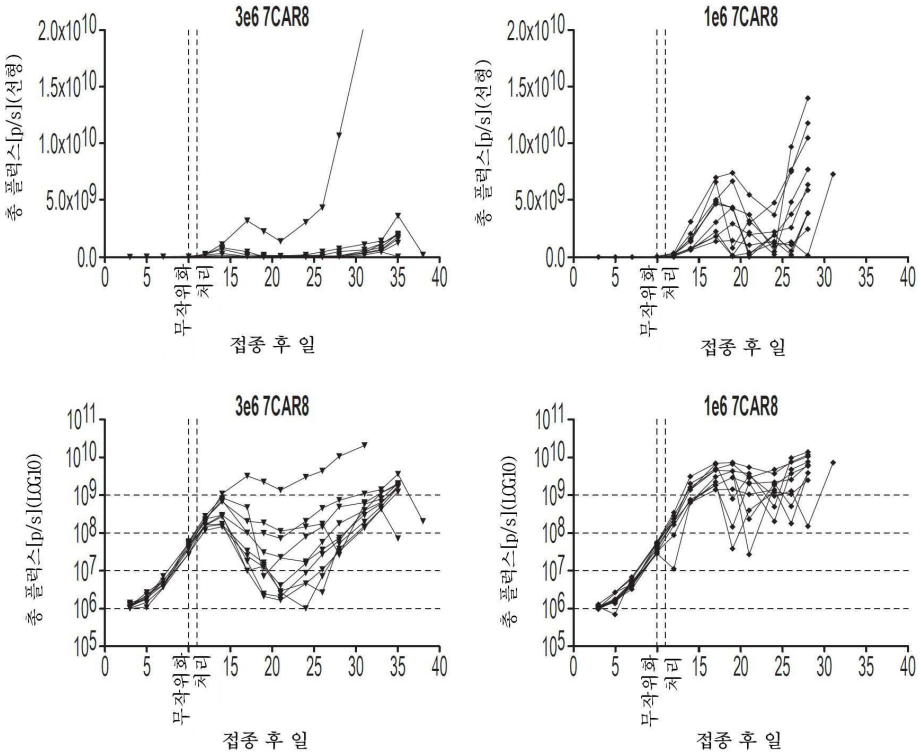
도면51


a
b
종양 부담
비히클
UTD
1e7 7CAR8
3e6 7CAR8
1e6 7CAR8
총 플럭스[p/s](선형)
총 플럭스[p/s](LOG10)
무작위화
처리
접종 후 일

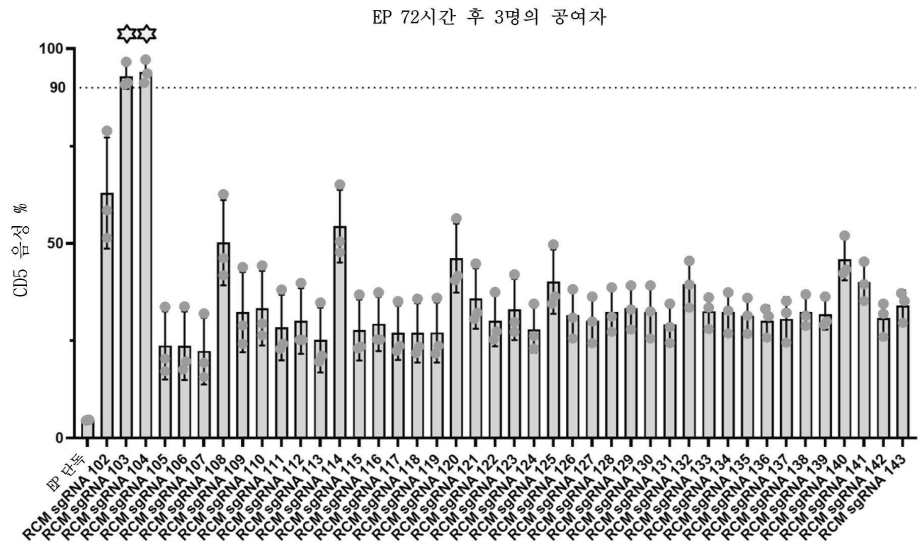
도면52a


비히클
UTD
1e7 7CAR8
총 플럭스[p/s](선형)
총 플럭스[p/s](LOG10)
무작위화 처리
접종 후 일

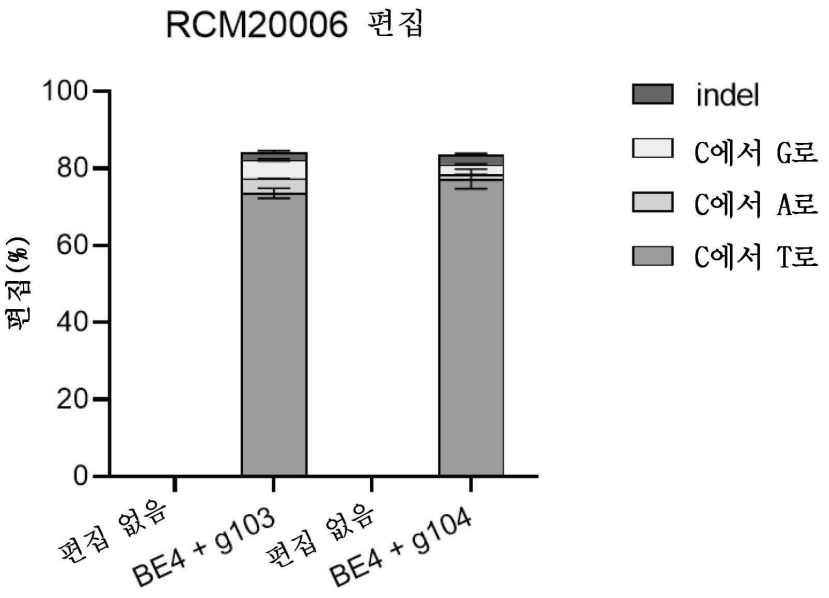
도면52b


3e6 7CAR8
1e6 7CAR8
총 플럭스[p/s](선형)
총 플럭스[p/s](LOG10)
무작위화 처리
접종 후 일

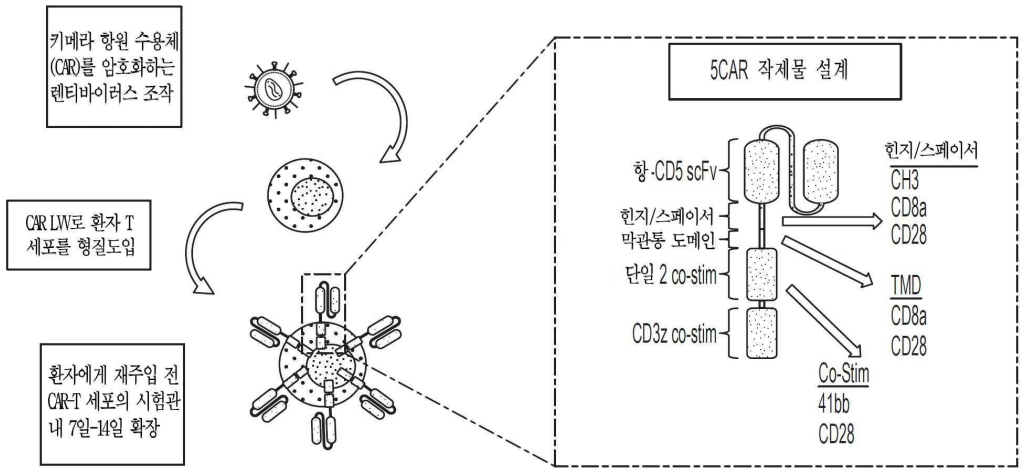
도면53


EP 72시간 후 3명의 공여자
CD5 음성 %
100
90
50
0
EP 단독
RCM sgRNA 102
RCM sgRNA 103
RCM sgRNA 104
RCM sgRNA 105
RCM sgRNA 106
RCM sgRNA 107
RCM sgRNA 108
RCM sgRNA 109
RCM sgRNA 110
RCM sgRNA 111
RCM sgRNA 112
RCM sgRNA 113
RCM sgRNA 114
RCM sgRNA 115
RCM sgRNA 116
RCM sgRNA 117
RCM sgRNA 118
RCM sgRNA 119
RCM sgRNA 120
RCM sgRNA 121
RCM sgRNA 122
RCM sgRNA 123
RCM sgRNA 124
RCM sgRNA 125
RCM sgRNA 126
RCM sgRNA 127
RCM sgRNA 128
RCM sgRNA 129
RCM sgRNA 130
RCM sgRNA 131
RCM sgRNA 132
RCM sgRNA 133
RCM sgRNA 134
RCM sgRNA 135
RCM sgRNA 136
RCM sgRNA 137
RCM sgRNA 138
RCM sgRNA 139
RCM sgRNA 140
RCM sgRNA 141
RCM sgRNA 142
RCM sgRNA 143

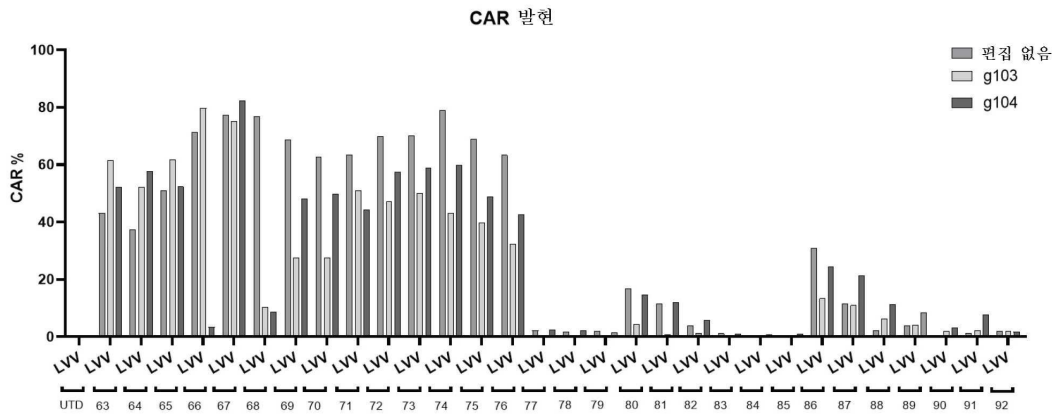
도면54


RCM20006 편집
편집 (%)
100
80
60
40
20
0
편집 없음
BE4 + g103
편집 없음
BE4 + g104
indel
C에서 G로
C에서 A로
C에서 T로

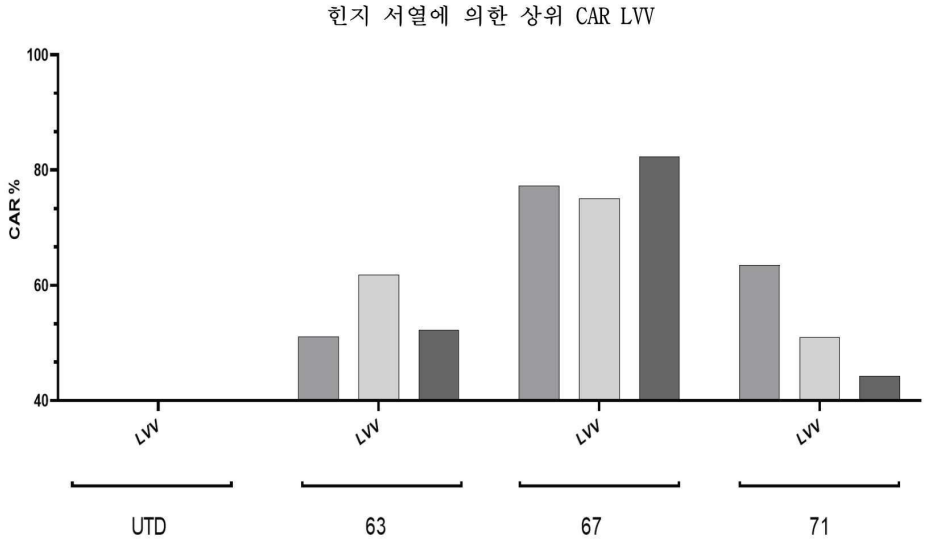
도면55



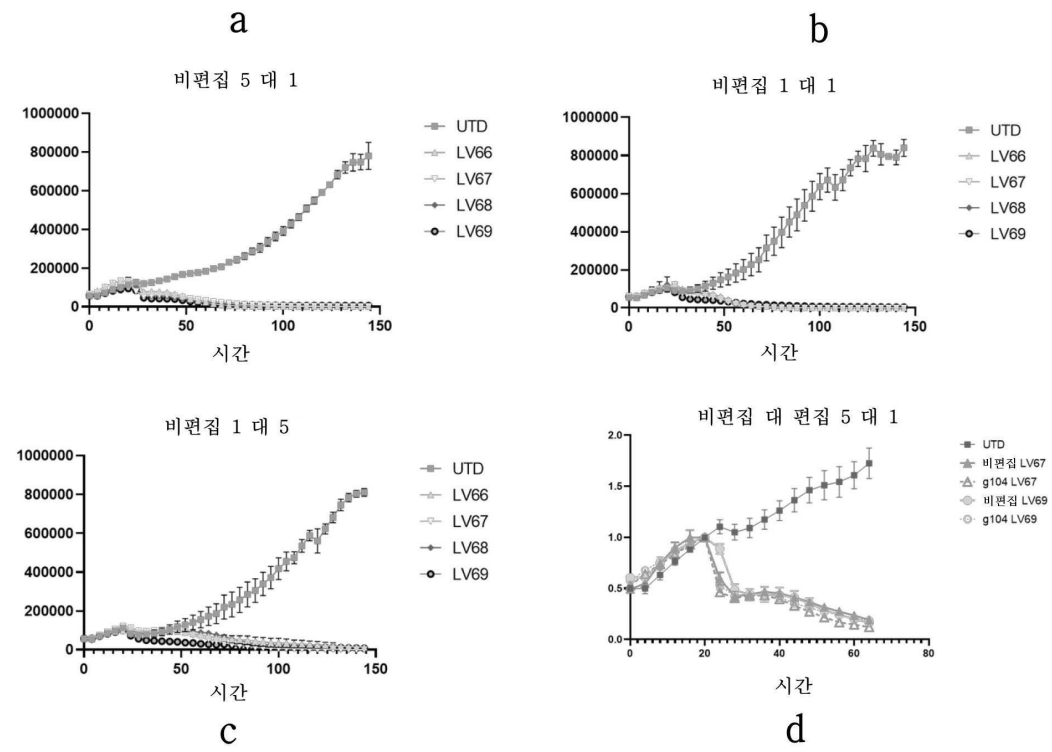

도면56



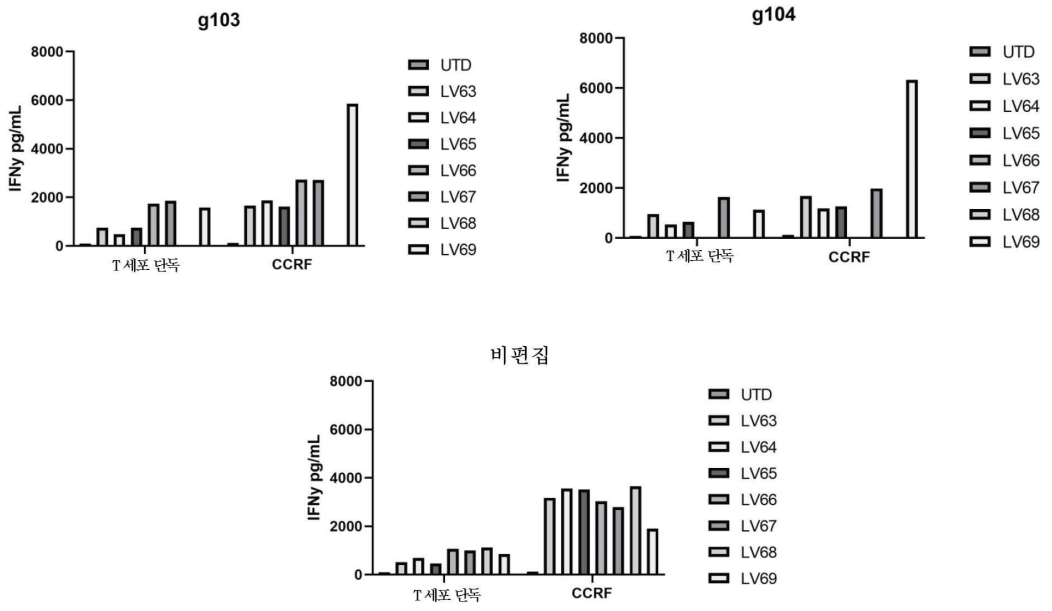

도면57



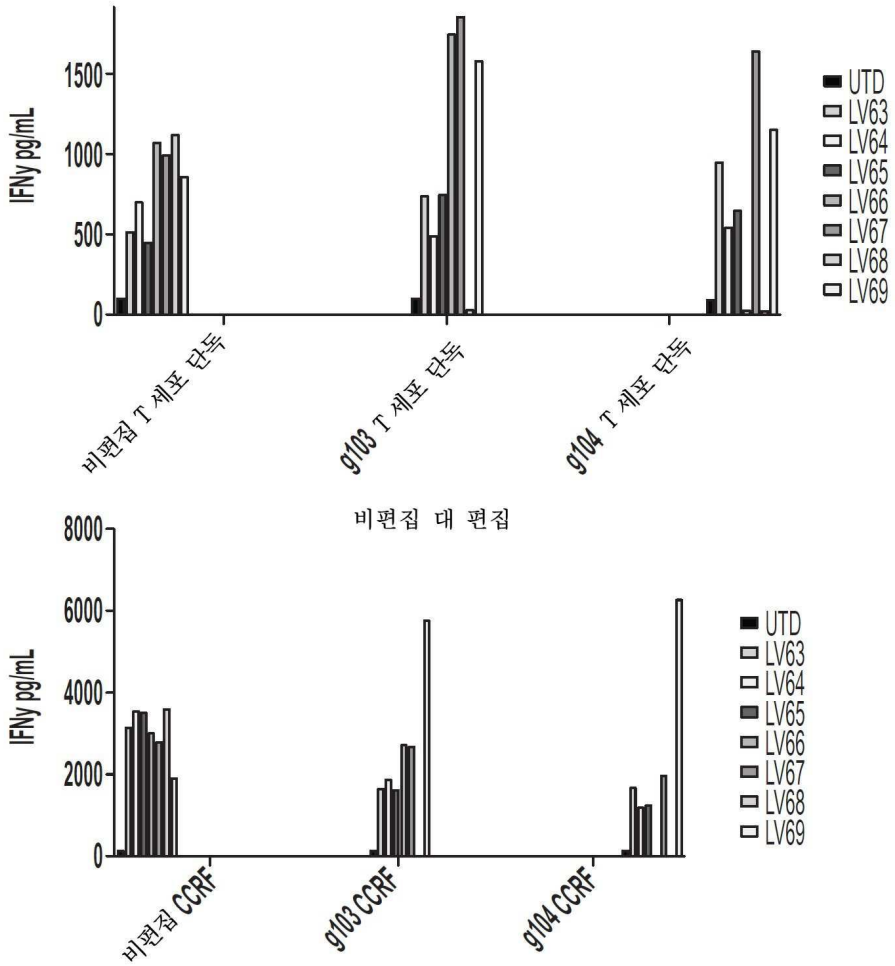
도면58


a
비편집 5 대 1
1000000
800000
600000
400000
200000
0
0
50
100
150
시간
UTD
LV66
LV67
LV68
LV69
b
비편집 1 대 1
1000000
800000
600000
400000
200000
0
0
50
100
150
시간
UTD
LV66
LV67
LV68
LV69
비편집 1 대 5
1000000
800000
600000
400000
200000
0
0
50
100
150
시간
UTD
LV66
LV67
LV68
LV69
c
비편집 대 편집 5 대 1
2.0
1.5
1.0
0.5
0.0
0
20
40
60
80
시간
UTD
비편집 LV67
g104 LV67
비편집 LV69
g104 LV69
d

도면59a


g103
IFNy pg/mL
8000
6000
4000
2000
0
T 세포 단독
CCRF
UTD
LV63
LV64
LV65
LV66
LV67
LV68
LV69
g104
IFNy pg/mL
8000
6000
4000
2000
0
T 세포 단독
CCRF
UTD
LV63
LV64
LV65
LV66
LV67
LV68
LV69
비편집
IFNy pg/mL
8000
6000
4000
2000
0
T 세포 단독
CCRF
UTD
LV63
LV64
LV65
LV66
LV67
LV68
LV69

도면59b




서 열 목 록

SEQUENCE LISTING

<110> BEAM THERAPEUTICS INC.

<120> COMPOSITIONS AND METHODS FOR TREATMENT OF LIQUID CANCERS

<130> WO2021062227

<140> PCT/US2020/052822

<141> 2020-09-25

<150> 62/907,254

<151> 2019-09-27

<160> 1512

<170> PatentIn version 3.5

<210> 1

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1

agcgacugca gaaagaagag 20

<210> 2

<211>
20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 2

cauaccagcu gagccguccg 20

<210> 3

<211> 573

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 3

Met Glu Phe Gly Leu Ser Trp Leu Phe Leu Val Ala Ile Leu Lys Gly
1 5 10 15

Val Gln Cys Ile Asp Ala Met Gly Asn Ile Gln Leu Val Gln Ser Gly

20 25 30

Pro Glu Leu Lys Lys Pro Gly Glu Thr Val Lys Ile Ser Cys Lys Ala
35 40 45

Ser Gly Tyr Thr Phe Thr Asn Tyr Gly Met Asn Trp Val Lys Gln Ala
50 55 60

Pro Gly Lys Gly Leu Arg Trp Met Gly Trp Ile Asn Thr His Thr Gly
65 70 75 80

Glu Pro Thr Tyr Ala Asp Asp Phe Lys Gly Arg Phe Ala Phe Ser Leu

85 90 95

Glu Thr Ser Ala Ser Thr Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn

100 105 110

Glu Asp Thr Ala Thr Tyr Phe Cys Thr Arg Arg Gly Tyr Asp Trp Tyr

115 120 125

Phe Asp Val Trp Gly Ala Gly Thr Thr Val Thr Val Ser Ser Gly Gly

130 135 140

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Lys

145 150 155 160

Met Thr Gln Ser Pro Ser Ser Met Tyr Ala Ser Leu Gly Glu Arg Val

165 170 175

Thr Ile Thr Cys Lys Ala Ser Gln Asp Ile Asn Ser Tyr Leu Ser Trp

180 185 190

Phe His His Lys Pro Gly Lys Ser Pro Lys Thr Leu Ile Tyr Arg Ala

195 200 205

Asn Arg Leu Val Asp Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser

210 215 220

Gly Gln Asp Tyr Ser Leu Thr Ile Ser Ser Leu Asp Tyr Glu Asp Met

225 230 235 240

Gly Ile Tyr Tyr Cys Gln Gln Tyr Asp Glu Ser Pro Trp Thr Phe Gly

245 250 255

Gly Gly Thr Lys Leu Glu Met Lys Gly Ser Gly Asp Pro Ala Glu Pro

260 265 270

Lys Ser Pro Asp Lys Thr His Thr Cys Pro Gly Gln Pro Arg Glu Pro

275 280 285

Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln

290 295 300

Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala

305 310 315 320

Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr

325 330 335

Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu

340 345 350

Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser

355 360 365

Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser

370 375 380

Leu Ser Pro Gly Lys Lys Asp Pro Lys Phe Trp Val Leu Val Val Val

385 390 395 400

Gly Gly Val Leu Ala Cys Tyr Ser Leu Leu Val Thr Val Ala Phe Ile

405 410 415

Ile Phe Trp Val Arg Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr

420 425 430

Met Asn Met Thr Pro Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln

435 440 445

Pro Tyr Ala Pro Pro Arg Asp Phe Ala Ala Tyr Arg Ser Arg Val Lys

450 455 460

Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln Gly Gln Asn Gln

465 470 475 480

Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr Asp Val Leu

485 490 495

Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys Pro Arg Arg

500 505 510

Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys Asp Lys Met

515 520 525

Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg Arg Arg Gly

530 535 540

Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala Thr Lys Asp

545 550 555 560

Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg

565 570

<210> 4

<211> 495

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 4

Met Glu Phe Gly Leu Ser Trp Leu Phe Leu Val Ala Ile Leu Lys Gly

1 5 10 15

Val Gln Cys Ile Asp Ala Met Gly Asn Ile Gln Leu Val Gln Ser Gly

20 25 30

Pro Glu Leu Lys Lys Pro Gly Glu Thr Val Lys Ile Ser Cys Lys Ala

35 40 45

Ser Gly Tyr Thr Phe Thr Asn Tyr Gly Met Asn Trp Val Lys Gln Ala

50 55 60

Pro Gly Lys Gly Leu Arg Trp Met Gly Trp Ile Asn Thr His Thr Gly

65 70 75 80

Glu Pro Thr Tyr Ala Asp Asp Phe Lys Gly Arg Phe Ala Phe Ser Leu

85 90 95

Glu Thr Ser Ala Ser Thr Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn

100 105 110

Glu Asp Thr Ala Thr Tyr Phe Cys Thr Arg Arg Gly Tyr Asp Trp Tyr

115 120 125

Phe Asp Val Trp Gly Ala Gly Thr Thr Val Thr Val Ser Ser Gly Gly

130 135 140

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Lys

145 150 155 160

Met Thr Gln Ser Pro Ser Ser Met Tyr Ala Ser Leu Gly Glu Arg Val

165 170 175

Thr Ile Thr Cys Lys Ala Ser Gln Asp Ile Asn Ser Tyr Leu Ser Trp

180 185 190

Phe His His Lys Pro Gly Lys Ser Pro Lys Thr Leu Ile Tyr Arg Ala

195 200 205

Asn Arg Leu Val Asp Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser

210 215 220

Gly Gln Asp Tyr Ser Leu Thr Ile Ser Ser Leu Asp Tyr Glu Asp Met

225 230 235 240

Gly Ile Tyr Tyr Cys Gln Gln Tyr Asp Glu Ser Pro Trp Thr Phe Gly

245 250 255

Gly Gly Thr Lys Leu Glu Met Lys Gly Ser Gly Asp Pro Ala Thr Thr

260 265 270

Thr Pro Ala Pro Arg Pro Pro Thr Pro Ala Pro Thr Ile Ala Ser Gln

275 280 285

Pro Leu Ser Leu Arg Pro Glu Ala Cys Arg Pro Ala Ala Gly Gly Ala

290 295 300

Val His Thr Arg Gly Leu Asp Phe Ala Cys Asp Phe Trp Val Leu Val

305 310 315 320

Val Val Gly Gly Val Leu Ala Cys Tyr Ser Leu Leu Val Thr Val Ala

325 330 335

Phe Ile Ile Phe Trp Val Arg Ser Lys Arg Ser Arg Leu Leu His Ser

340 345 350

Asp Tyr Met Asn Met Thr Pro Arg Arg Pro Gly Pro Thr Arg Lys His

355 360 365

Tyr Gln Pro Tyr Ala Pro Pro Arg Asp Phe Ala Ala Tyr Arg Ser Arg

370 375 380

Val Lys Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln Gly Gln

385 390 395 400

Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr Asp

405 410 415

Val Leu Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys Pro

420 425 430

Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys Asp

435 440 445

Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg Arg

450 455 460

Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala Thr

465 470 475 480

Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg

485 490 495

<210> 5

<211> 489

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 5

Met Glu Phe Gly Leu Ser Trp Leu Phe Leu Val Ala Ile Leu Lys Gly

1 5 10 15

Val Gln Cys Ile Asp Ala Met Gly Asn Ile Gln Leu Val Gln Ser Gly

20 25 30

Pro Glu Leu Lys Lys Pro Gly Glu Thr Val Lys Ile Ser Cys Lys Ala

35 40 45

Ser Gly Tyr Thr Phe Thr Asn Tyr Gly Met Asn Trp Val Lys Gln Ala

50 55 60

Pro Gly Lys Gly Leu Arg Trp Met Gly Trp Ile Asn Thr His Thr Gly

65 70 75 80

Glu Pro Thr Tyr Ala Asp Asp Phe Lys Gly Arg Phe Ala Phe Ser Leu

85 90 95

Glu Thr Ser Ala Ser Thr Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn

100 105 110

Glu Asp Thr Ala Thr Tyr Phe Cys Thr Arg Arg Gly Tyr Asp Trp Tyr

115 120 125

Phe Asp Val Trp Gly Ala Gly Thr Thr Val Thr Val Ser Ser Gly Gly

130 135 140

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Lys
145 150 155 160
Met Thr Gln Ser Pro Ser Ser Met Tyr Ala Ser Leu Gly Glu Arg Val
165 170 175
Thr Ile Thr Cys Lys Ala Ser Gln Asp Ile Asn Ser Tyr Leu Ser Trp
180 185 190
Phe His His Lys Pro Gly Lys Ser Pro Lys Thr Leu Ile Tyr Arg Ala

195 200 205
Asn Arg Leu Val Asp Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser
210 215 220
Gly Gln Asp Tyr Ser Leu Thr Ile Ser Ser Leu Asp Tyr Glu Asp Met
225 230 235 240
Gly Ile Tyr Tyr Cys Gln Gln Tyr Asp Glu Ser Pro Trp Thr Phe Gly
245 250 255
Gly Gly Thr Lys Leu Glu Met Lys Gly Ser Gly Asp Pro Ala Ile Glu

260 265 270
Val Met Tyr Pro Pro Pro Tyr Leu Asp Asn Glu Lys Ser Asn Gly Thr
275 280 285
Ile Ile His Val Lys Gly Lys His Leu Cys Pro Ser Pro Leu Phe Pro
290 295 300
Gly Pro Ser Lys Pro Phe Trp Val Leu Val Val Val Gly Gly Val Leu
305 310 315 320
Ala Cys Tyr Ser Leu Leu Val Thr Val Ala Phe Ile Ile Phe Trp Val

325 330 335
Arg Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr Met Asn Met Thr
340 345 350
Pro Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr Ala Pro
355 360 365
Pro Arg Asp Phe Ala Ala Tyr Arg Ser Arg Val Lys Phe Ser Arg Ser
370 375 380
Ala Asp Ala Pro Ala Tyr Gln Gln Gly Gln Asn Gln Leu Tyr Asn Glu

385 390 395 400

Leu Asn Leu Gly Arg Arg Glu Glu Tyr Asp Val Leu Asp Lys Arg Arg

405 410 415

Gly Arg Asp Pro Glu Met Gly Gly Lys Pro Arg Arg Lys Asn Pro Gln

420 425 430

Glu Gly Leu Tyr Asn Glu Leu Gln Lys Asp Lys Met Ala Glu Ala Tyr

435 440 445

Ser Glu Ile Gly Met Lys Gly Glu Arg Arg Arg Gly Lys Gly His Asp

450 455 460

Gly Leu Tyr Gln Gly Leu Ser Thr Ala Thr Lys Asp Thr Tyr Asp Ala

465 470 475 480

Leu His Met Gln Ala Leu Pro Pro Arg

485

<210> 6

<211> 617

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 6

Met Glu Phe Gly Leu Ser Trp Leu Phe Leu Val Ala Ile Leu Lys Gly

1 5 10 15

Val Gln Cys Ile Asp Ala Met Gly Asn Ile Gln Leu Val Gln Ser Gly

20 25 30

Pro Glu Leu Lys Lys Pro Gly Glu Thr Val Lys Ile Ser Cys Lys Ala

35 40 45

Ser Gly Tyr Thr Phe Thr Asn Tyr Gly Met Asn Trp Val Lys Gln Ala

50 55 60

Pro Gly Lys Gly Leu Arg Trp Met Gly Trp Ile Asn Thr His Thr Gly

65 70 75 80

Glu Pro Thr Tyr Ala Asp Asp Phe Lys Gly Arg Phe Ala Phe Ser Leu

85 90 95

Glu Thr Ser Ala Ser Thr Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn
100 105 110
Glu Asp Thr Ala Thr Tyr Phe Cys Thr Arg Arg Gly Tyr Asp Trp Tyr
115 120 125
Phe Asp Val Trp Gly Ala Gly Thr Thr Val Thr Val Ser Ser Gly Gly
130 135 140

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Lys
145 150 155 160
Met Thr Gln Ser Pro Ser Ser Met Tyr Ala Ser Leu Gly Glu Arg Val
165 170 175
Thr Ile Thr Cys Lys Ala Ser Gln Asp Ile Asn Ser Tyr Leu Ser Trp
180 185 190
Phe His His Lys Pro Gly Lys Ser Pro Lys Thr Leu Ile Tyr Arg Ala
195 200 205

Asn Arg Leu Val Asp Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser
210 215 220
Gly Gln Asp Tyr Ser Leu Thr Ile Ser Ser Leu Asp Tyr Glu Asp Met
225 230 235 240
Gly Ile Tyr Tyr Cys Gln Gln Tyr Asp Glu Ser Pro Trp Thr Phe Gly
245 250 255
Gly Gly Thr Lys Leu Glu Met Lys Gly Ser Gly Asp Pro Ala Glu Pro
260 265 270

Lys Ser Pro Asp Lys Thr His Thr Cys Pro Gly Gln Pro Arg Glu Pro
275 280 285
Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys Asn Gln
290 295 300
Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala
305 310 315 320
Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr
325 330 335

Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu

340 345 350

Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser

355 360 365

Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser

370 375 380

Leu Ser Pro Gly Lys Lys Asp Pro Lys Pro Thr Thr Thr Pro Ala Pro

385 390 395 400

Arg Pro Pro Thr Pro Ala Pro Thr Ile Ala Ser Gln Pro Leu Ser Leu

405 410 415

Arg Pro Glu Ala Cys Arg Pro Ala Ala Gly Gly Ala Val His Thr Arg

420 425 430

Gly Leu Asp Phe Ala Cys Asp Ile Tyr Ile Trp Ala Pro Leu Ala Gly

435 440 445

Thr Cys Gly Val Leu Leu Leu Ser Leu Val Ile Thr Leu Tyr Cys Lys

450 455 460

Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe Lys Gln Pro Phe Met Arg

465 470 475 480

Pro Val Gln Thr Thr Gln Glu Glu Asp Gly Cys Ser Cys Arg Phe Pro

485 490 495

Glu Glu Glu Glu Gly Gly Cys Glu Leu Arg Val Lys Phe Ser Arg Ser

500 505 510

Ala Asp Ala Pro Ala Tyr Gln Gln Gly Gln Asn Gln Leu Tyr Asn Glu

515 520 525

Leu Asn Leu Gly Arg Arg Glu Glu Tyr Asp Val Leu Asp Lys Arg Arg

530 535 540

Gly Arg Asp Pro Glu Met Gly Gly Lys Pro Arg Arg Lys Asn Pro Gln

545 550 555 560

Glu Gly Leu Tyr Asn Glu Leu Gln Lys Asp Lys Met Ala Glu Ala Tyr

565 570 575

Ser Glu Ile Gly Met Lys Gly Glu Arg Arg Arg Gly Lys Gly His Asp

580 585 590

Gly Leu Tyr Gln Gly Leu Ser Thr Ala Thr Lys Asp Thr Tyr Asp Ala
595 600 605

Leu His Met Gln Ala Leu Pro Pro Arg
610 615

<210> 7

<211> 490

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MOD_RES

<222> (490)..(490)

<223> Any amino acid

<400> 7

Met Glu Phe Gly Leu Ser Trp Leu Phe Leu Val Ala Ile Leu Lys Gly
1 5 10 15

Val Gln Cys Ile Asp Ala Met Gly Asn Ile Gln Leu Val Gln Ser Gly
20 25 30

Pro Glu Leu Lys Lys Pro Gly Glu Thr Val Lys Ile Ser Cys Lys Ala
35 40 45

Ser Gly Tyr Thr Phe Thr Asn Tyr Gly Met Asn Trp Val Lys Gln Ala
50 55 60

Pro Gly Lys Gly Leu Arg Trp Met Gly Trp Ile Asn Thr His Thr Gly
65 70 75 80

Glu Pro Thr Tyr Ala Asp Asp Phe Lys Gly Arg Phe Ala Phe Ser Leu
85 90 95

Glu Thr Ser Ala Ser Thr Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn
100 105 110

Glu Asp Thr Ala Thr Tyr Phe Cys Thr Arg Arg Gly Tyr Asp Trp Tyr
115 120 125

Phe Asp Val Trp Gly Ala Gly Thr Thr Val Thr Val Ser Ser Gly Gly
130 135 140

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Lys
145 150 155 160
Met Thr Gln Ser Pro Ser Ser Met Tyr Ala Ser Leu Gly Glu Arg Val
165 170 175
Thr Ile Thr Cys Lys Ala Ser Gln Asp Ile Asn Ser Tyr Leu Ser Trp
180 185 190
Phe His His Lys Pro Gly Lys Ser Pro Lys Thr Leu Ile Tyr Arg Ala
195 200 205

Asn Arg Leu Val Asp Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser
210 215 220
Gly Gln Asp Tyr Ser Leu Thr Ile Ser Ser Leu Asp Tyr Glu Asp Met
225 230 235 240
Gly Ile Tyr Tyr Cys Gln Gln Tyr Asp Glu Ser Pro Trp Thr Phe Gly
245 250 255
Gly Gly Thr Lys Leu Glu Met Lys Gly Ser Gly Asp Pro Ala Thr Thr
260 265 270

Thr Pro Ala Pro Arg Pro Pro Thr Pro Ala Pro Thr Ile Ala Ser Gln
275 280 285
Pro Leu Ser Leu Arg Pro Glu Ala Cys Arg Pro Ala Ala Gly Gly Ala
290 295 300
Val His Thr Arg Gly Leu Asp Phe Ala Cys Asp Ile Tyr Ile Trp Ala
305 310 315 320
Pro Leu Ala Gly Thr Cys Gly Val Leu Leu Leu Ser Leu Val Ile Thr
325 330 335

Leu Tyr Cys Lys Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe Lys Gln
340 345 350
Pro Phe Met Arg Pro Val Gln Thr Thr Gln Glu Glu Asp Gly Cys Ser
355 360 365
Cys Arg Phe Pro Glu Glu Glu Glu Gly Gly Cys Glu Leu Arg Val Lys
370 375 380
Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln Gly Gln Asn Gln

385 390 395 400

Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr Asp Val Leu
405 410 415
Asp Lys Arg Arg Asp Pro Glu Met Gly Gly Lys Pro Arg Arg Lys Asn
420 425 430
Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys Asp Lys Met Ala Glu
435 440 445
Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg Arg Arg Gly Lys Gly
450 455 460

His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala Thr Lys Asp Asp Ala
465 470 475 480
Leu His Met Gln Ala Leu Pro Pro Arg Xaa
485 490

<210> 8
<211> 167
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 8

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45
Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp

165

<210> 9

<211> 184

<212> PRT

<213> Homo sapiens

<400> 9

Met Leu Gln Met Ala Gly Gln Cys Ser Gln Asn Glu Tyr Phe Asp Ser

1 5 10 15

Leu Leu His Ala Cys Ile Pro Cys Gln Leu Arg Cys Ser Ser Asn Thr

20 25 30

Pro Pro Leu Thr Cys Gln Arg Tyr Cys Asn Ala Ser Val Thr Asn Ser

35 40 45

Val Lys Gly Thr Asn Ala Ile Leu Trp Thr Cys Leu Gly Leu Ser Leu

50 55 60

Ile Ile Ser Leu Ala Val Phe Val Leu Met Phe Leu Leu Arg Lys Ile

65 70 75 80

Asn Ser Glu Pro Leu Lys Asp Glu Phe Lys Asn Thr Gly Ser Gly Leu

85 90 95

Leu Gly Met Ala Asn Ile Asp Leu Glu Lys Ser Arg Thr Gly Asp Glu

100 105 110

Ile Ile Leu Pro Arg Gly Leu Glu Tyr Thr Val Glu Glu Cys Thr Cys

115 120 125

Glu Asp Cys Ile Lys Ser Lys Pro Lys Val Asp Ser Asp His Cys Phe

130 135 140

Pro Leu Pro Ala Met Glu Glu Gly Ala Thr Ile Leu Val Thr Thr Lys

145 150 155 160

Thr Asn Asp Tyr Cys Lys Ser Leu Pro Ala Ala Leu Ser Ala Thr Glu

165 170 175

Ile Glu Lys Ser Ile Ser Ala Arg

180

<210> 10

<211> 994

<212> DNA

<213> Homo sapiens

<400> 10

aagactcaaa cttagaaact tgaattagat gtggtattca aatccttagc tgccgcgaag 60

acacagacag cccccgtaag aacccacgaa gcaggcgaag ttcattgttc tcaacattct 120

agctgctctt gctgcatttg ctctggaatt cttgtagaga tattacttgt ccttccaggc 180

tgttctttct gtagctccct tgttttcttt ttgtgatcat gttgcagatg gctgggcagt 240

gctcccaaaa tgaatatttt gacagtttgt tgcatgcttg cataccttgt caacttcgat 300

gttcttctaa tactcctcct ctaacatgtc agcgttattg taatgcaagt gtgaccaatt 360

cagtgaaagg aacgaatgcg attctctgga cctgtttggg actgagctta ataatttctt 420

tggcagtttt cgtgctaatg tttttgctaa ggaagataaa ctctgaacca ttaaaggacg 480

agtttaaaaa cacaggatca ggtctcctgg gcatggctaa cattgacctg gaaaagagca 540

ggactggtga tgaaattatt cttccgagag gcctcgagta cacggtggaa gaatgcacct 600

gtgaagactg catcaagagc aaaccgaagg tcgactctga ccattgcttt ccactcccag 660

ctatggagga aggcgcaacc attcttgtca ccacgaaaac gaatgactat tgcaagagcc 720

tgccagctgc tttgagtgct acggagatag agaaatcaat ttctgctagg taattaacca 780

tttcgactcg agcagtgcca ctttaaaaat cttttgtcag aatagatgat gtgtcagatc 840

tctttaggat gactgtattt ttcagttgcc gatacagctt tttgtcctct aactgtggaa 900

actctttatg ttagatatat ttctctaggt tactgttggg agcttaatgg tagaaacttc 960

cttggtttca tgattaaact ctttttttc ctga 994

<210> 11

<211> 8811

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 11

atatgccaag tacgcccccct attgacgtca atgacggtaa atggcccgcc tggcattatg 60

cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg 120

ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact 180

cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa 240

atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta 300

ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct 360

agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat gaaacggaca 420

gccgacggaa gcgagttcga gtcaccaaag aagaagcgga aagtctctga agtcgagttt 480

agccacgagt attggatgag gcacgcactg accctggcaa agcgagcatg ggatgaaaga 540

gaagtccccg tgggcgccgt gctggtgcac aacaatagag tgatcggaga gggatggaac 600

aggccaatcg gccgccacga ccctaccgca cacgcagaga tcatggcact gaggcaggga 660

ggcctggtca tgcagaatta ccgcctgatc gatgccaccc tgtatgtgac actggagcca 720

tgcgtgatgt gcgcaggagc aatgatccac agcaggatcg gaagagtggt gttcggagca 780

cgggacgcca agaccggcgc agcaggctcc ctgatggatg tgctgcacca ccccggcatg 840

aaccaccggg tggagatcac agagggaatc ctggcagacg agtgcgccgc cctgctgagc 900

gatttcttta gaatgcggag acaggagatc aaggcccaga agaaggcaca gagctccacc 960

gactctggag gatctagcgg aggatcctct ggaagcgaga caccaggcac aagcgagtcc 1020

gccacaccag agagctccgg cggctcctcc ggaggatcct ctgaggtgga gttttcccac 1080

gagtactgga tgagacatgc cctgaccctg gccaagaggg cacgcgatga gagggaggtg 1140

cctgtgggag ccgtgctggt gctgaacaat agagtgatcg gcgagggctg gaacagagcc 1200

atcggcctgc acgacccaac agcccatgcc gaaattatgg ccctgagaca gggcggcctg 1260

gtcatgcaga actacagact gattgacgcc accctgtacg tgacattcga gccttgcgtg 1320

atgtgcgccg gcgccatgat ccactctagg atcggccgcg tggtgtttgg cgtgaggaac 1380

gcaaaaaccg gcgccgcagg ctccctgatg gacgtgctgc actaccccgg catgaatcac 1440

cgcgtcgaaa ttaccgaggg aatcctggca gatgaatgtg ccgccctgct gtgctatttc 1500

tttcggatgc ctagacaggt gttcaatgct cagaagaagg cccagagctc caccgactcc 1560

ggaggatcta gcggaggctc ctctggctct gagacacctg gcacaagcga gagcgcaaca 1620

cctgaaagca gcgggggcag cagcgggggg tcagacaaga agtacagcat cggcctggcc 1680

atcggcacca actctgtggg ctgggccgtg atcaccgacg agtacaaggt gcccagcaag 1740

aaattcaagg tgctgggcaa caccgaccgg cacagcatca agaagaacct gatcggagcc 1800

ctgctgttcg acagcggcga aacagccgag gccacccggc tgaagagaac cgccagaaga 1860

agatacacca gacggaagaa ccggatctgc tatctgcaag agatcttcag caacgagatg 1920

gccaaggtgg acgacagctt cttccacaga ctggaagagt ccttcctggt ggaagaggat 1980

aagaagcacg agcggcaccc catcttcggc aacatcgtgg acgaggtggc ctaccacgag 2040

aagtacccca ccatctacca cctgagaaag aaactggtgg acagcaccga caaggccgac 2100

ctgcggctga tctatctggc cctggcccac atgatcaagt tccggggcca cttcctgatc 2160

gagggcgacc tgaaccccga caacagcgac gtggacaagc tgttcatcca gctggtgcag 2220

acctacaacc agctgttcga ggaaaacccc atcaacgcca gcggcgtgga cgccaaggcc 2280

atcctgtctg ccagactgag caagagcaga cggctggaaa atctgatcgc ccagctgccc 2340

ggcgagaaga agaatggcct gttcggaaac ctgattgccc tgagcctggg cctgaccccc 2400

aacttcaaga gcaacttcga cctggccgag gatgccaaac tgcagctgag caaggacacc 2460

tacgacgacg acctggacaa cctgctggcc cagatcggcg accagtacgc cgacctgttt 2520

ctggccgcca agaacctgtc cgacgccatc ctgctgagcg acatcctgag agtgaacacc 2580

gagatcacca aggccccct gagcgcctct atgatcaaga gatacgacga gcaccaccag 2640

gacctgaccc tgctgaaagc tctcgtgcgg cagcagctgc ctgagaagta caaagagatt 2700

ttcttcgacc agagcaagaa cggctacgcc ggctacattg acggcggagc cagccaggaa 2760

gagttctaca agttcatcaa gcccatcctg gaaaagatgg acggcaccga ggaactgctc 2820

gtgaagctga acagagagga cctgctgcgg aagcagcgga ccttcgacaa cggcagcatc 2880

ccccaccaga tccacctggg agagctgcac gccattctgc ggcggcagga agatttttac 2940

ccattcctga aggacaaccg ggaaaagatc gagaagatcc tgaccttccg catcccctac 3000

tacgtgggcc ctctggccag gggaaacagc agattcgcct ggatgaccag aaagagcgag 3060

gaaaccatca ccccctggaa cttcgaggaa gtggtggaca agggcgcttc cgcccagagc 3120

ttcatcgagc ggatgaccaa cttcgataag aacctgccca acgagaaggt gctgcccaag 3180

cacagcctgc tgtacgagta cttcaccgtg tataacgagc tgaccaaagt gaaatacgtg 3240

accgagggaa tgagaaagcc cgccttcctg agcggcgagc agaaaaaggc catcgtggac 3300

ctgctgttca agaccaaccg gaaagtgacc gtgaagcagc tgaaagagga ctacttcaag 3360

aaaatcgagt gcttcgactc cgtggaaatc tccggcgtgg aagatcggtt caacgcctcc 3420
ctgggcacat accacgatct gctgaaaatt atcaaggaca aggacttcct ggacaatgag 3480
gaaaacgagg acattctgga agatatcgtg ctgaccctga cactgtttga ggacagagag 3540
atgatcgagg aacggctgaa aacctatgcc cacctgttcg acgacaaagt gatgaagcag 3600
ctgaagcggc ggagatacac cggctggggc aggctgagcc ggaagctgat caacggcatc 3660
cgggacaagc agtccggcaa gacaatcctg gatttcctga agtccgacgg cttcgccaac 3720
agaaacttca tgcagctgat ccacgacgac agcctgacct ttaaagagga catccagaaa 3780

gcccaggtgt ccggccaggg cgatagcctg cacgagcaca ttgccaatct ggccggcagc 3840
cccgccatta agaagggcat cctgcagaca gtgaaggtgg tggacgagct cgtgaaagtg 3900
atgggccggc acaagcccga gaacatcgtg atcgaaatgg ccagagagaa ccagaccacc 3960
cagaagggac agaagaacag ccgcgagaga atgaagcgga tcgaagaggg catcaaagag 4020
ctgggcagcc agatcctgaa agaacacccc gtggaaaaca cccagctgca gaacgagaag 4080
ctgtacctgt actacctgca gaatgggcgg gatatgtacg tggaccagga actggacatc 4140
aaccggctgt ccgactacga tgtggaccat atcgtgcctc agagctttct gaaggacgac 4200

tccatcgaca acaaggtgct gaccagaagc gacaagaacc ggggcaagag cgacaacgtg 4260
ccctccgaag aggtcgtgaa gaagatgaag aactactggc ggcagctgct gaacgccaag 4320
ctgattaccc agagaaagtt cgacaatctg accaaggccg agagaggcgg cctgagcgaa 4380
ctggataagg ccggcttcat caagagacag ctggtggaaa cccggcagat cacaaagcac 4440
gtggcacaga tcctggactc ccggatgaac actaagtacg acgagaatga caagctgatc 4500
cgggaagtga aagtgatcac cctgaagtcc aagctggtgt ccgatttccg gaaggatttc 4560
cagttttaca aagtgcgcga gatcaacaac taccaccacg cccacgacgc ctacctgaac 4620

gccgtcgtgg gaaccgccct gatcaaaaag taccctaagc tggaaagcga gttcgtgtac 4680
ggcgactaca aggtgtacga cgtgcggaag atgatcgcca agagcgagca ggaaatcggc 4740
aaggctaccg ccaagtactt cttctacagc aacatcatga acttttcaa gaccgagatt 4800
accctggcca acggcgagat ccggaagcgg cctctgatcg agacaaacgg cgaaaccggg 4860
gagatcgtgt gggataaggg ccgggatttt gccaccgtgc ggaaagtgct gagcatgccc 4920
caagtgaata tcgtgaaaaa gaccgaggtg cagacaggcg gcttcagcaa agagtctatc 4980
ctgcccaaga ggaacagcga taagctgatc gccagaaaga aggactggga ccctaagaag 5040

tacggcggct tcgacagccc caccgtggcc tattctgtgc tggtggtggc caaagtggaa 5100

aagggcaagt ccaagaaact gaagagtgtg aaagagctgc tggggatcac catcatggaa 5160

agaagcagct tcgagaagaa tcccatcgac tttctggaag ccaagggcta caaagaagtg 5220

aaaaaggacc tgatcatcaa gctgcctaag tactccctgt tcgagctgga aaacggccgg 5280

aagagaatgc tggcctctgc cggcgaactg cagaagggaa acgaactggc cctgccctcc 5340

aaatatgtga acttcctgta cctggccagc cactatgaga agctgaaggg ctcccccgag 5400

gataatgagc agaaacagct gtttgtggaa cagcacaagc actacctgga cgagatcatc 5460

gagcagatca gcgagttctc caagagagtg atcctggccg acgctaatct ggacaaagtg 5520

ctgtccgcct acaacaagca ccgggataag cccatcagag agcaggccga gaatatcatc 5580

cacctgttta ccctgaccaa tctgggagcc cctgccgcct tcaagtactt tgacaccacc 5640

atcgaccgga agaggtacac cagcaccaaa gaggtgctgg acgccaccct gatccaccag 5700

agcatcaccg gcctgtacga gacacggatc gacctgtctc agctgggagg tgactctggc 5760

ggctcaaaaa gaaccgccga cggcagcgaa ttcgagccca agaagaagag gaaagtctaa 5820

ccggtcatca tcaccatcac cattgagttt aaacccgctg atcagcctcg actgtgcctt 5880

ctagttgcca gccatctgtt gtttgcccct cccccgtgcc ttccttgacc ctggaaggtg 5940

ccactcccac tgtcctttcc taataaaatg aggaaattgc atcgcattgt ctgagtaggt 6000

gtcattctat tctgggggggt ggggtggggc aggacagcaa gggggaggat tgggaagaca 6060

atagcaggca tgctggggat gcggtgggct ctatggcttc tgaggcggaa agaaccagct 6120

ggggctcgat accgtcgacc tctagctaga gcttggcgta atcatggtca tagctgtttc 6180

ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat acgagccgga agcataaagt 6240

gtaaagccta gggtgcctaa tgagtgagct aactcacatt aattgcgttg cgctcactgc 6300

ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg 6360

ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc gctcactgac tcgctgcgct 6420

cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa ggcggtaata cggttatcca 6480

cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga 6540

accgtaaaaa ggccgcgttg ctggcgtttt tccataggct ccgcccccct gacgagcatc 6600

acaaaaatcg acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg 6660

cgtttccccc tggaagctcc ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat 6720

acctgtccgc ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt 6780

atctcagttc ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa ccccccgttc 6840

agcccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg 6900

acttatcgcc actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg 6960

gtgctacaga gttcttgaag tggtggccta actacggcta cactagaaga acagtatttg 7020

gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg 7080

gcaaacaaac caccgctggt agcggtggtt tttttgtttg caagcagcag attacgcgca 7140

gaaaaaaagg atctcaagaa gatcctttga tcttttctac ggggtctgac actcagtgga 7200

acgaaaactc acgttaaggg attttggtca tgagattatc aaaaaggatc ttcacctaga 7260

tccttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag taaacttggt 7320

ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcgtt 7380

catccatagt tgcctgactc cccgtcgtgt agataactac gatacgggag ggcttaccat 7440

ctggccccag tgctgcaatg ataccgcgag acccacgctc accggctcca gatttatcag 7500

caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgcct 7560

ccatccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt 7620

tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg 7680

cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatccccc atgttgtgca 7740

aaaaagcggt tagctccttc ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt 7800

tatcactcat ggttatggca gcactgcata attctcttac tgtcatgcca tccgtaagat 7860

gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac 7920

cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa 7980

aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt 8040

tgagatccag ttcgatgtaa cccactcgtg cacccaactg atcttcagca tcttttactt 8100

tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa 8160

gggcgacacg gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt 8220

atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa 8280

taggggttcc gcgcacattt ccccgaaaag tgccacctga cgtcgacgga tcgggagatc 8340

gatctcccga tcccctaggg tcgactctca gtacaatctg ctctgatgcc gcatagttaa 8400

gccagtatct gctccctgct tgtgtgttgg aggtcgctga gtagtgcgcg agcaaaattt 8460

aagctacaac aaggcaaggc ttgaccgaca attgcatgaa gaatctgctt agggttaggc 8520

gttttgcgct gcttcgcgat gtacgggcca gatatacgcg ttgacattga ttattgacta 8580

gttattaata gtaatcaatt acggggtcat tagttcatag cccatatatg gagttccgcg 8640

ttacataact tacggtaaat ggcccgcctg gctgaccgcc caacgacccc cgcccattga 8700

cgtcaataat gacgtatgtt cccatagtaa cgccaatagg gactttccat tgacgtcaat 8760

gggtggagta tttacggtaa actgcccact tggcagtaca tcaagtgtat c 8811

<210> 12

<211> 167

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 12

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Thr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp
165

<210> 13

<211> 8877

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 13

atatgccaag tacgcccccct attgacgtca atgacggtaa atggcccgcc tggcattatg 60

cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg 120

ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact 180

cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa 240

atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta 300

ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct 360

agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat gagctcagag 420

actggcccag tggctgtgga ccccacattg agacggcgga tcgagcccca tgagtttgag 480

gtattcttcg atccgagaga gctccgcaag gagacctgcc tgctttacga aattaattgg 540

gggggccggc actccatttg gcgacataca tcacagaaca ctaacaagca cgtcgaagtc 600

aacttcatcg agaagttcac gacagaaaga tatttctgtc cgaacacaag gtgcagcatt 660

acctggtttc tcagctggag cccatgcggc gaatgtagta gggccatcac tgaattcctg 720

tcaaggtatc cccacgtcac tctgtttatt tacatcgcaa ggctgtacca ccacgctgac 780

ccccgcaatc gacaaggcct gcgggatttg atctcttcag gtgtgactat ccaaattatg 840

actgagcagg agtcaggata ctgctggaga aactttgtga attatagccc gagtaatgaa 900

gcccactggc ctaggtatcc ccatctgtgg gtacgactgt acgttcttga actgtactgc 960

atcatactgg gcctgcctcc ttgtctcaac attctgagaa ggaagcagcc acagctgaca 1020

ttctttacca tcgctcttca gtcttgtcat taccagcgac tgcccccaca cattctctgg 1080

gccaccgggt tgaaatctgg tggttcttct ggtggttcta gcggcagcga gactcccggg 1140

acctcagagt ccgccacacc cgaaagttct ggtggttctt ctggtggttc tgataaaaag 1200

tattctattg gtttagccat cggcactaat tccgttggat gggctgtcat aaccgatgaa 1260

tacaaagtac cttcaaagaa atttaaggtg ttggggaaca cagaccgtca ttcgattaaa 1320

aagaatctta tcggtgccct cctattcgat agtggcgaaa cggcagaggc gactcgcctg 1380

aaacgaaccg ctcggagaag gtatacacgt cgcaagaacc gaatatgtta cttacaagaa 1440

atttttagca atgagatggc caaagttgac gattctttct ttcaccgttt ggaagagtcc 1500

ttccttgtcg aagaggacaa gaaacatgaa cggcacccca tctttggaaa catagtagat 1560

gaggtggcat atcatgaaaa gtacccaacg atttatcacc tcagaaaaaa gctagttgac 1620

tcaactgata aagcggacct gaggttaatc tacttggctc ttgcccatat gataaagttc 1680
cgtgggcact ttctcattga gggtgatcta aatccggaca actcggatgt cgacaaactg 1740
ttcatccagt tagtacaaac ctataatcag ttgtttgaag agaaccctat aaatgcaagt 1800
ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta aatcccgacg gctagaaaac 1860
ctgatcgcac aattacccgg agagaagaaa aatgggttgt tcggtaacct tatagcgctc 1920
tcactaggcc tgacaccaaa ttttaagtcg aacttcgact tagctgaaga tgccaaattg 1980
cagcttagta aggacacgta cgatgacgat ctcgacaatc tactggcaca aattggagat 2040

cagtatgcgg acttattttt ggctgccaaa aaccttagcg atgcaatcct cctatctgac 2100
atactgagag ttaatactga gattaccaag gcgccgttat ccgcttcaat gatcaaaagg 2160
tacgatgaac atcaccaaga cttgacactt ctcaaggccc tagtccgtca gcaactgcct 2220
gagaaatata aggaaatatt ctttgatcag tcgaaaaacg ggtacgcagg ttatattgac 2280
ggcggagcga gtcaagagga attctacaag tttatcaaac ccatattaga gaagatggat 2340
gggacggaag agttgcttgt aaaactcaat cgcgaagatc tactgcgaaa gcagcggact 2400
ttcgacaacg gtagcattcc acatcaaatc cacttaggcg aattgcatgc tatacttaga 2460

aggcaggagg atttttatcc gttcctcaaa gacaatcgtg aaaagattga gaaaatccta 2520
acctttcgca taccttacta tgtgggaccc ctggcccgag ggaactctcg gttcgcatgg 2580
atgacaagaa agtccgaaga aacgattact ccatggaatt ttgaggaagt tgtcgataaa 2640
ggtgcgtcag ctcaatcgtt catcgagagg atgaccaact ttgacaagaa tttaccgaac 2700
gaaaaagtat tgcctaagca cagtttactt tacgagtatt tcacagtgta caatgaactc 2760
acgaaagtta agtatgtcac tgagggcatg cgtaaacccg cctttctaag cggagaacag 2820
aagaaagcaa tagtagatct gttattcaag accaaccgca aagtgacagt taagcaattg 2880

aaagaggact actttaagaa aattgaatgc ttcgattctg tcgagatctc cggggtagaa 2940
gatcgattta atgcgtcact tggtacgtat catgacctcc taaagataat taaagataag 3000
gacttcctgg ataacgaaga gaatgaagat atcttagaag atatagtgtt gactcttacc 3060
ctctttgaag atcgggaaat gattgaggaa agactaaaaa catacgctca cctgttcgac 3120
gataaggtta tgaaacagtt aaagaggcgt cgctatacgg gctggggacg attgtcgcgg 3180
aaacttatca acgggataag agacaagcaa agtggtaaaa ctattctcga ttttctaaag 3240
agcgacggct tcgccaatag gaactttatg cagctgatcc atgatgactc tttaaccttc 3300

aaagaggata tacaaaaggc acaggtttcc ggacaagggg actcattgca cgaacatatt 3360

gcgaatcttg ctggttcgcc agccatcaaa aagggcatac tccagacagt caaagtagtg 3420

gatgagctag ttaaggtcat gggacgtcac aaaccggaaa acattgtaat cgagatggca 3480

cgcgaaaatc aaacgactca gaaggggcaa aaaaacagtc gagagcggat gaagagaata 3540

gaagagggta ttaaagaact gggcagccag atcttaaagg agcatcctgt ggaaaatacc 3600

caattgcaga acgagaaact ttacctctat tacctacaaa atggaaggga catgtatgtt 3660

gatcaggaac tggacataaa ccgtttatct gattacgacg tcgatcacat tgtaccccaa 3720

tccttttga aggacgattc aatcgacaat aaagtgctta cacgctcgga taagaaccga 3780

gggaaaagtg acaatgttcc aagcgaggaa gtcgtaaaga aaatgaagaa ctattggcgg 3840

cagctcctaa atgcgaaact gataacgcaa agaaagttcg ataacttaac taaagctgag 3900

aggggtggct tgtctgaact tgacaaggcc ggatttatta aacgtcagct cgtggaaacc 3960

cgccaaatca caaagcatgt tgcacagata ctagattccc gaatgaatac gaaatacgac 4020

gagaacgata agctgattcg ggaagtcaaa gtaatcactt taaagtcaaa attggtgtcg 4080

gacttcagaa aggattttca attctataaa gttagggaga taaataacta ccaccatgcg 4140

cacgacgctt atcttaatgc cgtcgtaggg accgcactca ttaagaaata cccgaagcta 4200

gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg tccgtaagat gatcgcgaaa 4260

agcgaacagg agataggcaa ggctacagcc aaatacttct tttattctaa cattatgaat 4320

ttctttaaga cggaaatcac tctggcaaac ggagagatac gcaaacgacc tttaattgaa 4380

accaatgggg agacaggtga aatcgtatgg gataagggcc gggacttcgc gacggtgaga 4440

aaagttttgt ccatgcccca agtcaacata gtaaagaaaa ctgaggtgca gaccggaggg 4500

ttttcaaagg aatcgattct tccaaaaagg aatagtgata agctcatcgc tcgtaaaaag 4560

gactgggacc cgaaaaagta cggtggcttc gatagcccta cagttgccta ttctgtccta 4620

gtagtggcaa aagttgagaa gggaaaatcc aagaaactga agtcagtcaa agaattattg 4680

gggataacga ttatggagcg ctcgtctttt gaaaagaacc ccatcgactt ccttgaggcg 4740

aaaggttaca aggaagtaaa aaaggatctc ataattaaac taccaaagta tagtctgttt 4800

gagttagaaa atggccgaaa acggatgttg gctagcgccg gagagcttca aaaggggaac 4860

gaactcgcac taccgtctaa atacgtgaat ttcctgtatt tagcgtccca ttacgagaag 4920

ttgaaaggtt cacctgaaga taacgaacag aagcaacttt ttgttgagca gcacaaacat 4980

tatctcgacg aaatcataga gcaaatttcg gaattcagta agagagtcat cctagctgat 5040

gccaatctgg acaaagtatt aagcgcatac aacaagcaca gggataaacc catacgtgag 5100

caggcggaaa atattatcca tttgtttact cttaccaacc tcggcgctcc agccgcattc 5160

aagtattttg acacaacgat agatcgcaaa cgatacactt ctaccaagga ggtgctagac 5220

gcgacactga ttcaccaatc catcacggga ttatatgaaa ctcggataga tttgtcacag 5280

cttgggggtg actctggtgg ttctggagga tctggtggtt ctactaatct gtcagatatt 5340

attgaaaagg agaccggtaa gcaactggtt atccaggaat ccatcctcat gctcccagag 5400

gaggtggaag aagtcattgg gaacaagccg gaaagcgata tactcgtgca caccgcctac 5460

gacgagagca ccgacgagaa tgtcatgctt ctgactagcg acgcccctga atacaagcct 5520

tgggctctgg tcatacagga tagcaacggt gagaacaaga ttaagatgct ctctggtggt 5580

tctggaggat ctggtggttc tactaatctg tcagatatta ttgaaaagga gaccggtaag 5640

caactggtta tccaggaatc catcctcatg ctcccagagg aggtggaaga agtcattggg 5700

aacaagccgg aaagcgatat actcgtgcac accgcctacg acgagagcac cgacgagaat 5760

gtcatgcttc tgactagcga cgcccctgaa tacaagcctt gggctctggt catacaggat 5820

agcaacggtg agaacaagat taagatgctc tctggtggtt ctcccaagaa gaagaggaaa 5880

gtctaaccgg tcatcatcac catcaccatt gagtttaaac ccgctgatca gcctcgactg 5940

tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg 6000

aaggtgccac tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga 6060

gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg 6120

aagacaatag caggcatgct ggggatgcgg tgggctctat ggcttctgag gcggaaagaa 6180

ccagctgggg ctcgataccg tcgacctcta gctagagctt ggcgtaatca tggtcatagc 6240

tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca 6300

taaagtgtaa agcctagggt gcctaatgag tgagctaact cacattaatt gcgttgcgct 6360

cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac 6420

gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc 6480

tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt 6540

tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg 6600

ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccctgacg 6660

agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat 6720

accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta 6780

ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct 6840

gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc 6900

ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa 6960

gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg 7020

taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact agaagaacag 7080

tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt 7140
gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta 7200
cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc 7260
agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa aggatcttca 7320
cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa 7380
cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat 7440
ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct 7500

taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt 7560
tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat 7620
ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta 7680
atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg 7740
gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt 7800
tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg 7860
cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg 7920

taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc 7980
ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa 8040
ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac 8100
cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt 8160
ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg 8220
gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa 8280
gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata 8340

aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc gacggatcgg 8400
gagatcgatc tcccgatccc ctagggtcga ctctcagtac aatctgctct gatgccgcat 8460
agttaagcca gtatctgctc cctgcttgtg tgttggaggt cgctgagtag tgcgcgagca 8520
aaatttaagc tacaacaagg caaggcttga ccgacaattg catgaagaat ctgcttaggg 8580
ttaggcgttt tgcgctgctt cgcgatgtac gggccagata tacgcgttga cattgattat 8640
tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt 8700
tccgcgttac ataacttacg gtaaatggcc cgcctggctg accgcccaac gacccccgcc 8760

cattgacgtc aataatgacg tatgttccca tagtaacgcc aatagggact ttccattgac 8820

gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatc 8877

<210> 14

<211> 5508

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 14

atgagctcag agactggccc agtggctgtg gaccccacat tgagacggcg gatcgagccc 60

catgagtttg aggtattctt cgatccgaga gagctccgca aggagacctg cctgctttac 120

gaaattaatt ggggggccg gcactccatt tggcgacata catcacagaa cactaacaag 180

cacgtcgaag tcaacttcat cgagaagttc acgacagaaa gatatttctg tccgaacaca 240

aggtgcagca ttacctggtt tctcagctgg agccgcgaat gtagtagggc catcactgaa 300

ttcctgtcaa ggtatcccca cgtcactctg tttatttaca tcgcaaggct gtaccaccac 360

gctgaccccc gcaatcgaca aggcctgcgg gatttgatct cttcaggtgt gactatccaa 420

attatgactg agcaggagtc aggatactgc tggagaaact ttgtgaatta tagcccgagt 480

aatgaagccc actggcctag gtatccccat ctgtgggtac gactgtacgt tcttgaactg 540

tactgcatca tactgggcct gcctccttgt ctcaacattc tgagaaggaa gcagccacag 600

ctgacattct ttaccatcgc tcttcagtct tgtcattacc agcgactgcc cccacacatt 660

ctctgggcca ccgggttgaa atctggtggt tcttctggtg gttctagcgg cagcgagact 720

cccgggacct cagagtccgc cacacccgaa agttctggtg gttcttctgg tggttctgat 780

aaaaagtatt ctattggttt agccatcggc actaattccg ttggatgggc tgtcataacc 840

gatgaataca aagtaccttc aaagaaattt aaggtgttgg ggaacacaga ccgtcattcg 900

attaaaaaga atcttatcgg tgccctccta ttcgatagtg gcgaaacggc agaggcgact 960

cgcctgaaac gaaccgctcg gagaaggtat acacgtcgca agaaccgaat atgttactta 1020

caagaaattt ttagcaatga gatggccaaa gttgacgatt ctttctttca ccgtttggaa 1080

gagtccttcc ttgtcgaaga ggacaagaaa catgaacggc accccatctt tggaaacata 1140

gtagatgagg tggcatatca tgaaaagtac ccaacgattt atcacctcag aaaaaagcta 1200

gttgactcaa ctgataaagc ggacctgagg ttaatctact tggctcttgc ccatatgata 1260

aagttccgtg ggcactttct cattgagggt gatctaaatc cggacaactc ggatgtcgac 1320

aaactgttca tccagttagt acaaacctat aatcagttgt ttgaagagaa ccctataaat 1380

gcaagtggcg tggatgcgaa ggctattctt agcgcccgcc tctctaaatc ccgacggcta 1440

gaaaacctga tcgcacaatt acccggagag aagaaaaatg ggttgttcgg taaccttata 1500
gcgctctcac taggcctgac accaaatttt aagtcgaact tcgacttagc tgaagatgcc 1560
aaattgcagc ttagtaagga cacgtacgat gacgatctcg acaatctact ggcacaaatt 1620
ggagatcagt atgcggactt atttttggct gccaaaaacc ttagcgatgc aatcctccta 1680
tctgacatac tgagagttaa tactgagatt accaaggcgc cgttatccgc ttcaatgatc 1740
aaaaggtacg atgaacatca ccaagacttg acacttctca aggccctagt ccgtcagcaa 1800
ctgcctgaga aatataagga aatattcttt gatcagtcga aaaacgggta cgcaggttat 1860

attgacggcg gagcgagtca agaggaattc tacaagttta tcaaacccat attagagaag 1920
atggatggga cggaagagtt gcttgtaaaa ctcaatcgcg aagatctact gcgaaagcag 1980
cggactttcg acaacggtag cattccacat caaatccact taggcgaatt gcatgctata 2040
cttagaaggc aggaggattt ttatccgttc ctcaaagaca atcgtgaaaa gattgagaaa 2100
atcctaacct ttcgcatacc ttactatgtg ggacccctgg cccgagggaa ctctcggttc 2160
gcatggatga caagaaagtc cgaagaaacg attactccat ggaattttga ggaagttgtc 2220
gataaaggtg cgtcagctca atcgttcatc gagaggatga ccaactttga caagaattta 2280

ccgaacgaaa aagtattgcc taagcacagt ttactttacg agtatttcac agtgtacaat 2340
gaactcacga aagttaagta tgtcactgag ggcatgcgta aacccgcctt tctaagcgga 2400
gaacagaaga aagcaatagt agatctgtta ttcaagacca accgcaaagt gacagttaag 2460
caattgaaag aggactactt taagaaaatt gaatgcttcg attctgtcga gatctccggg 2520
gtagaagatc gatttaatgc gtcacttggt acgtatcatg acctcctaaa gataattaaa 2580
gataaggact tcctggataa cgaagagaat gaagatatct tagaagatat agtgttgact 2640
cttaccctct ttgaagatcg ggaaatgatt gaggaaagac taaaaacata cgctcacctg 2700

ttcgacgata aggttatgaa acagttaaag aggcgtcgct atacgggctg gggacgattg 2760
tcgcggaaac ttatcaacgg gataagagac aagcaaagtg gtaaaactat tctcgatttt 2820
ctaaagagcg acggcttcgc caataggaac tttatgcagc tgatccatga tgactcttta 2880
accttcaaag aggatataca aaaggcacag gtttccggac aaggggactc attgcacgaa 2940
catattgcga atcttgctgg ttcgccagcc atcaaaaagg gcatactcca gacagtcaaa 3000
gtagtggatg agctagttaa ggtcatggga cgtcacaaac cggaaaacat tgtaatcgag 3060
atggcacgcg aaaatcaaac gactcagaag gggcaaaaaa acagtcgaga gcggatgaag 3120

agaatagaag agggtattaa agaactgggc agccagatct taaaggagca tcctgtggaa 3180

aatacccaat tgcagaacga gaaactttac ctctattacc tacaaaatgg aagggacatg 3240

tatgttgatc aggaactgga cataaaccgt ttatctgatt acgacgtcga tcacattgta 3300

ccccaatcct ttttgaagga cgattcaatc gacaataaag tgcttacacg ctcggataag 3360

aaccgaggga aaagtgacaa tgttccaagc gaggaagtcg taaagaaaat gaagaactat 3420

tggcggcagc tcctaaatgc gaaactgata acgcaaagaa agttcgataa cttaactaaa 3480

gctgagaggg gtggcttgtc tgaacttgac aaggccggat ttattaaacg tcagctcgtg 3540

gaaacccgcc aaatcacaaa gcatgttgca cagatactag attcccgaat gaatacgaaa 3600

tacgacgaga acgataagct gattcgggaa gtcaaagtaa tcactttaaa gtcaaaattg 3660

gtgtcggact tcagaaagga ttttcaattc tataaagtta gggagataaa taactaccac 3720

catgcgcacg acgcttatct taatgccgtc gtagggaccg cactcattaa gaaatacccg 3780

aagctagaaa gtgagtttgt gtatggtgat tacaaagttt atgacgtccg taagatgatc 3840

gcgaaaagcg aacaggagat aggcaaggct acagccaaat acttctttta ttctaacatt 3900

atgaatttct ttaagacgga aatcactctg gcaaacggag agatacgcaa acgaccttta 3960

attgaaacca atggggagac aggtgaaatc gtatgggata agggccggga cttcgcgacg 4020

gtgagaaaag ttttgtccat gccccaagtc aacatagtaa agaaaactga ggtgcagacc 4080

ggagggtttt caaaggaatc gattcttcca aaaaggaata gtgataagct catcgctcgt 4140

aaaaaggact gggacccgaa aaagtacggt ggcttcgata gccctacagt tgcctattct 4200

gtcctagtag tggcaaaagt tgagaaggga aaatccaaga aactgaagtc agtcaaagaa 4260

ttattgggga taacgattat ggagcgctcg tcttttgaaa agaaccccat cgacttcctt 4320

gaggcgaaag gttacaagga agtaaaaaag gatctcataa ttaaactacc aaagtatagt 4380

ctgtttgagt tagaaaatgg ccgaaaacgg atgttggcta gcgccggaga gcttcaaaag 4440

gggaacgaac tcgcactacc gtctaaatac gtgaatttcc tgtatttagc gtcccattac 4500

gagaagttga aaggttcacc tgaagataac gaacagaagc aacttttgt tgagcagcac 4560

aaacattatc tcgacgaaat catagagcaa atttcggaat tcagtaagag agtcatccta 4620

gctgatgcca atctggacaa agtattaagc gcatacaaca agcacaggga taaacccata 4680

cgtgagcagg cggaaaatat tatccatttg tttactctta ccaacctcgg cgctccagcc 4740

gcattcaagt attttgacac aacgatagat cgcaaacgat acacttctac caaggaggtg 4800

ctagacgcga cactgattca ccaatccatc acgggattat atgaaactcg gatagatttg 4860

tcacagcttg ggggtgactc tggtggttct ggaggatctg gtggttctac taatctgtca 4920

gatattattg aaaaggagac cggtaagcaa ctggttatcc aggaatccat cctcatgctc 4980

ccagaggagg tggaagaagt cattgggaac aagccggaaa gcgatatact cgtgcacacc 5040

gcctacgacg agagcaccga cgagaatgtc atgcttctga ctagcgacgc ccctgaatac 5100

aagccttggg ctctggtcat acaggatagc aacggtgaga acaagattaa gatgctctct 5160

ggtggttctg gaggatctgg tggttctact aatctgtcag atattattga aaaggagacc 5220

ggtaagcaac tggttatcca ggaatccatc ctcatgctcc cagaggaggt ggaagaagtc 5280

attgggaaca agccggaaag cgatatactc gtgcacaccg cctacgacga gagcaccgac 5340

gagaatgtca tgcttctgac tagcgacgcc cctgaataca agccttgggc tctggtcata 5400

caggatagca acggtgagaa caagattaag atgctctctg gtggttctaa aaggacggcg 5460

gacggatcag agttcgagag tccgaaaaaa aaacgaaagg tcgaataa 5508

<210> 15

<211> 5514

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 15

atgtcatccg aaaccgggcc agtggccgta gacccaacac tcaggaggcg gatagaaccc 60

catgagtttg aagtgttctt cgaccccaga gagctgcgca aagagacttg cctcctgtat 120

gaaataaatt gggggggtcg ccattcaatt tggaggcaca ctagccagaa tactaacaaa 180

cacgtggagg taaattttat cgagaagttt accaccgaaa gatacttttg ccccaataca 240

cggtgttcaa ttacctggtt tctgtcatgg agtccatgtg gagaatgtag tagagcgata 300

actgagttcc tgtctcgata tcctcacgtc acgttgttta tatacatcgc tcggctttat 360

caccatgcgg acccgcggaa caggcaaggt cttcgggacc tcatatcctc tggggtgacc 420

atccagataa tgacggagca agagagcgga tactgctggc gaaactttgt taactacagc 480

ccaagcaatg aggcacactg gcctagatat ccgcatctct gggttcgact gtatgtcctt 540

gaactgtact gcataattct gggacttccg ccatgcttga acattctgcg gcggaaacaa 600

ccacagctga ccttttcac gattgctctc caaagttgtc actaccagcg attgccaccc 660

cacatcttgt gggctactgg actcaagtct ggaggaagtt caggcggaag cagcgggtct 720

gaaacgcccg gaacctcaga gagcgcaacg cccgaaagct ctggagggtc aagtggtggt 780

agtgataaga aatactccat cggcctcgcc atcggtacga attctgtcgg ttgggccgtt 840

atcaccgatg agtacaaggt cccttctaag aaattcaagg ttttgggcaa tacagaccgc 900

cattctataa aaaaaaacct gatcggcgcc cttttgtttg acagtggtga gactgctgaa 960

gcgactcgcc tgaagcgaac tgccaggagg cggtatacga ggcgaaaaaa ccgaatttgt 1020

tacctccagg agattttctc aaatgaaatg gccaaggtag atgatagttt ttttcaccgc 1080

ttggaagaaa gttttctcgt tgaggaggac aaaaagcacg agaggcaccc aatctttggc 1140

aacatagtcg atgaggtcgc ataccatgag aaatatccta cgatctatca tctccgcaag 1200

aagctggtcg atagcacgga taaagctgac ctccggctga tctaccttgc tcttgctcac 1260

atgattaaat tcaggggcca tttcctgata gaaggagacc tcaatcccga caattctgat 1320

gtcgacaaac tgtttattca gctcgttcag acctataatc aactctttga ggagaacccc 1380

atcaatgctt caggggtgga cgcaaaggcc attttgtccg cgcgcttgag taaatcacga 1440

cgcctcgaga atttgatagc tcaactgccg ggtgagaaga aaaacgggtt gtttgggaat 1500

ctcatagcgt tgagtttggg acttacgcca aactttaagt ctaactttga tttggccgaa 1560

gatgccaaat tgcagctgtc caaagatacc tatgatgacg acttggataa ccttcttgcg 1620

cagattggtg accaatacgc ggatctgttt cttgccgcaa aaaatctgtc cgacgccata 1680

ctcttgtccg atatactgcg cgtcaatact gagataacta aggctcccct cagcgcgtcc 1740

atgattaaaa gatacgatga gcaccaccaa gatctcactc tgttgaaagc cctggttcgc 1800

cagcagcttc cagagaagta taaggagata tttttcgacc aatctaaaaa cggctatgcg 1860

ggttacattg acggtggcgc ctctcaagaa gaattctaca agtttataaa gccgatactt 1920

gagaaaatgg acggtacaga ggaattgttg gttaagctca atcgcgagga cttgttgaga 1980

aagcagcgca catttgacaa tggtagtatt ccacaccaga ttcatctggg cgagttgcat 2040

gccattctta gaagacaaga agatttttat ccgtttctga aagataacag agaaaagatt 2100

gaaaagatac ttacctttcg cataccgtat tatgtaggtc ccctggctag agggaacagt 2160

cgcttcgctt ggatgactcg aaaatcagaa gaaacaataa ccccctggaa ttttgaagaa 2220

gtggtagata aaggtgcgag tgcccaatct tttattgagc ggatgacaaa ttttgacaag 2280

aatctgccta acgaaaaggt gcttcccaag cattcccttt tgtatgaata ctttacagta 2340

tataatgaac tgactaaagt gaagtacgtt accgagggga tgcgaaagcc agcttttctc 2400

agtggcgagc agaaaaaagc aatagttgac ctgctgttca agacgaatag gaaggttacc 2460

gtcaaacagc tcaaagaaga ttactttaaa aagatcgaat gttttgattc agttgagata 2520

agcggagtag aggatagatt taacgcaagt cttggaactt atcatgacct tttgaagatc 2580

atcaaggata aagatttttt ggacaacgag gagaatgaag atatcctgga agatatagta 2640

cttaccttga cgctttttga agatcgagag atgatcgagg agcgacttaa gacgtacgca 2700

catctctttg acgataaggt tatgaaacaa ttgaaacgcc ggcggtatac tggctggggc 2760

aggctttctc gaaagctgat taatggtatc cgcgataagc agtctggaaa gacaatcctt 2820

gactttctga aaagtgatgg atttgcaaat agaaacttta tgcagcttat acatgatgac 2880

tctttgacgt tcaaggaaga catccagaag gcacaggtat ccggccaagg ggatagcctc 2940

catgaacaca tagccaacct ggccggctca ccagctatta aaaagggaat attgcaaacc 3000

gttaaggttg ttgacgaact cgttaaggtt atgggccgac acaaaccaga gaatatcgtg 3060

attgagatgg ctagggagaa tcagaccact caaaaaggtc agaaaaattc tcgcgaaagg 3120

atgaagcgaa ttgaagaggg aatcaaagaa cttggctctc aaattttgaa agagcacccg 3180

gtagaaaaca ctcagctgca gaatgaaaag ctgtatctgt attatctgca gaatggtcga 3240

gatatgtacg ttgatcagga gctggatatc aataggctca gtgactacga tgtcgaccac 3300

atcgttcctc aatctttcct gaaagatgac tctatcgaca acaaagtgtt gacgcgatca 3360

gataagaacc ggggaaaatc cgacaatgta ccctcagaag aagttgtcaa gaagatgaaa 3420

aactattgga gacaattgct gaacgccaag ctcataacac aacgcaagtt cgataacttg 3480

acgaaagccg aaagaggtgg gttgtcagaa ttggacaaag ctggctttat taagcgccaa 3540

ttggtggaga cccggcagat tacgaaacac gtagcacaaa ttttggattc acgaatgaat 3600

accaaatacg acgaaaacga caaattgata cgcgaggtga aagtgattac gcttaagagt 3660

aagttggttt ccgatttcag gaaggatttt cagttttaca aagtaagaga aataaacaac 3720

taccaccacg cccatgatgc ttacctcaac gcggtagttg gcacagctct tatcaaaaaa 3780

tatccaaagc tggaaagcga gttcgtttac ggtgactata aagtatacga cgttcggaag 3840

atgatagcca aatcagagca ggaaattggg aaggcaaccg caaaatactt cttctattca 3900

aacatcatga acttctttaa gacggagatt acgctcgcga acggcgaaat acgcaagagg 3960

cccctcatag agactaacgg cgaaaccggg gagatcgtat gggacaaagg acgggacttt 4020

gcgaccgtta gaaaagtact ttcaatgcca caagtgaata ttgttaaaaa gacagaagta 4080

caaacagggg ggttcagtaa ggaatccatt ttgcccaagc ggaacagtga taaattgata 4140

gcaaggaaaa aagattggga ccctaagaag tacggtggtt tcgactctcc taccgttgca 4200

tattcagtcc ttgtagttgc gaaagtggaa aaggggaaaa gtaagaagct taagagtgtt 4260

aaagagcttc tgggcataac cataatggaa cggtctagct tcgagaaaaa tccaattgac 4320

tttctcgagg ctaaaggtta caaggaggta aaaaaggacc tgataattaa actcccaaag 4380

tacagtctct tcgagttgga gaatgggagg aagagaatgt tggcatctgc aggggagctc 4440

caaaagggga acgagctggc tctgccttca aaatacgtga actttctgta cctggccagc 4500

cactacgaga aactcaaggg ttctcctgag gataacgagc agaaacagct gtttgtagag 4560

cagcacaagc attacctgga cgagataatt gagcaaatta gtgagttctc aaaaagagta 4620

atccttgcag acgcgaatct ggataaagtt ctttccgcct ataataagca ccgggacaag 4680
cctatacgag aacaagccga gaacatcatt cacctcttta cccttactaa tctgggcgcg 4740
ccggccgcct tcaaatactt cgacaccacg atagacagga aaaggtatac gagtaccaaa 4800
gaagtacttg acgccactct catccaccag tctataacag ggttgtacga aacgaggata 4860
gatttgtccc agctcggcgg cgactcagga gggtcaggcg gctccggtgg atcaacgaat 4920
ctttccgaca taatcgagaa agaaaccggc aaacagttgg tgatccaaga atcaatcctg 4980
atgctgcctg aagaagtaga agaggtgatt ggcaacaaac ctgagtctga cattcttgtc 5040

cacaccgcgt atgacgagag cacggacgag aacgttatgc ttctcactag cgacgcccct 5100
gagtataaac catgggcgct ggtcatccaa gattccaatg gggaaaacaa gattaagatg 5160
cttagtggtg ggtctggagg gagcggtggg tccacgaacc tcagcgacat tattgaaaaa 5220
gagactggta aacaacttgt aatacaagag tctattctga tgttgcctga agaggtggag 5280
gaggtgattg ggaacaaacc ggagtctgat atacttgttc ataccgccta tgacgaatct 5340
actgatgaga atgtgatgct tttaacgtca gacgctcccg agtacaaacc ctgggctctg 5400
gtgattcagg acagcaatgg tgagaataag attaaaatgt tgagtggggg ctcaaagcgc 5460

acggctgacg gtagcgaatt tgagagcccc aaaaaaaaac gaaaggtcga ataa 5514

<210> 16
<211> 5514
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 16

atgagcagcg agacaggccc tgtggctgtg gatcctacac tgcggagaag aatcgagccc 60
cacgagttcg aggtgttctt cgaccccaga gagctgcgga aagagacatg cctgctgtac 120
gagatcaact ggggcggcag acactctatc tggcggcaca caagccagaa caccaacaag 180
cacgtggaag tgaactttat cgagaagttt acgaccgagc ggtacttctg ccccaacacc 240

agatgcagca tcacctggtt tctgagctgg tccccttgcg gcgagtgcag cagagccatc 300
accgagtttc tgtccagata tccccacgtg accctgttca tctatatcgc ccggctgtac 360
caccacgccg atcctagaaa tagacaggga ctgcgcgacc tgatcagcag cggagtgacc 420
atccagatca tgaccgagca agagagcggc tactgctggc ggaacttcgt gaactacagc 480

cccagcaacg aagcccactg gcctagatat cctcacctgt gggtccgact gtacgtgctg 540

gaactgtact gcatcatcct gggcctgcct ccatgcctga acatcctgag aagaaagcag 600

cctcagctga ccttcttcac aatcgccctg cagagctgcc actaccagag actgcctcca 660

cacatcctgt gggccaccgg acttaagagc ggaggatcta gcggcggctc tagcggatct 720

gagacacctg gcacaagcga gtctgccaca cctgagagta gcggcggatc ttctggcggc 780

tccgacaaga agtactctat cggactggcc atcggcacca actctgttgg atgggccgtg 840

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 900

cacagcatca agaagaatct gatcggcgcc ctgctgttcg actctggcga aacagccgaa 960

gccaccagac tgaagagaac cgccaggcgg agatacaccc ggcggaagaa ccggatctgc 1020

tacctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 1080

ctggaagagt ccttcctggt ggaagaggac aagaagcacg agcggcaccc catcttcggc 1140

aacatcgtgg atgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 1200

aaactggtgg acagcaccga caaggccgac ctgagactga tctacctggc tctggcccac 1260

atgatcaagt tccggggcca ctttctgatc gagggcgatc tgaaccccga caacagcgac 1320

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 1380

atcaacgcct ctggcgtgga cgccaaggct atcctgtctg ccagactgag caagagcaga 1440

aggctggaaa acctgatcgc ccagctgcct ggcgagaaga agaatggcct gttcggcaac 1500

ctgattgccc tgagcctggg actgacccct aacttcaaga gcaacttcga cctggccgag 1560

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa tctgctggcc 1620

cagatcggcg atcagtacgc cgacttgttt ctggccgcca agaacctgtc cgacgccatc 1680

ctgctgagcg atatcctgag agtgaacacc gagatcacaa aggcccctct gagcgcctct 1740

atgatcaaga gatacgacga gcaccaccag gatctgaccc tgctgaaggc cctcgttaga 1800

cagcagctgc cagagaagta caaagagatt ttcttcgatc agtccaagaa cggctacgcc 1860

ggctacattg atggcggagc cagccaagag gaattctaca agttcatcaa gcccatcctg 1920

gaaaagatgg acggcaccga ggaactgctg gtcaagctga acagagagga cctgctgcgg 1980

aagcagcgga ccttcgacaa tggctctatc cctcaccaga tccacctggg agagctgcac 2040

gccattctgc ggagacaaga ggacttttac ccattcctga aggacaaccg ggaaaagatc 2100

gagaagatcc tgaccttcag gatcccctac tacgtgggac cactggccag aggcaatagc 2160

agattcgcct ggatgaccag aaagagcgag gaaaccatca caccctggaa cttcgaggaa 2220

gtggtggaca agggcgccag cgctcagtcc ttcatcgagc ggatgaccaa cttcgataag 2280

aacctgccta acgagaaggt gctgcccaag cactccctgc tgtatgagta cttcaccgtg 2340

tacaacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgcctttctg 2400
agcggcgagc agaaaaaggc cattgtggat ctgctgttca agaccaaccg gaaagtgacc 2460
gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgacag cgtggaaatc 2520
agcggcgtgg aagatcggtt caatgccagc ctgggcacat accacgacct gctgaaaatt 2580
atcaaggaca aggacttcct ggacaacgaa gagaacgagg acattctcga ggacatcgtg 2640
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacatacgcc 2700
cacctgttcg acgacaaagt gatgaagcaa ctgaagcgga ggcggtacac aggctggggc 2760

agactgtctc ggaagctgat caacggcatc cgggataagc agtccggcaa gacaatcctg 2820
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2880
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaagg cgattctctg 2940
cacgagcaca ttgccaacct ggccggatct cccgccatta agaagggcat cctgcagaca 3000
gtgaaggtgg tggacgagct tgtgaaagtg atgggcagac acaagcccga gaacatcgtg 3060
atcgaaatgg ccagagagaa ccagaccaca cagaagggcc agaagaacag ccgcgagaga 3120
atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 3180

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatggacgg 3240
gatatgtacg tggaccaaga gctggacatc aaccggctga gcgactacga tgtggaccat 3300
atcgtgcccc agagctttct gaaggacgac tccatcgata acaaggtcct gaccagaagc 3360
gacaagaacc ggggcaagag cgataacgtg ccctccgaag aggtggtcaa gaagatgaag 3420
aactactggc gacagctgct gaacgccaag ctgattaccc agcggaagtt cgataacctg 3480
accaaggccg agagaggcgg cctgagcgaa cttgataagg ccggcttcat taagcggcag 3540
ctggtggaaa cccggcagat caccaaacac gtggcacaga ttctggactc ccggatgaac 3600

actaagtacg acgagaatga caagctgatc cgggaagtga aagtcatcac cctgaagtct 3660
aagctggtgt ccgatttccg gaaggatttc cagttctaca aagtgcggga aatcaacaac 3720
taccatcacg cccacgacgc ctacctgaat gccgttgttg gaacagccct gatcaagaag 3780
tatcccaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3840
atgatcgcca agagcgaaca agagatcggc aaggctaccg ccaagtactt tttctacagc 3900
aacatcatga acttttttcaa gacagagatc accctggcca acggcgagat ccggaaaaga 3960
cccctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg cagagatttt 4020

gccacagtgc ggaaagtgct gagcatgccc caagtgaata tcgtgaagaa aaccgaggtg 4080

cagacaggcg gcttcagcaa agagtctatc ctgcctaagc ggaacagcga taagctgatc 4140

gccagaaaga aggactggga ccctaagaag tacggcggct tcgatagccc taccgtggcc 4200

tattctgtgc tggtggtggc caaagtggaa aagggcaagt ccaaaaagct caagagcgtg 4260

aaagagctgc tggggatcac catcatggaa agaagcagct ttgagaagaa cccgatcgac 4320

tttctggaag ccaagggcta caaagaagtc aagaaggacc tcatcatcaa gctccccaag 4380

tacagcctgt tcgagctgga aaatggccgg aagcggatgc tggcctcagc aggcgaactg 4440

cagaaaggca atgaactggc cctgcctagc aaatacgtca acttcctgta cctggccagc 4500

cactatgaga agctgaaggg cagccccgag gacaatgagc aaaagcagct gtttgtggaa 4560

cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttctc caagagagtg 4620

atcctggccg acgctaacct ggataaggtg ctgtctgcct ataacaagca ccgggacaag 4680

cctatcagag agcaggccga gaatatcatc cacctgttta ccctgaccaa cctgggagcc 4740

cctgccgcct tcaagtactt cgacaccacc atcgaccgga agaggtacac cagcaccaaa 4800

gaggtgctgg acgccacact gatccaccag tctatcaccg gcctgtacga aacccggatc 4860

gacctgtctc agctcggcgg cgattctggt ggttctggcg gaagtggcgg atccaccaat 4920

ctgagcgaca tcatcgaaaa agagacaggc aagcagctcg tgatccaaga atccatcctg 4980

atgctgcctg aagaggttga ggaagtgatc ggcaacaagc ctgagtccga catcctggtg 5040

cacaccgcct acgatgagag caccgatgag aacgtcatgc tgctgacaag cgacgccccc 5100

gagtacaagc cttgggctct cgtgattcag gacagcaatg gggagaacaa gatcaagatg 5160

ctgagcggag gtagcggagg cagtggcgga agcacaaacc tgtctgatat cattgaaaaa 5220

gaaaccggga agcaactggt cattcaagag tccattctca tgctcccgga agaagtcgag 5280

gaagtcattg gaaacaaacc cgagagcgat attctggtcc acacagccta tgacgagtct 5340

acagacgaaa acgtgatgct cctgacctct gacgctcccg agtataagcc ctgggcactt 5400

gttatccagg actctaacgg ggaaaacaaa atcaaaatgt tgtccggcgg cagcaagcgg 5460

acagccgatg gatctgagtt cgagagcccc aagaagaaac ggaaggtgga gtaa 5514

<210> 17

<211> 119

<212> PRT

<213> Homo sapiens

<400> 17

Met Ser Arg Ser Val Ala Leu Ala Val Leu Ala Leu Leu Ser Leu Ser

1 5 10 15

Gly Leu Glu Ala Ile Gln Arg Thr Pro Lys Ile Gln Val Tyr Ser Arg

20 25 30

His Pro Ala Glu Asn Gly Lys Ser Asn Phe Leu Asn Cys Tyr Val Ser

35 40 45

Gly Phe His Pro Ser Asp Ile Glu Val Asp Leu Leu Lys Asn Gly Glu

50 55 60

Arg Ile Glu Lys Val Glu His Ser Asp Leu Ser Phe Ser Lys Asp Trp

65 70 75 80

Ser Phe Tyr Leu Leu Tyr Tyr Thr Glu Phe Thr Pro Thr Glu Lys Asp

85 90 95

Glu Tyr Ala Cys Arg Val Asn His Val Thr Leu Ser Gln Pro Lys Ile

100 105 110

Val Lys Trp Asp Arg Asp Met

115

<210> 18

<211> 9199

<212> DNA

<213> Homo sapiens

<400> 18

catgtcataa atggtaagtc caagaaaaat acaggtattc ccccccaaag aaaactgtaa 60

aatcgacttt tttctatctg tactgttttt tattggtttt taaattggtt ttccaagtga 120

gtaaatcaga atctatctgt aatggatttt aaatttagtg tttctctgtg atgtagtaaa 180

caagaaacta gaggcaaaaa tagccctgtc ccttgctaaa cttctaaggc acttttctag 240

tacaactcaa cactaacatt tcaggccttt agtgccttat atgagttttt aaaaggggga 300

aaagggaggg agcaagagtg tcttaactca tacatttagg cataacaatt attctcatat 360

tttagttatt gagagggctg gtagaaaaac taggtaaata atattaataa ttatagcgct 420

tattaaacac tacagaacac ttactatgta ccaggcattg tgggaggctc tctcttgtgc 480

attatctcat ttcattaggt ccatggagag tattgcattt tcttagttta ggcatggcct 540

ccacaataaa gattatcaaa agcctaaaaa tatgtaaaag aaacctagaa gttatttgtt 600

gtgctccttg gggaagctag gcaaatcctt tcaactgaaa accatggtga cttccaagat 660

ctctgcccct ccccatcgcc atggtccact tcctcttctc actgttcctc ttagaaaaga 720

tctgtggact ccaccaccac gaaatggcgg caccttattt atggtcactt tagagggtag 780

gttttcttaa tgggtctgcc tgtcatgttt aacgtccttg gctgggtcca aggcagatgc 840

agtccaaact ctcactaaaa ttgccgagcc ctttgtcttc cagtgtctaa aatattaatg 900

tcaatggaat caggccagag tttgaattct agtctcttag cctttgtttc ccctgtccat 960

aaaatgaatg ggggtaattc tttcctccta cagtttattt atatattcac taattcattc 1020

attcatccat ccattcgttc attcggttta ctgagtacct actatgtgcc agcccctgtt 1080

ctagggtgga aactaagaga atgatgtacc tagagggcgc tggaagctct aaagccctag 1140

cagttactgc ttttactatt agtggtcgtt ttttttctccc ccccgccccc cgacaaatca 1200

acagaacaaa gaaaattacc taaacagcaa ggacataggg aggaacttct tggcacagaa 1260

ctttccaaac actttttcct gaagggatac aagaagcaag aaaggtactc tttcactagg 1320

accttctctg agctgtcctc aggatgcttt tgggactatt tttcttaccc agagaatgga 1380

gaaaccctgc agggaattcc caagctgtag ttataaacag aagttctcct tctgctaggt 1440

agcattcaaa gatcttaatc ttctgggttt ccgttttctc gaatgaaaaa tgcaggtccg 1500

agcagttaac tggctggggc accattagca agtcacttag catctctggg gccagtctgc 1560

aaagcgaggg ggcagcctta atgtgcctcc agcctgaagt cctagaatga gcgcccggtg 1620

tcccaagctg gggcgcgcac cccagatcgg agggcgccga tgtacagaca gcaaactcac 1680

ccagtctagt gcatgccttc ttaaacatca cgagactcta agaaaaggaa actgaaaacg 1740

ggaaagtccc tctctctaac ctggcactgc gtcgctggct tggagacagg tgacggtccc 1800

tgcgggcctt gtcctgattg gctgggcacg cgtttaatat aagtggaggc gtcgcgctgg 1860

cgggcattcc tgaagctgac agcattcggg ccgagatgtc tcgctccgtg gccttagctg 1920

tgctcgcgct actctctctt tctggcctgg aggctatcca gcgtgagtct ctcctaccct 1980

cccgctctgg tccttcctct cccgctctgc accctctgtg gccctcgctg tgctctctcg 2040

ctccgtgact tcccttctcc aagttctcct tggtggcccg ccgtggggct agtccagggc 2100

tggatctcgg ggaagcggcg gggtggcctg ggagtgggga agggggtgcg cacccgggac 2160

gcgcgctact tgcccctttc ggcggggagc aggggagacc tttggcctac ggcgacggga 2220

gggtcgggac aaagtttagg gcgtcgataa gcgtcagagc gccgaggttg ggggagggtt 2280

tctcttccgc tctttcgcgg ggcctctggc tcccccagcg cagctggagt gggggacggg 2340

taggctcgtc ccaaaggcgc ggcgctgagg tttgtgaacg cgtggagggg cgcttggggt 2400

ctgggggagg cgtcgcccgg gtaagcctgt ctgctgcggc tctgcttccc ttagactgga 2460

gagctgtgga cttcgtctag gcgcccgcta agttcgcatg tcctagcacc tctgggtcta 2520

tgtggggcca caccgtgggg aggaaacagc acgcgacgtt tgtagaatgc ttggctgtga 2580

tacaaagcgg tttcgaataa ttaacttatt tgttcccatc acatgtcact tttaaaaaat 2640

tataagaact acccgttatt gacatctttc tgtgtgccaa ggactttatg tgctttgcgt 2700

catttaattt tgaaaacagt tatcttccgc catagataac tactatggtt atcttctgcc 2760

tctcacagat gaagaaacta aggcaccgag attttaagaa acttaattac acaggggata 2820

aatggcagca atcgagattg aagtcaagcc taaccagggc ttttgcggga gcgcatgcct 2880

tttggctgta attcgtgcat tttttttaa gaaaaacgcc tgccttctgc gtgagattct 2940

ccagagcaaa ctgggcggca tgggccctgt ggtcttttcg tacagagggc ttcctctttg 3000

gctctttgcc tggttgtttc caagatgtac tgtgcctctt actttcggtt ttgaaaacat 3060

gagggggttg ggcgtggtag cttacgcctg taatcccagc acttagggag gccgaggcgg 3120

gaggatggct tgaggtccgt agttgagacc agcctggcca acatggtgaa gcctggtctc 3180

tacaaaaaat aataacaaaa attagccggg tgtggtggct cgtgcctgtg gtcccagctg 3240

ctccggtggc tgaggcggga ggatctcttg agcttaggct tttgagctat catggcgcca 3300

gtgcactcca gcgtgggcaa cagagcgaga ccctgtctct caaaaaagaa aaaaaaaaaa 3360

aaagaaagag aaaagaaaag aaagaaagaa gtgaaggttt gtcagtcagg ggagctgtaa 3420

aaccattaat aaagataatc caagatggtt accaagactg ttgaggacgc cagagatctt 3480

gagcactttc taagtacctg gcaatacact aagcgcgctc acctttcct ctggcaaaac 3540

atgatcgaaa gcagaatgtt ttgatcatga gaaaattgca tttaatttga atacaattta 3600

tttacaacat aaaggataat gtatatatca ccaccattac tggtatttgc tggttatgtt 3660

agatgtcatt ttaaaaaata acaatctgat atttaaaaaa aaatcttatt ttgaaaattt 3720

ccaaagtaat acatgccatg catagaccat ttctggaaga taccacaaga aacatgtaat 3780

gatgattgcc tctgaaggtc tattttcctc ctctgacctg tgtgtgggtt ttgtttttgt 3840

tttactgtgg gcataaatta atttttcagt taagttttgg aagcttaaat aactctccaa 3900

aagtcataaa gccagtaact ggttgagccc aaattcaaac ccagcctgtc tgatacttgt 3960

cctcttctta gaaaagatta cagtgatgct ctcacaaaat cttgccgcct tccctcaaac 4020

agagagttcc aggcaggatg aatctgtgct ctgatccctg aggcatttaa tatgttctta 4080

ttattagaag ctcagatgca aagagctctc ttagctttta atgttatgaa aaaaatcagg 4140

tcttcattag attccccaat ccacctcttg atggggctag tagcctttcc ttaatgatag 4200

ggtgtttcta gagagatata tctggtcaag gtggcctggt actcctcctt ctccccacag 4260

cctcccagac aaggaggagt agctgccttt tagtgatcat gtaccctgaa tataagtgta 4320

tttaaaagaa ttttatacac atatatttag tgtcaatctg tatatttagt agcactaaca 4380

cttctcttca ttttcaatga aaaatataga gtttataata ttttcttccc acttccccat 4440
ggatggtcta gtcatgcctc tcattttgga aagtactgtt tctgaaacat taggcaatat 4500
attcccaacc tggctagttt acagcaatca cctgtggatg ctaattaaaa cgcaaatccc 4560
actgtcacat gcattactcc atttgatcat aatggaaagt atgttctgtc ccatttgcca 4620
tagtcctcac ctatccctgt tgtattttat cgggtccaac tcaaccattt aaggtatttg 4680
ccagctcttg tatgcattta ggttttgttt ctttgttttt tagctcatga aattaggtac 4740
aaagtcagag aggggtctgg catataaaac ctcagcagaa ataaagaggt tttgttgttt 4800

ggtaagaaca taccttgggt tggttgggca cggtggctcg tgcctgtaat cccaacactt 4860
tgggaggcca aggcaggctg atcacttgaa gttgggagtt caagaccagc ctggccaaca 4920
tggtgaaatc ccgtctctac tgaaaataca aaaattaacc aggcatggtg gtgtgtgcct 4980
gtagtcccag gaatcacttg aacccaggag gcggaggttg cagtgagctg agatctcacc 5040
actgcacact gcactccagc ctgggcaatg gaatgagatt ccatcccaaa aaataaaaaa 5100
ataaaaaaat aaagaacata ccttgggttg atccacttag gaacctcaga taataacatc 5160
tgccacgtat agagcaattg ctatgtccca ggcactctac tagacacttc atacagttta 5220

gaaaatcaga tgggtgtaga tcaaggcagg agcaggaacc aaaaagaaag gcataaacat 5280
aagaaaaaaa atggaagggg tggaaacaga gtacaataac atgagtaatt tgatgggggc 5340
tattatgaac tgagaaatga actttgaaaa gtatcttggg gccaaatcat gtagactctt 5400
gagtgatgtg ttaaggaatg ctatgagtgc tgagagggca tcagaagtcc ttgagagcct 5460
ccagagaaag gctcttaaaa atgcagcgca atctccagtg acagaagata ctgctagaaa 5520
tctgctagaa aaaaaacaaa aaaggcatgt atagaggaat tatgagggaa agataccaag 5580
tcacggttta ttcttcaaaa tggaggtggc ttgttgggaa ggtggaagct catttggcca 5640

gagtggaaat ggaattggga gaaatcgatg accaaatgta aacacttggt gcctgatata 5700
gcttgacacc aagttagccc caagtgaaat accctggcaa tattaatgtg tcttttcccg 5760
atattcctca ggtactccaa agattcaggt ttactcacgt catccagcag agaatggaaa 5820
gtcaaatttc ctgaattgct atgtgtctgg gtttcatcca tccgacattg aagttgactt 5880
actgaagaat ggagagagaa ttgaaaaagt ggagcattca gacttgtctt tcagcaagga 5940
ctggtctttc tatctcttgt actacactga attcaccccc actgaaaaag atgagtatgc 6000
ctgccgtgtg aaccatgtga ctttgtcaca gcccaagata gttaagtggg gtaagtctta 6060

cattcttttg taagctgctg aaagttgtgt atgagtagtc atatcataaa gctgctttga 6120

tataaaaaag gtctatggcc atactaccct gaatgagtcc catcccatct gatataaaca 6180

atctgcatat tgggattgtc agggaatgtt cttaaagatc agattagtgg cacctgctga 6240

gatactgatg cacagcatgg tttctgaacc agtagtttcc ctgcagttga gcagggagca 6300

gcagcagcac ttgcacaaat acatatacac tcttaacact tcttacctac tggcttcctc 6360

tagcttttgt ggcagcttca ggtatattta gcactgaacg aacatctcaa gaaggtatag 6420

gcctttgttt gtaagtcctg ctgtcctagc atcctataat cctggacttc tccagtactt 6480

tctggctgga ttggtatctg aggctagtag gaagggcttg ttcctgctgg gtagctctaa 6540

acaatgtatt catgggtagg aacagcagcc tattctgcca gccttatttc taaccatttt 6600

agacatttgt tagtacatgg tattttaaaa gtaaaactta atgtcttcct tttttttctc 6660

cactgtcttt ttcatagatc gagacatgta agcagcatca tggaggtaag tttttgacct 6720

tgagaaaatg tttttgtttc actgtcctga ggactattta tagacagctc taacatgata 6780

accctcacta tgtggagaac attgacagag taacatttta gcagggaaag aagaatccta 6840

cagggtcatg ttcccttctc ctgtggagtg gcatgaagaa ggtgtatggc cccaggtatg 6900

gccatattac tgaccctcta cagagagggc aaaggaactg ccagtatggt attgcaggat 6960

aaaggcaggt ggttacccac attacctgca aggctttgat ctttcttctg ccatttccac 7020

attggacatc tctgctgagg agagaaaatg aaccactctt ttcctttgta taatgttgtt 7080

ttattcttca gacagaagag aggagttata cagctctgca gacatcccat tcctgtatgg 7140

ggactgtgtt tgcctcttag aggttcccag gccactagag gagataaagg gaaacagatt 7200

gttataactt gatataatga tactataata gatgtaacta caaggagctc cagaagcaag 7260

agagagggag gaacttggac ttctctgcat ctttagttgg agtccaaagg cttttcaatg 7320

aaattctact gcccagggta cattgatgct gaaaccccat tcaaatctcc tgttatattc 7380

tagaacaggg aattgatttg ggagagcatc aggaaggtgg atgatctgcc cagtcacact 7440

gttagtaaat tgtagagcca ggacctgaac tctaatatag tcatgtgtta cttaatgacg 7500

gggacatgtt ctgagaaatg cttacacaaa cctaggtgtt gtagcctact acacgcatag 7560

gctacatggt atagcctatt gctcctagac tacaaacctg tacagcctgt tactgtactg 7620

aatactgtgg gcagttgtaa cacaatggta agtatttgtg tatctaaaca tagaagttgc 7680

agtaaaaata tgctatttta atcttatgag accactgtca tatatacagt ccatcattga 7740

ccaaaacatc atatcagcat tttttcttct aagattttgg gagcaccaaa gggatacact 7800

aacaggatat actctttata atgggtttgg agaactgtct gcagctactt cttttaaaaa 7860

ggtgatctac acagtagaaa ttagacaagt ttggtaatga gatctgcaat ccaaataaaa 7920

taaattcatt gctaaccttt ttcttttctt ttcaggtttg aagatgccgc atttggattg 7980

gatgaattcc aaattctgct tgcttgcttt ttaatattga tatgcttata cacttacact 8040

ttatgcacaa aatgtagggt tataataatg ttaacatgga catgatcttc tttataattc 8100

tactttgagt gctgtctcca tgtttgatgt atctgagcag gttgctccac aggtagctct 8160

aggagggctg gcaacttaga ggtggggagc agagaattct cttatccaac atcaacatct 8220

tggtcagatt tgaactcttc aatctcttgc actcaaagct tgttaagata gttaagcgtg 8280

cataagttaa cttccaattt acatactctg cttagaattt gggggaaaat ttagaaatat 8340

aattgacagg attattggaa atttgttata atgaatgaaa cattttgtca tataagattc 8400

atatttactt cttatacatt tgataaagta aggcatggtt gtggttaatc tggtttattt 8460

ttgttccaca agttaaataa atcataaaac ttgatgtgtt atctcttata tctcactccc 8520

actattaccc ctttattttc aaacagggaa acagtcttca agttccactt ggtaaaaaat 8580

gtgaacccct tgtatataga gtttggctca cagtgtaaag ggcctcagtg attcacattt 8640

tccagattag gaatctgatg ctcaaagaag ttaaatggca tagttggggt gacacagctg 8700

tctagtggga ggccagcctt ctatatttta gccagcgttc tttcctgcgg gccaggtcat 8760

gaggagtatg cagactctaa gagggagcaa aagtatctga aggatttaat attttagcaa 8820

ggaatagata tacaatcatc ccttggtctc cctgggggat tggtttcagg accccttctt 8880

ggacaccaaa tctatggata tttaagtccc ttctataaaa tggtatagta tttgcatata 8940

acctatccac atcctcctgt atactttaaa tcatttctag attacttgta atacctaata 9000

caatgtaaat gctatgcaaa tagttgttat tgtttaagga ataatgacaa gaaaaaaaag 9060

tctgtacatg ctcagtaaag acacaaccat ccctttttt ccccagtgtt tttgatccat 9120

ggtttgctga atccacagat gtggagcccc tggatacgga aggcccgctg tactttgaat 9180

gacaaataac agatttaaa 9199

<210> 19

<211> 982

<212> PRT

<213> Homo sapiens

<400> 19

Met Ala Asn Ser Met Asn Gly Arg Asn Pro Gly Gly Arg Gly Gly Asn

1 5 10 15

Pro Arg Lys Gly Arg Ile Leu Gly Ile Ile Asp Ala Ile Gln Asp Ala

20 25 30

Val Gly Pro Pro Lys Gln Ala Ala Ala Asp Arg Arg Thr Val Glu Lys
35 40 45
Thr Trp Lys Leu Met Asp Lys Val Val Arg Leu Cys Gln Asn Pro Lys
50 55 60
Leu Gln Leu Lys Asn Ser Pro Pro Tyr Ile Leu Asp Ile Leu Pro Asp
65 70 75 80
Thr Tyr Gln His Leu Arg Leu Ile Leu Ser Lys Tyr Asp Asp Asn Gln

85 90 95
Lys Leu Ala Gln Leu Ser Glu Asn Glu Tyr Phe Lys Ile Tyr Ile Asp
100 105 110
Ser Leu Met Lys Lys Ser Lys Arg Ala Ile Arg Leu Phe Lys Glu Gly
115 120 125
Lys Glu Arg Met Tyr Glu Glu Gln Ser Gln Asp Arg Arg Asn Leu Thr
130 135 140
Lys Leu Ser Leu Ile Phe Ser His Met Leu Ala Glu Ile Lys Ala Ile

145 150 155 160
Phe Pro Asn Gly Gln Phe Gln Gly Asp Asn Phe Arg Ile Thr Lys Ala
165 170 175
Asp Ala Ala Glu Phe Trp Arg Lys Phe Phe Gly Asp Lys Thr Ile Val
180 185 190
Pro Trp Lys Val Phe Arg Gln Cys Leu His Glu Val His Gln Ile Ser
195 200 205
Ser Gly Leu Glu Ala Met Ala Leu Lys Ser Thr Ile Asp Leu Thr Cys

210 215 220
Asn Asp Tyr Ile Ser Val Phe Glu Phe Asp Ile Phe Thr Arg Leu Phe
225 230 235 240
Gln Pro Trp Gly Ser Ile Leu Arg Asn Trp Asn Phe Leu Ala Val Thr
245 250 255
His Pro Gly Tyr Met Ala Phe Leu Thr Tyr Asp Glu Val Lys Ala Arg
260 265 270
Leu Gln Lys Tyr Ser Thr Lys Pro Gly Ser Tyr Ile Phe Arg Leu Ser

275 280 285

Cys Thr Arg Leu Gly Gln Trp Ala Ile Gly Tyr Val Thr Gly Asp Gly

290 295 300

Asn Ile Leu Gln Thr Ile Pro His Asn Lys Pro Leu Phe Gln Ala Leu

305 310 315 320

Ile Asp Gly Ser Arg Glu Gly Phe Tyr Leu Tyr Pro Asp Gly Arg Ser

325 330 335

Tyr Asn Pro Asp Leu Thr Gly Leu Cys Glu Pro Thr Pro His Asp His

340 345 350

Ile Lys Val Thr Gln Glu Gln Tyr Glu Leu Tyr Cys Glu Met Gly Ser

355 360 365

Thr Phe Gln Leu Cys Lys Ile Cys Ala Glu Asn Asp Lys Asp Val Lys

370 375 380

Ile Glu Pro Cys Gly His Leu Met Cys Thr Ser Cys Leu Thr Ala Trp

385 390 395 400

Gln Glu Ser Asp Gly Gln Gly Cys Pro Phe Cys Arg Cys Glu Ile Lys

405 410 415

Gly Thr Glu Pro Ile Ile Val Asp Pro Phe Asp Pro Arg Asp Glu Gly

420 425 430

Ser Arg Cys Cys Ser Ile Ile Asp Pro Phe Gly Met Pro Met Leu Asp

435 440 445

Leu Asp Asp Asp Asp Asp Arg Glu Glu Ser Leu Met Met Asn Arg Leu

450 455 460

Ala Asn Val Arg Lys Cys Thr Asp Arg Gln Asn Ser Pro Val Thr Ser

465 470 475 480

Pro Gly Ser Ser Pro Leu Ala Gln Arg Arg Lys Pro Gln Pro Asp Pro

485 490 495

Leu Gln Ile Pro His Leu Ser Leu Pro Pro Val Pro Pro Arg Leu Asp

500 505 510

Leu Ile Gln Lys Gly Ile Val Arg Ser Pro Cys Gly Ser Pro Thr Gly

515 520 525

Ser Pro Lys Ser Ser Pro Cys Met Val Arg Lys Gln Asp Lys Pro Leu

530 535 540

Pro Ala Pro Pro Pro Pro Leu Arg Asp Pro Pro Pro Pro Pro Pro Glu

545 550 555 560

Arg Pro Pro Pro Ile Pro Pro Asp Asn Arg Leu Ser Arg His Ile His

565 570 575

His Val Glu Ser Val Pro Ser Arg Asp Pro Pro Met Pro Leu Glu Ala

580 585 590

Trp Cys Pro Arg Asp Val Phe Gly Thr Asn Gln Leu Val Gly Cys Arg

595 600 605

Leu Leu Gly Glu Gly Ser Pro Lys Pro Gly Ile Thr Ala Ser Ser Asn

610 615 620

Val Asn Gly Arg His Ser Arg Val Gly Ser Asp Pro Val Leu Met Arg

625 630 635 640

Lys His Arg Arg His Asp Leu Pro Leu Glu Gly Ala Lys Val Phe Ser

645 650 655

Asn Gly His Leu Gly Ser Glu Glu Tyr Asp Val Pro Pro Arg Leu Ser

660 665 670

Pro Pro Pro Pro Val Thr Thr Leu Leu Pro Ser Ile Lys Cys Thr Gly

675 680 685

Pro Leu Ala Asn Ser Leu Ser Glu Lys Thr Arg Asp Pro Val Glu Glu

690 695 700

Asp Asp Asp Glu Tyr Lys Ile Pro Ser Ser His Pro Val Ser Leu Asn

705 710 715 720

Ser Gln Pro Ser His Cys His Asn Val Lys Pro Pro Val Arg Ser Cys

725 730 735

Asp Asn Gly His Cys Met Leu Asn Gly Thr His Gly Pro Ser Ser Glu

740 745 750

Lys Lys Ser Asn Ile Pro Asp Leu Ser Ile Tyr Leu Lys Gly Asp Val

755 760 765

Phe Asp Ser Ala Ser Asp Pro Val Pro Leu Pro Pro Ala Arg Pro Pro

770 775 780

Thr Arg Asp Asn Pro Lys His Gly Ser Ser Leu Asn Arg Thr Pro Ser

785 790 795 800

Asp Tyr Asp Leu Leu Ile Pro Pro Leu Gly Glu Asp Ala Phe Asp Ala

805 810 815

Leu Pro Pro Ser Leu Pro Pro Pro Pro Pro Pro Ala Arg His Ser Leu

820 825 830

Ile Glu His Ser Lys Pro Pro Gly Ser Ser Ser Arg Pro Ser Ser Gly

835 840 845

Gln Asp Leu Phe Leu Leu Pro Ser Asp Pro Phe Val Asp Leu Ala Ser

850 855 860

Gly Gln Val Pro Leu Pro Pro Ala Arg Arg Leu Pro Gly Glu Asn Val

865 870 875 880

Lys Thr Asn Arg Thr Ser Gln Asp Tyr Asp Gln Leu Pro Ser Cys Ser

885 890 895

Asp Gly Ser Gln Ala Pro Ala Arg Pro Pro Lys Pro Arg Pro Arg Arg

900 905 910

Thr Ala Pro Glu Ile His His Arg Lys Pro His Gly Pro Glu Ala Ala

915 920 925

Leu Glu Asn Val Asp Ala Lys Ile Ala Lys Leu Met Gly Glu Gly Tyr

930 935 940

Ala Phe Glu Glu Val Lys Arg Ala Leu Glu Ile Ala Gln Asn Asn Val

945 950 955 960

Glu Val Ala Arg Ser Ile Leu Arg Glu Phe Ala Phe Pro Pro Pro Val

965 970 975

Ser Pro Arg Leu Asn Leu

980

<210> 20

<211> 2949

<212> DNA

<213> Homo sapiens

<400> 20

atggcaaact caatgaatgg cagaaaccct ggtggtcgag gaggaaatcc ccgaaaaggt 60

cgaattttgg gtattattga tgctattcag gatgcagttg gacccccctaa gcaagctgcc 120

gcagatcgca ggaccgtgga gaagacttgg aagctcatgg acaaagtggt aagactgtgc 180

caaaatccca aacttcagtt gaaaaatagc ccaccatata tacttgatat tttgcctgat 240

acatatcagc atttacgact tatattgagt aaatatgatg acaaccagaa acttgcccaa 300

ctcagtgaga atgagtactt taaaatctac attgatagcc ttatgaaaaa gtcaaaacgg 360

gcaataagac tctttaaaga aggcaaggag agaatgtatg aagaacagtc acaggacaga 420

cgaaatctca caaaactgtc ccttatcttc agtcacatgc tggcagaaat caaagcaatc 480

tttcccaatg gtcaattcca gggagataac tttcgtatca caaaagcaga tgctgctgaa 540

ttctggagaa agttttttgg agacaaaact atcgtaccat ggaaagtatt cagacagtgc 600

cttcatgagg tccaccagat tagctctggc ctggaagcaa tggctctaaa atcaacaatt 660

gatttaactt gcaatgatta catttcagtt tttgaatttg atatttttac caggctgttt 720

cagccttggg gctctatttt gcggaattgg aatttcttag ctgtgacaca tccaggttac 780

atggcatttc tcacatatga tgaagttaaa gcacgactac agaaatatag caccaaaccc 840

ggaagctata ttttccggtt aagttgcact cgattgggac agtgggccat tggctatgtg 900

actggggatg ggaatatctt acagaccata cctcataaca agcccttatt tcaagccctg 960

attgatggca gcagggaagg attttatctt tatcctgatg ggaggagtta taatcctgat 1020

ttaactggat tatgtgaacc tacacctcat gaccatataa aagttacaca ggaacaatat 1080

gaattatatt gtgaaatggg ctccactttt cagctctgta agatttgtgc agagaatgac 1140

aaagatgtca agattgagcc ttgtgggcat ttgatgtgca cctcttgcct tacggcatgg 1200

caggagtcgg atggtcaggg ctgccctttc tgtcgttgtg aaataaaagg aactgagccc 1260

ataatcgtgg acccctttga tccaagagat gaaggctcca ggtgttgcag catcattgac 1320

ccctttggca tgccgatgct agacttggac gacgatgatg atcgtgagga gtccttgatg 1380

atgaatcggt tggcaaacgt ccgaaagtgc actgacaggc agaactcacc agtcacatca 1440

ccaggatcct ctccccttgc ccagagaaga aagccacagc ctgacccact ccagatccca 1500

catctaagcc tgccacccgt gcctcctcgc ctggatctaa ttcagaaagg catagttaga 1560

tctccctgtg gcagcccaac gggttcacca aagtcttctc cttgcatggt gagaaaacaa 1620

gataaaccac tcccagcacc acctcctccc ttaagagatc ctcctccacc gccacctgaa 1680

agacctccac caatcccacc agacaataga ctgagtagac acatccatca tgtggaaagc 1740

gtgccttcca gagacccgcc aatgcctctt gaagcatggt gccctcggga tgtgtttggg 1800

actaatcagc ttgtgggatg tcgactccta ggggagggct ctccaaaacc tggaatcaca 1860

gcgagttcaa atgtcaatgg aaggcacagt agagtgggct ctgacccagt gcttatgcgg 1920

aaacacagac gccatgattt gcctttagaa ggagctaagg tcttttccaa tggtcacctt 1980

ggaagtgaag aatatgatgt tcctccccgg ctttctcctc ctcctccagt taccaccctc 2040

ctccctagca taaagtgtac tggtccgtta gcaaattctc tttcagagaa aacaagagac 2100

ccagtagagg aagatgatga tgaatacaag attccttcat cccaccctgt ttccctgaat 2160

tcacaaccat ctcattgtca taatgtaaaa cctcctgttc ggtcttgtga taatggtcac 2220

tgtatgctga atggaacaca tggtccatct tcagagaaga aatcaaacat ccctgactta 2280

agcatatatt taaagggaga tgtttttgat tcagcctctg atcccgtgcc attaccacct 2340

gccaggcctc caactcggga caatccaaag catggttctt cactcaacag gacgccctct 2400

gattatgatc ttctcatccc tccattaggt gaagatgctt ttgatgccct ccctccatct 2460

ctcccacctc ccccacctcc tgcaaggcat agtctcattg aacattcaaa acctcctggc 2520

tccagtagcc ggccatcctc aggacaggat ctttttcttc ttccttcaga tccctttgtt 2580

gatctagcaa gtggccaagt tcctttgcct cctgctagaa ggttaccagg tgaaaatgtc 2640

aaaactaaca gaacatcaca ggactatgat cagcttcctt catgttcaga tggttcacag 2700

gcaccagcca gacccccta accacgaccg cgcaggactg caccagaaat tcaccacaga 2760

aaaccccatg ggcctgaggc ggcattggaa aatgtcgatg caaaaattgc aaaactcatg 2820

ggagagggtt atgcctttga agaggtgaag agagccttag agatagccca gaataatgtc 2880

gaagttgccc ggagcatcct ccgagaattt gccttccctc ctccagtatc cccacgtcta 2940

aatctatag 2949

<210> 21

<211> 1131

<212> PRT

<213> Homo sapiens

<400> 21

Met Arg Cys Leu Ala Pro Arg Pro Ala Gly Ser Tyr Leu Ser Glu Pro

1 5 10 15

Gln Gly Ser Ser Gln Cys Ala Thr Met Glu Leu Gly Pro Leu Glu Gly

20 25 30

Gly Tyr Leu Glu Leu Leu Asn Ser Asp Ala Asp Pro Leu Cys Leu Tyr

35 40 45

His Phe Tyr Asp Gln Met Asp Leu Ala Gly Glu Glu Glu Ile Glu Leu
50 55 60
Tyr Ser Glu Pro Asp Thr Asp Thr Ile Asn Cys Asp Gln Phe Ser Arg
65 70 75 80
Leu Leu Cys Asp Met Glu Gly Asp Glu Glu Thr Arg Glu Ala Tyr Ala
85 90 95
Asn Ile Ala Glu Leu Asp Gln Tyr Val Phe Gln Asp Ser Gln Leu Glu

100 105 110
Gly Leu Ser Lys Asp Ile Phe Ile Glu His Ile Gly Pro Asp Glu Val
115 120 125
Ile Gly Glu Ser Met Glu Met Pro Ala Glu Val Gly Gln Lys Ser Gln
130 135 140
Lys Arg Pro Phe Pro Glu Glu Leu Pro Ala Asp Leu Lys His Trp Lys
145 150 155 160
Pro Ala Glu Pro Pro Thr Val Val Thr Gly Ser Leu Leu Val Gly Pro

165 170 175
Val Ser Asp Cys Ser Thr Leu Pro Cys Leu Pro Leu Pro Ala Leu Phe
180 185 190
Asn Gln Glu Pro Ala Ser Gly Gln Met Arg Leu Glu Lys Thr Asp Gln
195 200 205
Ile Pro Met Pro Phe Ser Ser Ser Ser Leu Ser Cys Leu Asn Leu Pro
210 215 220
Glu Gly Pro Ile Gln Phe Val Pro Thr Ile Ser Thr Leu Pro His Gly

225 230 235 240
Leu Trp Gln Ile Ser Glu Ala Gly Thr Gly Val Ser Ser Ile Phe Ile
245 250 255
Tyr His Gly Glu Val Pro Gln Ala Ser Gln Val Pro Pro Pro Ser Gly
260 265 270
Phe Thr Val His Gly Leu Pro Thr Ser Pro Asp Arg Pro Gly Ser Thr
275 280 285
Ser Pro Phe Ala Pro Ser Ala Thr Asp Leu Pro Ser Met Pro Glu Pro

290 295 300

Ala Leu Thr Ser Arg Ala Asn Met Thr Glu His Lys Thr Ser Pro Thr

305 310 315 320

Gln Cys Pro Ala Ala Gly Glu Val Ser Asn Lys Leu Pro Lys Trp Pro

325 330 335

Glu Pro Val Glu Gln Phe Tyr Arg Ser Leu Gln Asp Thr Tyr Gly Ala

340 345 350

Glu Pro Ala Gly Pro Asp Gly Ile Leu Val Glu Val Asp Leu Val Gln

355 360 365

Ala Arg Leu Glu Arg Ser Ser Ser Lys Ser Leu Glu Arg Glu Leu Ala

370 375 380

Thr Pro Asp Trp Ala Glu Arg Gln Leu Ala Gln Gly Gly Leu Ala Glu

385 390 395 400

Val Leu Leu Ala Ala Lys Glu His Arg Arg Pro Arg Glu Thr Arg Val

405 410 415

Ile Ala Val Leu Gly Lys Ala Gly Gln Gly Lys Ser Tyr Trp Ala Gly

420 425 430

Ala Val Ser Arg Ala Trp Ala Cys Gly Arg Leu Pro Gln Tyr Asp Phe

435 440 445

Val Phe Ser Val Pro Cys His Cys Leu Asn Arg Pro Gly Asp Ala Tyr

450 455 460

Gly Leu Gln Asp Leu Leu Phe Ser Leu Gly Pro Gln Pro Leu Val Ala

465 470 475 480

Ala Asp Glu Val Phe Ser His Ile Leu Lys Arg Pro Asp Arg Val Leu

485 490 495

Leu Ile Leu Asp Gly Phe Glu Glu Leu Glu Ala Gln Asp Gly Phe Leu

500 505 510

His Ser Thr Cys Gly Pro Ala Pro Ala Glu Pro Cys Ser Leu Arg Gly

515 520 525

Leu Leu Ala Gly Leu Phe Gln Lys Lys Leu Leu Arg Gly Cys Thr Leu

530 535 540

Leu Leu Thr Ala Arg Pro Arg Gly Arg Leu Val Gln Ser Leu Ser Lys

545 550 555 560

Ala Asp Ala Leu Phe Glu Leu Ser Gly Phe Ser Met Glu Gln Ala Gln

565 570 575

Ala Tyr Val Met Arg Tyr Phe Glu Ser Ser Gly Met Thr Glu His Gln

580 585 590

Asp Arg Ala Leu Thr Leu Leu Arg Asp Arg Pro Leu Leu Leu Ser His

595 600 605

Ser His Ser Pro Thr Leu Cys Arg Ala Val Cys Gln Leu Ser Glu Ala

610 615 620

Leu Leu Glu Leu Gly Glu Asp Ala Lys Leu Pro Ser Thr Leu Thr Gly

625 630 635 640

Leu Tyr Val Gly Leu Leu Gly Arg Ala Ala Leu Asp Ser Pro Pro Gly

645 650 655

Ala Leu Ala Glu Leu Ala Lys Leu Ala Trp Glu Leu Gly Arg Arg His

660 665 670

Gln Ser Thr Leu Gln Glu Asp Gln Phe Pro Ser Ala Asp Val Arg Thr

675 680 685

Trp Ala Met Ala Lys Gly Leu Val Gln His Pro Pro Arg Ala Ala Glu

690 695 700

Ser Glu Leu Ala Phe Pro Ser Phe Leu Leu Gln Cys Phe Leu Gly Ala

705 710 715 720

Leu Trp Leu Ala Leu Ser Gly Glu Ile Lys Asp Lys Glu Leu Pro Gln

725 730 735

Tyr Leu Ala Leu Thr Pro Arg Lys Lys Arg Pro Tyr Asp Asn Trp Leu

740 745 750

Glu Gly Val Pro Arg Phe Leu Ala Gly Leu Ile Phe Gln Pro Pro Ala

755 760 765

Arg Cys Leu Gly Ala Leu Leu Gly Pro Ser Ala Ala Ala Ser Val Asp

770 775 780

Arg Lys Gln Lys Val Leu Ala Arg Tyr Leu Lys Arg Leu Gln Pro Gly

785 790 795 800

Thr Leu Arg Ala Arg Gln Leu Leu Glu Leu Leu His Cys Ala His Glu

805 810 815

Ala Glu Glu Ala Gly Ile Trp Gln His Val Val Gln Glu Leu Pro Gly

820 825 830

Arg Leu Ser Phe Leu Gly Thr Arg Leu Thr Pro Pro Asp Ala His Val

835 840 845

Leu Gly Lys Ala Leu Glu Ala Ala Gly Gln Asp Phe Ser Leu Asp Leu

850 855 860

Arg Ser Thr Gly Ile Cys Pro Ser Gly Leu Gly Ser Leu Val Gly Leu

865 870 875 880

Ser Cys Val Thr Arg Phe Arg Ala Ala Leu Ser Asp Thr Val Ala Leu

885 890 895

Trp Glu Ser Leu Gln Gln His Gly Glu Thr Lys Leu Leu Gln Ala Ala

900 905 910

Glu Glu Lys Phe Thr Ile Glu Pro Phe Lys Ala Lys Ser Leu Lys Asp

915 920 925

Val Glu Asp Leu Gly Lys Leu Val Gln Thr Gln Arg Thr Arg Ser Ser

930 935 940

Ser Glu Asp Thr Ala Gly Glu Leu Pro Ala Val Arg Asp Leu Lys Lys

945 950 955 960

Leu Glu Phe Ala Leu Gly Pro Val Ser Gly Pro Gln Ala Phe Pro Lys

965 970 975

Leu Val Arg Ile Leu Thr Ala Phe Ser Ser Leu Gln His Leu Asp Leu

980 985 990

Asp Ala Leu Ser Glu Asn Lys Ile Gly Asp Glu Gly Val Ser Gln Leu

995 1000 1005

Ser Ala Thr Phe Pro Gln Leu Lys Ser Leu Glu Thr Leu Asn Leu

1010 1015 1020

Ser Gln Asn Asn Ile Thr Asp Leu Gly Ala Tyr Lys Leu Ala Glu

1025 1030 1035

Ala Leu Pro Ser Leu Ala Ala Ser Leu Leu Arg Leu Ser Leu Tyr
1040 1045 1050
Asn Asn Cys Ile Cys Asp Val Gly Ala Glu Ser Leu Ala Arg Val

1055 1060 1065
Leu Pro Asp Met Val Ser Leu Arg Val Met Asp Val Gln Tyr Asn
1070 1075 1080
Lys Phe Thr Ala Ala Gly Ala Gln Gln Leu Ala Ala Ser Leu Arg
1085 1090 1095
Arg Cys Pro His Val Glu Thr Leu Ala Met Trp Thr Pro Thr Ile
1100 1105 1110
Pro Phe Ser Val Gln Glu His Leu Gln Gln Gln Asp Ser Arg Ile

1115 1120 1125
Ser Leu Arg
1130

<210> 22
<211> 4657
<212> DNA
<213> Homo sapiens
<400> 22

ggttagtgat gaggctagtg atgaggctgt gtgcttctga gctgggcatc cgaaggcatc 60
cttggggaag ctgagggcac gaggaggggc tgccagactc cgggagctgc tgcctggctg 120
ggattcctac acaatgcgtt gcctggctcc acgccctgct gggtcctacc tgtcagagcc 180
ccaaggcagc tcacagtgtg ccaccatgga gttggggccc ctagaaggtg gctacctgga 240
gcttcttaac agcgatgctg acccccctgtg cctctaccac ttctatgacc agatggacct 300

ggctggagaa gaagagattg agctctactc agaacccgac acagacacca tcaactgcga 360
ccagttcagc aggctgttgt gtgacatgga aggtgatgaa gagaccaggg aggcttatgc 420
caatatcgcg gaactggacc agtatgtctt ccaggactcc cagctggagg gcctgagcaa 480
ggacattttc atagagcaca taggaccaga tgaagtgatc ggtgagagta tggagatgcc 540
agcagaagtt gggcagaaaa gtcagaaaag acccttccca gaggagcttc cggcagacct 600
gaagcactgg aagccagctg agccccccac tgtggtgact ggcagtctcc tagtgggacc 660
agtgagcgac tgctccaccc tgccctgcct gccactgcct gcgctgttca accaggagcc 720

agcctccggc cagatgcgcc tggagaaaac cgaccagatt cccatgcctt tctccagttc 780

ctcgttgagc tgcctgaatc tccctgaggg acccatccag tttgtcccca ccatctccac 840

tctgccccat gggctctggc aaatctctga ggctggaaca ggggtctcca gtatattcat 900

ctaccatggt gaggtgcccc aggccagcca agtacccccc cccagtggat tcactgtcca 960

cggcctccca acatctccag accggccagg ctccaccagc cccttcgctc catcagccac 1020

tgacctgccc agcatgcctg aacctgccct gacctcccga gcaaacatga cagagcacaa 1080

gacgtccccc acccaatgcc cggcagctgg agaggtctcc aacaagcttc caaaatggcc 1140

tgagccggtg gagcagttct accgctcact gcaggacacg tatggtgccg agcccgcagg 1200

cccggatggc atcctagtgg aggtggatct ggtgcaggcc aggctggaga ggagcagcag 1260

caagagcctg gagcgggaac tggccacccc ggactgggca gaacggcagc tggcccaagg 1320

aggcctggct gaggtgctgt tggctgccaa ggagcaccgg cggccgcgtg agacacgagt 1380

gattgctgtg ctgggcaaag ctggtcaggg caagagctat tgggctgggg cagtgagccg 1440

ggcctgggct tgtggccggc ttccccagta cgactttgtc ttctctgtcc cctgccattg 1500

cttgaaccgt ccgggggatg cctatggcct gcaggatctg ctcttctccc tgggcccaca 1560

gccactcgtg gcggccgatg aggttttcag ccacatcttg aagagacctg accgcgttct 1620

gctcatccta gacggcttcg aggagctgga agcgcaagat ggcttcctgc acagcacgtg 1680

cggaccggca ccggcggagc cctgctccct ccgggggctg ctggccggcc ttttccagaa 1740

gaagctgctc cgaggttgca ccctcctcct cacagcccgg ccccggggcc gcctggtcca 1800

gagcctgagc aaggccgacg ccctatttga gctgtccggc ttctccatgg agcaggccca 1860

ggcatacgtg atgcgctact ttgagagctc agggatgaca gagcaccaag acagagccct 1920

gacgctcctc cgggaccggc cacttcttct cagtcacagc cacagcccta ctttgtgccg 1980

ggcagtgtgc cagctctcag aggccctgct ggagcttggg gaggacgcca agctgccctc 2040

cacgctcacg ggactctatg tcggcctgct gggccgtgca gccctcgaca gccccccggg 2100

ggccctggca gagctggcca agctggcctg ggagctgggc cgcagacatc aaagtaccct 2160

acaggaggac cagttcccat ccgcagacgt gaggacctgg gcgatggcca aaggcttagt 2220

ccaacaccca ccgcgggccg cagagtccga gctggccttc cccagcttcc tcctgcaatg 2280

cttcctgggg gccctgtggc tggctctgag tggcgaaatc aaggacaagg agctcccgca 2340

gtacctagca ttgaccccaa ggaagaagag gccctatgac aactggctgg agggcgtgcc 2400

acgctttctg gctgggctga tcttccagcc tcccgcccgc tgcctgggag ccctactcgg 2460

gccatcggcg gctgcctcgg tggacaggaa gcagaaggtg cttgcgaggt acctgaagcg 2520

gctgcagccg gggacactgc gggcgcggca gctgctggag ctgctgcact gcgcccacga 2580

ggccgaggag gctggaattt ggcagcacgt ggtacaggag ctccccggcc gcctctcttt 2640

tctgggcacc cgcctcacgc ctcctgatgc acatgtactg ggcaaggcct tggaggcggc 2700

gggccaagac ttctccctgg acctccgcag cactggcatt tgcccctctg gattggggag 2760

cctcgtggga ctcagctgtg tcacccgttt cagggctgcc ttgagcgaca cggtggcgct 2820

gtgggagtcc ctgcagcagc atggggagac caagctactt caggcagcag aggagaagtt 2880

caccatcgag cctttcaaag ccaagtccct gaaggatgtg gaagacctgg gaaagcttgt 2940

gcagactcag aggacgagaa gttcctcgga agacacagct ggggagctcc ctgctgttcg 3000

ggacctaaag aaactggagt ttgcgctggg ccctgtctca ggcccccagg ctttccccaa 3060

actggtgcgg atcctcacgg ccttttcctc cctgcagcat ctggacctgg atgcgctgag 3120

tgagaacaag atcggggacg agggtgtctc gcagctctca gccaccttcc cccagctgaa 3180

gtccttggaa accctcaatc tgtcccagaa caacatcact gacctgggtg cctacaaact 3240

cgccgaggcc ctgccttcgc tcgctgcatc cctgctcagg ctaagcttgt acaataactg 3300

catctgcgac gtgggagccg agagcttggc tcgtgtgctt ccggacatgg tgtccctccg 3360

ggtgatggac gtccagtaca acaagttcac ggctgccggg gcccagcagc tcgctgccag 3420

ccttcggagg tgtcctcatg tggagacgct ggcgatgtgg acgcccacca tcccattcag 3480

tgtccaggaa cacctgcaac aacaggattc acggatcagc ctgagatgat cccagctgtg 3540

ctctggacag gcatgttctc tgaggacact aaccacgctg gaccttgaac tgggtacttg 3600

tggacacagc tcttctccag gctgtatccc atgagcctca gcatcctggc acccggcccc 3660

tgctggttca gggttggccc ctgcccggct gcggaatgaa ccacatcttg ctctgctgac 3720

agacacaggc ccggctccag gctcctttag cgcccagttg ggtggatgcc tggtggcagc 3780

tgcggtccac ccaggagccc cgaggccttc tctgaaggac attgcggaca gccacggcca 3840

ggccagaggg agtgacagag gcagccccat tctgcctgcc caggcccctg ccaccctggg 3900

gagaaagtac ttctttttt ttatttttag acagagtctc actgttgccc aggctggcgt 3960

gcagtggtgc gatctgggtt cactgcaacc tccgcctctt gggttcaagc gattcttctg 4020

cttcagcctc ccgagtagct gggactacag gcacccacca tcatgtctgg ctaatttttc 4080

atttttagta gagacagggt tttgccatgt tggccaggct ggtctcaaac tcttgacctc 4140

aggtgatcca cccacctcag cctcccaaag tgctgggatt acaagcgtga gccactgcac 4200

cgggccacag agaaagtact tctccaccct gctctccgac cagacacctt gacagggcac 4260

accgggcact cagaagacac tgatgggcaa cccccagcct gctaattccc cagattgcaa 4320

caggctgggc ttcagtggca gctgcttttg tctatgggac tcaatgcact gacattgttg 4380

gccaaagcca aagctaggcc tggccagatg caccagccct tagcagggaa acagctaatg 4440

ggacactaat ggggcggtga gaggggaaca gactggaagc acagcttcat ttcctgtgtc 4500

tttttttcact acattataaa tgtctcttta atgtcacagg caggtccagg gtttgagttc 4560

ataccctgtt accattttgg ggtacccact gctctggtta tctaatatgt aacaagccac 4620

cccaaatcat agtggcttaa aacaacactc acattta 4657

<210> 23

<211> 223

<212> PRT

<213> Homo sapiens

<400> 23

Met Ala Cys Leu Gly Phe Gln Arg His Lys Ala Gln Leu Asn Leu Ala
1 5 10 15

Thr Arg Thr Trp Pro Cys Thr Leu Leu Phe Phe Leu Leu Phe Ile Pro
20 25 30

Val Phe Cys Lys Ala Met His Val Ala Gln Pro Ala Val Val Leu Ala
35 40 45

Ser Ser Arg Gly Ile Ala Ser Phe Val Cys Glu Tyr Ala Ser Pro Gly
50 55 60

Lys Ala Thr Glu Val Arg Val Thr Val Leu Arg Gln Ala Asp Ser Gln
65 70 75 80

Val Thr Glu Val Cys Ala Ala Thr Tyr Met Met Gly Asn Glu Leu Thr
85 90 95

Phe Leu Asp Asp Ser Ile Cys Thr Gly Thr Ser Ser Gly Asn Gln Val
100 105 110

Asn Leu Thr Ile Gln Gly Leu Arg Ala Met Asp Thr Gly Leu Tyr Ile
115 120 125

Cys Lys Val Glu Leu Met Tyr Pro Pro Pro Tyr Tyr Leu Gly Ile Gly
130 135 140

Asn Gly Thr Gln Ile Tyr Val Ile Asp Pro Glu Pro Cys Pro Asp Ser
145 150 155 160

Asp Phe Leu Leu Trp Ile Leu Ala Ala Val Ser Ser Gly Leu Phe Phe
165 170 175

Tyr Ser Phe Leu Leu Thr Ala Val Ser Leu Ser Lys Met Leu Lys Lys
180 185 190

Arg Ser Pro Leu Thr Thr Gly Val Tyr Val Lys Met Pro Pro Thr Glu
195 200 205

Pro Glu Cys Glu Lys Gln Phe Gln Pro Tyr Phe Ile Pro Ile Asn
210 215 220

<210> 24

<211> 748

<212> DNA

<213> Homo sapiens

<400> 24

gacctgaaca ccgctcccat aaagccatgg cttgccttgg atttcagcgg cacaaggctc 60

agctgaacct ggctaccagg acctggccct gcactctcct gttttttctt ctcttcatcc 120

ctgtcttctg caaagcaatg cacgtggccc agcctgctgt ggtactggcc agcagccgag 180

gcatcgccag ctttgtgtgt gagtatgcat ctccaggcaa agccactgag gtccgggtga 240

cagtgcttcg gcaggctgac agccaggtga ctgaagtctg tgcggcaacc tacatgatgg 300

ggaatgagtt gaccttccta gatgattcca tctgcacggg cacctccagt ggaaatcaag 360

tgaacctcac tatccaagga ctgagggcca tggacacggg actctacatc tgcaaggtgg 420

agctcatgta cccaccgcca tactacctgg gcataggcaa cggaacccag atttatgtaa 480

ttgatccaga accgtgccca gattctgact tcctcctctg gatccttgca gcagttagtt 540

cggggttgtt tttttatagc tttctcctca cagctgtttc tttgagcaaa atgctaaaga 600

aaagaagccc tcttacaaca ggggtctatg tgaaaatgcc cccaacagag ccagaatgtg 660

aaaagcaatt tcagccttat tttattccca tcaattgaga aaccattatg aagaagagag 720

tccatatttc aatttccaag agctgagg 748

<210> 25

<211> 351

<212> PRT

<213> Homo sapiens

<400> 25

Met Ser Phe Pro Cys Lys Phe Val Ala Ser Phe Leu Leu Ile Phe Asn
1 5 10 15

Val Ser Ser Lys Gly Ala Val Ser Lys Glu Ile Thr Asn Ala Leu Glu

```
            20                  25                  30
Thr Trp Gly Ala Leu Gly Gln Asp Ile Asn Leu Asp Ile Pro Ser Phe
        35                  40                  45
Gln Met Ser Asp Asp Ile Asp Asp Ile Lys Trp Glu Lys Thr Ser Asp

    50                  55                  60
Lys Lys Lys Ile Ala Gln Phe Arg Lys Glu Lys Glu Thr Phe Lys Glu
65                  70                  75                  80
Lys Asp Thr Tyr Lys Leu Phe Lys Asn Gly Thr Leu Lys Ile Lys His
                85                  90                  95
Leu Lys Thr Asp Asp Gln Asp Ile Tyr Lys Val Ser Ile Tyr Asp Thr
            100                 105                 110
Lys Gly Lys Asn Val Leu Glu Lys Ile Phe Asp Leu Lys Ile Gln Glu

        115                 120                 125
Arg Val Ser Lys Pro Lys Ile Ser Trp Thr Cys Ile Asn Thr Thr Leu
    130                 135                 140
Thr Cys Glu Val Met Asn Gly Thr Asp Pro Glu Leu Asn Leu Tyr Gln
145                 150                 155                 160
Asp Gly Lys His Leu Lys Leu Ser Gln Arg Val Ile Thr His Lys Trp
                165                 170                 175
Thr Thr Ser Leu Ser Ala Lys Phe Lys Cys Thr Ala Gly Asn Lys Val

            180                 185                 190
Ser Lys Glu Ser Ser Val Glu Pro Val Ser Cys Pro Glu Lys Gly Leu
        195                 200                 205
Asp Ile Tyr Leu Ile Ile Gly Ile Cys Gly Gly Gly Ser Leu Leu Met
    210                 215                 220
Val Phe Val Ala Leu Leu Val Phe Tyr Ile Thr Lys Arg Lys Lys Gln
225                 230                 235                 240
Arg Ser Arg Arg Asn Asp Glu Glu Leu Glu Thr Arg Ala His Arg Val

                245                 250                 255
Ala Thr Glu Glu Arg Gly Arg Lys Pro His Gln Ile Pro Ala Ser Thr
            260                 265                 270
```

Pro Gln Asn Pro Ala Thr Ser Gln His Pro Pro Pro Pro Pro Gly His

275 280 285

Arg Ser Gln Ala Pro Ser His Arg Pro Pro Pro Pro Gly His Arg Val

290 295 300

Gln His Gln Pro Gln Lys Arg Pro Pro Ala Pro Ser Gly Thr Gln Val

305 310 315 320

His Gln Gln Lys Gly Pro Pro Leu Pro Arg Pro Arg Val Gln Pro Lys

325 330 335

Pro Pro His Gly Ala Ala Glu Asn Ser Leu Ser Pro Ser Ser Asn

340 345 350

<210> 26

<211> 1565

<212> DNA

<213> Homo sapiens

<400> 26

agtctcactt cagttccttt tgcatgaaga gctcagaatc aaaagaggaa accaacccct 60

aagatgagct ttccatgtaa atttgtagcc agcttccttc tgattttcaa tgtttcttcc 120

aaaggtgcag tctccaaaga gattacgaat gccttggaaa cctggggtgc cttgggtcag 180

gacatcaact tggacattcc tagttttcaa atgagtgatg atattgacga tataaaatgg 240

gaaaaaactt cagacaagaa aaagattgca caattcagaa aagagaaaga gactttcaag 300

gaaaaagata catataagct atttaaaaat ggaactctga aaattaagca tctgaagacc 360

gatgatcagg atatctacaa ggtatcaata tatgatacaa aaggaaaaaa tgtgttggaa 420

aaaatatttg atttgaagat tcaagagagg gtctcaaaac caaagatctc ctggacttgt 480

atcaacacaa ccctgacctg tgaggtaatg aatggaactg accccgaatt aaacctgtat 540

caagatggga aacatctaaa actttctcag agggtcatca cacacaagtg gaccaccagc 600

ctgagtgcaa aattcaagtg cacagcaggg aacaaagtca gcaaggaatc cagtgtcgag 660

cctgtcagct gtccagagaa aggtctggac atctatctca tcattggcat atgtggagga 720

ggcagcctct tgatggtctt tgtggcactg ctcgttttct atatcaccaa aaggaaaaaa 780

cagaggagtc ggagaaatga tgaggagctg gagacaagag cccacagagt agctactgaa 840

gaaaggggcc ggaagcccca ccaaattcca gcttcaaccc ctcagaatcc agcaacttcc 900

caacatcctc ctccaccacc tggtcatcgt tcccaggcac ctagtcatcg tccccccgcct 960

cctggacacc gtgttcagca ccagcctcag aagaggcctc ctgctccgtc gggcacacaa 1020

gttcaccagc agaaaggccc gcccctcccc agacctcgag ttcagccaaa acctccccat 1080

ggggcagcag aaaactcatt gtccccttcc tctaattaaa aaagatagaa actgtctttt 1140

tcaataaaaa gcactgtgga tttctgccct cctgatgtgc atatccgtac ttccatgagg 1200

tgttttctgt gtgcagaaca ttgtcacctc ctgaggctgt gggccacagc cacctctgca 1260

tcttcgaact cagccatgtg gtcaacatct ggagtttttg gtctcctcag agagctccat 1320

cacaccagta aggagaagca atataagtgt gattgcaaga atggtagagg accgagcaca 1380

gaaatcttag agatttcttg tcccctctca ggtcatgtgt agatgcgata aatcaagtga 1440

ttggtgtgcc tgggtctcac tacaagcagc ctatctgctt aagagactct ggagtttctt 1500

atgtgccctg gtggacactt gcccaccatc ctgtgagtaa aagtgaaata aaagctttga 1560

ctaga 1565

<210> 27

<211> 438

<212> PRT

<213> Homo sapiens

<400> 27

Met Val Cys Ser Gln Ser Trp Gly Arg Ser Ser Lys Gln Trp Glu Asp

1 5 10 15

Pro Ser Gln Ala Ser Lys Val Cys Gln Arg Leu Asn Cys Gly Val Pro

20 25 30

Leu Ser Leu Gly Pro Phe Leu Val Thr Tyr Thr Pro Gln Ser Ser Ile

35 40 45

Ile Cys Tyr Gly Gln Leu Gly Ser Phe Ser Asn Cys Ser His Ser Arg

50 55 60

Asn Asp Met Cys His Ser Leu Gly Leu Thr Cys Leu Glu Pro Gln Lys

65 70 75 80

Thr Thr Pro Pro Thr Thr Arg Pro Pro Pro Thr Thr Thr Pro Glu Pro

85 90 95

Thr Ala Pro Pro Arg Leu Gln Leu Val Ala Gln Ser Gly Gly Gln His

100 105 110

Cys Ala Gly Val Val Glu Phe Tyr Ser Gly Ser Leu Gly Gly Thr Ile

115 120 125

Ser Tyr Glu Ala Gln Asp Lys Thr Gln Asp Leu Glu Asn Phe Leu Cys

130 135 140

Asn Asn Leu Gln Cys Gly Ser Phe Leu Lys His Leu Pro Glu Thr Glu

145 150 155 160

Ala Gly Arg Ala Gln Asp Pro Gly Glu Pro Arg Glu His Gln Pro Leu

165 170 175

Pro Ile Gln Trp Lys Ile Gln Asn Ser Ser Cys Thr Ser Leu Glu His

180 185 190

Cys Phe Arg Lys Ile Lys Pro Gln Lys Ser Gly Arg Val Leu Ala Leu

195 200 205

Leu Cys Ser Gly Phe Gln Pro Lys Val Gln Ser Arg Leu Val Gly Gly

210 215 220

Ser Ser Ile Cys Glu Gly Thr Val Glu Val Arg Gln Gly Ala Gln Trp

225 230 235 240

Ala Ala Leu Cys Asp Ser Ser Ser Ala Arg Ser Ser Leu Arg Trp Glu

245 250 255

Glu Val Cys Arg Glu Gln Gln Cys Gly Ser Val Asn Ser Tyr Arg Val

260 265 270

Leu Asp Ala Gly Asp Pro Thr Ser Arg Gly Leu Phe Cys Pro His Gln

275 280 285

Lys Leu Ser Gln Cys His Glu Leu Trp Glu Arg Asn Ser Tyr Cys Lys

290 295 300

Lys Val Phe Val Thr Cys Gln Asp Pro Asn Pro Ala Gly Leu Ala Ala

305 310 315 320

Gly Thr Val Ala Ser Ile Ile Leu Ala Leu Val Leu Leu Val Val Leu

325 330 335

Leu Val Val Cys Gly Pro Leu Ala Tyr Lys Lys Leu Val Lys Lys Phe

340 345 350

Arg Gln Lys Lys Gln Arg Gln Trp Ile Gly Pro Thr Gly Met Asn Gln

355 360 365

Asn Met Ser Phe His Arg Asn His Thr Ala Thr Val Arg Ser His Ala

370 375 380

Glu Asn Pro Thr Ala Ser His Val Asp Asn Glu Tyr Ser Gln Pro Pro

385 390 395 400

Arg Asn Ser His Leu Ser Ala Tyr Pro Ala Leu Glu Gly Ala Leu His

405 410 415

Arg Ser Ser Met Gln Pro Asp Asn Ser Ser Asp Ser Asp Tyr Asp Leu

420 425 430

His Gly Ala Gln Arg Leu

435

<210> 28

<211> 3376

<212> DNA

<213> Homo sapiens

<400> 28

gagtcttgct gatgctcccg gctgaataaa ccccttcctt ctttaacttg gtgtctgagg 60

ggttttgtct gtggcttgtc ctgctacatt tcttggttcc ctgaccagga agcaaagtga 120

ttaacggaca gttgaggcag ccccttaggc agcttaggcc tgccttgtgg agcatccccg 180

cggggaactc tggccagctt gagcgacacg gatcctcaga gcgctcccag gtaggcaatt 240

gccccagtgg aatgcctcgt cagagcagtg catggcaggc ccctgtggag gatcaacgca 300

gtggctgaac acagggaagg aactggcact tggagtccgg acaactgaaa cttgtcgctt 360

cctgcctcgg acggctcagc tggtatgacc cagatttcca ggcaaggctc acccgttcca 420

actcgaagtg ccagggccag ctggaggtct acctcaagga cggatggcac atggtttgca 480

gccagagctg gggccggagc tccaagcagt gggaggaccc cagtcaagcg tcaaaagtct 540

gccagcggct gaactgtggg gtgcccttaa gccttggccc cttccttgtc acctacacac 600

ctcagagctc aatcatctgc tacggacaac tgggctcctt ctccaactgc agccacagca 660

gaaatgacat gtgtcactct ctgggcctga cctgcttaga accccagaag acaacacctc 720

caacgacaag gccccgccc accacaactc cagagcccac agctcctccc aggctgcagc 780

tggtggcaca gtctggcggc cagcactgtg ccggcgtggt ggagttctac agcggcagcc 840

tgggggtac catcagctat gaggcccagg acaagaccca ggacctggag aacttcctct 900

gcaacaacct ccagtgtggc tccttcttga agcatctgcc agagactgag gcaggcagag 960

cccaagaccc aggggagcca cgggaacacc agcccttgcc aatccaatgg aagatccaga 1020

actcaagctg tacctccctg gagcattgct tcaggaaaat caagccccag aaaagtggcc 1080

gagttcttgc cctcctttgc tcaggtttcc agcccaaggt gcagagccgt ctggtggggg 1140

gcagcagcat ctgtgaaggc accgtggagg tgcgccaggg ggctcagtgg gcagccctgt 1200

gtgacagctc ttcagccagg agctcgctgc ggtgggagga ggtgtgccgg gagcagcagt 1260

gtggcagcgt caactcctat cgagtgctgg acgctggtga cccaacatcc cgggggctct 1320

tctgtcccca tcagaagctg tcccagtgcc acgaactttg ggagagaaat tcctactgca 1380

agaaggtgtt tgtcacatgc caggatccaa accccgcagg cctggccgca ggcacggtgg 1440

caagcatcat cctggccctg gtgctcctgg tggtgctgct ggtcgtgtgc ggcccccttg 1500

cctacaagaa gctagtgaag aaattccgcc agaagaagca gcgccagtgg attggcccaa 1560

cgggaatgaa ccaaaaacatg tctttccatc gcaaccacac ggcaaccgtc cgatcccatg 1620

ctgagaaccc cacagcctcc cacgtggata acgaatacag ccaacctccc aggaactccc 1680

acctgtcagc ttatccagct ctggaagggg ctctgcatcg ctcctccatg cagcctgaca 1740

actcctccga cagtgactat gatctgcatg gggctcagag gctgtaaaga actgggatcc 1800

atgagcaaaa agccgagagc cagacctgtt tgtcctgaga aaactgtccg ctcttcactt 1860

gaaatcatgt ccctatttct accccggcca gaacatggac agaggccaga agccttccgg 1920

acaggcgctg ctgccccgag tggcaggcca gctcacactc tgctgcacaa cagctcggcc 1980

gcccctccac ttgtggaagc tgtggtgggc agagccccaa aacaagcagc cttccaacta 2040

gagactcggg ggtgtctgaa gggggccccc tttccctgcc cgctggggag cggcgtctca 2100

gtgaaatcgg ctttctcctc agactctgtc cctggtaagg agtgacaagg aagctcacag 2160

ctgggcgagt gcattttgaa tagtttttg taagtagtgc ttttcctcct tcctgacaaa 2220

tcgagcgctt tggcctcttc tgtgcagcat ccacccctgc ggatccctct ggggaggaca 2280

ggaaggggac tcccggagac ctctgcagcc gtggtggtca gaggctgctc acctgagcac 2340

aaagacagct ctgcacattc accgcagctg ccagccaggg gtctgggtgg gcaccaccct 2400

gacccacagc gtcaccccac tccctctgtc ttatgactcc cctccccaac cccctcatct 2460

aaagacacct tcctttccac tggctgtcaa gcccacaggg caccagtgcc acccagggcc 2520

cggcacaaag gggcgcctag taaaccttaa ccaacttggt tttttgcttc acccagcaat 2580

taaaagtccc aagctgaggt agtttcagtc catcacagtt catcttctaa cccaagagtc 2640

agagatgggg ctggtcatgt tcctttggtt tgaataactc ccttgacgaa aacagactcc 2700

tctagtactt ggagatcttg gacgtacacc taatcccatg gggcctcggc ttccttaact 2760

gcaagtgaga agaggaggtc tacccaggag cctcgggtct gatcaaggga gaggccaggc 2820

gcagctcact gcggcggctc cctaagaagg tgaagcaaca tgggaacaca tcctaagaca 2880

ggtcctttct ccacgccatt tgatgctgta tctcctggga gcacaggcat caatggtcca 2940

agccgcataa taagtctgga agagcaaaag ggagttacta ggatatgggg tgggctgctc 3000

ccagaatctg ctcagctttc tgcccccacc aacaccctcc aaccaggcct tgccttctga 3060

gagcccccgt ggccaagccc aggtcacaga tcttcccccg accatgctgg gaatccagaa 3120

acagggaccc catttgtctt cccatatctg gtggaggtga gggggctcct caaaagggaa 3180

ctgagaggct gctcttaggg agggcaaagg ttcgggggca gccagtgtct cccatcagtg 3240

ccttttttaa taaaagctct ttcatctata gtttggccac catacagtgg cctcaaagca 3300

accatggcct acttaaaaac caaaccaaaa ataaagagtt tagttgagga gaaaaaaaaa 3360

aaaaaaaaaa aaaaaa 3376

<210> 29

<211> 240

<212> PRT

<213> Homo sapiens

<400> 29

Met Ala Gly Pro Pro Arg Leu Leu Leu Leu Pro Leu Leu Leu Ala Leu

1 5 10 15

Ala Arg Gly Leu Pro Gly Ala Leu Ala Ala Gln Glu Val Gln Gln Ser

20 25 30

Pro His Cys Thr Thr Val Pro Val Gly Ala Ser Val Asn Ile Thr Cys

35 40 45

Ser Thr Ser Gly Gly Leu Arg Gly Ile Tyr Leu Arg Gln Leu Gly Pro

50 55 60

Gln Pro Gln Asp Ile Ile Tyr Tyr Glu Asp Gly Val Val Pro Thr Thr

65 70 75 80

Asp Arg Arg Phe Arg Gly Arg Ile Asp Phe Ser Gly Ser Gln Asp Asn

85 90 95

Leu Thr Ile Thr Met His Arg Leu Gln Leu Ser Asp Thr Gly Thr Tyr

100 105 110

Thr Cys Gln Ala Ile Thr Glu Val Asn Val Tyr Gly Ser Gly Thr Leu

115 120 125

Val Leu Val Thr Glu Glu Gln Ser Gln Gly Trp His Arg Cys Ser Asp

130 135 140

Ala Pro Pro Arg Ala Ser Ala Leu Pro Ala Pro Pro Thr Gly Ser Ala

145 150 155 160

Leu Pro Asp Pro Gln Thr Ala Ser Ala Leu Pro Asp Pro Pro Ala Ala

165 170 175

Ser Ala Leu Pro Ala Ala Leu Ala Val Ile Ser Phe Leu Leu Gly Leu

180 185 190

Gly Leu Gly Val Ala Cys Val Leu Ala Arg Thr Gln Ile Lys Lys Leu

195 200 205

Cys Ser Trp Arg Asp Lys Asn Ser Ala Ala Cys Val Val Tyr Glu Asp

210 215 220

Met Ser His Ser Arg Cys Asn Thr Leu Ser Ser Pro Asn Gln Tyr Gln

225 230 235 240

<210> 30

<211> 1284

<212> DNA

<213> Homo sapiens

<400> 30

ctctctgagc tctgagcgcc tgcggtctcc tgtgtgctgc tctctgtggg gtcctgtaga 60

cccagagagg ctcagctgca ctcgcccggc tgggagagct gggtgtgggg aacatggccg 120

ggcctccgag gctcctgctg ctgcccctgc ttctggcgct ggctcgcggc ctgcctgggg 180

ccctggctgc ccaagaggtg cagcagtctc cccactgcac gactgtcccc gtgggagcct 240

ccgtcaacat cacctgctcc accagcgggg gcctgcgtgg gatctacctg aggcagctcg 300

ggccacagcc ccaagacatc atttactacg aggacggggt ggtgcccact acggacagac 360

ggttccgggg ccgcatcgac ttctcagggt cccaggacaa cctgactatc accatgcacc 420

gcctgcagct gtcggacact ggcacctaca cctgccaggc catcacggag gtcaatgtct 480

acggctccgg caccctggtc ctggtgacag aggaacagtc ccaaggatgg cacagatgct 540

cggacgcccc accaagggcc tctgccctcc ctgccccacc gacaggctcc gccctccctg 600

acccgcagac agcctctgcc ctccctgacc cgccagcagc ctctgccctc cctgcggccc 660

tggcggtgat ctccttcctc ctcgggctgg gcctgggggt ggcgtgtgtg ctggcgagga 720

cacagataaa gaaactgtgc tcgtggcggg ataagaattc ggcggcatgt gtggtgtacg 780

aggacatgtc gcacagccgc tgcaacacgc tgtcctcccc caaccagtac cagtgaccca 840

gtgggcccct gcacgtcccg cctgtggtcc ccccagcacc ttccctgccc caccatgccc 900

cccaccctgc cacacccctc accctgctgt cctcccacgg ctgcagcaga gtttgaaggg 960

cccagccgtg cccagctcca agcagacaca caggcagtgg ccaggcccca cggtgcttct 1020

cagtggacaa tgatgcctcc tccgggaagc cttccctgcc cagcccacgc cgccaccggg 1080

aggaagcctg actgtccttt ggctgcatct cccgaccatg gccaaggagg gcttttctgt 1140

gggatgggcc tgggcacgcg gccctctcct gtcagtgccg gcccacccac cagcaggccc 1200

ccaaccccca ggcagcccgg cagaggacgg gaggagacca gtcccccacc cagccgtacc 1260

agaaataaag gcttctgtgc ttcc 1284

<210> 31

<211> 364

<212> PRT

<213> Homo sapiens

<400> 31

Met Pro Leu Leu Leu Leu Leu Pro Leu Leu Trp Ala Gly Ala Leu Ala
1 5 10 15

Met Asp Pro Asn Phe Trp Leu Gln Val Gln Glu Ser Val Thr Val Gln
20 25 30

Glu Gly Leu Cys Val Leu Val Pro Cys Thr Phe Phe His Pro Ile Pro
35 40 45

Tyr Tyr Asp Lys Asn Ser Pro Val His Gly Tyr Trp Phe Arg Glu Gly
50 55 60

Ala Ile Ile Ser Arg Asp Ser Pro Val Ala Thr Asn Lys Leu Asp Gln
65 70 75 80

Glu Val Gln Glu Glu Thr Gln Gly Arg Phe Arg Leu Leu Gly Asp Pro
85 90 95

Ser Arg Asn Asn Cys Ser Leu Ser Ile Val Asp Ala Arg Arg Arg Asp
100 105 110

Asn Gly Ser Tyr Phe Phe Arg Met Glu Arg Gly Ser Thr Lys Tyr Ser
115 120 125

Tyr Lys Ser Pro Gln Leu Ser Val His Val Thr Asp Leu Thr His Arg
130 135 140

Pro Lys Ile Leu Ile Pro Gly Thr Leu Glu Pro Gly His Ser Lys Asn
145 150 155 160
Leu Thr Cys Ser Val Ser Trp Ala Cys Glu Gln Gly Thr Pro Pro Ile
165 170 175
Phe Ser Trp Leu Ser Ala Ala Pro Thr Ser Leu Gly Pro Arg Thr Thr
180 185 190

His Ser Ser Val Leu Ile Ile Thr Pro Arg Pro Gln Asp His Gly Thr
195 200 205
Asn Leu Thr Cys Gln Val Lys Phe Ala Gly Ala Gly Val Thr Thr Glu
210 215 220
Arg Thr Ile Gln Leu Asn Val Thr Tyr Val Pro Gln Asn Pro Thr Thr
225 230 235 240
Gly Ile Phe Pro Gly Asp Gly Ser Gly Lys Gln Glu Thr Arg Ala Gly
245 250 255

Val Val His Gly Ala Ile Gly Gly Ala Gly Val Thr Ala Leu Leu Ala
260 265 270
Leu Cys Leu Cys Leu Ile Phe Phe Ile Val Lys Thr His Arg Arg Lys
275 280 285
Ala Ala Arg Thr Ala Val Gly Arg Asn Asp Thr His Pro Thr Thr Gly
290 295 300
Ser Ala Ser Pro Lys His Gln Lys Lys Ser Lys Leu His Gly Pro Thr
305 310 315 320

Glu Thr Ser Ser Cys Ser Gly Ala Ala Pro Thr Val Glu Met Asp Glu
325 330 335
Glu Leu His Tyr Ala Ser Leu Asn Phe His Gly Met Asn Pro Ser Lys
340 345 350
Asp Thr Ser Thr Glu Tyr Ser Glu Val Arg Thr Gln
355 360

<210> 32

<211> 1462

<212> DNA

<213> Homo sapiens

<400> 32

ctgctcacac aggaagccct ggaagctgct tcctcagaca tgccgctgct gctactgctg 60

cccctgctgt gggcaggggc cctggctatg gatccaaatt tctggctgca agtgcaggag 120

tcagtgacgg tacaggaggg tttgtgcgtc ctcgtgccct gcactttctt ccatcccata 180

ccctactacg acaagaactc cccagttcat ggttactggt tccgggaagg agccattata 240

tccagggact ctccagtggc cacaaacaag ctagatcaag aagtacagga ggagactcag 300

ggcagattcc gcctccttgg ggatcccagt aggaacaact gctccctgag catcgtagac 360

gccaggagga gggataatgg ttcatacttc tttcggatgg agagaggaag taccaaatac 420

agttacaaat ctccccagct ctctgtgcat gtgacagact tgacccacag gcccaaaatc 480

ctcatccctg gcactctaga acccggccac tccaaaaacc tgacctgctc tgtgtcctgg 540

gcctgtgagc agggaacacc cccgatcttc tcctggttgt cagctgcccc cacctccctg 600

ggccccagga ctactcactc ctcggtgctc ataatcaccc cacggcccca ggaccacggc 660

accaacctga cctgtcaggt gaagttcgct ggagctggtg tgactacgga gagaaccatc 720

cagctcaacg tcacctatgt tccacagaac ccaacaactg gtatctttcc aggagatggc 780

tcagggaaac aagagaccag agcaggagtg gttcatgggg ccattggagg agctggtgtt 840

acagccctgc tcgctctttg tctctgcctc atcttcttca tagtgaagac ccacaggagg 900

aaagcagcca ggacagcagt gggcaggaat gacacccacc ctaccacagg gtcagcctcc 960

ccgaaacacc agaagaagtc caagttacat ggccccactg aaacctcaag ctgttcaggt 1020

gccgccccta ctgtggagat ggatgaggag ctgcattatg cttccctcaa ctttcatggg 1080

atgaatcctt ccaaggacac ctccaccgaa tactcagagg tcaggaccca gtgaggaacc 1140

cacaagagca tcaggctcag ctagaagatc cacatcctct acaggtcggg gaccaaaggc 1200

tgattcttgg agatttaaca ccccacaggc aatgggttta tagacattat gtgagtttcc 1260

tgctatatta acatcatctt agactttgca agcagagagt cgtggaatca aatctgtgct 1320

ctttcatttg ctaagtgtat gatgtcacac aagctcctta accttccatg tctccatttt 1380

cttctctgtg aagtaggtat aagaagtcct atctcatagg gatgctgtga gcattaaata 1440

aaggtacaca tggaaaacac ca 1462

<210> 33

<211> 61

<212> PRT

<213> Homo sapiens

<400> 33

Met Lys Arg Phe Leu Phe Leu Leu Leu Thr Ile Ser Leu Leu Val Met

1 5 10 15

Val Gln Ile Gln Thr Gly Leu Ser Gly Gln Asn Asp Thr Ser Gln Thr

20 25 30

Ser Ser Pro Ser Ala Ser Ser Asn Ile Ser Gly Gly Ile Phe Leu Phe

35 40 45

Phe Val Ala Asn Ala Ile Ile His Leu Phe Cys Phe Ser

50 55 60

<210> 34

<211> 467

<212> DNA

<213> Homo sapiens

<400> 34

agacagccct gagatcacct aaaaagctgc taccaagaca gccacgaaga tcctaccaaa 60

atgaagcgct tcctcttcct cctactcacc atcagcctcc tggttatggt acagatacaa 120

actggactct caggacaaaa cgacaccagc caaaccagca gcccctcagc atccagcaac 180

ataagcggag gcattttcct tttcttcgtg gccaatgcca taatccacct cttctgcttc 240

agttgaggtg acacgtctca gccttagccc tgtgcccccc gaaacagctg ccaccatcac 300

tcgcaagaga atcccctcca tctttgggag gggttgatgc cagacatcac caggttgtag 360

aagttgacag gcagtgccat gggggcaaca gccaaaatag gggggtaatg atgtaggggc 420

caagcagtgc ccagctgggg gtcaataaag ttacccttgt acttgca 467

<210> 35

<211> 378

<212> PRT

<213> Homo sapiens

<400> 35

Met Val Leu Leu Trp Leu Thr Leu Leu Leu Ile Ala Leu Pro Cys Leu

1 5 10 15

Leu Gln Thr Lys Glu Asp Pro Asn Pro Pro Ile Thr Asn Leu Arg Met

20 25 30

Lys Ala Lys Ala Gln Gln Leu Thr Trp Asp Leu Asn Arg Asn Val Thr

35 40 45

```
Asp Ile Glu Cys Val Lys Asp Ala Asp Tyr Ser Met Pro Ala Val Asn
    50                  55                  60
Asn Ser Tyr Cys Gln Phe Gly Ala Ile Ser Leu Cys Glu Val Thr Asn
65                  70                  75                  80
Tyr Thr Val Arg Val Ala Asn Pro Pro Phe Ser Thr Trp Ile Leu Phe

                85                  90                  95
Pro Glu Asn Ser Gly Lys Pro Trp Ala Gly Ala Glu Asn Leu Thr Cys
            100                 105                 110
Trp Ile His Asp Val Asp Phe Leu Ser Cys Ser Trp Ala Val Gly Pro
        115                 120                 125
Gly Ala Pro Ala Asp Val Gln Tyr Asp Leu Tyr Leu Asn Val Ala Asn
    130                 135                 140
Arg Arg Gln Gln Tyr Glu Cys Leu His Tyr Lys Thr Asp Ala Gln Gly

145                 150                 155                 160
Thr Arg Ile Gly Cys Arg Phe Asp Asp Ile Ser Arg Leu Ser Ser Gly
                165                 170                 175
Ser Gln Ser Ser His Ile Leu Val Arg Gly Arg Ser Ala Ala Phe Gly
            180                 185                 190
Ile Pro Cys Thr Asp Lys Phe Val Val Phe Ser Gln Ile Glu Ile Leu
        195                 200                 205
Thr Pro Pro Asn Met Thr Ala Lys Cys Asn Lys Thr His Ser Phe Met

    210                 215                 220
His Trp Lys Met Arg Ser His Phe Asn Arg Lys Phe Arg Tyr Glu Leu
225                 230                 235                 240
Gln Ile Gln Lys Arg Met Gln Pro Val Ile Thr Glu Gln Val Arg Asp
                245                 250                 255
Arg Thr Ser Phe Gln Leu Leu Asn Pro Gly Thr Tyr Thr Val Gln Ile
            260                 265                 270
Arg Ala Arg Glu Arg Val Tyr Glu Phe Leu Ser Ala Trp Ser Thr Pro

        275                 280                 285
Gln Arg Phe Glu Cys Asp Gln Glu Glu Gly Ala Asn Thr Arg Ala Trp
```

290 295 300

Arg Thr Ser Leu Leu Ile Ala Leu Gly Thr Leu Leu Ala Leu Val Cys

305 310 315 320

Val Phe Val Ile Cys Arg Arg Tyr Leu Val Met Gln Arg Leu Phe Pro

325 330 335

Arg Ile Pro His Met Lys Asp Pro Ile Gly Asp Ser Phe Gln Asn Asp

340 345 350

Lys Leu Val Val Trp Glu Ala Gly Lys Ala Gly Leu Glu Glu Cys Leu

355 360 365

Val Thr Glu Val Gln Val Val Gln Lys Thr

370 375

<210> 36

<211> 1541

<212> DNA

<213> Homo sapiens

<400> 36

cttcggtttc tcttcgggga aagctgcttt cagcgcacac gggaagatat cagaaacatc 60

ctaggatcag gacaccccag atcttctcaa ctggaaccac gaaggctgtt tcttccacac 120

agtactttga tctccattta agcaggcacc tctgtcctgc gttccggagc tgcgttcccg 180

atggtcctcc tttggctcac gctgctcctg atcgccctgc cctgtctcct gcaaacgaag 240

gaagatccaa acccaccaat cacgaaccta aggatgaaag caaaggctca gcagttgacc 300

tgggacctta acagaaatgt gaccgatatc gagtgtgtta aagacgccga ctattctatg 360

ccggcagtga acaatagcta ttgccagttt ggagcaattt ccttatgtga agtgaccaac 420

tacaccgtcc gagtggccaa cccaccattc tccacgtgga tcctcttccc tgagaacagt 480

gggaagcctt gggcaggtgc ggagaatctg acctgctgga ttcatgacgt ggatttcttg 540

agctgcagct gggcggtagg cccgggggcc cccgcggacg tccagtacga cctgtacttg 600

aacgttgcca acaggcgtca acagtacgag tgtcttcact acaaaacgga tgctcaggga 660

acacgtatcg ggtgtcgttt cgatgacatc tctcgactct ccagcggttc tcaaagttcc 720

cacatcctgg tgcggggcag gagcgcagcc ttcggtatcc cctgcacaga taagtttgtc 780

gtcttttcac agattgagat attaactcca cccaacatga ctgcaaagtg taataagaca 840

cattccttta tgcactggaa aatgagaagt catttcaatc gcaaatttcg ctatgagctt 900

cagatacaaa agagaatgca gcctgtaatc acagaacagg tcagagacag aacctccttc 960

cagctactca atcctggaac gtacacagta caaataagag cccgggaaag agtgtatgaa 1020

ttcttgagcg cctggagcac cccccagcgc ttcgagtgcg accaggagga gggcgcaaac 1080

acacgtgcct ggcggacgtc gctgctgatc gcgctgggga cgctgctggc cctggtctgt 1140

gtcttcgtga tctgcagaag gtatctggtg atgcagagac tctttccccg catccctcac 1200

atgaaagacc ccatcggtga cagcttccaa aacgacaagc tggtggtctg ggaggcgggc 1260

aaagccggcc tggaggagtg tctggtgact gaagtacagg tcgtgcagaa aacttgagac 1320

tggggttcag ggcttgtggg ggtctgcctc aatctccctg gccgggccag gcgcctgcac 1380

agactggctg ctggacctgc gcacgcagcc caggaatgga cattcctaac gggtggtggg 1440

catgggagat gcctgtgtaa tttcgtccga agctgccagg aagaagaaca gaactttgtg 1500

tgtttatttc atgataaagt gatttttttt ttttaaccc a 1541

<210> 37

<211> 255

<212> PRT

<213> Homo sapiens

<400> 37

Met Gly Asn Ser Cys Tyr Asn Ile Val Ala Thr Leu Leu Leu Val Leu
1 5 10 15

Asn Phe Glu Arg Thr Arg Ser Leu Gln Asp Pro Cys Ser Asn Cys Pro
20 25 30

Ala Gly Thr Phe Cys Asp Asn Asn Arg Asn Gln Ile Cys Ser Pro Cys
35 40 45

Pro Pro Asn Ser Phe Ser Ser Ala Gly Gly Gln Arg Thr Cys Asp Ile
50 55 60

Cys Arg Gln Cys Lys Gly Val Phe Arg Thr Arg Lys Glu Cys Ser Ser
65 70 75 80

Thr Ser Asn Ala Glu Cys Asp Cys Thr Pro Gly Phe His Cys Leu Gly
85 90 95

Ala Gly Cys Ser Met Cys Glu Gln Asp Cys Lys Gln Gly Gln Glu Leu
100 105 110

Thr Lys Lys Gly Cys Lys Asp Cys Cys Phe Gly Thr Phe Asn Asp Gln
115 120 125

Lys Arg Gly Ile Cys Arg Pro Trp Thr Asn Cys Ser Leu Asp Gly Lys
130 135 140
Ser Val Leu Val Asn Gly Thr Lys Glu Arg Asp Val Val Cys Gly Pro
145 150 155 160
Ser Pro Ala Asp Leu Ser Pro Gly Ala Ser Ser Val Thr Pro Pro Ala

165 170 175
Pro Ala Arg Glu Pro Gly His Ser Pro Gln Ile Ile Ser Phe Phe Leu
180 185 190
Ala Leu Thr Ser Thr Ala Leu Leu Phe Leu Leu Phe Phe Leu Thr Leu
195 200 205
Arg Phe Ser Val Val Lys Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe
210 215 220
Lys Gln Pro Phe Met Arg Pro Val Gln Thr Thr Gln Glu Glu Asp Gly

225 230 235 240
Cys Ser Cys Arg Phe Pro Glu Glu Glu Glu Gly Gly Cys Glu Leu
245 250 255

<210> 38

<211> 5872

<212> DNA

<213> Homo sapiens

<400> 38

gcagaagcct gaagaccaag gagtggaaag ttctccggca gccctgagat ctcaagagtg 60

acatttgtga gaccagctaa tttgattaaa attctcttgg aatcagcttt gctagtatca 120

tacctgtgcc agatttcatc atgggaaaca gctgttacaa catagtagcc actctgttgc 180

tggtcctcaa ctttgagagg acaagatcat tgcaggatcc ttgtagtaac tgcccagctg 240

gtacattctg tgataataac aggaatcaga tttgcagtcc ctgtcctcca aatagtttct 300

ccagcgcagg tggacaaagg acctgtgaca tatgcaggca gtgtaaaggt gttttcagga 360

ccaggaagga gtgttcctcc accagcaatg cagagtgtga ctgcactcca gggtttcact 420

gcctgggggc aggatgcagc atgtgtgaac aggattgtaa acaaggtcaa gaactgacaa 480

aaaaaggttg taaagactgt tgctttggga catttaacga tcagaaacgt ggcatctgtc 540

gaccctggac aaactgttct ttggatggaa agtctgtgct tgtgaatggg acgaaggaga 600

gggacgtggt ctgtggacca tctccagccg acctctctcc gggagcatcc tctgtgaccc 660

cgcctgcccc tgcgagagag ccaggacact ctccgcagat catctccttc tttcttgcgc 720
tgacgtcgac tgcgttgctc ttcctgctgt tcttcctcac gctccgtttc tctgttgtta 780
aacggggcag aaagaaactc ctgtatatat tcaaacaacc atttatgaga ccagtacaaa 840
ctactcaaga ggaagatggc tgtagctgcc gatttccaga agaagaagaa ggaggatgtg 900
aactgtgaaa tggaagtcaa tagggctgtt gggactttct tgaaaagaag caaggaaata 960
tgagtcatcc gctatcacag ctttcaaaag caagaacacc atcctacata atacccagga 1020
ttcccccaac acacgttctt ttctaaatgc caatgagttg gcctttaaaa atgcaccact 1080

ttttttttt ttttgacagg gtctcactct gtcacccagg ctggagtgca gtggcaccac 1140
catggctctc tgcagccttg acctctggga gctcaagtga tcctcctgcc tcagtctcct 1200
gagtagctgg aactacaagg aagggccacc acacctgact aacttttttg ttttttgttt 1260
ggtaaagatg gcatttcacc atgttgtaca ggctggtctc aaactcctag gttcactttg 1320
gcctcccaaa gtgctgggat tacagacatg aactgccagg cccggccaaa ataatgcacc 1380
acttttaaca gaacagacag atgaggacag agctggtgat aaaaaaaaaa aaaaaaaagc 1440
attttctaga taccacttaa caggtttgag ctagtttttt tgaaatccaa agaaaattat 1500

agtttaaatt caattacata gtccagtggt ccaactataa ttataatcaa aatcaatgca 1560
ggtttgtttt ttggtgctaa tatgacatat gacaataagc cacgaggtgc agtaagtacc 1620
cgactaaagt ttccgtgggt tctgtcatgt aacacgacat gctccaccgt caggggggag 1680
tatgagcaga gtgcctgagt ttagggtcaa ggacaaaaaa cctcaggcct ggaggaagtt 1740
ttggaaagag ttcaagtgtc tgtatatcct atggtcttct ccatcctcac accttctgcc 1800
tttgtcctgc tccctttaa gccaggttac attctaaaaa ttcttaactt ttaacataat 1860
attttatacc aaagccaata aatgaactgc atatgatagg tatgaagtac agtgagaaaa 1920

ttaacacctg tgagctcatt gtcctaccac agcactagag tgggggccgc caaactccca 1980
tggccaaacc tggtgcacca tttgcctttg tttgtctgtt ggtttgcttg agacagtctt 2040
gctctgttgc ccaggctgga atggagtggc tattcacagg cacaatcata gcacacttta 2100
gccttaaact cctgggctca agtgatccac ccgcctcagt ctcccaagta gctgggatta 2160
caggtgcaaa cctggcatgc ctgccattgt ttggcttatg atctaaggat agctttttaa 2220
attttattca ttttattttt ttttgagaca gtgtctcact ctgtctccca ggctggagta 2280
cagtggtaca atcttggatc accgcctccc agtttcaagt gatctccctg cctcagcctc 2340

ctaagtagct gggactacag gtatgtgcca ccacgcctgg ctaattttta tatttttagt 2400

agagacgggg tttcaccatg ttgtccaggc tggtctcaaa ctcctgacct caggtgatct 2460

gcccacctct gcctcccaaa gtgctgggat tacaggcatg agccaccatg cctggccatt 2520

tcttacactt ttgtatgaca tgcctattgc aagcttgcgt gcctctgtcc catgttattt 2580

tactctggga tttaggtgga gggagcagct tctatttgga acattggcca tcgcatggca 2640

aatgggtatc tgtcacttct gctcctattt agttggttct actataacct ttagagcaaa 2700

tcctgcagcc aagccaggca tcaatagggc agaaaagtat attctgtaaa taggggtgag 2760

gagaagatat ttctgaacaa tagtctactg cagtaccaaa ttgcttttca aagtggctgt 2820

tctaatgtac tcccgtcagt catataagtg tcatgtaagt atcccattga tccacatcct 2880

tgctaccctc tggtactatc aggtgccctt aattttgcca agccagtggg tatagaatga 2940

gatctcactg tggtcttagt ttgcatttgc ttggttactg atgagcacct tgtcaaatat 3000

ttatatacca tttgtgttta ttttttaaa taaaatgctt gctcatgctt ttttgcccat 3060

ttgcaaaaaa acttggggcc gggtgcagtg gctcatgcct gtagtcccag ctctttggga 3120

ggccaaggtg ggcagatcgc ttgagcccag gagttcgaga ccagccttgg caacatggcg 3180

aaaccctgtc tttacaaaaa atacaaaaat tagccgggtg tggtggtgtg cacctgaagt 3240

cccagctact cagtaggttc gctttgagcc tgggaggcag aggttgcagt gagctgggac 3300

cgcatcacta cacttcagcc tgggcaacag agaaaaacct tttctcagaa acaaacaaac 3360

ccaaatgtgg ttgtttgtcc tgattcctaa aaggtcttta tgtattctag ataataatct 3420

ttggtcagtt atatgtgtta aaaaatatct tctttgtggc caggcacggt agctcacacc 3480

tgtaatccca gcactttgcg gggctgaggt gggtggatca tctgaggtca agagttcaag 3540

atcagcctgg ccaacacagt gaaaccccat ctctactaaa catgtacaaa acttagctgg 3600

gtatggtggc gggtgcctgt aaccccagct gctccagagg ctgtggcaga agaatcgctt 3660

gaacccagga ggcagaggtt gcagcgagcc aagattgtgc cattgcactc cagactgggt 3720

gacaagagtg aaattctgcc tatctatcta tctatctatc tatatctata tatatatata 3780

tatatatcct ttgtaattta tttttccctt tttaaaattt tttataaaat tctttttttat 3840

ttttattttt agcagaggtg aggtttctga ggtttcatta tgttgcccag gctggtcttg 3900

aactcctgag ctcaagtgat cctcccacct cagccttcca aagtgctgga attgcagaca 3960

tgagccaccg cgcccctcct gtttttctct aattaatggt gtctttcttt gtctttctgg 4020

taataagcaa aaagttcttc atttgatttg gttaaattta taactgtttt ctcatatggt 4080

taacattttt tcttgcctgg ctaaagaaat ccttttctgc ccaatactat aaagaggttt 4140

gcccacattt tattccaaaa gttttaagtt ttgtctttca tcttgaagtc taatgtatca 4200

ggaactggct tttgtgcctg ttgggaggta gtgatccaat tccatgtctt gcatgtaggt 4260

aaccactggt ccctgcgcca tgtattcaat acgtcgtctt tctcctgcgg gtctgcaatc 4320

tcacctacca tccatcaagt ttccataggg ccatgggtct gcttctgggc tccctgttct 4380

gttccattgt caatttgtct atcctgtgcc agtatcacac tgtgtttatt acaatagctt 4440

tgtaacagct ctcgatatcc ggtaggacat ctccctccac cttctttttc tacttcagaa 4500

gtgtcttagc taggtcaggc acggtggctc acgcctgtaa tcccagcact ttgggaggcc 4560

gacgcggatg gatcacctga ggtcaggagt tttgagacag cctggccaac atggtgaaac 4620

cccatctcta ctaaaaaata caaaaattag tcaggcatgg tggcatgtgc ctgtaatccc 4680

agctatttgg gaggctgagg ccggagaatt gcttgaaccc gggggggcgga ggttgcagtg 4740

agccgagatc gtaccattgc actccagcct gggtgacaga gcgaaactct gtctcaggaa 4800

aaaaaagaaa agagatgtct tggttattct tggttcttta ttattcaata taaattttag 4860

aagctgaatt tgaaaagatt tggattggaa tttcattaaa tctacaggtc aatttaggga 4920

gagttgataa ttttacagaa ttgagtcatc tggtgttcca ataagaataa gagaacaatt 4980

attggctgta caattcttgc caaatagtag gcaaagcaaa gcttaggaag tatactggtg 5040

ccatttcagg aacaaagcta ggtgcgaata tttttgtctt tctgaatcat gatgctgtaa 5100

gttctaaagt gatttctcct cttggctttg gacacatggt gtttaattac ctactgctga 5160

ctatccacaa acagaaagag actggtcatg ccccacaggg ttggggtatc caagataatg 5220

gagcgaggct ctcatgtgtc ctaggttaca caccgaaaat ccacagttta ttctgtgaag 5280

aaaggaggct atgtttatga tacagactgt gatattttta tcatagccta ttctggtatc 5340

atgtgcaaaa gctataaatg aaaaacacag gaacttggca tgtgagtcat tgctcccct 5400

aaatgacaat taataaggaa ggaacattga gacagaataa aatgatcccc ttctgggttt 5460

aatttagaaa gttccataat taggtttaat agaaataaat gtaaatttct atgattaaaa 5520

ataaattagc acatttaggg atacacaaat tataaatcat tttctaaatg ctaaaaacaa 5580

gctcaggttt ttttcagaag aaagttttaa tttttttct ttagtggaag atatcactct 5640

gacggaaagt tttgatgtga ggggcggatg actataaagt gggcatcttc ccccacagga 5700

agatgtttcc atctgtgggt gagaggtgcc caccgcagct agggcaggtt acatgtgccc 5760

tgtgtgtggt aggacttgga gagtgatctt tatcaacgtt tttatttaaa agactatcta 5820

ataaaacaca aaactatgat gttcacagga aaaaaagaat aagaaaaaaa ga 5872

<210> 39

<211> 164

<212> PRT

<213> Homo sapiens

<400> 39

Met Lys Trp Lys Ala Leu Phe Thr Ala Ala Ile Leu Gln Ala Gln Leu
1 5 10 15

Pro Ile Thr Glu Ala Gln Ser Phe Gly Leu Leu Asp Pro Lys Leu Cys
20 25 30

Tyr Leu Leu Asp Gly Ile Leu Phe Ile Tyr Gly Val Ile Leu Thr Ala
35 40 45

Leu Phe Leu Arg Val Lys Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr
50 55 60

Gln Gln Gly Gln Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg
65 70 75 80

Glu Glu Tyr Asp Val Leu Asp Lys Arg Arg Gly Arg Asp Pro Glu Met
85 90 95

Gly Gly Lys Pro Gln Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn
100 105 110

Glu Leu Gln Lys Asp Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met
115 120 125

Lys Gly Glu Arg Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly
130 135 140

Leu Ser Thr Ala Thr Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala
145 150 155 160

Leu Pro Pro Arg

<210> 40

<211> 1600

<212> DNA

<213> Homo sapiens

<400> 40

aaccgtcccg gccaccgctg cctcagcctc tgcctcccag cctctttctg agggaaagga 60

caagatgaag tggaaggcgc ttttcaccgc ggccatcctg caggcacagt tgccgattac 120

agaggcacag agctttggcc tgctggatcc caaactctgc tacctgctgg atggaatcct 180

cttcatctat ggtgtcattc tcactgcctt gttcctgaga gtgaagttca gcaggagcgc 240

agacgccccc gcgtaccagc agggccagaa ccagctctat aacgagctca atctaggacg 300

aagagaggag tacgatgttt tggacaagag acgtggccgg gaccctgaga tgggggggaaa 360

gccgcagaga aggaagaacc ctcaggaagg cctgtacaat gaactgcaga aagataagat 420

ggcggaggcc tacagtgaga ttgggatgaa aggcgagcgc cggaggggca aggggcacga 480

tggcctttac cagggtctca gtacagccac caaggacacc tacgacgccc ttcacatgca 540

ggccctgccc cctcgctaac agccaggggа tttcaccact caaaggccag acctgcagac 600

gcccagatta tgagacacag gatgaagcat ttacaacccg gttcactctt ctcagccact 660

gaagtattcc cctttatgta caggatgctt tggttatatt tagctccaaa ccttcacaca 720

cagactgttg tccctgcact ctttaaggga gtgtactccc agggcttacg gccctggcct 780

tgggccctct ggtttgccgg tggtgcaggt agacctgtct cctggcggtt cctcgttctc 840

cctgggaggc gggcgcactg cctctcacag ctgagttgtt gagtctgttt tgtaaagtcc 900

ccagagaaag cgcagatgct agcacatgcc ctaatgtctg tatcactctg tgtctgagtg 960

gcttcactcc tgctgtaaat ttggcttctg ttgtcacctt cacctccttt caaggtaact 1020

gtactgggcc atgttgtgcc tccctggtga gagggccggg cagaggggca gatggaaagg 1080

agcctaggcc aggtgcaacc agggagctgc aggggcatgg gaaggtgggc gggcagggga 1140

gggtcagcca gggcctgcga gggcagcggg agcctccctg cctcaggcct ctgtgccgca 1200

ccattgaact gtaccatgtg ctacaggggc cagaagatga acagactgac cttgatgagc 1260

tgtgcacaaa gtggcataaa aaacatgtgg ttacacagtg tgaataaagt gctgcggagc 1320

aagaggaggc cgttgattca cttcacgctt tcagcgaatg acaaaatcat ctttgtgaag 1380

gcctcgcagg aagacccaac acatgggacc tataactgcc cagcggacag tggcaggaca 1440

ggaaaaaccc gtcaatgtac taggatactg ctgcgtcatt acagggcaca ggccatggat 1500

ggaaaacgct ctctactctg ctttttttct actgttttaa tttatactgg catgctaaag 1560

ccttcctatt ttgcataata aatgcttcag tgaaaatgca 1600

<210> 41

<211> 208

<212> PRT

<213> Petromyzon marinus

<400> 41

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr

1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
20 25 30
Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45
Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly

50 55 60
Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65 70 75 80
Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
85 90 95
Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110
Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr

115 120 125
Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
130 135 140
Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145 150 155 160
Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175
Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser

180 185 190
Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205
<210> 42
<211> 766
<212> DNA
<213> Petromyzon marinus
<400> 42
tgacacgaca cagccgtgta tatgaggaag ggtagctgga tgggggggggg gggaatacgt 60
tcagagagga cattagcgag cgtcttgttg gtggccttga gtctagacac ctgcagacat 120
gaccgacgct gagtacgtga gaatccatga gaagttggac atctacacgt ttaagaaaca 180

gtttttcaac aacaaaaaat ccgtgtcgca tagatgctac gttctctttg aattaaaacg 240

acggggtgaa cgtagagcgt gtttttgggg ctatgctgtg aataaaccac agagcgggac 300
agaacgtgga attcacgccg aaatctttag cattagaaaa gtcgaagaat acctgcgcga 360
caaccccgga caattcacga taaattggta ctcatcctgg agtccttgtg cagattgcgc 420
tgaaaagatc ttagaatggt ataaccagga gctgcggggg aacggccaca ctttgaaaat 480
ctgggcttgc aaactctatt acgagaaaaa tgcgaggaat caaattgggc tgtggaacct 540
cagagataac ggggttgggt tgaatgtaat ggtaagtgaa cactaccaat gttgcaggaa 600
aatattcatc caatcgtcgc acaatcaatt gaatgagaat agatggcttg agaagacttt 660

gaagcgagct gaaaaacgac ggagcgagtt gtccattatg attcaggtaa aaatactcca 720
caccactaag agtcctgctg tttaagaggc tatgcggatg gttttc 766

<210> 43
<211> 145
<212> PRT
<213> Homo sapiens
<400> 43

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45
Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95
Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110
Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Ala Pro

130 135 140

Val

145

<210> 44

<211> 10681

<212> DNA

<213> Homo sapiens

<400> 44

agagaaccat cattaattga agtgagattt ttctggcctg agacttgcag ggaggcaaga 60

agacactctg gacaccacta tggacaggta aagaggcagt cttctcgtgg gtgattgcac 120

tggccttcct ctcagagcaa atctgagtaa tgagactggt agctatccct ttctctcatg 180

taactgtctg actgataaga tcagcttgat caatatgcat atatattttt tgatctgtct 240

ccttttcttc tattcagatc ttatacgctg tcagcccaat tctttctgtt tcagacttct 300

cttgatttcc ctctttttca tgtggcaaaa gaagtagtgc gtacaatgta ctgattcgtc 360

ctgagatttg taccatggtt gaaactaatt tatggtaata atattaacat agcaaatctt 420

tagagactca aatcatgaaa aggtaatagc agtactgtac taaaaacggt agtgctaatt 480

ttcgtaataa ttttgtaaat attcaacagt aaaacaactt gaagacacac tttcctaggg 540

aggcgttact gaaataattt agctatagta agaaaatttg taattttaga aatgccaagc 600

attctaaatt aattgcttga aagtcactat gattgtgtcc attataagga gacaaattca 660

ttcaagcaag ttatttaatg ttaaaggccc aattgttagg cagttaatgg cacttttact 720

attaactaat ctttccattt gttcagacgt agcttaactt acctcttagg tgtgaatttg 780

gttaaggtcc tcataatgtc tttatgtgca gtttttgata ggttattgtc atagaactta 840

ttctattcct acatttatga ttactatgga tgtatgagaa taacacctaa tccttatact 900

ttacctcaat ttaactcctt tataaagaac ttacattaca gaataaagat tttttaaaaa 960

tatatttttt tgtagagaca gggtcttagc ccagccgagg ctggtctcta agtcctggcc 1020

caagcgatcc tcctgcctgg gcctcctaaa gtgctggaat tatagacatg agccatcaca 1080

tccaatatac agaataaaga tttttaatgg aggatttaat gttcttcaga aaattttctt 1140

gaggtcagac aatgtcaaat gtctcctcag tttacactga gattttgaaa acaagtctga 1200

gctataggtc cttgtgaagg gtccattgga aatacttgtt caaagtaaaa tggaaagcaa 1260

aggtaaaatc agcagttgaa attcagagaa agacagaaaa ggagaaaaga tgaaattcaa 1320

caggacagaa gggaaatata ttatcattaa ggaggacagt atctgtagag ctcattagtg 1380

atggcaaaat gacttggtca ggattatttt taacccgctt gtttctggtt tgcacggctg 1440
gggatgcagc tagggttctg cctcagggag cacagctgtc cagagcagct gtcagcctgc 1500
aagcctgaaa cactccctcg gtaaagtcct tcctactcag gacagaaatg acgagaacag 1560
ggagctggaa acaggcccct aaccagagaa gggaagtaat ggatcaacaa agttaactag 1620
caggtcagga tcacgcaatt catttcactc tgactggtaa catgtgacag aaacagtgta 1680
ggcttattgt attttcatgt agagtaggac ccaaaaatcc acccaaagtc ctttatctat 1740
gccacatcct tcttatctat acttccagga cacttttct tccttatgat aaggctctct 1800

ctctctccac acacacacac acacacacac acacacacac acacacacac acaaacacac 1860
accccgccaa ccaaggtgca tgtaaaaaga tgtagattcc tctgcctttc tcatctacac 1920
agcccaggag ggtaagttaa tataagaggg atttattggt aagagatgat gcttaatctg 1980
tttaacactg ggcctcaaag agagaatttc ttttcttctg tacttattaa gcacctatta 2040
tgtgttgagc ttatatatac aaagggttat tatatgctaa tatagtaata gtaatggtgg 2100
ttggtactat ggtaattacc ataaaaatta ttatcctttt aaaataaagc taattattat 2160
tggatctttt ttagtattca ttttatgttt tttatgtttt tgatttttta aaagacaatc 2220

tcaccctgtt acccaggctg gagtgcagtg gtgcaatcat agctttctgc agtcttgaac 2280
tcctgggctc aagcaatcct cctgccttgg cctcccaaag tgttgggata cagtcatgag 2340
ccactgcatc tggcctagga tccatttaga ttaaaatatg cattttaaat tttaaaataa 2400
tatggctaat ttttacctta tgtaatgtgt atactggcaa taaatctagt ttgctgccta 2460
aagtttaaag tgctttccag taagcttcat gtacgtgagg ggagacattt aaagtgaaac 2520
agacagccag gtgtggtggc tcacgcctgt aatcccagca ctctgggagg ctgaggtggg 2580
tggatcgctt gagccctgga gttcaagacc agcctgagca acatggcaaa acgctgtttc 2640

tataacaaaa attagccggg catggtggca tgtgcctgtg gtcccagcta ctaggggggct 2700
gaggcaggag aatcgttgga gcccaggagg tcaaggctgc actgagcagt gcttgcgcca 2760
ctgcactcca gcctgggtga caggaccaga ccttgcctca aaaaaataag aagaaaaatt 2820
aaaaataaat ggaaacaact acaaagagct gttgtcctag atgagctact tagttaggct 2880
gatattttgg tatttaactt ttaaagtcag ggtctgtcac ctgcactaca ttattaaaat 2940
atcaattctc aatgtatatc cacacaaaga ctggtacgtg aatgttcata gtacctttat 3000
tcacaaaacc ccaaagtaga gactatccaa atatccatca acaagtgaac aaataaacaa 3060

aatgtgctat atccatgcaa tggaatacca ccctgcagta caaagaagct acttggggat 3120

gaatcccaaa gtcatgacgc taaatgaaag agtcagacat gaaggaggag ataatgtatg 3180

ccatacgaaa ttctagaaaa tgaaagtaac ttatagttac agaaagcaaa tcagggcagg 3240

catagaggct cacacctgta atcccagcac tttgagaggc cacgtgggaa gattgctaga 3300

actcaggagt tcaagaccag cctgggcaac acagtgaaac tccattctcc acaaaaatgg 3360

gaaaaaaaga aagcaaatca gtggttgtcc tgtggggagg ggaaggactg caaagaggga 3420

agaagctctg gtggggtgag ggtggtgatt caggttctgt atcctgactg tggtagcagt 3480

ttggggtgtt tacatccaaa aatattcgta gaattatgca tcttaaatgg gtggagttta 3540

ctgtatgtaa attatacctc aatgtaagaa aaaataatgt gtaagaaaac tttcaattct 3600

cttgccagca aacgttattc aaattcctga gccctttact tcgcaaattc tctgcacttc 3660

tgccccgtac cattaggtga cagcactagc tccacaaatt ggataaatgc atttctggaa 3720

aagactaggg acaaaatcca ggcatcactt gtgctttcat atcaaccatg ctgtacagct 3780

tgtgttgctg tctgcagctg caatggggac tcttgatttc tttaaggaaa cttgggttac 3840

cagagtattt ccacaaatgc tattcaaatt agtgcttatg atatgcaaga cactgtgcta 3900

ggagccagaa aacaaagagg aggagaaatc agtcattatg tgggaacaac atagcaagat 3960

atttagatca ttttgactag ttaaaaaagc agcagagtac aaaatcacac atgcaatcag 4020

tataatccaa atcatgtaaa tatgtgcctg tagaaagact agaggaataa acacaagaat 4080

cttaacagtc attgtcatta gacactaagt ctaattatta ttattagaca ctatgatatt 4140

tgagatttaa aaaatcttta atattttaaa atttagagct cttctatttt tccatagtat 4200

tcaagtttga caatgatcaa gtattactct ttctttttttt ttttttttttt tttttttga 4260

gatggagttt tggtcttgtt gcccatgctg gagtggaatg gcatgaccat agctcactgc 4320

aacctccacc tcctgggttc aagcaaagct gtcgcctcag cctcccgggt agatgggatt 4380

acaggcgccc accaccacac tcggctaatg tttgtatttt tagtagagat ggggtttcac 4440

catgttggcc aggctggtct caaactcctg acctcagagg atccacctgc ctcagcctcc 4500

caaagtgctg ggattacaga tgtaggccac tgcgcccggc caagtattgc tcttatacat 4560

taaaaaacag gtgtgagcca ctgcgcccag ccaggtattg ctcttataca ttaaaaaata 4620

ggccggtgca gtggctcacg cctgtaatcc cagcactttg ggaagccaag gcgggcagaa 4680

cacccgaggt caggagtcca aggccagcct ggccaagatg gtgaaacccc gtctctatta 4740

aaaatacaaa cattacctgg gcatgatggt gggcgcctgt aatcccagct actcaggagg 4800

ctgaggcagg aggatccgcg gagcctggca gatctgcctg agcctgggag gttgaggcta 4860

cagtaagcca agatcatgcc agtatacttc agcctgggcg acaaagtgag accgtaacaa 4920

aaaaaaaaaa atttaaaaaa agaaatttag atcaagatcc aactgtaaaa agtggcctaa 4980

acaccacatt aaagagtttg gagtttattc tgcaggcaga agagaaccat caggggggtct 5040

tcagcatggg aatggcatgg tgcacctggt ttttgtgaga tcatggtggt gacagtgtgg 5100

ggaatgttat tttggaggga ctggaggcag acagaccggt taaaaggcca gcacaacaga 5160

taaggaggaa gaagatgagg gcttggaccg aagcagagaa gagcaaacag ggaaggtaca 5220

aattcaagaa atattggggg gtttgaatca acacatttag atgattaatt aaatatgagg 5280

actgaggaat aagaaatgag tcaaggatgg ttccaggctg ctaggctgct tacctgaggt 5340

ggcaaagtcg ggaggagtgg cagtttagga cagggggcag ttgaggaata ttgttttgat 5400

cattttgagt ttgaggtaca agttggacac ttaggtaaag actggagggg aaatctgaat 5460

atacaattat gggactgagg aacaagttta ttttatttttt tgtttcgttt tcttgttgaa 5520

gaacaaattt aattgtaatc ccaagtcatc agcatctaga agacagtggc aggaggtgac 5580

tgtcttgtgg gtaagggttt ggggtccttg atgagtatct ctcaattggc cttaaatata 5640

agcaggaaaa ggagtttatg atggattcca ggctcagcag ggctcaggag ggctcaggca 5700

gccagcagag gaagtcagag catcttcttt ggtttagccc aagtaatgac ttccttaaaa 5760

agctgaagga aaatccagag tgaccagatt ataaactgta ctcttgcatt ttctctccct 5820

cctctcaccc acagcctctt gatgaaccgg aggaagtttc tttaccaatt caaaaatgtc 5880

cgctgggcta agggtcggcg tgagacctac ctgtgctacg tagtgaagag gcgtgacagt 5940

gctacatcct tttcactgga ctttggttat cttcgcaata aggtatcaat taaagtcggc 6000

tttgcaagca gtttaatggt caactgtgag tgcttttaga gccacctgct gatggtatta 6060

cttccatcct tttttggcat ttgtgtctct atcacattcc tcaaatcctt ttttttattt 6120

ctttttccat gtccatgcac ccatattaga catggcccaa aatatgtgat ttaattcctc 6180

cccagtaatg ctgggcaccc taataccact ccttccttca gtgccaagaa caactgctcc 6240

caaactgttt accagctttc ctcagcatct gaattgcctt tgagattaat taagctaaaa 6300

gcatttttat atgggagaat attatcagct tgtccaagca aaaattttaa atgtgaaaaa 6360

caaattgtgt cttaagcatt tttgaaaatt aaggaagaag aatttgggaa aaaattaacg 6420

gtggctcaat tctgtcttcc aaatgatttc ttttccctcc tactcacatg ggtcgtaggc 6480

cagtgaatac attcaacatg gtgatcccca gaaaactcag agaagcctcg gctgatgatt 6540

aattaaattg atctttcggc tacccgagag aattacattt ccaagagact tcttcaccaa 6600

aatccagatg ggtttacata aacttctgcc cacgggtatc tcctctctcc taacacgctg 6660

tgacgtctgg gcttggtgga atctcaggga agcatccgtg gggtggaagg tcatcgtctg 6720

gctcgttgtt tgatggttat attaccatgc aattttcttt gcctacattt gtattgaata 6780

catcccaatc tccttcctat tcggtgacat gacacattct atttcagaag gctttgattt 6840

tatcaagcac tttcatttac ttctcatggc agtgcctatt acttctctta caatacccat 6900
ctgtctgctt taccaaaatc tatttcccct tttcagatcc tcccaaatgg tcctcataaa 6960
ctgtcctgcc tccacctagt ggtccaggta tatttccaca atgttacatc aacaggcact 7020
tctagccatt ttccttctca aaaggtgcaa aaagcaactt cataaacaca aattaaatct 7080
tcggtgaggt agtgtgatgc tgcttcctcc caactcagcg cacttcgtct tcctcattcc 7140
acaaaaaccc atagccttcc ttcactctgc aggactagtg ctgccaaggg ttcagctcta 7200
cctactggtg tgctcttttg agcaagttgc ttagcctctc tgtaacacaa ggacaatagc 7260

tgcaagcatc cccaaagatc attgcaggag acaatgacta aggctaccag agccgcaata 7320
aaagtcagtg aattttagcg tggtcctctc tgtctctcca gaacggctgc cacgtggaat 7380
tgctcttcct ccgctacatc tcggactggg acctagaccc tggccgctgc taccgcgtca 7440
cctggttcac ctcctggagc ccctgctacg actgtgcccg acatgtggcc gactttctgc 7500
gagggaaccc caacctcagt ctgaggatct tcaccgcgcg cctctacttc tgtgaggacc 7560
gcaaggctga gcccgagggg ctgcggcggc tgcaccgcgc cggggtgcaa atagccatca 7620
tgaccttcaa aggtgcgaaa gggccttccg cgcaggcgca gtgcagcagc ccgcattcgg 7680

gattgcgatg cggaatgaat gagttagtgg ggaagctcga ggggaagaag tgggcgggga 7740
ttctggttca cctctggagc cgaaattaaa gattagaagc agagaaaaga gtgaatggct 7800
cagagacaag gccccgagga aatgagaaaa tggggccagg gttgcttctt tcccctcgat 7860
ttggaacctg aactgtcttc tacccccata tccccgcctt tttttccttt ttttttttt 7920
gaagattatt tttactgctg gaatactttt gtagaaaacc acgaaagaac tttcaaagcc 7980
tgggaagggc tgcatgaaaa ttcagttcgt ctctccagac agcttcggcg catccttttg 8040
gtaaggggct tcctcgcttt ttaaattttc tttctttctc tacagtcttt tttggagttt 8100

cgtatatttc ttatattttc ttattgttca atcactctca gttttcatct gatgaaaact 8160
ttatttctcc tccacatcag ctttttcttc tgctgtttca ccattcagag ccctctgcta 8220
aggttccttt tccctccctt ttctttcttt tgttgtttca catctttaaa tttctgtctc 8280
tccccagggt tgcgtttcct tcctggtcag aattcttttc tccttttttt ttttttttt 8340
ttttttttt aaacaaacaa acaaaaaacc caaaaaaact ctttcccaat ttactttctt 8400
ccaacatgtt acaaagccat ccactcagtt tagaagactc tccggcccca ccgaccccca 8460
acctcgtttt gaagccattc actcaatttg cttctctctt tctctacagc ccctgtatga 8520

ggttgatgac ttacgagacg catttcgtac tttgggactt tgatagcaac ttccaggaat 8580

gtcacacacg atgaaatatc tctgctgaag acagtggata aaaaacagtc cttcaagtct 8640

tctctgtttt tattcttcaa ctctcacttt cttagagttt acagaaaaaa tatttatata 8700

cgactcttta aaaagatcta tgtcttgaaa atagagaagg aacacaggtc tggccaggga 8760

cgtgctgcaa ttggtgcagt tttgaatgca acattgtccc ctactgggaa taacagaact 8820

gcaggacctg ggagcatcct aaagtgtcaa cgtttttcta tgacttttag gtaggatgag 8880

agcagaaggt agatcctaaa aagcatggtg agaggatcaa atgtttttat atcaacatcc 8940

tttattattt gattcatttg agttaacagt ggtgttagtg atagattttt ctattctttt 9000

cccttgacgt ttactttcaa gtaacacaaa ctcttccatc aggccatgat ctataggacc 9060

tcctaatgag agtatctggg tgattgtgac cccaaaccat ctctccaaag cattaatatc 9120

caatcatgcg ctgtatgttt taatcagcag aagcatgttt ttatgtttgt acaaaagaag 9180

attgttatgg gtggggatgg aggtatagac catgcatggt caccttcaag ctactttaat 9240

aaaggatctt aaaatgggca ggaggactgt gaacaagaca ccctaataat gggttgatgt 9300

ctgaagtagc aaatcttctg gaaacgcaaa ctcttttaag gaagtcccta atttagaaac 9360

acccacaaac ttcacatatc ataattagca aacaattgga aggaagttgc ttgaatgttg 9420

gggagaggaa aatctattgg ctctcgtggg tctcttcatc tcagaaatgc caatcaggtc 9480

aaggtttgct acattttgta tgtgtgtgat gcttctccca aaggtatatt aactatataa 9540

gagagttgtg acaaaacaga atgataaagc tgcgaaccgt ggcacacgct catagttcta 9600

gctgcttggg aggttgagga gggaggatgg cttgaacaca ggtgttcaag gccagcctgg 9660

gcaacataac aagatcctgt ctctcaaaaa aaaaaaaaaa aaaaagaaag agagagggcc 9720

gggcgtggtg gctcacgcct gtaatcccag cactttggga ggccgagccg ggcggatcac 9780

ctgtggtcag gagtttgaga ccagcctggc caacatggca aaaccccgtc tgtactcaaa 9840

atgcaaaaat tagccaggcg tggtagcagg cacctgtaat cccagctact tgggaggctg 9900

aggcaggaga atcgcttgaa cccaggaggt ggaggttgca gtaagctgag atcgtgccgt 9960

tgcactccag cctgggcgac aagagcaaga ctctgtctca gaaaaaaaaa aaaaaaagag 10020

agagagagag aaagagaaca atatttggga gagaaggatg gggaagcatt gcaaggaaat 10080

tgtgctttat ccaacaaaat gtaaggagcc aataagggat ccctatttgt ctcttttggt 10140

gtctatttgt ccctaacaac tgtctttgac agtgagaaaa atattcagaa taaccatatc 10200

cctgtgccgt tattacctag caacccttgc aatgaagatg agcagatcca caggaaaact 10260

tgaatgcaca actgtcttat tttaatctta ttgtacataa gtttgtaaaa gagttaaaaa 10320

ttgttacttc atgtattcat ttatatttta tattattttg cgtctaatga ttttttatta 10380

acatgatttc cttttctgat atattgaaat ggagtctcaa agcttcataa atttataact 10440

ttagaaatga ttctaataac aacgtatgta attgtaacat tgcagtaatg gtgctacgaa 10500

gccatttctc ttgattttta gtaaactttt atgacagcaa atttgcttct ggctcacttt 10560

caatcagtta aataaatgat aaataatttt ggaagctgtg aagataaaat accaaataaa 10620

ataatataaa agtgatttat atgaagttaa aataaaaaat cagtatgatg gaataaactt 10680

g 10681

<210> 45

<211> 335

<212> PRT

<213> Homo sapiens

<400> 45

Met Leu Gly Ile Trp Thr Leu Leu Pro Leu Val Leu Thr Ser Val Ala

1 5 10 15

Arg Leu Ser Ser Lys Ser Val Asn Ala Gln Val Thr Asp Ile Asn Ser

20 25 30

Lys Gly Leu Glu Leu Arg Lys Thr Val Thr Thr Val Glu Thr Gln Asn

35 40 45

Leu Glu Gly Leu His His Asp Gly Gln Phe Cys His Lys Pro Cys Pro

50 55 60

Pro Gly Glu Arg Lys Ala Arg Asp Cys Thr Val Asn Gly Asp Glu Pro

65 70 75 80

Asp Cys Val Pro Cys Gln Glu Gly Lys Glu Tyr Thr Asp Lys Ala His

85 90 95

Phe Ser Ser Lys Cys Arg Arg Cys Arg Leu Cys Asp Glu Gly His Gly

100 105 110

Leu Glu Val Glu Ile Asn Cys Thr Arg Thr Gln Asn Thr Lys Cys Arg

115 120 125

Cys Lys Pro Asn Phe Phe Cys Asn Ser Thr Val Cys Glu His Cys Asp

130 135 140

Pro Cys Thr Lys Cys Glu His Gly Ile Ile Lys Glu Cys Thr Leu Thr

145 150 155 160

Ser Asn Thr Lys Cys Lys Glu Glu Gly Ser Arg Ser Asn Leu Gly Trp

165 170 175

Leu Cys Leu Leu Leu Leu Pro Ile Pro Leu Ile Val Trp Val Lys Arg

180 185 190

Lys Glu Val Gln Lys Thr Cys Arg Lys His Arg Lys Glu Asn Gln Gly

195 200 205

Ser His Glu Ser Pro Thr Leu Asn Pro Glu Thr Val Ala Ile Asn Leu

210 215 220

Ser Asp Val Asp Leu Ser Lys Tyr Ile Thr Thr Ile Ala Gly Val Met

225 230 235 240

Thr Leu Ser Gln Val Lys Gly Phe Val Arg Lys Asn Gly Val Asn Glu

245 250 255

Ala Lys Ile Asp Glu Ile Lys Asn Asp Asn Val Gln Asp Thr Ala Glu

260 265 270

Gln Lys Val Gln Leu Leu Arg Asn Trp His Gln Leu His Gly Lys Lys

275 280 285

Glu Ala Tyr Asp Thr Leu Ile Lys Asp Leu Lys Lys Ala Asn Leu Cys

290 295 300

Thr Leu Ala Glu Lys Ile Gln Thr Ile Ile Leu Lys Asp Ile Thr Ser

305 310 315 320

Asp Ser Glu Asn Ser Asn Phe Arg Asn Glu Ile Gln Ser Leu Val

325 330 335

<210> 46

<211> 3696

<212> DNA

<213> Homo sapiens

<400> 46

ctcttctccc gcgggttggt ggacccgctc agtacggagt tggggaagct ctttcacttc 60

ggaggattgc tcaacaacca tgctgggcat ctggaccctc ctacctctgg ttcttacgtc 120

tgttgctaga ttatcgtcca aaagtgttaa tgcccaagtg actgacatca actccaaggg 180

attggaattg aggaagactg ttactacagt tgagactcag aacttggaag gcctgcatca 240

tgatggccaa ttctgccata agccctgtcc tccaggtgaa aggaaagcta gggactgcac 300

agtcaatggg gatgaaccag actgcgtgcc ctgccaagaa gggaaggagt acacagacaa 360

agcccatttt tcttccaaat gcagaagatg tagattgtgt gatgaaggac atggcttaga 420

agtggaaata aactgcaccc ggacccagaa taccaagtgc agatgtaaac caaacttttt 480

ttgtaactct actgtatgtg aacactgtga cccttgcacc aaatgtgaac atggaatcat 540

caaggaatgc acactcacca gcaacaccaa gtgcaaagag gaaggatcca gatctaactt 600

ggggtggctt tgtcttcttc ttttgccaat tccactaatt gtttgggtga agagaaagga 660

agtacagaaa acatgcagaa agcacagaaa ggaaaaccaa ggttctcatg aatctccaac 720

tttaaatcct gaaacagtgg caataaattt atctgatgtt gacttgagta aatatatcac 780

cactattgct ggagtcatga cactaagtca agttaaaggc tttgttcgaa agaatggtgt 840

caatgaagcc aaaatagatg agatcaagaa tgacaatgtc caagacacag cagaacagaa 900

agttcaactg cttcgtaatt ggcatcaact tcatggaaag aaagaagcgt atgacacatt 960

gattaaagat ctcaaaaaag ccaatctttg tactcttgca gagaaaattc agactatcat 1020

cctcaaggac attactagtg actcagaaaa ttcaaacttc agaaatgaaa tccaaagctt 1080

ggtctagagt gaaaaacaac aaattcagtt ctgagtatat gcaattagtg tttgaaaaga 1140

ttcttaatag ctggctgtaa atactgcttg gttttttact gggtacattt tatcatttat 1200

tagcgctgaa gagccaacat atttgtagat ttttaatatc tcatgattct gcctccaagg 1260

atgtttaaaa tctagttggg aaaacaaact tcatcaagag taaatgcagt ggcatgctaa 1320

gtacccaaat aggagtgtat gcagaggatg aaagattaag attatgctct ggcatctaac 1380

atatgattct gtagtatgaa tgtaatcagt gtatgttagt acaaatgtct atccacaggc 1440

taaccccact ctatgaatca atagaagaag ctatgacctt ttgctgaaat atcagttact 1500

gaacaggcag gccactttgc ctctaaatta cctctgataa ttctagagat tttaccatat 1560

ttctaaactt tgtttataac tctgagaaga tcatatttat gtaaagtata tgtatttgag 1620

tgcagaattt aaataaggct ctacctcaaa gacctttgca cagtttattg gtgtcatatt 1680

atacaatatt tcaattgtga attcacatag aaaacattaa attataatgt ttgactatta 1740

tatatgtgta tgcattttac tggctcaaaa ctacctactt ctttctcagg catcaaaagc 1800

attttgagca ggagagtatt actagagctt tgccacctct ccatttttgc cttggtgctc 1860

atcttaatgg cctaatgcac ccccaaacat ggaaatatca ccaaaaaata cttaatagtc 1920

caccaaaagg caagactgcc cttagaaatt ctagcctggt ttggagatac taactgctct 1980

cagagaaagt agctttgtga catgtcatga acccatgttt gcaatcaaag atgataaaat 2040

agattcttat ttttccccca cccccgaaaa tgttcaataa tgtcccatgt aaaacctgct 2100

acaaatggca gcttatacat agcaatggta aaatcatcat ctggatttag gaattgctct 2160

tgtcataccc ccaagtttct aagatttaag attctcctta ctactatcct acgtttaaat 2220

atctttgaaa gtttgtatta aatgtgaatt ttaagaaata atatttatat ttctgtaaat 2280

gtaaactgtg aagatagtta taaactgaag cagatacctg gaaccaccta aagaacttcc 2340

atttatggag gatttttttg ccccttgtgt ttggaattat aaaatatagg taaaagtacg 2400

taattaaata atgtttttgg tatttctggt tttctctttt ttggtagggg cttgcttttt 2460

ggttttgtct tccttttctc taactgatgc taaatataac ttgtctttaa tgcttcttgg 2520

atcccttaga aggtacttcc tttttaacct taacccttttt agtagttaaa taattatttc 2580

cataggttgc tattgccaag aagacctctt ccaaacagca catgattatt cgtcaaacag 2640

tttcgtattc cagatactgg aatgtggata agaaagtata catttcaagg ggtaggtttt 2700

attattaaga aagccaaatg aggattttga aatattcttt cctgcatatt atccattcta 2760

gctacatgct ggccagtggg ccacctttct tttctgcaat ttaatgctag taatatattc 2820

tatttaaccc atgagtccca aagtattagc atttcaacat gtaagcatgt cggtaagata 2880

gttgtgcttt gcttagggtt ccctcctgtg ttatggtctg gaaagtgtct ttaggcagaa 2940

agtctgagtg atcacagggt tcactcatta atttctcttt tctgagccat catagtctgt 3000

gctgtctgct ctccagtttt ctatttctag acagaagtag ggcaagttag gtactagtta 3060

ttcttcatgg ccagaagtgc aagttctact ttgcaagaca agattaagtt agagaacacc 3120

ctattccact ttggtgaact cagagcaaga actttgagtt cctttgggag gaagacagtg 3180

gagaagtctt tgtacttggt gatgtggttt ttttcctcat ggcttcacct agtggcccca 3240

agcatgactt ctcccatgtc aatgagcaca gccacattcc cgagttgagg tgaccccacg 3300

gtccagaatc atcctcattc tggtgaacct ggttctcttt gtggtgggca tactgggtag 3360

gagaatcacc caaaggtcac ccatgagctg cagaaaaaaa ggctatttgc agaaggagct 3420

cacagatcac attgaaagca ttgcatattc aaacatcttg gtcttcttta ttggcatgcc 3480

cacagggtct tctgacctct gattagatca gacacttttt agatattgaa tcatcagttt 3540

ctgtacaact atctgaataa ggtatataat caatgaaatt tagaattttt ttctatgctt 3600

actcctgatt ggtaatttgt ttgggtttag aattctatac aaggccattt gtaattttcc 3660

tcagcacttt aaaaatatta aaccatgttt tcttaa 3696

<210> 47

<211> 306

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 47

tgcctgtcat acgaaaccga gatactgaca gtagaatatg gccttctgcc aatcgggaag 60

attgtggaga aacggataga atgcacagtt tactctgtcg ataacaatgg taacatttat 120

actcagccag ttgcccagtg gcacgaccgg ggagagcagg aagtattcga atactgtctg 180

gaggatggaa gtctcattag ggccactaag gaccacaaat ttatgacagt cgatggccag 240

atgctgccta tagacgaaat ctttgagcga gagttggacc tcatgcgagt tgacaacctt 300

cctaat 306

<210> 48

<211> 102

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 48

Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu
1 5 10 15

Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser
20 25 30

Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His
35 40 45

Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
50 55 60

Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln
65 70 75 80

Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg
85 90 95

Val Asp Asn Leu Pro Asn
100

<210> 49

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 49

atgatcaaga tagctacaag gaagtatctt ggcaaacaaa acgtttatga tattggagtc 60

gaaagagatc acaactttgc tctgaagaac ggattcatag cttctaat 108

<210> 50

<211> 36

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 50

Met Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn Val Tyr
1 5 10 15

Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn Gly Phe
20 25 30

Ile Ala Ser Asn
35

<210> 51

<211> 303

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 51

tgcctgtctt atgataccga gatacttacc gttgaatatg gcttcttgcc tattggaaag 60

attgtcgaag agagaattga atgcacagta tatactgtag acaagaatgg tttcgtttac 120

acacagccca ttgctcaatg gcacaatcgc ggcgaacaag aagtatttga gtactgtctc 180

gaggatggaa gcatcatacg agcaactaaa gatcataaat tcatgaccac tgacgggcag 240

atgttgccaa tagatgagat attcgagcgg ggcttggatc tcaaacaagt ggatggattg 300

cca 303

<210> 52

<211> 101

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 52

Cys Leu Ser Tyr Asp Thr Glu Ile Leu Thr Val Glu Tyr Gly Phe Leu
1 5 10 15

Pro Ile Gly Lys Ile Val Glu Glu Arg Ile Glu Cys Thr Val Tyr Thr
20 25 30

Val Asp Lys Asn Gly Phe Val Tyr Thr Gln Pro Ile Ala Gln Trp His

35 40 45

Asn Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
50 55 60

Ile Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Thr Asp Gly Gln
65 70 75 80

Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Gly Leu Asp Leu Lys Gln
85 90 95

Val Asp Gly Leu Pro
100

<210

> 53

<211> 159

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 53

atgaagagga ctgccgatgg atcagagttt gaatctccca agaagaagag gaaagtaaag 60

ataatatctc gaaaaagtct tggtacccaa aatgtctatg atattggagt ggagaaagat 120

cacaacttcc ttctcaagaa cggtctcgta gccagcaac 159

<210> 54

<211> 53

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 54

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Lys Ile Ile Ser Arg Lys Ser Leu Gly Thr Gln Asn Val
20 25 30

Tyr Asp Ile Gly Val Glu Lys Asp His Asn Phe Leu Leu Lys Asn Gly
35 40 45

Leu Val Ala Ser Asn
50

<210> 55

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 55

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly
20

<210> 56

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 56

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1 5 10 15

<210> 57

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 57

Ser Gly Gly Ser

1

<210> 58

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (1)..(120)

<223> This sequence may encompass 1-30 "Ser Gly Gly Ser" repeating units

<400> 58

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

1 5 10 15

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

20 25 30

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

35 40 45

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

50 55 60

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

65 70 75 80

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

85 90 95

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

100 105 110

Ser Gly Gly Ser Ser Gly Gly Ser

115 120

<210> 59

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<220><221> MISC_FEATURE

<222> (1)..(120)

<223> This sequence may encompass 1-30 "Gly Gly Gly Ser"

repeating units

<400> 59

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

20 25 30

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

35 40 45

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

50 55 60

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

65 70 75 80

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

85 90 95

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser

100 105 110

Gly Gly Gly Ser Gly Gly Gly Ser

115 120

<210> 60

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (1)..(150)

<223> This sequence may encompass 1-30 "Gly Gly Gly Gly Ser" repeating units

<400> 60

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
50 55 60

Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
65 70 75 80

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
85 90 95

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
100 105 110

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
115 120 125

Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
130 135 140

Ser Gly Gly Gly Gly Ser
145 150

<210> 61

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (1)..(30)

<223> This sequence may encompass 1-30 residues

<400> 61

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
1 5 10 15

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
20 25 30

<210> 62

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (1)..(150)

<223> This sequence may encompass 1-30 "Glu Ala Ala Ala Lys" repeating units

<400> 62

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu
1 5 10 15

Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala
20 25 30

Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala
35 40 45

Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala
50 55 60

Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys

65 70 75 80

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu

85 90 95

Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala

100 105 110

Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala

115 120 125

Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala

130 135 140

Lys Glu Ala Ala Ala Lys

145 150

<210> 63

<211> 90

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (1)..(90)

<223> This sequence may encompass 1-30 "Gly Gly Ser" repeating units

<400> 63

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

1 5 10 15

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

20 25 30

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

35 40 45

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

50 55 60

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

65 70 75 80

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

85 90

<210> 64

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<220><221> MISC_FEATURE

<222> (1)..(60)

<223> This sequence may encompass 1-30 "Xaa Pro"

repeating units

<220><221> MOD_RES

<222> (1)..(1)

<223> Any amino acid

<220><221> MOD_RES

<222> (3)..(3)

<223> Any amino acid

<220><221> MOD_RES

<222> (5)..(5)

<223> Any amino acid

<220><221> MOD_RES

<222> (7)..(7)

<223> Any amino acid

<220><221> MOD_RES

<222> (9)..(9)

<223> Any amino acid

<220><221> MOD_RES

<222> (11)..(11)

<223> Any amino acid

<220><

221> MOD_RES

<222> (13)..(13)

<223> Any amino acid

<220><221> MOD_RES

<222> (15)..(15)

<223> Any amino acid

<220><221> MOD_RES

<222> (17)..(17)

<223> Any amino acid

<220><221> MOD_RES

<222> (19)..(19)

<223> Any amino acid

<220><221> MOD_RES

<222> (21)..(21)

<223> Any amino acid

<220><221> MOD_RES

<222> (23)..(23)

<223> Any amino acid

<220><221> MOD_RES

<222> (25)..(25)

<223> Any amino acid

<220><221> MOD_RES

<222> (27)..(27)

<223> Any amino acid

<220><221> MOD_RES

<222> (29)..(29)

<223> Any amino acid

<220><221> MOD_RES

<222> (31)..(31)

<223> Any amino acid

<220><221> MOD_RES

<222> (33)..(33)

<223> Any amino acid

<220><221> MOD_RES

<222> (35)..(35)

<223> Any amino acid

<220><221> MOD_RES

<222> (37)..(37)

<223> Any amino acid

<220><221> MOD_RES

<222> (39)..(39)

<223> Any amino acid

<220><221> MOD_RES

<222> (41)..(41)

<223> Any amino acid

<220><221> MOD_RES

<222> (43)..(43)

<223> Any amino acid

<220><221> MOD_RES

<222> (45)..(45)

<223> Any amino acid

<220><221> MOD_RES

<222> (47)..(47)

<223> Any amino acid

<220><221> MOD_RES

<222> (49)..(49)

<223> Any amino acid

<220><221> MOD_RES

<222> (51)..(51)

<223> Any amino acid

<220><221> MOD_RES

<222> (53)..(53)

<223> Any amino acid

<220><221> MOD_RES

<222> (55)..(55)

<223> Any amino acid

<220><221> MOD_RES

<222> (57)..(57)

<223> Any amino acid

<220><221> MOD_RES

<222> (59)..(59)

<223> Any amino acid

<400> 64

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

1 5 10 15

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

20 25 30

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

35 40 45

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

50 55 60

<210> 65

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 65

Pro Ala Pro Ala Pro

1 5

<210> 66

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 66

Pro Ala Pro Ala Pro Ala

1 5

<210> 67

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 67

Pro Ala Pro Ala Pro Ala Pro

1 5

<210> 68

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 68

Pro Ala Pro Ala Pro Ala Pro Ala

1 5

<210> 69

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 69

Pro Ala Pro Ala Pro Ala Pro Ala Pro

1 5

<210> 70

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 70

Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

1 5 10 15

<210> 71

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 71

Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala

1 5 10 15

Pro Ala Pro Ala Pro

20

<210> 72

<211> 24

<212> PRT

<

213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 72

Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala

1 5 10 15

Thr Pro Glu Ser Ser Gly Gly Ser

20

<210> 73

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 73

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser

20 25 30

<210> 74

<211> 104

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 74

Gly Gly Ser Gly Gly Ser Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser

1 5 10 15

Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly

20 25 30

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly

35 40 45

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly

50 55 60

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly

65 70 75 80

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala

85 90 95

Thr Ser Gly Gly Ser Gly Gly Ser

100

<210> 75

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 75

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser

20

<210> 76

<211> 40

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 76

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

Ser Gly Gly Ser Ser Gly Gly Ser
35 40

<210> 77

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 77

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
35 40 45

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser

50 55 60

<210> 78

<211> 92

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 78

Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser

1 5 10 15

Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser

20 25 30

Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu

35 40 45

Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser

50 55 60

Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr

65 70 75 80

Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser

85 90

<210> 79

<211> 525

<212> PRT

<213> Homo sapiens

<400> 79

Met Trp Glu Ala Gln Phe Leu Gly Leu Leu Phe Leu Gln Pro Leu Trp

1 5 10 15

Val Ala Pro Val Lys Pro Leu Gln Pro Gly Ala Glu Val Pro Val Val

20 25 30

Trp Ala Gln Glu Gly Ala Pro Ala Gln Leu Pro Cys Ser Pro Thr Ile

35 40 45

Pro Leu Gln Asp Leu Ser Leu Leu Arg Arg Ala Gly Val Thr Trp Gln

50 55 60

His Gln Pro Asp Ser Gly Pro Pro Ala Ala Ala Pro Gly His Pro Leu

65 70 75 80

Ala Pro Gly Pro His Pro Ala Ala Pro Ser Ser Trp Gly Pro Arg Pro

85 90 95

Arg Arg Tyr Thr Val Leu Ser Val Gly Pro Gly Gly Leu Arg Ser Gly

100 105 110

Arg Leu Pro Leu Gln Pro Arg Val Gln Leu Asp Glu Arg Gly Arg Gln

115 120 125

Arg Gly Asp Phe Ser Leu Trp Leu Arg Pro Ala Arg Arg Ala Asp Ala

130 135 140

Gly Glu Tyr Arg Ala Ala Val His Leu Arg Asp Arg Ala Leu Ser Cys

145 150 155 160

Arg Leu Arg Leu Arg Leu Gly Gln Ala Ser Met Thr Ala Ser Pro Pro

165 170 175

Gly Ser Leu Arg Ala Ser Asp Trp Val Ile Leu Asn Cys Ser Phe Ser

180 185 190

Arg Pro Asp Arg Pro Ala Ser Val His Trp Phe Arg Asn Arg Gly Gln

195 200 205

Gly Arg Val Pro Val Arg Glu Ser Pro His His His Leu Ala Glu Ser

210 215 220

Phe Leu Phe Leu Pro Gln Val Ser Pro Met Asp Ser Gly Pro Trp Gly

225 230 235 240

Cys Ile Leu Thr Tyr Arg Asp Gly Phe Asn Val Ser Ile Met Tyr Asn

245 250 255

Leu Thr Val Leu Gly Leu Glu Pro Pro Thr Pro Leu Thr Val Tyr Ala

260 265 270

Gly Ala Gly Ser Arg Val Gly Leu Pro Cys Arg Leu Pro Ala Gly Val

275 280 285

Gly Thr Arg Ser Phe Leu Thr Ala Lys Trp Thr Pro Pro Gly Gly Gly

290 295 300

Pro Asp Leu Leu Val Thr Gly Asp Asn Gly Asp Phe Thr Leu Arg Leu

305 310 315 320

Glu Asp Val Ser Gln Ala Gln Ala Gly Thr Tyr Thr Cys His Ile His

325 330 335

Leu Gln Glu Gln Gln Leu Asn Ala Thr Val Thr Leu Ala Ile Ile Thr

340 345 350

Val Thr Pro Lys Ser Phe Gly Ser Pro Gly Ser Leu Gly Lys Leu Leu

355 360 365

Cys Glu Val Thr Pro Val Ser Gly Gln Glu Arg Phe Val Trp Ser Ser

370 375 380

Leu Asp Thr Pro Ser Gln Arg Ser Phe Ser Gly Pro Trp Leu Glu Ala

385 390 395 400

Gln Glu Ala Gln Leu Leu Ser Gln Pro Trp Gln Cys Gln Leu Tyr Gln

405 410 415

Gly Glu Arg Leu Leu Gly Ala Ala Val Tyr Phe Thr Glu Leu Ser Ser

420 425 430

Pro Gly Ala Gln Arg Ser Gly Arg Ala Pro Gly Ala Leu Pro Ala Gly

435 440 445

His Leu Leu Leu Phe Leu Ile Leu Gly Val Leu Ser Leu Leu Leu Leu

450 455 460

Val Thr Gly Ala Phe Gly Phe His Leu Trp Arg Arg Gln Trp Arg Pro

465 470 475 480

Arg Arg Phe Ser Ala Leu Glu Gln Gly Ile His Pro Pro Gln Ala Gln

485 490 495

Ser Lys Ile Glu Glu Leu Glu Gln Glu Pro Glu Pro Glu Pro Glu Pro

500 505 510

Glu Pro Glu Pro Glu Pro Glu Pro Glu Pro Glu Gln Leu

515 520 525

<210> 80

<211> 1976

<212> DNA

<213> Homo sapiens

<400> 80

agagaccagc agaacggcat cccagccacg acggccactt tgctctgtct gctctccgcc 60

acggccctgc tctgttccct gggacacccc cgcccccacc tcctcaggct gcctgatctg 120

cccagctttc cagctttcct ctggattccg gcctctggtc atccctcccc accctctctc 180

caaggccctc tcctggtctc ccttcttcta gaacccccttc ctccacctcc ctctctgcag 240

aacttctcct ttaccccca ccccccacca ctgccccctt tccttttctg acctcctttt 300

ggagggctca gcgctgccca gaccatagga gagatgtggg aggctcagtt cctgggcttg 360

ctgtttctgc agccgctttg ggtggctcca gtgaagcctc tccagccagg ggctgaggtc 420

ccggtggtgt gggcccagga gggggctcct gcccagctcc cctgcagccc cacaatcccc 480

ctccaggatc tcagccttct gcgaagagca ggggtcactt ggcagcatca gccagacagt 540

ggcccgcccg ctgccgcccc cggccatccc ctggcccccg gccctcaccc ggcggcgccc 600

tcctcctggg ggcccaggcc ccgccgctac acggtgctga gcgtgggtcc cggaggcctg 660

cgcagcggga ggctgcccct gcagccccgc gtccagctgg atgagcgcgg ccggcagcgc 720

ggggacttct cgctatggct gcgcccagcc cggcgcgcgg acgccggcga gtaccgcgcc 780

gcggtgcacc tcagggaccg cgccctctcc tgccgcctcc gtctgcgcct gggccaggcc 840

tcgatgactg ccagcccccc aggatctctc agagcctccg actgggtcat tttgaactgc 900

tccttcagcc gccctgaccg cccagcctct gtgcattggt tccggaaccg gggccagggc 960

cgagtccctg tccgggagtc cccccatcac cacttagcgg aaagcttcct cttcctgccc 1020

caagtcagcc ccatggactc tgggccctgg ggctgcatcc tcacctacag agatggcttc 1080

aacgtctcca tcatgtataa cctcactgtt ctgggtctgg agcccccaac tcccttgaca 1140

gtgtacgctg gagcaggttc cagggtgggg ctgccctgcc gcctgcctgc tggtgtgggg 1200

acccggtctt tcctcactgc caagtggact cctcctgggg gaggccctga cctcctggtg 1260

actggagaca atggcgactt taccccttcga ctagaggatg tgagccaggc ccaggctggg 1320

acctacacct gccatatcca tctgcaggaa cagcagctca atgccactgt cacattggca 1380

atcatcacag tgactcccaa atcctttggg tcacctggat ccctggggaa gctgctttgt 1440

gaggtgactc cagtatctgg acaagaacgc tttgtgtgga gctctctgga caccccatcc 1500

cagaggagtt tctcaggacc ttggctggag gcacaggagg cccagctcct ttcccagcct 1560

tggcaatgcc agctgtacca gggggagagg cttcttggag cagcagtgta cttcacagag 1620

ctgtctagcc caggtgccca acgctctggg agagccccag gtgccctccc agcaggccac 1680

ctcctgctgt ttctcatcct tggtgtcctt tctctgctcc ttttggtgac tggagccttt 1740

ggctttcacc tttggagaag acagtggcga ccaagacgat tttctgcctt agagcaaggg 1800

attcaccctc cgcaggctca gagcaagata gaggagctgg agcaagaacc ggagccggag 1860

ccggagccgg aaccggagcc cgagcccgag cccgagccgg agcagctctg acctggagct 1920

gaggcagcca gcagatctca gcagcccagt ccaaataaac tccctgtcag cagcaa 1976

<210> 81

<211> 716

<212> PRT

<213> Homo sapiens

<400> 81

```
Met Pro Ser Thr Ser Phe Pro Val Pro Ser Lys Phe Pro Leu Gly Pro
1               5                  10                  15
Ala Ala Ala Val Phe Gly Arg Gly Glu Thr Leu Gly Pro Ala Pro Arg
            20                  25                  30
Ala Gly Gly Thr Met Lys Ser Ala Glu Glu Glu His Tyr Gly Tyr Ala
        35                  40                  45
Ser Ser Asn Val Ser Pro Ala Leu Pro Leu Pro Thr Ala His Ser Thr
    50                  55                  60

Leu Pro Ala Pro Cys His Asn Leu Gln Thr Ser Thr Pro Gly Ile Ile
65                  70                  75                  80
Pro Pro Ala Asp His Pro Ser Gly Tyr Gly Ala Ala Leu Asp Gly Gly
                85                  90                  95
Pro Ala Gly Tyr Phe Leu Ser Ser Gly His Thr Arg Pro Asp Gly Ala
            100                 105                 110
Pro Ala Leu Glu Ser Pro Arg Ile Glu Ile Thr Ser Cys Leu Gly Leu
        115                 120                 125

Tyr His Asn Asn Asn Gln Phe Phe His Asp Val Glu Val Glu Asp Val
    130                 135                 140
Leu Pro Ser Ser Lys Arg Ser Pro Ser Thr Ala Thr Leu Ser Leu Pro
145                 150                 155                 160
Ser Leu Glu Ala Tyr Arg Asp Pro Ser Cys Leu Ser Pro Ala Ser Ser
                165                 170                 175
Leu Ser Ser Arg Ser Cys Asn Ser Glu Ala Ser Ser Tyr Glu Ser Asn
            180                 185                 190

Tyr Ser Tyr Pro Tyr Ala Ser Pro Gln Thr Ser Pro Trp Gln Ser Pro
        195                 200                 205
Cys Val Ser Pro Lys Thr Thr Asp Pro Glu Glu Gly Phe Pro Arg Gly
    210                 215                 220
Leu Gly Ala Cys Thr Leu Leu Gly Ser Pro Arg His Ser Pro Ser Thr
225                 230                 235                 240
Ser Pro Arg Ala Ser Val Thr Glu Glu Ser Trp Leu Gly Ala Arg Ser
```

245 250 255

Ser Arg Pro Ala Ser Pro Cys Asn Lys Arg Lys Tyr Ser Leu Asn Gly
260 265 270
Arg Gln Pro Pro Tyr Ser Pro His His Ser Pro Thr Pro Ser Pro His
275 280 285
Gly Ser Pro Arg Val Ser Val Thr Asp Asp Ser Trp Leu Gly Asn Thr
290 295 300
Thr Gln Tyr Thr Ser Ser Ala Ile Val Ala Ala Ile Asn Ala Leu Thr
305 310 315 320

Thr Asp Ser Ser Leu Asp Leu Gly Asp Gly Val Pro Val Lys Ser Arg
325 330 335
Lys Thr Thr Leu Glu Gln Pro Pro Ser Val Ala Leu Lys Val Glu Pro
340 345 350
Val Gly Glu Asp Leu Gly Ser Pro Pro Pro Pro Ala Asp Phe Ala Pro
355 360 365
Glu Asp Tyr Ser Ser Phe Gln His Ile Arg Lys Gly Gly Phe Cys Asp
370 375 380

Gln Tyr Leu Ala Val Pro Gln His Pro Tyr Gln Trp Ala Lys Pro Lys
385 390 395 400
Pro Leu Ser Pro Thr Ser Tyr Met Ser Pro Thr Leu Pro Ala Leu Asp
405 410 415
Trp Gln Leu Pro Ser His Ser Gly Pro Tyr Glu Leu Arg Ile Glu Val
420 425 430
Gln Pro Lys Ser His His Arg Ala His Tyr Glu Thr Glu Gly Ser Arg
435 440 445

Gly Ala Val Lys Ala Ser Ala Gly Gly His Pro Ile Val Gln Leu His
450 455 460
Gly Tyr Leu Glu Asn Glu Pro Leu Met Leu Gln Leu Phe Ile Gly Thr
465 470 475 480
Ala Asp Asp Arg Leu Leu Arg Pro His Ala Phe Tyr Gln Val His Arg
485 490 495

Ile Thr Gly Lys Thr Val Ser Thr Thr Ser His Glu Ala Ile Leu Ser
500 505 510

Asn Thr Lys Val Leu Glu Ile Pro Leu Leu Pro Glu Asn Ser Met Arg
515 520 525

Ala Val Ile Asp Cys Ala Gly Ile Leu Lys Leu Arg Asn Ser Asp Ile
530 535 540

Glu Leu Arg Lys Gly Glu Thr Asp Ile Gly Arg Lys Asn Thr Arg Val
545 550 555 560

Arg Leu Val Phe Arg Val His Val Pro Gln Pro Ser Gly Arg Thr Leu
565 570 575

Ser Leu Gln Val Ala Ser Asn Pro Ile Glu Cys Ser Gln Arg Ser Ala
580 585 590

Gln Glu Leu Pro Leu Val Glu Lys Gln Ser Thr Asp Ser Tyr Pro Val
595 600 605

Val Gly Gly Lys Lys Met Val Leu Ser Gly His Asn Phe Leu Gln Asp
610 615 620

Ser Lys Val Ile Phe Val Glu Lys Ala Pro Asp Gly His His Val Trp
625 630 635 640

Glu Met Glu Ala Lys Thr Asp Arg Asp Leu Cys Lys Pro Asn Ser Leu
645 650 655

Val Val Glu Ile Pro Pro Phe Arg Asn Gln Arg Ile Thr Ser Pro Val
660 665 670

His Val Ser Phe Tyr Val Cys Asn Gly Lys Arg Lys Arg Ser Gln Tyr
675 680 685

Gln Arg Phe Thr Tyr Leu Pro Ala Asn Gly Asn Ala Ile Phe Leu Thr
690 695 700

Val Ser Arg Glu His Glu Arg Val Gly Cys Phe Phe
705 710 715

<210> 82

<211> 2978

<212> DNA

<213> Homo sapiens

<400> 82

ggcgggcgct cggcgactcg tccccggggc cccgcgcggg cccgggcagc aggggcgtga 60

tgtcacggca gggaggggggc gcgggagccg ccgggccggc ggggaggcgg gggaggtgtt 120

ttccagcttt aaaaaggcag gaggcagagc gcggccctgc gtcagagcga gactcagagg 180

ctccgaactc gccggcggag tcgccgcgcc agatcccagc agcagggcgc gggcaccggg 240

gcgcgggcag ggctcggagc caccgcgcag gtcctagggc cgcggccggg ccccgccacg 300

cgcgcacacg cccctcgatg actttcctcc ggggcgcgcg gcgctgagcc cggggcgagg 360

gctgtcttcc cggagacccg accccggcag cgcggggcgg ccgcttctcc tgtgcctccg 420

cccgccgctc cactccccgc cgccgccgcg cggatgccaa gcaccagctt tccagtccct 480

tccaagtttc cacttggccc tgcggctgcg gtcttcggga gaggagaaac tttggggccc 540

gcgccgcgcg ccggcggcac catgaagtca gcggaggaag aacactatgg ctatgcatcc 600

tccaacgtca gccccgccct gccgctcccc acggcgcact ccaccctgcc ggccccgtgc 660

cacaaccttc agacctccac accgggcatc atcccgccgg cggatcaccc ctcggggtac 720

ggagcagctt tggacggtgg gcccgcgggc tacttcctct cctccggcca caccaggcct 780

gatggggccc ctgccctgga gagtcctcgc atcgagataa cctcgtgctt gggcctgtac 840

cacaacaata accagttttt ccacgatgtg gaggtggaag acgtcctccc tagctccaaa 900

cggtcccccct ccacggccac gctgagtctg cccagcctgg aggcctacag agacccctcg 960

tgcctgagcc cggccagcag cctgtcctcc cggagctgca actcagaggc ctcctcctac 1020

gagtccaact actcgtaccc gtacgcgtcc ccccagacgt cgccatggca gtctccctgc 1080

gtgtctccca agaccacgga ccccgaggag ggctttcccc gcgggctggg ggcctgcaca 1140

ctgctgggtt ccccgcggca ctccccctcc acctcgcccc gcgccagcgt cactgaggag 1200

agctggctgg gtgcccgctc ctccagaccc gcgtcccctt gcaacaagag gaagtacagc 1260

ctcaacggcc ggcagccgcc ctactcaccc caccactcgc ccacgccgtc cccgcacggc 1320

tccccgcggg tcagcgtgac cgacgactcg tggttgggca acaccaccca gtacaccagc 1380

tcggccatcg tggccgccat caacgcgctg accaccgaca gcagcctgga cctgggagat 1440

ggcgtccctg tcaagtcccg caagaccacc ctggagcagc cgccctcagt ggcgctcaag 1500

gtggagcccg tcggggagga cctgggcagc cccccgcccc cggccgactt cgcgcccgaa 1560

gactactcct ctttccagca catcaggaag ggcggcttct gcgaccagta cctggcggtg 1620

ccgcagcacc cctaccagtg ggcgaagccc aagcccctgt cccctacgtc ctacatgagc 1680

ccgaccctgc ccgccctgga ctggcagctg ccgtcccact caggcccgta tgagcttcgg 1740

attgaggtgc agcccaagtc ccaccaccga gcccactacg agacggaggg cagccggggg 1800

gccgtgaagg cgtcggccgg aggacacccc atcgtgcagc tgcatggcta cttggagaat 1860

gagccgctga tgctgcagct tttcattggg acggcggacg accgcctgct gcgcccgcac 1920

gccttctacc aggtgcaccg catcacaggg aagaccgtgt ccaccaccag ccacgaggcc 1980

atcctctcca acaccaaagt cctggagatc ccactcctgc cggagaacag catgcgagcc 2040

gtcattgact gtgccggaat cctgaaactc agaaactccg acattgaact tcggaaagga 2100

gagacggaca tcgggaggaa gaacacacgg gtacggctgg tgttccgcgt tcacgtcccg 2160

caacccagcg gccgcacgct gtccctgcag gtggcctcca accccatcga atgctcccag 2220

cgctcagctc aggagctgcc tctggtggag aagcagagca cggacagcta tccggtcgtg 2280

ggcgggaaga agatggtcct gtctggccac aacttcctgc aggactccaa ggtcattttc 2340

gtggagaaag ccccagatgg ccaccatgtc tgggagatgg aagcgaaaac tgaccgggac 2400

ctgtgcaagc cgaattctct ggtggttgag atcccgccat ttcggaatca gaggataacc 2460

agccccgttc acgtcagttt ctacgtctgc aacgggaaga gaaagcgaag ccagtaccag 2520

cgtttcacct accttcccgc caacggtaac gccatctttc taaccgtaag ccgtgaacat 2580

gagcgcgtgg ggtgcttttt ctaaagacgc agaaacgacg tcgccgtaaa gcagcgtggc 2640

gtgttgcaca tttaactgtg tgatgtcccg ttagtgagac cgagccatcg atgccctgaa 2700

aaggaaagga aaagggaagc ttcggatgca ttttccttga tccctgttgg gggtgggggg 2760

cgggggttgc atactcagat agtcacggtt attttgcttc ttgcgaatgt ataacagcca 2820

aggggaaaac atggctcttc tgctccaaaa aactgagggg gtcctggtgt gcatttgcac 2880

cctaaagctg cttacggtga aaaggcaaat aggtatagct attttgcagg cacctttagg 2940

aataaacttt gcttttaagc ctgtaaaaaa aaaaaaaa 2978

<210> 83

<211> 29

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 83

Pro Lys Lys Lys Arg Lys Val Glu Gly Ala Asp Lys Arg Thr Ala Asp

1 5 10 15

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

20 25

<210> 84

<211> 18

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 84

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg

1 5 10 15

Lys Val

<210> 85

<211> 16

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 85

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 86

<211> 17

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 86

Lys Lys Thr Glu Leu Gln Thr Thr Asn Ala Glu Asn Lys Thr Lys Lys

1 5 10 15

Leu

<210> 87

<211> 19

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 87

Lys Arg Gly Ile Asn Asp Arg Asn Phe Trp Arg Gly Glu Asn Gly Arg

1 5 10 15

Lys Thr Arg

<210> 88

<211> 16

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 88

Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys

1 5 10 15

<210> 89

<211> 7

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 89

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 90

<211> 30

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization sequence

<400> 90

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys

20 25 30

<210> 91

<211> 288

<212> PRT

<213> Homo sapiens

<400> 91

Met Gln Ile Pro Gln Ala Pro Trp Pro Val Val Trp Ala Val Leu Gln

1 5 10 15

Leu Gly Trp Arg Pro Gly Trp Phe Leu Asp Ser Pro Asp Arg Pro Trp

20 25 30

Asn Pro Pro Thr Phe Ser Pro Ala Leu Leu Val Val Thr Glu Gly Asp

35 40 45

Asn Ala Thr Phe Thr Cys Ser Phe Ser Asn Thr Ser Glu Ser Phe Val

50 55 60

Leu Asn Trp Tyr Arg Met Ser Pro Ser Asn Gln Thr Asp Lys Leu Ala

65 70 75 80

Ala Phe Pro Glu Asp Arg Ser Gln Pro Gly Gln Asp Cys Arg Phe Arg

85 90 95

Val Thr Gln Leu Pro Asn Gly Arg Asp Phe His Met Ser Val Val Arg

100 105 110

Ala Arg Arg Asn Asp Ser Gly Thr Tyr Leu Cys Gly Ala Ile Ser Leu

115 120 125

Ala Pro Lys Ala Gln Ile Lys Glu Ser Leu Arg Ala Glu Leu Arg Val

130 135 140

Thr Glu Arg Arg Ala Glu Val Pro Thr Ala His Pro Ser Pro Ser Pro

145 150 155 160

Arg Pro Ala Gly Gln Phe Gln Thr Leu Val Val Gly Val Val Gly Gly

165 170 175

Leu Leu Gly Ser Leu Val Leu Leu Val Trp Val Leu Ala Val Ile Cys
180 185 190
Ser Arg Ala Ala Arg Gly Thr Ile Gly Ala Arg Arg Thr Gly Gln Pro

195 200 205
Leu Lys Glu Asp Pro Ser Ala Val Pro Val Phe Ser Val Asp Tyr Gly
210 215 220
Glu Leu Asp Phe Gln Trp Arg Glu Lys Thr Pro Glu Pro Pro Val Pro
225 230 235 240
Cys Val Pro Glu Gln Thr Glu Tyr Ala Thr Ile Val Phe Pro Ser Gly
245 250 255
Met Gly Thr Ser Ser Pro Ala Arg Arg Gly Ser Ala Asp Gly Pro Arg

260 265 270
Ser Ala Gln Pro Leu Arg Pro Glu Asp Gly His Cys Ser Trp Pro Leu
275 280 285

<210> 92
<211> 867
<212> DNA
<213> Homo sapiens
<400> 92

atgcagatcc cacaggcgcc ctggccagtc gtctgggcgg tgctacaact gggctggcgg 60
ccaggatggt tcttagactc cccagacagg ccctggaacc cccccacctt ctccccagcc 120
ctgctcgtgg tgaccgaagg ggacaacgcc accttcacct gcagcttctc caacacatcg 180
gagagcttcg tgctaaactg gtaccgcatg agccccagca accagacgga caagctggcc 240

gccttccccg aggaccgcag ccagcccggc caggactgcc gcttccgtgt cacacaactg 300
cccaacgggc gtgacttcca catgagcgtg gtcagggccc ggcgcaatga cagcggcacc 360
tacctctgtg gggccatctc cctggccccc aaggcgcaga tcaaagagag cctgcgggca 420
gagctcaggg tgacagagag aagggcagaa gtgcccacag cccaccccag cccctcaccc 480
aggccagccg gccagttcca aaccctggtg gttggtgtcg tgggcggcct gctgggcagc 540
ctggtgctgc tagtctgggt cctggccgtc atctgctccc gggccgcacg agggacaata 600
ggagccaggc gcaccggcca gcccctgaag gaggacccct cagccgtgcc tgtgttctct 660

gtggactatg gggagctgga tttccagtgg cgagagaaga ccccggagcc ccccgtgccc 720

tgtgtccctg agcagacgga gtatgccacc attgtctttc ctagcggaat gggcacctca 780

tcccccgccc gcaggggctc agctgacggc cctcggagtg cccagccact gaggcctgag 840

gatggacact gctcttggcc cctctga 867

<210> 93

<211> 140

<212> PRT

<213> Homo sapiens

<400> 93

Ile Gln Asn Pro Asp Pro Ala Val Tyr Gln Leu Arg Asp Ser Lys Ser

1 5 10 15

Ser Asp Lys Ser Val Cys Leu Phe Thr Asp Phe Asp Ser Gln Thr Asn

20 25 30

Val Ser Gln Ser Lys Asp Ser Asp Val Tyr Ile Thr Asp Lys Thr Val

35 40 45

Leu Asp Met Arg Ser Met Asp Phe Lys Ser Asn Ser Ala Val Ala Trp

50 55 60

Ser Asn Lys Ser Asp Phe Ala Cys Ala Asn Ala Phe Asn Asn Ser Ile

65 70 75 80

Ile Pro Glu Asp Thr Phe Phe Pro Ser Pro Glu Ser Ser Cys Asp Val

85 90 95

Lys Leu Val Glu Lys Ser Phe Glu Thr Asp Thr Asn Leu Asn Phe Gln

100 105 110

Asn Leu Ser Val Ile Gly Phe Arg Ile Leu Leu Leu Lys Val Ala Gly

115 120 125

Phe Asn Leu Leu Met Thr Leu Arg Leu Trp Ser Ser

130 135 140

<210> 94

<211> 5651

<212> DNA

<213> Homo sapiens

<400> 94

catgctaatc ctccggcaaa cctctgtttc ctcctcaaaa ggcaggaggt cggaaagaat 60

aaacaatgag agtcacatta aaaacacaaa atcctacgga aatactgaag aatgagtctc 120

agcactaagg aaaagcctcc agcagctcct gctttctgag ggtgaaggat agacgctgtg 180

gctctgcatg actcactagc actctatcac ggccatattc tggcagggtc agtggctcca 240

actaacattt gtttggtact ttacagttta ttaaatagat gtttatatgg agaagctctc 300

atttctttct cagaagagcc tggctaggaa ggtggatgag gcaccatatt cattttgcag 360

gtgaaattcc tgagatgtaa ggagctgctg tgacttgctc aaggccttat atcgagtaaa 420

cggtagtgct ggggcttaga cgcaggtgtt ctgatttata gttcaaaacc tctatcaatg 480

agagagcaat ctcctggtaa tgtgatagat ttcccaactt aatgccaaca taccataaac 540

ctcccattct gctaatgccc agcctaagtt ggggagacca ctccagattc caagatgtac 600

agtttgcttt gctgggcctt tttcccatgc ctgcctttac tctgccagag ttatattgct 660

ggggttttga agaagatcct attaaataaa agaataagca gtattattaa gtagccctgc 720

atttcaggtt tccttgagtg gcaggccagg cctggccgtg aacgttcact gaaatcatgg 780

cctcttggcc aagattgata gcttgtgcct gtccctgagt cccagtccat cacgagcagc 840

tggtttctaa gatgctattt cccgtataaa gcatgagacc gtgacttgcc agccccacag 900

agccccgccc ttgtccatca ctggcatctg gactccagcc tgggttgggg caaagaggga 960

aatgagatca tgtcctaacc ctgatcctct tgtcccacag atatccagaa ccctgaccct 1020

gccgtgtacc agctgagaga ctctaaatcc agtgacaagt ctgtctgcct attcaccgat 1080

tttgattctc aaacaaatgt gtcacaaagt aaggattctg atgtgtatat cacagacaaa 1140

actgtgctag acatgaggtc tatggacttc aagagcaaca gtgctgtggc ctggagcaac 1200

aaatctgact ttgcatgtgc aaacgccttc aacaacagca ttattccaga agacaccttc 1260

ttccccagcc caggtaaggg cagctttggt gccttcgcag gctgtttcct tgcttcagga 1320

atggccaggt tctgcccaga gctctggtca atgatgtcta aaactcctct gattggtggt 1380

ctcggcctta tccattgcca ccaaaaccct ctttttacta agaaacagtg agccttgttc 1440

tggcagtcca gagaatgaca cgggaaaaaa gcagatgaag agaaggtggc aggagagggc 1500

acgtggccca gcctcagtct ctccaactga gttcctgcct gcctgccttt gctcagactg 1560

tttgcccctt actgctcttc taggcctcat tctaagcccc ttctccaagt tgcctctcct 1620

tatttctccc tgtctgccaa aaaatctttc ccagctcact aagtcagtct cacgcagtca 1680

ctcattaacc caccaatcac tgattgtgcc ggcacatgaa tgcaccaggt gttgaagtgg 1740

aggaattaaa aagtcagatg aggggtgtgc ccagaggaag caccattcta gttgggggag 1800

cccatctgtc agctgggaaa agtccaaata acttcagatt ggaatgtgtt ttaactcagg 1860

gttgagaaaa cagctacctt caggacaaaa gtcagggaag ggctctctga agaaatgcta 1920

cttgaagata ccagccctac caagggcagg gagaggaccc tatagaggcc tgggacagga 1980

gctcaatgag aaaggagaag agcagcaggc atgagttgaa tgaaggaggc agggccgggt 2040

cacagggcct tctaggccat gagagggtag acagtattct aaggacgcca gaaagctgtt 2100

gatcggcttc aagcagggga gggacaccta atttgctttt ctttttttt ttttttttt 2160

ttttttttt tgagatggag ttttgctctt gttgcccagg ctggagtgca atggtgcatc 2220

ttggctcact gcaacctccg cctcccaggt tcaagtgatt ctcctgcctc agcctcccga 2280

gtagctgaga ttacaggcac ccgccaccat gcctggctaa ttttttgtat ttttagtaga 2340

gacagggttt cactatgttg gccaggctgg tctcgaactc ctgacctcag gtgatccacc 2400

cgcttcagcc tcccaaagtg ctgggattac aggcgtgagc caccacaccc ggcctgcttt 2460

tcttaaagat caatctgagt gctgtacgga gagtgggttg taagccaaga gtagaagcag 2520

aaagggagca gttgcagcag agagatgatg gaggcctggg cagggtggtg gcagggaggt 2580

aaccaacacc attcaggttt caaaggtaga accatgcagg gatgagaaag caaagagggg 2640

atcaaggaag gcagctggat tttggcctga gcagctgagt caatgatagt gccgtttact 2700

aagaagaaac caaggaaaaa atttggggtg cagggatcaa aacttttgg aacatatgaa 2760

agtacgtgtt tatactcttt atggcccttg tcactatgta tgcctcgctg cctccattgg 2820

actctagaat gaagccaggc aagagcaggg tctatgtgtg atggcacatg tggccagggt 2880

catgcaacat gtactttgta caaacagtgt atattgagta aatagaaatg gtgtccagga 2940

gccgaggtat cggtcctgcc agggccaggg gctctcccta gcaggtgctc atatgctgta 3000

agttccctcc agatctctcc acaaggaggc atggaaaggc tgtagttgtt cacctgccca 3060

agaactagga ggtctggggt gggagagtca gcctgctctg gatgctgaaa gaatgtctgt 3120

ttttcctttt agaaagttcc tgtgatgtca agctggtcga gaaaagcttt gaaacaggta 3180

agacaggggt ctagcctggg tttgcacagg attgcggaag tgatgaaccc gcaataaccc 3240

tgcctggatg agggagtggg aagaaattag tagatgtggg aatgaatgat gaggaatgga 3300

aacagcggtt caagacctgc ccagagctgg gtggggtctc tcctgaatcc ctctcaccat 3360

ctctgacttt ccattctaag cactttgagg atgagtttct agcttcaata gaccaaggac 3420

tctctcctag gcctctgtat tcctttcaac agctccactg tcaagagagc cagagagagc 3480

ttctgggtgg cccagctgtg aaatttctga gtcccttagg gatagcccta aacgaaccag 3540

atcatcctga ggacagccaa gaggttttgc cttctttcaa gacaagcaac agtactcaca 3600

taggctgtgg gcaatggtcc tgtctctcaa gaatcccctg ccactcctca cacccaccct 3660

gggcccatat tcatttccat ttgagttgtt cttattgagt catccttcct gtggtagcgg 3720

aactcactaa ggggcccatc tggacccgag gtattgtgat gataaattct gagcacctac 3780

cccatcccca gaagggctca gaaataaaat aagagccaag tctagtcggt gtttcctgtc 3840

ttgaaacaca atactgttgg ccctggaaga atgcacagaa tctgtttgta aggggatatg 3900
cacagaagct gcaagggaca ggaggtgcag gagctgcagg cctcccccac ccagcctgct 3960
ctgccttggg gaaaaccgtg ggtgtgtcct gcaggccatg caggcctggg acatgcaagc 4020
ccataaccgc tgtggcctct tggttttaca gatacgaacc taaactttca aaacctgtca 4080
gtgattgggt tccgaatcct cctcctgaaa gtggccgggt ttaatctgct catgacgctg 4140
cggctgtggt ccagctgagg tgaggggcct tgaagctggg agtggggttt agggacgcgg 4200
gtctctgggt gcatcctaag ctctgagagc aaacctccct gcagggtctt gcttttaagt 4260

ccaaagcctg agcccaccaa actctcctac ttcttcctgt tacaaattcc tcttgtgcaa 4320
taataatggc ctgaaacgct gtaaaatatc ctcatttcag ccgcctcagt tgcacttctc 4380
ccctatgagg taggaagaac agttgtttag aaacgaagaa actgaggccc cacagctaat 4440
gagtggagga agagagacac ttgtgtacac cacatgcctt gtgttgtact tctctcaccg 4500
tgtaacctcc tcatgtcctc tctccccagt acggctctct tagctcagta gaaagaagac 4560
attacactca tattacaccc caatcctggc tagagtctcc gcaccctcct cccccagggt 4620
ccccagtcgt cttgctgaca actgcatcct gttccatcac catcaaaaaa aaactccagg 4680

ctgggtgcgg gggctcacac ctgtaatccc agcactttgg gaggcagagg caggaggagc 4740
acaggagctg gagaccagcc tgggcaacac agggagaccc cgcctctaca aaaagtgaaa 4800
aaattaacca ggtgtggtgc tgcacacctg tagtcccagc tacttaagag gctgagatgg 4860
gaggatcgct tgagccctgg aatgttgagg ctacaatgag ctgtgattgc gtcactgcac 4920
tccagcctgg aagacaaagc aagatcctgt ctcaaataat aaaaaaaata agaactccag 4980
ggtacatttg ctcctagaac tctaccacat agccccaaac agagccatca ccatcacatc 5040
cctaacagtc ctgggtcttc ctcagtgtcc agcctgactt ctgttcttcc tcattccaga 5100

tctgcaagat tgtaagacag cctgtgctcc ctcgctcctt cctctgcatt gcccctcttc 5160
tccctctcca aacagaggga actctcctac ccccaaggag gtgaaagctg ctaccacctc 5220
tgtgcccccc cggcaatgcc accaactgga tcctacccga atttatgatt aagattgctg 5280
aagagctgcc aaacactgct gccacccccc ctgttccctt attgctgctt gtcactgcct 5340
gacattcacg gcagaggcaa ggctgctgca gcctcccctg gctgtgcaca ttccctcctg 5400
ctccccagag actgcctccg ccatcccaca gatgatggat cttcagtggg ttctcttggg 5460
ctctaggtcc tgcagaatgt tgtgaggggt ttatttttt ttaatagtgt tcataaagaa 5520

atacatagta ttcttcttct caagacgtgg ggggaaatta tctcattatc gaggccctgc 5580

tatgctgtgt atctgggcgt gttgtatgtc ctgctgccga tgccttcatt aaaatgattt 5640

ggaagagcag a 5651

<210> 95

<211> 1508

<212> DNA

<213> Homo sapiens

<400> 95

ttttgaaacc cttcaaaggc agagacttgt ccagcctaac ctgcctgctg ctcctagctc 60

ctgaggctca gggcccttgg cttctgtccg ctctgctcag ggccctccag cgtggccact 120

gctcagccat gctcctgctg ctcgtcccag tgctcgaggt gatttttacc ctgggaggaa 180

ccagagccca gtcggtgacc cagcttggca gccacgtctc tgtctctgaa ggagccctgg 240

ttctgctgag gtgcaactac tcatcgtctg ttccaccata tctcttctgg tatgtgcaat 300

accccaacca aggactccag cttctcctga agtacacatc agcggccacc ctggttaaag 360

gcatcaacgg ttttgaggct gaatttaaga agagtgaaac ctccttccac ctgacgaaac 420

cctcagccca tatgagcgac gcggctgagt acttctgtgc tgtgagtgat ctcgaaccga 480

acagcagtgc ttccaagata atctttggat cagggaccag actcagcatc cggccaaata 540

tccagaaccc tgaccctgcc gtgtaccagc tgagagactc taaatccagt gacaagtctg 600

tctgcctatt caccgatttt gattctcaaa caaatgtgtc acaaagtaag gattctgatg 660

tgtatatcac agacaaaact gtgctagaca tgaggtctat ggacttcaag agcaacagtg 720

ctgtggcctg gagcaacaaa tctgactttg catgtgcaaa cgccttcaac aacagcatta 780

ttccagaaga caccttcttc cccagcccag aaagttcctg tgatgtcaag ctggtcgaga 840

aaagctttga aacagatacg aacctaaact ttcaaaacct gtcagtgatt gggttccgaa 900

tcctcctcct gaaagtggcc gggtttaatc tgctcatgac gctgcggctg tggtccagct 960

gagatctgca agattgtaag acagcctgtg ctccctcgct ccttcctctg cattgcccct 1020

cttctccctc tccaaacaga gggaactctc ctacccccaa ggaggtgaaa gctgctacca 1080

cctctgtgcc cccccggtaa tgccaccaac tggatcctac ccgaatttat gattaagatt 1140

gctgaagagc tgccaaacac tgctgccacc ccctctgttc ccttattgct gcttgtcact 1200

gcctgacatt cacggcagag gcaaggctgc tgcagcctcc cctggctgtg cacattccct 1260

cctgctcccc agagactgcc tccgccatcc cacagatgat ggatcttcag tgggttctct 1320

tgggctctag gtcctggaga atgttgtgag gggtttattt ttttttaata gtgttcataa 1380

agaaatacat agtattcttc ttctcaagac gtgggggggaa attatctcat tatcgaggcc 1440

ctgctatgct gtgtgtctgg gcgtgttgta tgtcctgctg ccgatgcctt cattaaaatg 1500

atttggaa 1508

<210> 96

<211> 176

<212> PRT

<213> Homo sapiens

<400> 96

Asp Leu Asn Lys Val Phe Pro Pro Glu Val Ala Val Phe Glu Pro Ser
1 5 10 15

Glu Ala Glu Ile Ser His Thr Gln Lys Ala Thr Leu Val Cys Leu Ala
20 25 30

Thr Gly Phe Phe Pro Asp His Val Glu Leu Ser Trp Trp Val Asn Gly
35 40 45

Lys Glu Val His Ser Gly Val Ser Thr Asp Pro Gln Pro Leu Lys Glu
50 55 60

Gln Pro Ala Leu Asn Asp Ser Arg Tyr Cys Leu Ser Ser Arg Leu Arg
65 70 75 80

Val Ser Ala Thr Phe Trp Gln Asn Pro Arg Asn His Phe Arg Cys Gln
85 90 95

Val Gln Phe Tyr Gly Leu Ser Glu Asn Asp Glu Trp Thr Gln Asp Arg
100 105 110

Ala Lys Pro Val Thr Gln Ile Val Ser Ala Glu Ala Trp Gly Arg Ala
115 120 125

Asp Cys Gly Phe Thr Ser Val Ser Tyr Gln Gln Gly Val Leu Ser Ala
130 135 140

Thr Ile Leu Tyr Glu Ile Leu Leu Gly Lys Ala Thr Leu Tyr Ala Val
145 150 155 160

Leu Val Ser Ala Leu Val Leu Met Ala Met Val Lys Arg Lys Asp Phe
165 170 175

<210> 97

<211> 1151

<212> DNA

<213> Homo sapiens

<400> 97

ctggtctaga atattccaca tctgctctca ctctgccatg gactcctgga ccttctgctg 60

tgtgtccctt tgcatcctgg tagcgaagca tacagatgct ggagttatcc agtcaccccg 120

ccatgaggtg acagagatgg gacaagaagt gactctgaga tgtaaaccaa tttcaggcca 180

caactccctt ttctggtaca gacagaccat gatgcgggga ctggagttgc tcatttactt 240

taacaacaac gttccgatag atgattcagg gatgcccgag gatcgattct cagctaagat 300

gcctaatgca tcattctcca ctctgaagat ccagccctca gaacccaggg actcagctgt 360

gtacttctgt gccagcagtt tctcgacctg ttcggctaac tatggctaca ccttcggttc 420

ggggaccagg ttaaccgttg tagaggacct gaacaaggtg ttcccacccg aggtcgctgt 480

gtttgagcca tcagaagcag agatctccca cacccaaaag gccacactgg tgtgcctggc 540

cacaggcttc ttccccgacc acgtggagct gagctggtgg gtgaatggga aggaggtgca 600

cagtggggtc agcacagacc cgcagccccc caaggagcag cccgccctca atgactccag 660

atactgcctg agcagccgcc tgagggtctc ggccaccttc tggcagaacc cccgcaacca 720

cttccgctgt caagtccagt tctacgggct ctcggagaat gacgagtgga cccaggatag 780

ggccaaaccc gtcacccaga tcgtcagcgc cgaggcctgg ggtagagcag actgtggctt 840

tacctcggtg tcctaccagc aaggggtcct gtctgccacc atcctctatg agatcctgct 900

agggaaggcc accctgtatg ctgtgctggt cagcgccctt gtgttgatgg ccatggtcaa 960

gagaaaggat ttctgaaggc agccctggaa gtggagttag gagcttctaa cccgtcatgg 1020

ttcaatacac attcttcttt tgccagcgct tctgaagagc tgctctcacc tctctgcatc 1080

ccaatagata tccccctatg tgcatgcaca cctgcacact cacggctgaa atctccctaa 1140

cccaggggga c 1151

<210> 98

<211> 178

<212> PRT

<213> Homo sapiens

<400> 98

Asp Leu Lys Asn Val Phe Pro Pro Lys Val Ala Val Phe Glu Pro Ser

1 5 10 15

Glu Ala Glu Ile Ser His Thr Gln Lys Ala Thr Leu Val Cys Leu Ala

20 25 30

Thr Gly Phe Tyr Pro Asp His Val Glu Leu Ser Trp Trp Val Asn Gly

35 40 45

Lys Glu Val His Ser Gly Val Ser Thr Asp Pro Gln Pro Leu Lys Glu

50 55 60

Gln Pro Ala Leu Asn Asp Ser Arg Tyr Cys Leu Ser Ser Arg Leu Arg

65 70 75 80

Val Ser Ala Thr Phe Trp Gln Asn Pro Arg Asn His Phe Arg Cys Gln

85 90 95

Val Gln Phe Tyr Gly Leu Ser Glu Asn Asp Glu Trp Thr Gln Asp Arg

100 105 110

Ala Lys Pro Val Thr Gln Ile Val Ser Ala Glu Ala Trp Gly Arg Ala

115 120 125

Asp Cys Gly Phe Thr Ser Glu Ser Tyr Gln Gln Gly Val Leu Ser Ala

130 135 140

Thr Ile Leu Tyr Glu Ile Leu Leu Gly Lys Ala Thr Leu Tyr Ala Val

145 150 155 160

Leu Val Ser Ala Leu Val Leu Met Ala Met Val Lys Arg Lys Asp Ser

165 170 175

Arg Gly

<210> 99

<211> 1489

<212> DNA

<213> Homo sapiens

<400> 99

aggacctgaa aaacgtgttc ccacccgagg tcgctgtgtt tgagccatca gaagcagaga 60

tctcccacac ccaaaaggcc acactggtat gcctggccac aggcttctac cccgaccacg 120

tggagctgag ctggtgggtg aatgggaagg aggtgcacag tggggtcagc acagacccgc 180

agcccctcaa ggagcagccc gccctcaatg actccagata ctgcctgagc agccgcctga 240

gggtctcggc caccttctgg cagaaccccc gcaaccactt ccgctgtcaa gtccagttct 300

acgggctctc ggagaatgac gagtggaccc aggatagggc caaacccgtc acccagatcg 360

tcagcgccga ggcctggggt agagcaggtg agtggggcct ggggagatgc ctggaggaga 420

ttaggtgaga ccagctacca gggaaaatgg aaagatccag gtagcggaca agactagatc 480

cagaagaaag ccagagtgga caaggtggga tgatcaaggt tcacagggtc agcaaagcac 540

ggtgtgcact tcccccacca agaagcatag aggctgaatg gagcacctca agctcattct 600

tccttcagat cctgacacct tagagctaag ctttcaagtc tccctgagga ccagccatac 660

agctcagcat ctgagtggtg tgcatcccat tctcttctgg ggtcctggtt tcctaagatc 720

atagtgacca cttcgctggc actggagcag catgagggag acagaaccag ggctatcaaa 780

ggaggctgac tttgtactat ctgatatgca tgtgtttgtg gcctgtgagt ctgtgatgta 840

aggctcaatg tccttacaaa gcagcattct ctcatccatt tttcttcccc tgttttcttt 900

cagactgtgg cttcacctcc ggtaagtgag tctctccttt ttctctctat ctttcgccgt 960

ctctgctctc gaaccagggc atggagaatc cacggacaca ggggcgtgag ggaggccaga 1020

gccacctgtg cacaggtgcc tacatgctct gttcttgtca acagagtctt accagcaagg 1080

ggtcctgtct gccaccatcc tctatgagat cttgctaggg aaggccacct tgtatgccgt 1140

gctggtcagt gccctcgtgc tgatggccat ggtaaggagg agggtgggat agggcagatg 1200

atgggggcag gggatggaac atcacacatg ggcataaagg aatctcagag ccagagcaca 1260

gcctaatata tcctatcacc tcaatgaaac cataatgaag ccagactggg gagaaaatgc 1320

agggaatatc acagaatgca tcatgggagg atggagacaa ccagcgagcc ctactcaaat 1380

taggcctcag agcccgcctc ccctgcccta ctcctgctgt gccatagccc ctgaaaccct 1440

gaaaatgttc tctcttccac aggtcaagag aaaggattcc agaggctag 1489

<210> 100

<211> 2002

<212> PRT

<213> Homo sapiens

<400> 100

Met Glu Gln Asp Arg Thr Asn His Val Glu Gly Asn Arg Leu Ser Pro

1 5 10 15

Phe Leu Ile Pro Ser Pro Pro Ile Cys Gln Thr Glu Pro Leu Ala Thr

20 25 30

Lys Leu Gln Asn Gly Ser Pro Leu Pro Glu Arg Ala His Pro Glu Val

35 40 45

Asn Gly Asp Thr Lys Trp His Ser Phe Lys Ser Tyr Tyr Gly Ile Pro

50 55 60

Cys Met Lys Gly Ser Gln Asn Ser Arg Val Ser Pro Asp Phe Thr Gln

65 70 75 80

Glu Ser Arg Gly Tyr Ser Lys Cys Leu Gln Asn Gly Gly Ile Lys Arg

85 90 95

Thr Val Ser Glu Pro Ser Leu Ser Gly Leu Leu Gln Ile Lys Lys Leu

100 105 110

Lys Gln Asp Gln Lys Ala Asn Gly Glu Arg Arg Asn Phe Gly Val Ser

115 120 125

Gln Glu Arg Asn Pro Gly Glu Ser Ser Gln Pro Asn Val Ser Asp Leu

130 135 140

Ser Asp Lys Lys Glu Ser Val Ser Ser Val Ala Gln Glu Asn Ala Val

145 150 155 160

Lys Asp Phe Thr Ser Phe Ser Thr His Asn Cys Ser Gly Pro Glu Asn

165 170 175

Pro Glu Leu Gln Ile Leu Asn Glu Gln Glu Gly Lys Ser Ala Asn Tyr

180 185 190

His Asp Lys Asn Ile Val Leu Leu Lys Asn Lys Ala Val Leu Met Pro

195 200 205

Asn Gly Ala Thr Val Ser Ala Ser Ser Val Glu His Thr His Gly Glu

210 215 220

Leu Leu Glu Lys Thr Leu Ser Gln Tyr Tyr Pro Asp Cys Val Ser Ile

225 230 235 240

Ala Val Gln Lys Thr Thr Ser His Ile Asn Ala Ile Asn Ser Gln Ala

245 250 255

Thr Asn Glu Leu Ser Cys Glu Ile Thr His Pro Ser His Thr Ser Gly

260 265 270

Gln Ile Asn Ser Ala Gln Thr Ser Asn Ser Glu Leu Pro Pro Lys Pro

275 280 285

Ala Ala Val Val Ser Glu Ala Cys Asp Ala Asp Asp Ala Asp Asn Ala

290 295 300

Ser Lys Leu Ala Ala Met Leu Asn Thr Cys Ser Phe Gln Lys Pro Glu

305 310 315 320

Gln Leu Gln Gln Gln Lys Ser Val Phe Glu Ile Cys Pro Ser Pro Ala

325 330 335

Glu Asn Asn Ile Gln Gly Thr Thr Lys Leu Ala Ser Gly Glu Glu Phe

340 345 350

Cys Ser Gly Ser Ser Ser Asn Leu Gln Ala Pro Gly Gly Ser Ser Glu

355 360 365

Arg Tyr Leu Lys Gln Asn Glu Met Asn Gly Ala Tyr Phe Lys Gln Ser

370 375 380

Ser Val Phe Thr Lys Asp Ser Phe Ser Ala Thr Thr Thr Pro Pro Pro

385 390 395 400

Pro Ser Gln Leu Leu Leu Ser Pro Pro Pro Pro Leu Pro Gln Val Pro

405 410 415

Gln Leu Pro Ser Glu Gly Lys Ser Thr Leu Asn Gly Gly Val Leu Glu

420 425 430

Glu His His His Tyr Pro Asn Gln Ser Asn Thr Thr Leu Leu Arg Glu

435 440 445

Val Lys Ile Glu Gly Lys Pro Glu Ala Pro Pro Ser Gln Ser Pro Asn

450 455 460

Pro Ser Thr His Val Cys Ser Pro Ser Pro Met Leu Ser Glu Arg Pro

465 470 475 480

Gln Asn Asn Cys Val Asn Arg Asn Asp Ile Gln Thr Ala Gly Thr Met

485 490 495

Thr Val Pro Leu Cys Ser Glu Lys Thr Arg Pro Met Ser Glu His Leu

500 505 510

Lys His Asn Pro Pro Ile Phe Gly Ser Ser Gly Glu Leu Gln Asp Asn

515 520 525

Cys Gln Gln Leu Met Arg Asn Lys Glu Gln Glu Ile Leu Lys Gly Arg

530 535 540

Asp Lys Glu Gln Thr Arg Asp Leu Val Pro Pro Thr Gln His Tyr Leu
545 550 555 560
Lys Pro Gly Trp Ile Glu Leu Lys Ala Pro Arg Phe His Gln Ala Glu
565 570 575
Ser His Leu Lys Arg Asn Glu Ala Ser Leu Pro Ser Ile Leu Gln Tyr

580 585 590
Gln Pro Asn Leu Ser Asn Gln Met Thr Ser Lys Gln Tyr Thr Gly Asn
595 600 605
Ser Asn Met Pro Gly Gly Leu Pro Arg Gln Ala Tyr Thr Gln Lys Thr
610 615 620
Thr Gln Leu Glu His Lys Ser Gln Met Tyr Gln Val Glu Met Asn Gln
625 630 635 640
Gly Gln Ser Gln Gly Thr Val Asp Gln His Leu Gln Phe Gln Lys Pro

645 650 655
Ser His Gln Val His Phe Ser Lys Thr Asp His Leu Pro Lys Ala His
660 665 670
Val Gln Ser Leu Cys Gly Thr Arg Phe His Phe Gln Gln Arg Ala Asp
675 680 685
Ser Gln Thr Glu Lys Leu Met Ser Pro Val Leu Lys Gln His Leu Asn
690 695 700
Gln Gln Ala Ser Glu Thr Glu Pro Phe Ser Asn Ser His Leu Leu Gln

705 710 715 720
His Lys Pro His Lys Gln Ala Ala Gln Thr Gln Pro Ser Gln Ser Ser
725 730 735
His Leu Pro Gln Asn Gln Gln Gln Gln Gln Lys Leu Gln Ile Lys Asn
740 745 750
Lys Glu Glu Ile Leu Gln Thr Phe Pro His Pro Gln Ser Asn Asn Asp
755 760 765
Gln Gln Arg Glu Gly Ser Phe Phe Gly Gln Thr Lys Val Glu Glu Cys

770 775 780
Phe His Gly Glu Asn Gln Tyr Ser Lys Ser Ser Glu Phe Glu Thr His

785 790 795 800

Asn Val Gln Met Gly Leu Glu Glu Val Gln Asn Ile Asn Arg Arg Asn

805 810 815

Ser Pro Tyr Ser Gln Thr Met Lys Ser Ser Ala Cys Lys Ile Gln Val

820 825 830

Ser Cys Ser Asn Asn Thr His Leu Val Ser Glu Asn Lys Glu Gln Thr

835 840 845

Thr His Pro Glu Leu Phe Ala Gly Asn Lys Thr Gln Asn Leu His His

850 855 860

Met Gln Tyr Phe Pro Asn Asn Val Ile Pro Lys Gln Asp Leu Leu His

865 870 875 880

Arg Cys Phe Gln Glu Gln Glu Gln Lys Ser Gln Gln Ala Ser Val Leu

885 890 895

Gln Gly Tyr Lys Asn Arg Asn Gln Asp Met Ser Gly Gln Gln Ala Ala

900 905 910

Gln Leu Ala Gln Gln Arg Tyr Leu Ile His Asn His Ala Asn Val Phe

915 920 925

Pro Val Pro Asp Gln Gly Gly Ser His Thr Gln Thr Pro Pro Gln Lys

930 935 940

Asp Thr Gln Lys His Ala Ala Leu Arg Trp His Leu Leu Gln Lys Gln

945 950 955 960

Glu Gln Gln Gln Thr Gln Gln Pro Gln Thr Glu Ser Cys His Ser Gln

965 970 975

Met His Arg Pro Ile Lys Val Glu Pro Gly Cys Lys Pro His Ala Cys

980 985 990

Met His Thr Ala Pro Pro Glu Asn Lys Thr Trp Lys Lys Val Thr Lys

995 1000 1005

Gln Glu Asn Pro Pro Ala Ser Cys Asp Asn Val Gln Gln Lys Ser

1010 1015 1020

Ile Ile Glu Thr Met Glu Gln His Leu Lys Gln Phe His Ala Lys

1025 1030 1035

Ser Leu Phe Asp His Lys Ala Leu Thr Leu Lys Ser Gln Lys Gln

1040 1045 1050

Val Lys Val Glu Met Ser Gly Pro Val Thr Val Leu Thr Arg Gln

1055 1060 1065

Thr Thr Ala Ala Glu Leu Asp Ser His Thr Pro Ala Leu Glu Gln

1070 1075 1080

Gln Thr Thr Ser Ser Glu Lys Thr Pro Thr Lys Arg Thr Ala Ala

1085 1090 1095

Ser Val Leu Asn Asn Phe Ile Glu Ser Pro Ser Lys Leu Leu Asp

1100 1105 1110

Thr Pro Ile Lys Asn Leu Leu Asp Thr Pro Val Lys Thr Gln Tyr

1115 1120 1125

Asp Phe Pro Ser Cys Arg Cys Val Glu Gln Ile Ile Glu Lys Asp

1130 1135 1140

Glu Gly Pro Phe Tyr Thr His Leu Gly Ala Gly Pro Asn Val Ala

1145 1150 1155

Ala Ile Arg Glu Ile Met Glu Glu Arg Phe Gly Gln Lys Gly Lys

1160 1165 1170

Ala Ile Arg Ile Glu Arg Val Ile Tyr Thr Gly Lys Glu Gly Lys

1175 1180 1185

Ser Ser Gln Gly Cys Pro Ile Ala Lys Trp Val Val Arg Arg Ser

1190 1195 1200

Ser Ser Glu Glu Lys Leu Leu Cys Leu Val Arg Glu Arg Ala Gly

1205 1210 1215

His Thr Cys Glu Ala Ala Val Ile Val Ile Leu Ile Leu Val Trp

1220 1225 1230

Glu Gly Ile Pro Leu Ser Leu Ala Asp Lys Leu Tyr Ser Glu Leu

1235 1240 1245

Thr Glu Thr Leu Arg Lys Tyr Gly Thr Leu Thr Asn Arg Arg Cys

1250 1255 1260

Ala Leu Asn Glu Glu Arg Thr Cys Ala Cys Gln Gly Leu Asp Pro

1265 1270 1275

Glu Thr Cys Gly Ala Ser Phe Ser Phe Gly Cys Ser Trp Ser Met

1280 1285 1290

Tyr Tyr Asn Gly Cys Lys Phe Ala Arg Ser Lys Ile Pro Arg Lys

1295 1300 1305

Phe Lys Leu Leu Gly Asp Asp Pro Lys Glu Glu Glu Lys Leu Glu

1310 1315 1320

Ser His Leu Gln Asn Leu Ser Thr Leu Met Ala Pro Thr Tyr Lys

1325 1330 1335

Lys Leu Ala Pro Asp Ala Tyr Asn Asn Gln Ile Glu Tyr Glu His

1340 1345 1350

Arg Ala Pro Glu Cys Arg Leu Gly Leu Lys Glu Gly Arg Pro Phe

1355 1360 1365

Ser Gly Val Thr Ala Cys Leu Asp Phe Cys Ala His Ala His Arg

1370 1375 1380

Asp Leu His Asn Met Gln Asn Gly Ser Thr Leu Val Cys Thr Leu

1385 1390 1395

Thr Arg Glu Asp Asn Arg Glu Phe Gly Gly Lys Pro Glu Asp Glu

1400 1405 1410

Gln Leu His Val Leu Pro Leu Tyr Lys Val Ser Asp Val Asp Glu

1415 1420 1425

Phe Gly Ser Val Glu Ala Gln Glu Glu Lys Lys Arg Ser Gly Ala

1430 1435 1440

Ile Gln Val Leu Ser Ser Phe Arg Arg Lys Val Arg Met Leu Ala

1445 1450 1455

Glu Pro Val Lys Thr Cys Arg Gln Arg Lys Leu Glu Ala Lys Lys

1460 1465 1470

Ala Ala Ala Glu Lys Leu Ser Ser Leu Glu Asn Ser Ser Asn Lys

1475 1480 1485

Asn Glu Lys Glu Lys Ser Ala Pro Ser Arg Thr Lys Gln Thr Glu

1490 1495 1500

Asn Ala Ser Gln Ala Lys Gln Leu Ala Glu Leu Leu Arg Leu Ser

1505 1510 1515

Gly Pro Val Met Gln Gln Ser Gln Gln Pro Gln Pro Leu Gln Lys

1520 1525 1530

Gln Pro Pro Gln Pro Gln Gln Gln Gln Arg Pro Gln Gln Gln Gln

1535 1540 1545

Pro His His Pro Gln Thr Glu Ser Val Asn Ser Tyr Ser Ala Ser

1550 1555 1560

Gly Ser Thr Asn Pro Tyr Met Arg Arg Pro Asn Pro Val Ser Pro

1565 1570 1575

Tyr Pro Asn Ser Ser His Thr Ser Asp Ile Tyr Gly Ser Thr Ser

1580 1585 1590

Pro Met Asn Phe Tyr Ser Thr Ser Ser Gln Ala Ala Gly Ser Tyr

1595 1600 1605

Leu Asn Ser Ser Asn Pro Met Asn Pro Tyr Pro Gly Leu Leu Asn

1610 1615 1620

Gln Asn Thr Gln Tyr Pro Ser Tyr Gln Cys Asn Gly Asn Leu Ser

1625 1630 1635

Val Asp Asn Cys Ser Pro Tyr Leu Gly Ser Tyr Ser Pro Gln Ser

1640 1645 1650

Gln Pro Met Asp Leu Tyr Arg Tyr Pro Ser Gln Asp Pro Leu Ser

1655 1660 1665

Lys Leu Ser Leu Pro Pro Ile His Thr Leu Tyr Gln Pro Arg Phe

1670 1675 1680

Gly Asn Ser Gln Ser Phe Thr Ser Lys Tyr Leu Gly Tyr Gly Asn

1685 1690 1695

Gln Asn Met Gln Gly Asp Gly Phe Ser Ser Cys Thr Ile Arg Pro

1700 1705 1710

Asn Val His His Val Gly Lys Leu Pro Pro Tyr Pro Thr His Glu

1715 1720 1725

Met Asp Gly His Phe Met Gly Ala Thr Ser Arg Leu Pro Pro Asn

1730 1735 1740

Leu Ser Asn Pro Asn Met Asp Tyr Lys Asn Gly Glu His His Ser

1745 1750 1755

Pro Ser His Ile Ile His Asn Tyr Ser Ala Ala Pro Gly Met Phe

1760 1765 1770

Asn Ser Ser Leu His Ala Leu His Leu Gln Asn Lys Glu Asn Asp

1775 1780 1785

Met Leu Ser His Thr Ala Asn Gly Leu Ser Lys Met Leu Pro Ala

1790 1795 1800

Leu Asn His Asp Arg Thr Ala Cys Val Gln Gly Gly Leu His Lys

1805 1810 1815

Leu Ser Asp Ala Asn Gly Gln Glu Lys Gln Pro Leu Ala Leu Val

1820 1825 1830

Gln Gly Val Ala Ser Gly Ala Glu Asp Asn Asp Glu Val Trp Ser

1835 1840 1845

Asp Ser Glu Gln Ser Phe Leu Asp Pro Asp Ile Gly Gly Val Ala

1850 1855 1860

Val Ala Pro Thr His Gly Ser Ile Leu Ile Glu Cys Ala Lys Arg

1865 1870 1875

Glu Leu His Ala Thr Thr Pro Leu Lys Asn Pro Asn Arg Asn His

1880 1885 1890

Pro Thr Arg Ile Ser Leu Val Phe Tyr Gln His Lys Ser Met Asn

1895 1900 1905

Glu Pro Lys His Gly Leu Ala Leu Trp Glu Ala Lys Met Ala Glu

1910 1915 1920

Lys Ala Arg Glu Lys Glu Glu Glu Cys Glu Lys Tyr Gly Pro Asp

1925 1930 1935

Tyr Val Pro Gln Lys Ser His Gly Lys Lys Val Lys Arg Glu Pro

1940 1945 1950

Ala Glu Pro His Glu Thr Ser Glu Pro Thr Tyr Leu Arg Phe Ile

1955 1960 1965

Lys Ser Leu Ala Glu Arg Thr Met Ser Val Thr Thr Asp Ser Thr

1970 1975 1980

Val Thr Thr Ser Pro Tyr Ala Phe Thr Arg Val Thr Gly Pro Tyr

1985 1990 1995

Asn Arg Tyr Ile

2000

<210> 101

<211> 10178

<212> DNA

<213> Homo sapiens

<400> 101

ccgtgccatc ccaacctccc acctcgcccc caaccttcgc gcttgctctg cttcttctcc 60

caggggtgga gacccgccga ggtccccggg gttcccgagg gctgcaccct tccccgcgct 120

cgccagccct ggcccctact ccgcgctggt ccgggcgcac cactcccccc gcgccactgc 180

acggcgtgag ggcagcccag gtctccactg cgcgccccgc tgtacggccc caggtgccgc 240

cggcctttgt gctggacgcc cggtgcgggg ggctaattcc ctgggagccg gggctgaggg 300

ccccagggcg gcggcgcagg ccggggcgga gcgggaggag gccggggcgg agcaggagga 360

ggcccgggcg gaggaggaga gccggcggta gcggcagtgg cagcggcgag agcttgggcg 420

gccgccgccg cctcctcgcg agcgccgcgc gcccgggtcc cgctcgcatg caagtcacgt 480

ccgccccctc ggcgcggccg ccccgagacg ccggccccgc tgagtgatga gaacagacgt 540

caaactgcct tatgaatatt gatgcggagg ctaggctgct ttcgtagaga agcagaagga 600

agcaagatgg ctgcccttta ggatttgtta gaaaggagac ccgactgcaa ctgctggatt 660

gctgcaaggc tgagggacga gaacgaggct ggcaaacatt cagcagcaca ccctctcaag 720

attgtttact tgcctttgct cctgttgagt tacaacgctt ggaagcagga gatgggctca 780

gcagcagcca ataggacatg atccaggaag agcaaattca actagagggc agccttgtgg 840

atggccccga agcaagcctg atggaacagg atagaaccaa ccatgttgag ggcaacagac 900

taagtccatt cctgatacca tcacctccca tttgccagac agaacctctg gctacaaagc 960

tccagaatgg aagcccactg cctgagagag ctcatccaga agtaaatgga gacaccaagt 1020

ggcactcttt caaaagttat tatggaatac cctgtatgaa gggaagccag aatagtcgtg 1080

tgagtcctga ctttacacaa gaaagtagag ggtattccaa gtgtttgcaa aatggaggaa 1140

taaaacgcac agttagtgaa ccttctctct ctgggctcct tcagatcaag aaattgaaac 1200

aagaccaaaa ggctaatgga gaaagacgta acttcggggt aagccaagaa agaaatccag 1260

gtgaaagcag tcaaccaaat gtctccgatt tgagtgataa gaaagaatct gtgagttctg 1320

tagcccaaga aaatgcagtt aaagatttca ccagtttttc aacacataac tgcagtgggc 1380

ctgaaaatcc agagcttcag attctgaatg agcaggaggg gaaaagtgct aattaccatg 1440

acaagaacat tgtattactt aaaaacaagg cagtgctaat gcctaatggt gctacagttt 1500

ctgcctcttc cgtggaacac acacatggtg aactcctgga aaaaacactg tctcaatatt 1560

atccagattg tgtttccatt gcggtgcaga aaaccacatc tcacataaat gccattaaca 1620

gtcaggctac taatgagttg tcctgtgaga tcactcaccc atcgcatacc tcagggcaga 1680

tcaattccgc acagacctct aactctgagc tgcctccaaa gccagctgca gtggtgagtg 1740

aggcctgtga tgctgatgat gctgataatg ccagtaaact agctgcaatg ctaaatacct 1800

gttcctttca gaaaccagaa caactacaac aacaaaaatc agtttttgag atatgcccat 1860

ctcctgcaga aaataacatc cagggaacca caaagctagc gtctggtgaa gaattctgtt 1920

caggttccag cagcaatttg caagctcctg gtggcagctc tgaacggtat ttaaaacaaa 1980

atgaaatgaa tggtgcttac ttcaagcaaa gctcagtgtt cactaaggat tccttttctg 2040

ccactaccac accaccacca ccatcacaat tgcttctttc tccccctcct cctcttccac 2100

aggttcctca gcttccttca gaaggaaaaa gcactctgaa tggtggagtt ttagaagaac 2160

accaccacta ccccaaccaa agtaacacaa cacttttaag ggaagtgaaa atagagggta 2220

aacctgaggc accaccttcc cagagtccta atccatctac acatgtatgc agcccttctc 2280

cgatgctttc tgaaaggcct cagaataatt gtgtgaacag gaatgacata cagactgcag 2340

ggacaatgac tgttccattg tgttctgaga aaacaagacc aatgtcagaa cacctcaagc 2400

ataacccacc aatttttggt agcagtggag agctacagga caactgccag cagttgatga 2460

gaaacaaaga gcaagagatt ctgaagggtc gagacaagga gcaaacacga gatcttgtgc 2520

ccccaacaca gcactatctg aaaccaggat ggattgaatt gaaggcccct cgttttcacc 2580

aagcggaatc ccatctaaaa cgtaatgagg catcactgcc atcaattctt cagtatcaac 2640

ccaatctctc caatcaaatg acctccaaac aatacactgg aaattccaac atgcctgggg 2700

ggctcccaag gcaagcttac acccagaaaa caacacagct ggagcacaag tcacaaatgt 2760

accaagttga aatgaatcaa gggcagtccc aaggtacagt ggaccaacat ctccagttcc 2820

aaaaaccctc acaccaggtg cacttctcca aaacagacca tttaccaaaa gctcatgtgc 2880

agtcactgtg tggcactaga tttcattttc aacaaagagc agattcccaa actgaaaaac 2940

ttatgtcccc agtgttgaaa cagcacttga atcaacaggc ttcagagact gagccatttt 3000

caaactcaca ccttttgcaa cataagcctc ataaacaggc agcacaaaca caaccatccc 3060

agagttcaca tctccctcaa aaccagcaac agcagcaaaa attacaaata aagaataaag 3120

aggaaatact ccagactttt cctcaccccc aaagcaacaa tgatcagcaa agagaaggat 3180

cattctttgg ccagactaaa gtggaagaat gttttcatgg tgaaaatcag tattcaaaat 3240

caagcgagtt cgagactcat aatgtccaaa tgggactgga ggaagtacag aatataaatc 3300

gtagaaattc cccttatagt cagaccatga aatcaagtgc atgcaaaata caggtttctt 3360

gttcaaacaa tacacaccta gtttcagaga ataaagaaca gactacacat cctgaacttt 3420

ttgcaggaaa caagacccaa aacttgcatc acatgcaata ttttccaaat aatgtgatcc 3480

caaagcaaga tcttcttcac aggtgctttc aagaacagga gcagaagtca caacaagctt 3540

cagttctaca gggatataaa aatagaaacc aagatatgtc tggtcaacaa gctgcgcaac 3600

ttgctcagca aaggtacttg atacataacc atgcaaatgt ttttcctgtg cctgaccagg 3660

gaggaagtca cactcagacc cctccccaga aggacactca aaagcatgct gctctaaggt 3720

ggcatctctt acagaagcaa gaacagcagc aaacacagca accccaaact gagtcttgcc 3780

atagtcagat gcacaggcca attaaggtgg aacctggatg caagccacat gcctgtatgc 3840

acacagcacc accagaaaac aaaacatgga aaaaggtaac taagcaagag aatccacctg 3900

caagctgtga taatgtgcag caaaagagca tcattgagac catggagcag catctgaagc 3960

agtttcacgc caagtcgtta tttgaccata aggctcttac tctcaaatca cagaagcaag 4020

taaaagttga aatgtcaggg ccagtcacag ttttgactag acaaaccact gctgcagaac 4080

ttgatagcca caccccagct ttagagcagc aaacaacttc ttcagaaaag acaccaacca 4140

aaagaacagc tgcttctgtt ctcaataatt ttatagagtc accttccaaa ttactagata 4200

ctcctataaa aaatttattg gatacacctg tcaagactca atatgatttc ccatcttgca 4260

gatgtgtaga gcaaattatt gaaaaagatg aaggtccttt ttatacccat ctaggagcag 4320

gtcctaatgt ggcagctatt agagaaatca tggaagaaag gtttggacag aagggtaaag 4380

ctattaggat tgaaagagtc atctatactg gtaaagaagg caaaagttct cagggatgtc 4440

ctattgctaa gtgggtggtt cgcagaagca gcagtgaaga gaagctactg tgtttggtgc 4500

gggagcgagc tggccacacc tgtgaggctg cagtgattgt gattctcatc ctggtgtggg 4560

aaggaatccc gctgtctctg gctgacaaac tctactcgga gcttaccgag acgctgagga 4620

aatacggcac gctcaccaat cgccggtgtg ccttgaatga agagagaact tgcgcctgtc 4680

aggggctgga tccagaaacc tgtggtgcct ccttctcttt tggttgttca tggagcatgt 4740

actacaatgg atgtaagttt gccagaagca agatcccaag gaagtttaag ctgcttgggg 4800

atgacccaaa agaggaagag aaactggagt ctcatttgca aaacctgtcc actcttatgg 4860

caccaacata taagaaactt gcacctgatg catataataa tcagattgaa tatgaacaca 4920

gagcaccaga gtgccgtctg ggtctgaagg aaggccgtcc attctcaggg gtcactgcat 4980
gtttggactt ctgtgctcat gcccacagag acttgcacaa catgcagaat ggcagcacat 5040
tggtatgcac tctcactaga gaagacaatc gagaatttgg aggaaaacct gaggatgagc 5100
agcttcacgt tctgccttta tacaaagtct ctgacgtgga tgagtttggg agtgtggaag 5160
ctcaggagga gaaaaaacgg agtggtgcca ttcaggtact gagttctttt cggcgaaaag 5220
tcaggatgtt agcagagcca gtcaagactt gccgacaaag gaaactagaa gccaagaaag 5280
ctgcagctga aaagctttcc tccctggaga acagctcaaa taaaaatgaa aaggaaaagt 5340

cagccccatc acgtacaaaa caaactgaaa acgcaagcca ggctaaacag ttggcagaac 5400
ttttgcgact ttcaggacca gtcatgcagc agtcccagca gccccagcct ctacagaagc 5460
agccaccaca gccccagcag cagcagagac cccagcagca gcagccacat caccctcaga 5520
cagagtctgt caactcttat tctgcttctg gatccaccaa tccatacatg agacggccca 5580
atccagttag tccttatcca aactcttcac acacttcaga tatctatgga agcaccagcc 5640
ctatgaactt ctattccacc tcatctcaag ctgcaggttc atatttgaat tcttctaatc 5700
ccatgaaccc ttaccctggg cttttgaatc agaataccca atatccatca tatcaatgca 5760

atggaaacct atcagtggac aactgctccc catatctggg ttcctattct ccccagtctc 5820
agccgatgga tctgtatagg tatccaagcc aagaccctct gtctaagctc agtctaccac 5880
ccatccatac actttaccag ccaaggtttg gaaatagcca gagttttaca tctaaatact 5940
taggttatgg aaaccaaaat atgcagggag atggtttcag cagttgtacc attagaccaa 6000
atgtacatca tgtagggaaa ttgcctcctt atcccactca tgagatggat ggccacttca 6060
tgggagccac ctctagatta ccacccaatc tgagcaatcc aaacatggac tataaaaatg 6120
gtgaacatca ttcaccttct cacataatcc ataactacag tgcagctccg ggcatgttca 6180

acagctctct tcatgccctg catctccaaa acaaggagaa tgacatgctt tcccacacag 6240
ctaatgggtt atcaaagatg cttccagctc ttaaccatga tagaactgct tgtgtccaag 6300
gaggcttaca caaattaagt gatgctaatg gtcaggaaaa gcagccattg gcactagtcc 6360
agggtgtggc ttctggtgca gaggacaacg atgaggtctg gtcagacagc gagcagagct 6420
ttctggatcc tgacattggg ggagtggccg tggctccaac tcatgggtca attctcattg 6480
agtgtgcaaa gcgtgagctg catgccacaa ccccctttaaa gaatcccaat aggaatcacc 6540
ccaccaggat ctccctcgtc ttttaccagc ataagagcat gaatgagcca aaacatggct 6600

tggctctttg ggaagccaaa atggctgaaa aagcccgtga gaaagaggaa gagtgtgaaa 6660

agtatggccc agactatgtg cctcagaaat cccatggcaa aaaagtgaaa cgggagcctg 6720

ctgagccaca tgaaacttca gagcccactt acctgcgttt catcaagtct cttgccgaaa 6780

ggaccatgtc cgtgaccaca gactccacag taactacatc tccatatgcc ttcactcggg 6840

tcacagggcc ttacaacaga tatatatgaa gatatatatg atatcacccc cttttgttgg 6900

ttacctcact tgaaaagacc acaaccaacc tgtcagtagt atagttctca tgacgtgggc 6960

agtggggaaa ggtcacagta ttcatgacaa atgtggtggg aaaaacctca gctcaccagc 7020

aacaaaagag gttatcttac catagcactt aattttcact ggctcccaag tggtcacaga 7080

tggcatctag gaaaagacca aagcattcta tgcaaaaaga aggtggggaa gaaagtgttc 7140

cgcaatttac atttttaaac actggttcta ttattggacg agatgatatg taaatgtgat 7200

ccccccccc cgcttacaac tctacacatc tgtgaccact tttaataata tcaagtttgc 7260

atagtcatgg aacacaaatc aaacaagtac tgtagtatta cagtgacagg aatcttaaaa 7320

taccatctgg tgctgaatat atgatgtact gaaatactgg aattatggct ttttgaaatg 7380

cagtttttac tgtaatctta acttttattt atcaaaatag ctacaggaaa catgaatagc 7440

aggaaaacac tgaatttgtt tggatgttct aagaaatggt gctaagaaaa tggtgtcttt 7500

aatagctaaa aatttaatgc ctttatatca tcaagatgct atcagtgtac tccagtgccc 7560

ttgaataata ggggtacctt ttcattcaag tttttatcat aattacctat tcttacacaa 7620

gcttagtttt taaaatgtgg acattttaaa ggcctctgga ttttgctcat ccagtgaagt 7680

ccttgtagga caataaacgt atatatgtac atatatacac aaacatgtat atgtgcacac 7740

acatgtatat gtataaatat tttaaatggt gttttagaag cactttgtct acctaagctt 7800

tgacaacttg aacaatgcta aggtactgag atgtttaaaa aacaagttta ctttcatttt 7860

agaatgcaaa gttgattttt ttaaggaaac aaagaaagct tttaaaatat ttttgctttt 7920

agccatgcat ctgctgatga gcaattgtgt ccatttttaa cacagccagt taaatccacc 7980

atggggctta ctggattcaa gggaatacgt tagtccacaa aacatgtttt ctggtgctca 8040

tctcacatgc tatactgtaa aacagtttta tacaaaattg tatgacaagt tcattgctca 8100

aaaatgtaca gttttaagaa ttttctatta actgcaggta ataattagct gcatgctgca 8160

gactcaacaa agctagttca ctgaagccta tgctatttta tggatcatag gctcttcaga 8220

gaactgaatg gcagtctgcc tttgtgttga taattatgta cattgtgacg ttgtcatttc 8280

ttagcttaag tgtcctcttt aacaagagga ttgagcagac tgatgcctgc ataagatgaa 8340

taaacagggt tagttccatg tgaatctgtc agttaaaaag aaacaaaaac aggcagctgg 8400

tttgctgtgg tggttttaaa tcattaattt gtataaagaa gtgaaagagt tgtatagtaa 8460

attaaattgt aaacaaaact tttttaatgc aatgctttag tattttagta ctgtaaaaaa 8520

attaaatata tacatatata tatatatata tatatatata tatatgagtt tgaagcagaa 8580

ttcacatcat gatggtgcta ctcagcctgc tacaaatata tcataatgtg agctaagaat 8640

tcattaaatg tttgagtgat gttcctactt gtcatatacc tcaacactag tttggcaata 8700

ggatattgaa ctgagagtga aagcattgtg taccatcatt tttttccaag tccttttttt 8760

tattgttaaa aaaaaaagca tacctttttt caatacttga tttcttagca agtataactt 8820

gaacttcaac ctttttgttc taaaaattca gggatatttc agctcatgct ctccctatgc 8880

caacatgtca cctgtgttta tgtaaaattg ttgtaggtta ataaatatat tctttgtcag 8940

ggatttaacc cttttatttt gaatcccttc tattttactt gtacatgtgc tgatgtaact 9000

aaaactaatt ttgtaaatct gttggctctt tttattgtaa agaaaagcat tttaaaagtt 9060

tgaggaatct tttgactgtt tcaagcagga aaaaaaaatt acatgaaaat agaatgcact 9120

gagttgataa agggaaaaat tgtaaggcag gagtttggca agtggctgtt ggccagagac 9180

ttacttgtaa ctctctaaat gaagtttttt tgatcctgta atcactgaag gtacatactc 9240

catgtggact tcccttaaac aggcaaacac ctacaggtat ggtgtgcaac agattgtaca 9300

attacatttt ggcctaaata cattttgct tactagtatt taaaataaat tcttaatcag 9360

aggaggcctt tgggttttat tggtcaaatc tttgtaagct ggcttttgtc tttttaaaaa 9420

atttcttgaa tttgtggttg tgtccaattt gcaaacattt ccaaaaatgt ttgctttgct 9480

tacaaaccac atgattttaa tgttttttgt ataccataat atctagcccc aaacatttga 9540

ttactacatg tgcattggtg attttgatca tccattctta atatttgatt tctgtgtcac 9600

ctactgtcat ttgttaaact gctggccaac aagaacagga agtatagttt ggggggttgg 9660

ggagagttta cataaggaag agaagaaatt gagtggcata ttgtaaatat cagatctata 9720

attgtaaata taaaacctgc ctcagttaga atgaatggaa agcagatcta caatttgcta 9780

atataggaat atcaggttga ctatatagcc atacttgaaa atgcttctga gtggtgtcaa 9840

ctttacttga atgaattttt catcttgatt gacgcacagt gatgtacagt tcacttctga 9900

agctagtggt taacttgtgt aggaaacttt tgcagtttga cactaagata acttctgtgt 9960

gcatttttct atgctttttt aaaaactagt ttcatttcat tttcatgaga tgtttggttt 10020

ataagatctg aggatggtta taaatactgt aagtattgta atgttatgaa tgcaggttat 10080

ttgaaagctg tttattatta tatcattcct gataatgcta tgtgagtgtt tttaataaaa 10140

tttatattta tttaatgcac tctaagtgtt gtcttcct 10178

<210> 102

<211> 567

<212> PRT

<213> Homo sapiens

<400> 102

Met Gly Arg Gly Leu Leu Arg Gly Leu Trp Pro Leu His Ile Val Leu
1 5 10 15

Trp Thr Arg Ile Ala Ser Thr Ile Pro Pro His Val Gln Lys Ser Val
20 25 30

Asn Asn Asp Met Ile Val Thr Asp Asn Asn Gly Ala Val Lys Phe Pro
35 40 45

Gln Leu Cys Lys Phe Cys Asp Val Arg Phe Ser Thr Cys Asp Asn Gln
50 55 60

Lys Ser Cys Met Ser Asn Cys Ser Ile Thr Ser Ile Cys Glu Lys Pro
65 70 75 80

Gln Glu Val Cys Val Ala Val Trp Arg Lys Asn Asp Glu Asn Ile Thr
85 90 95

Leu Glu Thr Val Cys His Asp Pro Lys Leu Pro Tyr His Asp Phe Ile
100 105 110

Leu Glu Asp Ala Ala Ser Pro Lys Cys Ile Met Lys Glu Lys Lys Lys
115 120 125

Pro Gly Glu Thr Phe Phe Met Cys Ser Cys Ser Ser Asp Glu Cys Asn
130 135 140

Asp Asn Ile Ile Phe Ser Glu Glu Tyr Asn Thr Ser Asn Pro Asp Leu
145 150 155 160

Leu Leu Val Ile Phe Gln Val Thr Gly Ile Ser Leu Leu Pro Pro Leu
165 170 175

Gly Val Ala Ile Ser Val Ile Ile Ile Phe Tyr Cys Tyr Arg Val Asn
180 185 190

Arg Gln Gln Lys Leu Ser Ser Thr Trp Glu Thr Gly Lys Thr Arg Lys
195 200 205

Leu Met Glu Phe Ser Glu His Cys Ala Ile Ile Leu Glu Asp Asp Arg
210 215 220

Ser Asp Ile Ser Ser Thr Cys Ala Asn Asn Ile Asn His Asn Thr Glu
225 230 235 240
Leu Leu Pro Ile Glu Leu Asp Thr Leu Val Gly Lys Gly Arg Phe Ala
245 250 255
Glu Val Tyr Lys Ala Lys Leu Lys Gln Asn Thr Ser Glu Gln Phe Glu
260 265 270
Thr Val Ala Val Lys Ile Phe Pro Tyr Glu Glu Tyr Ala Ser Trp Lys

275 280 285
Thr Glu Lys Asp Ile Phe Ser Asp Ile Asn Leu Lys His Glu Asn Ile
290 295 300
Leu Gln Phe Leu Thr Ala Glu Glu Arg Lys Thr Glu Leu Gly Lys Gln
305 310 315 320
Tyr Trp Leu Ile Thr Ala Phe His Ala Lys Gly Asn Leu Gln Glu Tyr
325 330 335
Leu Thr Arg His Val Ile Ser Trp Glu Asp Leu Arg Lys Leu Gly Ser

340 345 350
Ser Leu Ala Arg Gly Ile Ala His Leu His Ser Asp His Thr Pro Cys
355 360 365
Gly Arg Pro Lys Met Pro Ile Val His Arg Asp Leu Lys Ser Ser Asn
370 375 380
Ile Leu Val Lys Asn Asp Leu Thr Cys Cys Leu Cys Asp Phe Gly Leu
385 390 395 400
Ser Leu Arg Leu Asp Pro Thr Leu Ser Val Asp Asp Leu Ala Asn Ser

405 410 415
Gly Gln Val Gly Thr Ala Arg Tyr Met Ala Pro Glu Val Leu Glu Ser
420 425 430
Arg Met Asn Leu Glu Asn Val Glu Ser Phe Lys Gln Thr Asp Val Tyr
435 440 445
Ser Met Ala Leu Val Leu Trp Glu Met Thr Ser Arg Cys Asn Ala Val
450 455 460
Gly Glu Val Lys Asp Tyr Glu Pro Pro Phe Gly Ser Lys Val Arg Glu

465 470 475 480

His Pro Cys Val Glu Ser Met Lys Asp Asn Val Leu Arg Asp Arg Gly

485 490 495

Arg Pro Glu Ile Pro Ser Phe Trp Leu Asn His Gln Gly Ile Gln Met

500 505 510

Val Cys Glu Thr Leu Thr Glu Cys Trp Asp His Asp Pro Glu Ala Arg

515 520 525

Leu Thr Ala Gln Cys Val Ala Glu Arg Phe Ser Glu Leu Glu His Leu

530 535 540

Asp Arg Leu Ser Gly Arg Ser Cys Ser Glu Glu Lys Ile Pro Glu Asp

545 550 555 560

Gly Ser Leu Asn Thr Thr Lys

565

<210> 103

<211> 2090

<212> DNA

<213> Homo sapiens

<400> 103

gttggcgagg agtttcctgt ttcccccgca gcgctgagtt gaagttgagt gagtcactcg 60

cgcgcacgga gcgacgacac ccccgcgcgt gcacccgctc gggacaggag ccggactcct 120

gtgcagcttc cctcggccgc cgggggcctc cccgcgcctc gccggcctcc aggcccctcc 180

tggctggcga gcgggcgcca catctggccc gcacatctgc gctgccggcc cggcgcgggg 240

tccggagagg gcgcggcgcg gagcgcagcc aggggtccgg gaaggcgccg tccgtgcgct 300

gggggctcgg tctatgacga gcagcggggt ctgccatggg tcgggggctg ctcaggggcc 360

tgtggccgct gcacatcgtc ctgtggacgc gtatcgccag cacgatccca ccgcacgttc 420

agaagtcggt taataacgac atgatagtca ctgacaacaa cggtgcagtc aagtttccac 480

aactgtgtaa attttgtgat gtgagatttt ccacctgtga caaccagaaa tcctgcatga 540

gcaactgcag catcacctcc atctgtgaga agccacagga agtctgtgtg gctgtatgga 600

gaaagaatga cgagaacata acactagaga cagtttgcca tgaccccaag ctcccctacc 660

atgactttat tctggaagat gctgcttctc caaagtgcat tatgaaggaa aaaaaaaagc 720

ctggtgagac tttcttcatg tgttcctgta gctctgatga gtgcaatgac aacatcatct 780

tctcagaaga atataacacc agcaatcctg acttgttgct agtcatattt caagtgacag 840

gcatcagcct cctgccacca ctgggagttg ccatatctgt catcatcatc ttctactgct 900

accgcgttaa ccggcagcag aagctgagtt caacctggga aaccggcaag acgcggaagc 960

tcatggagtt cagcgagcac tgtgccatca tcctggaaga tgaccgctct gacatcagct 1020

ccacgtgtgc caacaacatc aaccacaaca cagagctgct gcccattgag ctggacaccc 1080

tggtggggaa aggtcgcttt gctgaggtct ataaggccaa gctgaagcag aacacttcag 1140

agcagtttga gacagtggca gtcaagatct ttccctatga ggagtatgcc tcttggaaga 1200

cagagaagga catcttctca gacatcaatc tgaagcatga gaacatactc cagttcctga 1260

cggctgagga gcggaagacg gagttgggga aacaatactg gctgatcacc gccttccacg 1320

ccaagggcaa cctacaggag tacctgacgc ggcatgtcat cagctgggag gacctgcgca 1380

agctgggcag ctccctcgcc cgggggattg ctcacctcca cagtgatcac actccatgtg 1440

ggaggcccaa gatgcccatc gtgcacaggg acctcaagag ctccaatatc ctcgtgaaga 1500

acgacctaac ctgctgcctg tgtgactttg ggctttccct gcgtctggac cctactctgt 1560

ctgtggatga cctggctaac agtgggcagg tgggaactgc aagatacatg gctccagaag 1620

tcctagaatc caggatgaat ttggagaatg ctgagtcctt caagcagacc gatgtctact 1680

ccatggctct ggtgctctgg gaaatgacat ctcgctgtaa tgcagtggga gaagtaaaag 1740

attatgagcc tccatttggt tccaaggtgc gggagcaccc ctgtgtcgaa agcatgaagg 1800

acaacgtgtt gagagatcga gggcgaccag aaattcccag cttctggctc aaccaccagg 1860

gcatccagat ggtgtgtgag acgttgactg agtgctggga ccacgaccca gaggcccgtc 1920

tcacagccca gtgtgtggca gaacgcttca gtgagctgga gcatctggac aggctctcgg 1980

ggaggagctg ctcggaggag aagattcctg aagacggctc cctaaacact accaaatagc 2040

tcttatgggg caggctgggc atgtccaaag aggctgcccc tctcaccaaa 2090

<210> 104

<211> 244

<212> PRT

<213> Homo sapiens

<400> 104

Met Arg Trp Cys Leu Leu Leu Ile Trp Ala Gln Gly Leu Arg Gln Ala

1 5 10 15

Pro Leu Ala Ser Gly Met Met Thr Gly Thr Ile Glu Thr Thr Gly Asn

20 25 30

Ile Ser Ala Glu Lys Gly Gly Ser Ile Ile Leu Gln Cys His Leu Ser

35 40 45

Ser Thr Thr Ala Gln Val Thr Gln Val Asn Trp Glu Gln Gln Asp Gln

50 55 60

Leu Leu Ala Ile Cys Asn Ala Asp Leu Gly Trp His Ile Ser Pro Ser

65 70 75 80

Phe Lys Asp Arg Val Ala Pro Gly Pro Gly Leu Gly Leu Thr Leu Gln

85 90 95

Ser Leu Thr Val Asn Asp Thr Gly Glu Tyr Phe Cys Ile Tyr His Thr

100 105 110

Tyr Pro Asp Gly Thr Tyr Thr Gly Arg Ile Phe Leu Glu Val Leu Glu

115 120 125

Ser Ser Val Ala Glu His Gly Ala Arg Phe Gln Ile Pro Leu Leu Gly

130 135 140

Ala Met Ala Ala Thr Leu Val Val Ile Cys Thr Ala Val Ile Val Val

145 150 155 160

Val Ala Leu Thr Arg Lys Lys Lys Ala Leu Arg Ile His Ser Val Glu

165 170 175

Gly Asp Leu Arg Arg Lys Ser Ala Gly Gln Glu Glu Trp Ser Pro Ser

180 185 190

Ala Pro Ser Pro Pro Gly Ser Cys Val Gln Ala Glu Ala Ala Pro Ala

195 200 205

Gly Leu Cys Gly Glu Gln Arg Gly Glu Asp Cys Ala Glu Leu His Asp

210 215 220

Tyr Phe Asn Val Leu Ser Tyr Arg Ser Leu Gly Asn Cys Ser Phe Phe

225 230 235 240

Thr Glu Thr Gly

<210> 105

<211> 831

<212> DNA

<213> Homo sapiens

<400> 105

cgtcctatct gcagtcggct actttcagtg gcagaagagg ccacatctgc ttcctgtagg 60

ccctctgggc agaagcatgc gctggtgtct cctcctgatc tgggcccagg ggctgaggca 120

ggctcccctc gcctcaggaa tgatgacagg cacaatagaa acaacgggga acatttctgc 180

agagaaaggt ggctctatca tcttacaatg tcacctctcc tccaccacgg cacaagtgac 240

ccaggtcaac tgggagcagc aggaccagct tctggccatt tgtaatgctg acttggggtg 300

gcacatctcc ccatccttca aggatcgagt ggccccaggt cccggcctgg gcctcaccct 360

ccagtcgctg accgtgaacg atacagggga gtacttctgc atctatcaca cctaccctga 420

tgggacgtac actgggagaa tcttcctgga ggtcctagaa agctcagtgg ctgagcacgg 480

tgccaggttc cagattccat tgcttggagc catggccgcg acgctggtgg tcatctgcac 540

agcagtcatc gtggtggtcg cgttgactag aaagaagaaa gccctcagaa tccattctgt 600

ggaaggtgac ctcaggagaa aatcagctgg acaggaggaa tggagcccca gtgctccctc 660

acccccagga agctgtgtcc aggcagaagc tgcacctgct gggctctgtg gagagcagcg 720

gggagaggac tgtgccgagc tgcatgacta cttcaatgtc ctgagttaca gaagcctggg 780

taactgcagc ttcttcacag agactggtta gcaaccagag gcatcttctg g 831

<210> 106

<211> 84

<212> PRT

<213> Bacillus phage PBS2

<400> 106

Met Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu
1 5 10 15

Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val
20 25 30

Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp
35 40 45

Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu
50 55 60

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys
65 70 75 80

Ile Lys Met Leu

<210> 107

<211> 619

<212> PRT

<213> Homo sapiens

<400> 107

Met Pro Asp Pro Ala Ala His Leu Pro Phe Phe Tyr Gly Ser Ile Ser
1 5 10 15

Arg Ala Glu Ala Glu Glu His Leu Lys Leu Ala Gly Met Ala Asp Gly
20 25 30

Leu Phe Leu Leu Arg Gln Cys Leu Arg Ser Leu Gly Gly Tyr Val Leu
35 40 45

Ser Leu Val His Asp Val Arg Phe His His Phe Pro Ile Glu Arg Gln
50 55 60

Leu Asn Gly Thr Tyr Ala Ile Ala Gly Gly Lys Ala His Cys Gly Pro
65 70 75 80

Ala Glu Leu Cys Glu Phe Tyr Ser Arg Asp Pro Asp Gly Leu Pro Cys
85 90 95

Asn Leu Arg Lys Pro Cys Asn Arg Pro Ser Gly Leu Glu Pro Gln Pro
100 105 110

Gly Val Phe Asp Cys Leu Arg Asp Ala Met Val Arg Asp Tyr Val Arg
115 120 125

Gln Thr Trp Lys Leu Glu Gly Glu Ala Leu Glu Gln Ala Ile Ile Ser
130 135 140

Gln Ala Pro Gln Val Glu Lys Leu Ile Ala Thr Thr Ala His Glu Arg
145 150 155 160

Met Pro Trp Tyr His Ser Ser Leu Thr Arg Glu Glu Ala Glu Arg Lys
165 170 175

Leu Tyr Ser Gly Ala Gln Thr Asp Gly Lys Phe Leu Leu Arg Pro Arg
180 185 190

Lys Glu Gln Gly Thr Tyr Ala Leu Ser Leu Ile Tyr Gly Lys Thr Val
195 200 205

Tyr His Tyr Leu Ile Ser Gln Asp Lys Ala Gly Lys Tyr Cys Ile Pro

210 215 220

Glu Gly Thr Lys Phe Asp Thr Leu Trp Gln Leu Val Glu Tyr Leu Lys

225 230 235 240

Leu Lys Ala Asp Gly Leu Ile Tyr Cys Leu Lys Glu Ala Cys Pro Asn

245 250 255

Ser Ser Ala Ser Asn Ala Ser Gly Ala Ala Ala Pro Thr Leu Pro Ala

260 265 270

His Pro Ser Thr Leu Thr His Pro Gln Arg Arg Ile Asp Thr Leu Asn

275 280 285

Ser Asp Gly Tyr Thr Pro Glu Pro Ala Arg Ile Thr Ser Pro Asp Lys

290 295 300

Pro Arg Pro Met Pro Met Asp Thr Ser Val Tyr Glu Ser Pro Tyr Ser

305 310 315 320

Asp Pro Glu Glu Leu Lys Asp Lys Lys Leu Phe Leu Lys Arg Asp Asn

325 330 335

Leu Leu Ile Ala Asp Ile Glu Leu Gly Cys Gly Asn Phe Gly Ser Val

340 345 350

Arg Gln Gly Val Tyr Arg Met Arg Lys Lys Gln Ile Asp Val Ala Ile

355 360 365

Lys Val Leu Lys Gln Gly Thr Glu Lys Ala Asp Thr Glu Glu Met Met

370 375 380

Arg Glu Ala Gln Ile Met His Gln Leu Asp Asn Pro Tyr Ile Val Arg

385 390 395 400

Leu Ile Gly Val Cys Gln Ala Glu Ala Leu Met Leu Val Met Glu Met

405 410 415

Ala Gly Gly Gly Pro Leu His Lys Phe Leu Val Gly Lys Arg Glu Glu

420 425 430

Ile Pro Val Ser Asn Val Ala Glu Leu Leu His Gln Val Ser Met Gly

435 440 445

Met Lys Tyr Leu Glu Glu Lys Asn Phe Val His Arg Asp Leu Ala Ala

450 455 460

Arg Asn Val Leu Leu Val Asn Arg His Tyr Ala Lys Ile Ser Asp Phe
465 470 475 480
Gly Leu Ser Lys Ala Leu Gly Ala Asp Asp Ser Tyr Tyr Thr Ala Arg
485 490 495
Ser Ala Gly Lys Trp Pro Leu Lys Trp Tyr Ala Pro Glu Cys Ile Asn

500 505 510
Phe Arg Lys Phe Ser Ser Arg Ser Asp Val Trp Ser Tyr Gly Val Thr
515 520 525
Met Trp Glu Ala Leu Ser Tyr Gly Gln Lys Pro Tyr Lys Lys Met Lys
530 535 540
Gly Pro Glu Val Met Ala Phe Ile Glu Gln Gly Lys Arg Met Glu Cys
545 550 555 560
Pro Pro Glu Cys Pro Pro Glu Leu Tyr Ala Leu Met Ser Asp Cys Trp

565 570 575
Ile Tyr Lys Trp Glu Asp Arg Pro Asp Phe Leu Thr Val Glu Gln Arg
580 585 590
Met Arg Ala Cys Tyr Tyr Ser Leu Ala Ser Lys Val Glu Gly Pro Pro
595 600 605
Gly Ser Thr Gln Lys Ala Glu Ala Ala Cys Ala
610 615

<210> 108
<211> 2468
<212> DNA
<213> Homo sapiens
<400> 108
gcttgccgga gctcagcaga caccaggcct tccgggcagg cctggcccac cgtgggcctc 60

agagctgctg ctggggcatt cagaaccggc tctccattgg cattgggacc agagaccccg 120
caagtggcct gtttgcctgg acatccacct gtacgtcccc aggtttcggg aggcccaggg 180
gcgatgccag accccgcggc gcacctgccc ttcttctacg gcagcatctc gcgtgccgag 240
gccgaggagc acctgaagct ggcgggcatg gcggacgggc tcttcctgct gcgccagtgc 300
ctgcgctcgc tgggcggcta tgtgctgtcg ctcgtgcacg atgtgcgctt ccaccacttt 360
cccatcgagc gccagctcaa cggcacctac gccattgccg gcggcaaagc gcactgtgga 420

ccggcagagc tctgcgagtt ctactcgcgc gaccccgacg ggctgccctg caacctgcgc 480

aagccgtgca accggccgtc gggcctcgag ccgcagccgg gggtcttcga ctgcctgcga 540
gacgccatgg tgcgtgacta cgtgcgccag acgtggaagc tggagggcga ggccctggag 600
caggccatca tcagccaggc cccgcaggtg gagaagctca ttgctacgac ggcccacgag 660
cggatgccct ggtaccacag cagcctgacg cgtgaggagg ccgagcgcaa actttactct 720
ggggcgcaga ccgacggcaa gttcctgctg aggccgcgga aggagcaggg cacatacgcc 780
ctgtccctca tctatgggaa gacggtgtac cactacctca tcagccaaga caaggcgggc 840
aagtactgca ttcccgaggg caccaagttt gacacgctct ggcagctggt ggagtatctg 900

aagctgaagg cggacgggct catctactgc ctgaaggagg cctgccccaa cagcagtgcc 960
agcaacgcct caggggctgc tgctcccaca ctcccagccc acccatccac gttgactcat 1020
cctcagagac gaatcgacac cctcaactca gatggataca cccctgagcc agcacgcata 1080
acgtccccag acaaaccgcg gccgatgccc atggacacga gcgtgtatga gagcccctac 1140
agcgacccag aggagctcaa ggacaagaag ctcttcctga agcgcgataa cctcctcata 1200
gctgacattg aacttggctg cggcaacttt ggctcagtgc gccagggcgt gtaccgcatg 1260
cgcaagaagc agatcgacgt ggccatcaag gtgctgaagc agggcacgga gaaggcagac 1320

acggaagaga tgatgcgcga ggcgcagatc atgcaccagc tggacaaccc ctacatcgtg 1380
cggctcattg gcgtctgcca ggccgaggcc ctcatgctgg tcatggagat ggctgggggc 1440
gggccgctgc acaagttcct ggtcggcaag agggaggaga tccctgtgag caatgtggcc 1500
gagctgctgc accaggtgtc catggggatg aagtacctgg aggagaagaa ctttgtgcac 1560
cgtgacctgg cggcccgcaa cgtcctgctg gttaaccggc actacgccaa gatcagcgac 1620
tttggcctct ccaaagcact gggtgccgac gacagctact acactgcccg ctcagcaggg 1680
aagtggccgc tcaagtggta cgcacccgaa tgcatcaact tccgcaagtt ctccagccgc 1740

agcgatgtct ggagctatgg ggtcaccatg tgggaggcct tgtcctacgg ccagaagccc 1800
tacaagaaga tgaaagggcc ggaggtcatg gccttcatcg agcagggcaa gcggatggaa 1860
tgcccaccag agtgtccacc cgaactgtac gcactcatga gtgactgctg gatctacaag 1920
tgggaggatc gccccgactt cctgaccgtg gagcagcgca tgcgagcctg ttactacagc 1980
ctggccagca aggtggaagg gcccccaggc agcacacaga aggctgaggc tgcctgtgcc 2040
tgagctcccg ctgcccaggg gagccctcca caccggctct tccccaccct cagccccacc 2100
ccaggtcctg cagtctggct gagccctgct tggttgtctc cacacacagc tgggctgtgg 2160

taggggtgt ctcaggccac accggccttg cattgcctgc ctggcccct gtcctctctg 2220

gctggggagc agggaggtcc gggagggtgc ggctgtgcag cctgtcctgg gctggtggct 2280

cccggagggc cctgagctga gggcattgct tacacggatg ccttcccctg ggccctgaca 2340

ttggagcctg ggcatcctca ggtggtcagg cgtagatcac cagaataaac ccagcttccc 2400

tcttgaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2460

aaaaaaaa 2468

<210> 109

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 109

Gly Gly Gly Gly Ser

1 5

<210> 110

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 110

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

1 5 10 15

<210> 111

<211> 493

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 111

Met Ala Leu Pro Val Thr Ala Leu Leu Leu Pro Leu Ala Leu Leu Leu

1 5 10 15

His Ala Ala Arg Pro Asp Ile Val Leu Thr Gln Ser Pro Pro Ser Leu
20 25 30
Ala Met Ser Leu Gly Lys Arg Ala Thr Ile Ser Cys Arg Ala Ser Glu
35 40 45
Ser Val Thr Ile Leu Gly Ser His Leu Ile His Trp Tyr Gln Gln Lys
50 55 60

Pro Gly Gln Pro Pro Thr Leu Leu Ile Gln Leu Ala Ser Asn Val Gln
65 70 75 80
Thr Gly Val Pro Ala Arg Phe Ser Gly Ser Gly Ser Arg Thr Asp Phe
85 90 95
Thr Leu Thr Ile Asp Pro Val Glu Glu Asp Asp Val Ala Val Tyr Tyr
100 105 110
Cys Leu Gln Ser Arg Thr Ile Pro Arg Thr Phe Gly Gly Gly Thr Lys
115 120 125

Leu Glu Ile Lys Gly Ser Thr Ser Gly Ser Gly Lys Pro Gly Ser Gly
130 135 140
Glu Gly Ser Thr Lys Gly Gln Ile Gln Leu Val Gln Ser Gly Pro Glu
145 150 155 160
Leu Lys Lys Pro Gly Glu Thr Val Lys Ile Ser Cys Lys Ala Ser Gly
165 170 175
Tyr Thr Phe Thr Asp Tyr Ser Ile Asn Trp Val Lys Arg Ala Pro Gly
180 185 190

Lys Gly Leu Lys Trp Met Gly Trp Ile Asn Thr Glu Thr Arg Glu Pro
195 200 205
Ala Tyr Ala Tyr Asp Phe Arg Gly Arg Phe Ala Phe Ser Leu Glu Thr
210 215 220
Ser Ala Ser Thr Ala Tyr Leu Gln Ile Asn Asn Leu Lys Tyr Glu Asp
225 230 235 240
Thr Ala Thr Tyr Phe Cys Ala Leu Asp Tyr Ser Tyr Ala Met Asp Tyr
245 250 255

Trp Gly Gln Gly Thr Ser Val Thr Val Ser Ser Ala Ala Ala Thr Thr

260 265 270

Thr Pro Ala Pro Arg Pro Pro Thr Pro Ala Pro Thr Ile Ala Ser Gln

275 280 285

Pro Leu Ser Leu Arg Pro Glu Ala Cys Arg Pro Ala Ala Gly Gly Ala

290 295 300

Val His Thr Arg Gly Leu Asp Phe Ala Cys Asp Ile Tyr Ile Trp Ala

305 310 315 320

Pro Leu Ala Gly Thr Cys Gly Val Leu Leu Leu Ser Leu Val Ile Thr

325 330 335

Leu Tyr Cys Lys Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe Lys Gln

340 345 350

Pro Phe Met Arg Pro Val Gln Thr Thr Gln Glu Glu Asp Gly Cys Ser

355 360 365

Cys Arg Phe Pro Glu Glu Glu Glu Gly Gly Cys Glu Leu Arg Val Lys

370 375 380

Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln Gly Gln Asn Gln

385 390 395 400

Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr Asp Val Leu

405 410 415

Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys Pro Arg Arg

420 425 430

Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys Asp Lys Met

435 440 445

Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg Arg Arg Gly

450 455 460

Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala Thr Lys Asp

465 470 475 480

Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg

485 490

<210> 112

<211> 494

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MOD_RES

<222> (494)..(494)

<223> Any amino acid

<400> 112

Met Glu Phe Gly Leu Ser Trp Leu Phe Leu Val Ala Ile Leu Lys Gly
1 5 10 15

Val Gln Cys Ile Asp Ala Met Gly Asn Ile Gln Leu Val Gln Ser Gly
20 25 30

Pro Glu Leu Lys Lys Pro Gly Glu Thr Val Lys Ile Ser Cys Lys Ala
35 40 45

Ser Gly Tyr Thr Phe Thr Asn Tyr Gly Met Asn Trp Val Lys Gln Ala
50 55 60

Pro Gly Lys Gly Leu Arg Trp Met Gly Trp Ile Asn Thr His Thr Gly
65 70 75 80

Glu Pro Thr Tyr Ala Asp Asp Phe Lys Gly Arg Phe Ala Phe Ser Leu
85 90 95

Glu Thr Ser Ala Ser Thr Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn
100 105 110

Glu Asp Thr Ala Thr Tyr Phe Cys Thr Arg Arg Gly Tyr Asp Trp Tyr
115 120 125

Phe Asp Val Trp Gly Ala Gly Thr Thr Val Thr Val Ser Ser Gly Gly
130 135 140

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Lys
145 150 155 160

Met Thr Gln Ser Pro Ser Ser Met Tyr Ala Ser Leu Gly Glu Arg Val
165 170 175

Thr Ile Thr Cys Lys Ala Ser Gln Asp Ile Asn Ser Tyr Leu Ser Trp
180 185 190

Phe His His Lys Pro Gly Lys Ser Pro Lys Thr Leu Ile Tyr Arg Ala
195 200 205
Asn Arg Leu Val Asp Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser
210 215 220
Gly Gln Asp Tyr Ser Leu Thr Ile Ser Ser Leu Asp Tyr Glu Asp Met
225 230 235 240
Gly Ile Tyr Tyr Cys Gln Gln Tyr Asp Glu Ser Pro Trp Thr Phe Gly

245 250 255
Gly Gly Thr Lys Leu Glu Met Lys Gly Ser Gly Asp Pro Ala Thr Thr
260 265 270
Thr Pro Ala Pro Arg Pro Pro Thr Pro Ala Pro Thr Ile Ala Ser Gln
275 280 285
Pro Leu Ser Leu Arg Pro Glu Ala Cys Arg Pro Ala Ala Gly Gly Ala
290 295 300
Val His Thr Arg Gly Leu Asp Phe Ala Cys Asp Ile Tyr Ile Trp Ala

305 310 315 320
Pro Leu Ala Gly Thr Cys Gly Val Leu Leu Leu Ser Leu Val Ile Thr
325 330 335
Leu Tyr Cys Lys Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe Lys Gln
340 345 350
Pro Phe Met Arg Pro Val Gln Thr Thr Gln Glu Glu Asp Gly Cys Ser
355 360 365
Cys Arg Phe Pro Glu Glu Glu Glu Gly Gly Cys Glu Leu Arg Val Lys

370 375 380
Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln Gly Gln Asn Gln
385 390 395 400
Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr Asp Val Leu
405 410 415
Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys Pro Arg Arg
420 425 430
Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys Asp Lys Met

435 440 445

Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg Arg Arg Gly

450 455 460

Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala Thr Lys Asp

465 470 475 480

Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg Xaa

485 490

<210> 113

<211> 488

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 113

Met Ala Leu Pro Val Thr Ala Leu Leu Leu Pro Leu Ala Leu Leu Leu

1 5 10 15

His Ala Ala Arg Pro Gly Ser Asp Ile Glu Leu Thr Gln Ser Pro Ala

20 25 30

Ile Met Ser Ala Ser Leu Gly Glu Glu Ile Thr Leu Thr Cys Ser Ala

35 40 45

Ser Ser Ser Val Ser Tyr Met His Trp Tyr Gln Gln Lys Ser Gly Thr

50 55 60

Ser Pro Lys Leu Leu Ile Tyr Ser Thr Ser Asn Leu Ala Ser Gly Val

65 70 75 80

Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Phe Tyr Ser Leu Thr

85 90 95

Ile Ser Ser Val Glu Ala Glu Asp Ala Ala Asp Tyr Tyr Cys His Gln

100 105 110

Trp Ser Ser Tyr Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Arg

115 120 125

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gln

130 135 140

Val Lys Leu Gln Glu Ser Gly Gly Gly Leu Val Lys Pro Gly Gly Ser
145 150 155 160
Leu Lys Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr Ala
165 170 175
Met Ser Trp Val Arg Gln Thr Pro Glu Lys Arg Leu Glu Trp Val Ala

180 185 190
Thr Ile Ser Ser Gly Gly Ser Tyr Thr Tyr Tyr Pro Asp Ser Val Lys
195 200 205
Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Thr Leu Tyr Leu
210 215 220
Gln Met Ser Ser Leu Arg Ser Glu Asp Thr Ala Met Tyr Tyr Cys Ala
225 230 235 240
Arg Gln Asp Gly Tyr Tyr Pro Gly Trp Phe Ala Asn Trp Gly Gln Gly

245 250 255
Thr Thr Val Thr Val Ser Ser Ala Ala Ala Thr Thr Thr Pro Ala Pro
260 265 270
Arg Pro Pro Thr Pro Ala Pro Thr Ile Ala Ser Gln Pro Leu Ser Leu
275 280 285
Arg Pro Glu Ala Cys Arg Pro Ala Ala Gly Gly Ala Val His Thr Arg
290 295 300
Gly Leu Asp Phe Ala Cys Asp Ile Tyr Ile Trp Ala Pro Leu Ala Gly

305 310 315 320
Thr Cys Gly Val Leu Leu Leu Ser Leu Val Ile Thr Leu Tyr Cys Arg
325 330 335
Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr Met Asn Met Thr Pro
340 345 350
Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr Ala Pro Pro
355 360 365
Arg Asp Phe Ala Ala Tyr Arg Ser Arg Val Lys Phe Ser Arg Ser Ala

370 375 380
Asp Ala Pro Ala Tyr Gln Gln Gly Gln Asn Gln Leu Tyr Asn Glu Leu

385 390 395 400

Asn Leu Gly Arg Arg Glu Glu Tyr Asp Val Leu Asp Lys Arg Arg Gly

405 410 415

Arg Asp Pro Glu Met Gly Gly Lys Pro Arg Arg Lys Asn Pro Gln Glu

420 425 430

Gly Leu Tyr Asn Glu Leu Gln Lys Asp Lys Met Ala Glu Ala Tyr Ser

435 440 445

Glu Ile Gly Met Lys Gly Glu Arg Arg Arg Gly Lys Gly His Asp Gly

450 455 460

Leu Tyr Gln Gly Leu Ser Thr Ala Thr Lys Asp Thr Tyr Asp Ala Leu

465 470 475 480

His Met Gln Ala Leu Pro Pro Arg

485

<210> 114

<211> 50

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<220><221> MISC_FEATURE

<222> (1)..(50)

<223> This sequence may encompass 1-10 "Gly Gly Gly Gly Ser"

repeating units

<400> 114

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly

20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly

35 40 45

Gly Ser

50

<210> 115

<211> 72

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 115

Ser Asp Pro Thr Thr Thr Pro Ala Pro Arg Pro Pro Thr Pro Ala Pro
1 5 10 15

Thr Ile Ala Ser Gln Pro Leu Ser Leu Arg Pro Glu Ala Cys Arg Pro
20 25 30

Ala Ala Gly Gly Ala Val His Thr Arg Gly Leu Asp Phe Ala Cys Asp
35 40 45

Ile Tyr Ile Trp Ala Pro Leu Ala Gly Thr Cys Gly Val Leu Leu Leu
50 55 60

Ser Leu Val Ile Thr Leu Tyr Cys
65 70

<210> 116

<211> 112

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 116

Arg Val Lys Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln Gly
1 5 10 15

Gln Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu Tyr
20 25 30

Asp Val Leu Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly Lys
35 40 45

Pro Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln Lys
50 55 60

Asp Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu Arg

65 70 75 80

Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr Ala

85 90 95

Thr Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro Arg

100 105 110

<210> 117

<211> 37

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 117

Arg Arg Asp Gln Arg Leu Pro Pro Asp Ala His Lys Pro Pro Gly Gly

1 5 10 15

Gly Ser Phe Arg Thr Pro Ile Gln Glu Glu Gln Ala Asp Ala His Ser

20 25 30

Thr Leu Ala Lys Ile

35

<210> 118

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 118

Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr Met Asn Met Thr Pro

1 5 10 15

Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr Ala Pro Pro

20 25 30

Arg Asp Phe Ala Ala Tyr Arg Ser

35 40

<210> 119

<211> 42

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 119

Lys Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe Lys Gln Pro Phe Met
1 5 10 15

Arg Pro Val Gln Thr Thr Gln Glu Glu Asp Gly Cys Ser Cys Arg Phe
20 25 30

Pro Glu Glu Glu Glu Gly Gly Cys Glu Leu

35 40

<210> 120

<211> 47

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 120

Arg Phe Ser Val Val Lys Arg Gly Arg Lys Lys Leu Leu Tyr Ile Phe
1 5 10 15

Lys Gln Pro Phe Met Arg Pro Val Gln Thr Thr Gln Glu Glu Asp Gly
20 25 30

Cys Ser Cys Arg Phe Pro Glu Glu Glu Glu Gly Gly Cys Glu Leu

35 40 45

<210> 121

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 121

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacaga gucucucagc ugguacac 118

<210> 122

<211> 119

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 122

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugaggaaacu 60

ccuauugcug gacgaugucu cuuacgaggc auuagcacac cgauuuugau ucucaaaca 119

<210> 123

<211> 117

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 123

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacuca aacaaaugug cacaaag 117

<210> 124

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 124

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacuca aacaaaugug ucacaaag 118

<210> 125

<211> 117

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 125

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacuuu gagaaucaaa aucggua 117

<210> 126

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 126

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacuga uguguauauc acagacaa 118

<210> 127

<211> 116

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 127

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcaguug cuccaggcca cagcau 116

<210> 128

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 128

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacuuc cagaagacac cuucuucc 118

<210> 129

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 129

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcaccag aagacaccuu cuucccca 118

<210> 130

<211> 117

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 130

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagagaaac 60

uccuauugcu ggacgauguc ucuuacgagg cauuagcacg guuccgaauc cuccuga 117

<210> 131

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 131

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacgga acccaaucac ugacaggu 118

<210> 132

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 132

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacaga gtctctcagc tggtaca 117

<210> 133

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 133

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacacc gattttgatt ctcaaac 117

<210> 134

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 134

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcactga ttctcaaaca aatgtgt 117

<210> 135

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 135

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcactca aacaaatgtg tcacaaa 117

<210> 136

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 136

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacgtt tgagaatcaa aatcggt 117

<210> 137

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 137

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcactga tgtgtatatc acagaca 117

<210> 138

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 138

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacgtt gctccaggcc acagcac 117

<210> 139

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 139

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacttc cagaagacac cttcttc 117

<210> 140

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 140

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcaccag aagacacctt cttcccc 117

<210> 141

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 141

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacggt tccgaatcct cctcctg 117

<210> 142

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 142

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacgga acccaatcac tgacagg 117

<210> 143

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 143

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacag agtctctcag ctggtaca 108

<210> 144

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 144

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacac cgattttgat tctcaaac 108

<210> 145

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 145

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcactg attctcaaac aaatgtgt 108

<210> 146

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 146

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcactc aaacaaatgt gtcacaaa 108

<210> 147

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 147

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacgt ttgagaatca aaatcggt 108

<210> 148

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 148

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcactg atgtgtatat cacagaca 108

<210> 149

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 149

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacgt tgctccaggc cacagcac 108

<210> 150

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 150

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcactt ccagaagaca ccttcttc 108

<210> 151

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 151

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacca gaagacacct tcttcccc 108

<210> 152

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 152

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacgg ttccgaatcc tcctcctg 108

<210> 153

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 153

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacgg aacccaatca ctgacagg 108

<210> 154

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:
target sequence

<400> 154

ctcgatgcga ggactctcca 20

<210> 155

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 155

tctcgatgcg aggactctcc 20

<210> 156

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 156

catcgagata acctcgtgct 20

<210> 157

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 157

tggccgggct caggcacgag 20

<210> 158

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 158

gcccactggt aggggtgctg 20

<210> 159

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 159

tgggctcggt ggtgggactt 20

<210> 160

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 160

cgagcccact acgagacgga 20

<210> 161

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 161

ctcgtagtgg gctcggtggt 20

<210> 162

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 162

gccgtgaagg cgtcggccgg 20

<210> 163

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 163

gtttctgagt ttcaggattc 20

<210> 164

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 164

catcgggagg aagaacacac 20

<210> 165

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 165

ggaggaagaa cacacgggta 20

<210> 166

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 166

gagcgctggg ctgcatcaga 20

<210> 167

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 167

tgatctcgat ccgagggctc 20

<210> 168

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 168

acggagtgat ctcgatccga 20

<210> 169

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 169

gcggaggcat tcgtgcgccg 20

<210> 170

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 170

gccgcgctca gaaacttctg 20

<210> 171

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 171

gggcctcggg cctgagccct 20

<210> 172

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 172

cctcgggctg gcggccaccc 20

<210> 173

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 173

ccactcgccc gtgccccgtc 20

<210> 174

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 174

gcattcgtgc gccgaggcct 20

<210> 175

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 175

gagcctcacc ccagcgctcc 20

<210> 176

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 176

gaggggctcc gggagcgctg 20

<210> 177

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 177

agggctggtc ttccacatct 20

<210> 178

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 178

gcggggagcc caggccaaag 20

<210> 179

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 179

gccaccatga agacctttg 20

<210> 180

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 180

ttgcggcaca cctgagtacc 20

<210> 181

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 181

gtaggagaac tgggggaagt 20

<210> 182

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 182

ctcctactcg gccagcggca 20

<210> 183

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 183

gaaaaccttc tgtgggaccc 20

<210> 184

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 184

agtaggagaa ctgggggaag 20

<210> 185

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 185

gttgcaagtc tgacatttga 20

<210> 186

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 186

gaaacactac ctggtacaca 20

<210> 187

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 187

tgtggcagac ctacagtgta 20

<210> 188

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 188

gggcacacct tgcattggta 20

<210> 189

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 189

ctgcgcacct ggcattcatg 20

<210> 190

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 190

cctaccggtc cgcaagcgtc 20

<210> 191

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 191

actcacacat ctggataggt 20

<210> 192

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 192

gacttaccta gaccttcctg 20

<210> 193

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 193

aatcttacag agctgaaaag 20

<210> 194

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 194

catcatattc ttcacttcca 20

<210> 195

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 195

aagaatatga tgttcctccc 20

<210> 196

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 196

cccctaaacc acgaccgcgc 20

<210> 197

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 197

tcctgcgcgg tcgtggttta 20

<210> 198

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 198

ccctactctg tgaccaactg 20

<210> 199

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 199

cgcaggcgcg tcttccaggt 20

<210> 200

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 200

gcaccgcagg cgcgtcttcc 20

<210> 201

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 201

gaacaagagc aatgactttg 20

<210> 202

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 202

cacctaccta agaaccatcc 20

<210> 203

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 203

ggggttccag ggcctgtctg 20

<210> 204

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 204

gacttgcaca acatgcagaa 20

<210> 205

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 205

ttgccagaag caagatccca 20

<210> 206

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 206

ccatgaacaa ccaaaagaga 20

<210> 207

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 207

tcacccccat ccagaagccg 20

<210> 208

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 208

atctggctgg tctcgtccag 20

<210> 209

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 209

gatgagcgac ctcccggtga 20

<210> 210

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 210

agacagcgat gacgtctctg 20

<210> 211

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 211

acgtctctga ggtggacgcg 20

<210> 212

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 212

ttaggggaga gtttctcggc 20

<210> 213

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 213

ggagagtgag gaggaggaag 20

<210> 214

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 214

aaggctccga atccgaatct 20

<210> 215

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 215

atcgtcactc acgaccggct 20

<210> 216

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 216

tgacagtgag gaggaacaag 20

<210> 217

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 217

cacccgtggt tgttacactc 20

<210> 218

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 218

catcagccag gccccgcagg 20

<210> 219

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 219

ggtgtatcca tctgagttga 20

<210> 220

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 220

gggtgtatcc atctgagttg 20

<210> 221

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 221

gcgcaagaag cagatcgacg 20

<210> 222

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 222

ttactacagc ctggccagca 20

<210> 223

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 223

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 224

<211> 80

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 224

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 225

<211> 125

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 225

guuuuuguac ucucaagauu uaaguaacug uacaacgaaa cuuacacagu uacuuaaauc 60

uugcagaagc uacaaagaua aggcuucaug ccgaaaucaa cacccuguca uuuuauggca 120

gggug 125

<210> 226

<211> 76

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 226

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugc 76

<210> 227

<211> 84

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 227

guuuuaguac ucuguaauga aaauuacaga aucuacuaaa acaaggcaaa augccguguu 60

uaucucguca acuuguuggc gaga 84

<210> 228

<211> 97

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><223> Description of Combined DNA/RNA Molecule: Synthetic
oligonucleotide

<400> 228

guucugtcuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcac 97

<210> 229

<211> 88

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 229

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggagcaccug aaaacaggug cuuggcac 88

<210> 230

<211> 80

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 230

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

gggaccgagu cggugcuuuu 80

<210> 231

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 231

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa atctacaatc aattatttga agaaaaccct 600

attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat 720

ctcattgctt tgtcattggg attgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatagt gaaataacta aggctcccct atcagcttca 960

atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agataggggg atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta 2160

catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggtat tttacagact 2220

gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt 2280

gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg 2340

aaacgaatcg aagaaggtat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt 2400

gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctacaaaa tggaagagac 2460

atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt 2520

gttccacaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat 2580

aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac 2640

tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg 2700

aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg 2760

gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact 2820

aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa 2880

ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac 2940

catcatgccc atgatgcgta tctaaatgcc gtcgttggaa ctgctttgat taagaaatat 3000

ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg 3060

attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat 3120

atcatgaact tcttcaaaac agaaattaca cttgcaaatg gagagattcg caaacgccct 3180

ctaatcgaaa ctaatgggga aactggagaa attgtctggg ataaagggcg agattttgcc 3240

acagtgcgca aagtattgtc catgccccaa gtcaatattg tcaagaaaac agaagtacag 3300
acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360
cgtaaaaaag actgggatcc aaaaaaatat ggtggttttg atagtccaac ggtagcttat 3420
tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480
gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt 3540
ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600
agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660

aaaggaaatg agctggctct gccaagcaaa tatgtgaatt ttttatattt agctagtcat 3720
tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780
cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840
ttagcagatg ccaatttaga taaagttctt agtgcatata acaaacatag agacaaacca 3900
atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc 3960
gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc tacaaaagaa 4020
gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat 4080

ttgagtcagc taggaggtga ctga 4104

<210> 232
<211> 1367
<212> PRT
<213> Streptococcus pyogenes
<400> 232

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705 710 715 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 233
<211> 4212
<212> DNA
<213> Streptococcus pyogenes
<400> 233

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60
ataaccgatg aatacaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120
cattcgatta aaaagaatct tatcggtgcc ctcctattcg atagtggcga aacggcagag 180
gcgactcgcc tgaaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240
tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt 300
ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcaccc catctttgga 360

aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa 420
aagctagttg actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480
atgataaagt tccgtgggca ctttctcatt gagggtgatc taaatccgga caactcggat 540
gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct 600
ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga 660
cggctagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac 720
cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa 780

gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca 840
caaattggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc 900
ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca 960
atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc cctagtccgt 1020

cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca 1080

ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta 1140

gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcgaaga tctactgcga 1200

aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat 1260

gctatactta gaaggcagga ggatttttat ccgttcctca aagacaatcg tgaaaagatt 1320

gagaaaatcc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct 1380

cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa 1440

gttgtcgata aaggtgcgtc agctcaatcg ttcatcgaga ggatgaccaa ctttgacaag 1500

aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagtg 1560

tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta 1620

agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca 1680

gttaagcaat tgaaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc 1740

tccggggtag aagatcgatt taatgcgtca cttggtacgt atcatgacct cctaaagata 1800

attaaagata aggacttcct ggataacgaa gagaatgaag atatcttaga agatatagtg 1860

ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgct 1920

cacctgttcg acgataaggt tatgaaacag ttaaagaggc gtcgctatac gggctgggga 1980

cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc 2040

gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac 2100

tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg 2160

cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaagggcat actccagaca 2220

gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaaccgga aaacattgta 2280

atcgagatgg cacgcgaaaa tcaaacgact cagaaggggc aaaaaaacag tcgagagcgg 2340

atgaagagaa tagaagaggg tattaaagaa ctgggcagcc agatcttaaa ggagcatcct 2400

gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg 2460

gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac 2520

attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgct tacacgctcg 2580

gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag 2640

aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagtt cgataactta 2700

actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccggatttat taaacgtcag 2760

ctcgtggaaa cccgccaaat cacaaagcat gttgcacaga tactagattc ccgaatgaat 2820

acgaaatacg acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880

aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940
taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000
tacccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060
atgatcgcga aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120
aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180
cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240
gcgacggtga gaaaagtttt gtccatgccc caagtcaaca tagtaaagaa aactgaggtg 3300

cagaccggag ggttttcaaa ggaatcgatt cttccaaaaa ggaatagtga taagctcatc 3360
gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420
tattctgtcc tagtagtggc aaaagttgag aagggaaaat ccaagaaact gaagtcagtc 3480
aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540
ttccttgagg cgaaaggtta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600
tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660
caaaagggga acgaactcgc actaccgtct aaatacgtga atttcctgta tttagcgtcc 3720

cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact ttttgttgag 3780
cagcacaaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840
atcctagctg atgccaatct ggacaaagta ttaagcgcat acaacaagca cagggataaa 3900
cccatacgtg agcaggcgga aaatattatc catttgttta ctcttaccaa cctcggcgct 3960
ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020
gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080
gatttgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140

tacaaagacc atgacggtga ttataaagat catgacatcg attacaagga tgacgatgac 4200
aaggctgcag ga 4212
<210> 234
<211> 1368
<212> PRT
<213> Streptococcus pyogenes
<400> 234
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

```
        755                 760                 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

                805                 810                 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865                 870                 875                 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

    930                 935                 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe

        995                 1000                1005
```

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 235

<211> 4107

<212> DNA

<213> Streptococcus pyogenes

<400> 235

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600

attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720

ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca 960

atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta 2160

catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact 2220

gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt 2280

attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt 2340

atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct 2400

gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460
gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520
attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580
gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640
aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700
acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa 2760
ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820

actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct 2880
aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940
taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000
tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060
atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120
aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240

gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300
cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360
gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420
tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480
aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540
tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600
tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660

caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttata tttagctagt 3720
cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780
cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840
attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa 3900
ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960
cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080

gatttgagtc agctaggagg tgactga 4107

<210> 236

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 236

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 237

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 237

Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln
1 5 10 15
Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu
20 25 30
Ile Asp Asp Glu Glu Leu Arg Asn Val Leu Asn Leu Ala Tyr Lys Ile
35 40 45

Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys
50 55 60
Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65 70 75 80
Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys
85 90 95
Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
100 105 110

Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys
115 120 125
Pro Glu Phe Tyr Glu Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
130 135 140
Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Ile Ala Ala
145 150 155 160
Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
165 170 175

Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
180 185 190
Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
195 200 205
Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
210 215 220
Asn Pro Asn Val Ser Val Val Arg Ile Tyr Thr Ile Ser Asp Ala Val
225 230 235 240

Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr

245 250 255

Lys Leu Leu Glu Lys Arg Tyr Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270

Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
275 280 285

Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
290 295 300

Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305 310 315 320

Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
325 330 335

Asn Gly Glu Leu Ile Arg Gly Glu Gly
340 345

<210> 238

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 238

Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln
1 5 10 15

Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu
20 25 30

Ile Asp Asp Glu Glu Leu Arg Asn Val Leu Asn Leu Ala Tyr Lys Ile
35 40 45

Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys
50 55 60

Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65 70 75 80

Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys
85 90 95

Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
100 105 110

Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys
115 120 125
Pro Glu Phe Tyr Lys Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
130 135 140

Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Met Ala Ala
145 150 155 160
Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
165 170 175
Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
180 185 190
Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
195 200 205

Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
210 215 220
Asn Pro Asn Val Ser Val Val Ser Ile Tyr Thr Ile Ser Asp Ala Val
225 230 235 240
Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr
245 250 255
Lys Leu Leu Glu Lys Arg Asp Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270

Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
275 280 285
Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
290 295 300
Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305 310 315 320
Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
325 330 335

Asn Gly Glu Leu Ile Arg Gly Glu Gly
340 345

<210> 239

<211> 987

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Deltaproteobacteria CasX sequence

<400> 239

Met Glu Lys Arg Ile Asn Lys Ile Arg Lys Lys Leu Ser Ala Asp Asn
1 5 10 15
Ala Thr Lys Pro Val Ser Arg Ser Gly Pro Met Lys Thr Leu Leu Val
20 25 30
Arg Val Met Thr Asp Asp Leu Lys Lys Arg Leu Glu Lys Arg Arg Lys

35 40 45
Lys Pro Glu Val Met Pro Gln Val Ile Ser Asn Asn Ala Ala Asn Asn
50 55 60
Leu Arg Met Leu Leu Asp Asp Tyr Thr Lys Met Lys Glu Ala Ile Leu
65 70 75 80
Gln Val Tyr Trp Gln Glu Phe Lys Asp Asp His Val Gly Leu Met Cys
85 90 95
Lys Phe Ala Gln Pro Ala Ser Lys Lys Ile Asp Gln Asn Lys Leu Lys

100 105 110
Pro Glu Met Asp Glu Lys Gly Asn Leu Thr Thr Ala Gly Phe Ala Cys
115 120 125
Ser Gln Cys Gly Gln Pro Leu Phe Val Tyr Lys Leu Glu Gln Val Ser
130 135 140
Glu Lys Gly Lys Ala Tyr Thr Asn Tyr Phe Gly Arg Cys Asn Val Ala
145 150 155 160
Glu His Glu Lys Leu Ile Leu Leu Ala Gln Leu Lys Pro Val Lys Asp

165 170 175
Ser Asp Glu Ala Val Thr Tyr Ser Leu Gly Lys Phe Gly Gln Arg Ala
180 185 190
Leu Asp Phe Tyr Ser Ile His Val Thr Lys Glu Ser Thr His Pro Val
195 200 205

Lys Pro Leu Ala Gln Ile Ala Gly Asn Arg Tyr Ala Ser Gly Pro Val
210 215 220
Gly Lys Ala Leu Ser Asp Ala Cys Met Gly Thr Ile Ala Ser Phe Leu

225 230 235 240
Ser Lys Tyr Gln Asp Ile Ile Ile Glu His Gln Lys Val Val Lys Gly
245 250 255
Asn Gln Lys Arg Leu Glu Ser Leu Arg Glu Leu Ala Gly Lys Glu Asn
260 265 270
Leu Glu Tyr Pro Ser Val Thr Leu Pro Pro Gln Pro His Thr Lys Glu
275 280 285
Gly Val Asp Phe Ala Tyr Asn Glu Val Ile Ala Arg Val Arg Met Trp

290 295 300
Val Asn Leu Asn Leu Trp Gln Lys Leu Lys Leu Ser Arg Asp Asp Ala
305 310 315 320
Lys Pro Leu Leu Arg Leu Lys Gly Phe Pro Ser Phe Pro Val Val Glu
325 330 335
Arg Arg Glu Asn Glu Val Asp Trp Trp Asn Thr Ile Asn Glu Val Lys
340 345 350
Lys Leu Ile Asp Ala Lys Arg Asp Met Gly Arg Val Phe Trp Ser Gly

355 360 365
Val Thr Ala Glu Lys Arg Asn Thr Ile Leu Glu Gly Tyr Asn Tyr Leu
370 375 380
Pro Asn Glu Asn Asp His Lys Lys Arg Glu Gly Ser Leu Glu Asn Pro
385 390 395 400
Lys Lys Pro Ala Lys Arg Gln Phe Gly Asp Leu Leu Leu Tyr Leu Glu
405 410 415
Lys Lys Tyr Ala Gly Asp Trp Gly Lys Val Phe Asp Glu Ala Trp Glu

420 425 430
Arg Ile Asp Lys Lys Ile Ala Gly Leu Thr Ser His Ile Glu Arg Glu
435 440 445
Glu Ala Arg Asn Ala Glu Asp Ala Gln Ser Lys Ala Val Leu Thr Asp

450 455 460

Trp Leu Arg Ala Lys Ala Ser Phe Val Leu Glu Arg Leu Lys Glu Met

465 470 475 480

Asp Glu Lys Glu Phe Tyr Ala Cys Glu Ile Gln Leu Gln Lys Trp Tyr

485 490 495

Gly Asp Leu Arg Gly Asn Pro Phe Ala Val Glu Ala Glu Asn Arg Val

500 505 510

Val Asp Ile Ser Gly Phe Ser Ile Gly Ser Asp Gly His Ser Ile Gln

515 520 525

Tyr Arg Asn Leu Leu Ala Trp Lys Tyr Leu Glu Asn Gly Lys Arg Glu

530 535 540

Phe Tyr Leu Leu Met Asn Tyr Gly Lys Lys Gly Arg Ile Arg Phe Thr

545 550 555 560

Asp Gly Thr Asp Ile Lys Lys Ser Gly Lys Trp Gln Gly Leu Leu Tyr

565 570 575

Gly Gly Gly Lys Ala Lys Val Ile Asp Leu Thr Phe Asp Pro Asp Asp

580 585 590

Glu Gln Leu Ile Ile Leu Pro Leu Ala Phe Gly Thr Arg Gln Gly Arg

595 600 605

Glu Phe Ile Trp Asn Asp Leu Leu Ser Leu Glu Thr Gly Leu Ile Lys

610 615 620

Leu Ala Asn Gly Arg Val Ile Glu Lys Thr Ile Tyr Asn Lys Lys Ile

625 630 635 640

Gly Arg Asp Glu Pro Ala Leu Phe Val Ala Leu Thr Phe Glu Arg Arg

645 650 655

Glu Val Val Asp Pro Ser Asn Ile Lys Pro Val Asn Leu Ile Gly Val

660 665 670

Ala Arg Gly Glu Asn Ile Pro Ala Val Ile Ala Leu Thr Asp Pro Glu

675 680 685

Gly Cys Pro Leu Pro Glu Phe Lys Asp Ser Ser Gly Gly Pro Thr Asp

690 695 700

Ile Leu Arg Ile Gly Glu Gly Tyr Lys Glu Lys Gln Arg Ala Ile Gln
705 710 715 720
Ala Ala Lys Glu Val Glu Gln Arg Arg Ala Gly Gly Tyr Ser Arg Lys
725 730 735
Phe Ala Ser Lys Ser Arg Asn Leu Ala Asp Asp Met Val Arg Asn Ser

740 745 750
Ala Arg Asp Leu Phe Tyr His Ala Val Thr His Asp Ala Val Leu Val
755 760 765
Phe Ala Asn Leu Ser Arg Gly Phe Gly Arg Gln Gly Lys Arg Thr Phe
770 775 780
Met Thr Glu Arg Gln Tyr Thr Lys Met Glu Asp Trp Leu Thr Ala Lys
785 790 795 800
Leu Ala Tyr Glu Gly Leu Thr Ser Lys Thr Tyr Leu Ser Lys Thr Leu

805 810 815
Ala Gln Tyr Thr Ser Lys Thr Cys Ser Asn Cys Gly Phe Thr Ile Thr
820 825 830
Tyr Ala Asp Met Asp Val Met Leu Val Arg Leu Lys Lys Thr Ser Asp
835 840 845
Gly Trp Ala Thr Thr Leu Asn Asn Lys Glu Leu Lys Ala Glu Tyr Gln
850 855 860
Ile Thr Tyr Tyr Asn Arg Tyr Lys Arg Gln Thr Val Glu Lys Glu Leu

865 870 875 880
Ser Ala Glu Leu Asp Arg Leu Ser Glu Glu Ser Gly Asn Asn Asp Ile
885 890 895
Ser Lys Trp Thr Lys Gly Arg Arg Asp Glu Ala Leu Phe Leu Leu Lys
900 905 910
Lys Arg Phe Ser His Arg Pro Val Gln Glu Gln Phe Val Cys Leu Asp
915 920 925
Cys Gly His Glu Val His Ala Ala Glu Gln Ala Ala Leu Asn Ile Ala

930 935 940
Arg Ser Trp Leu Phe Leu Asn Ser Asn Ser Thr Glu Phe Lys Ser Tyr

945 950 955 960

Lys Ser Gly Lys Gln Pro Phe Val Gly Ala Trp Gln Ala Phe Tyr Lys

965 970 975

Arg Arg Leu Lys Glu Val Trp Lys Pro Asn Ala

980 985

<210> 240

<211> 1210

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

uncultured Parcubacteria group bacterium

<400> 240

Met Ser Lys Arg His Pro Arg Ile Ser Gly Val Lys Gly Tyr Arg Leu

1 5 10 15

His Ala Gln Arg Leu Glu Tyr Thr Gly Lys Ser Gly Ala Met Arg Thr

20 25 30

Ile Lys Tyr Pro Leu Tyr Ser Ser Pro Ser Gly Gly Arg Thr Val Pro

35 40 45

Arg Glu Ile Val Ser Ala Ile Asn Asp Asp Tyr Val Gly Leu Tyr Gly

50 55 60

Leu Ser Asn Phe Asp Asp Leu Tyr Asn Ala Glu Lys Arg Asn Glu Glu

65 70 75 80

Lys Val Tyr Ser Val Leu Asp Phe Trp Tyr Asp Cys Val Gln Tyr Gly

85 90 95

Ala Val Phe Ser Tyr Thr Ala Pro Gly Leu Leu Lys Asn Val Ala Glu

100 105 110

Val Arg Gly Gly Ser Tyr Glu Leu Thr Lys Thr Leu Lys Gly Ser His

115 120 125

Leu Tyr Asp Glu Leu Gln Ile Asp Lys Val Ile Lys Phe Leu Asn Lys

130 135 140

Lys Glu Ile Ser Arg Ala Asn Gly Ser Leu Asp Lys Leu Lys Lys Asp

145 150 155 160

Ile Ile Asp Cys Phe Lys Ala Glu Tyr Arg Glu Arg His Lys Asp Gln
165 170 175
Cys Asn Lys Leu Ala Asp Asp Ile Lys Asn Ala Lys Lys Asp Ala Gly

180 185 190
Ala Ser Leu Gly Glu Arg Gln Lys Lys Leu Phe Arg Asp Phe Phe Gly
195 200 205
Ile Ser Glu Gln Ser Glu Asn Asp Lys Pro Ser Phe Thr Asn Pro Leu
210 215 220
Asn Leu Thr Cys Cys Leu Leu Pro Phe Asp Thr Val Asn Asn Asn Arg
225 230 235 240
Asn Arg Gly Glu Val Leu Phe Asn Lys Leu Lys Glu Tyr Ala Gln Lys

245 250 255
Leu Asp Lys Asn Glu Gly Ser Leu Glu Met Trp Glu Tyr Ile Gly Ile
260 265 270
Gly Asn Ser Gly Thr Ala Phe Ser Asn Phe Leu Gly Glu Gly Phe Leu
275 280 285
Gly Arg Leu Arg Glu Asn Lys Ile Thr Glu Leu Lys Lys Ala Met Met
290 295 300
Asp Ile Thr Asp Ala Trp Arg Gly Gln Glu Gln Glu Glu Glu Leu Glu

305 310 315 320
Lys Arg Leu Arg Ile Leu Ala Ala Leu Thr Ile Lys Leu Arg Glu Pro
325 330 335
Lys Phe Asp Asn His Trp Gly Gly Tyr Arg Ser Asp Ile Asn Gly Lys
340 345 350
Leu Ser Ser Trp Leu Gln Asn Tyr Ile Asn Gln Thr Val Lys Ile Lys
355 360 365
Glu Asp Leu Lys Gly His Lys Lys Asp Leu Lys Lys Ala Lys Glu Met

370 375 380
Ile Asn Arg Phe Gly Glu Ser Asp Thr Lys Glu Glu Ala Val Val Ser
385 390 395 400
Ser Leu Leu Glu Ser Ile Glu Lys Ile Val Pro Asp Asp Ser Ala Asp

405 410 415

Asp Glu Lys Pro Asp Ile Pro Ala Ile Ala Ile Tyr Arg Arg Phe Leu

420 425 430

Ser Asp Gly Arg Leu Thr Leu Asn Arg Phe Val Gln Arg Glu Asp Val

435 440 445

Gln Glu Ala Leu Ile Lys Glu Arg Leu Glu Ala Glu Lys Lys Lys Lys

450 455 460

Pro Lys Lys Arg Lys Lys Lys Ser Asp Ala Glu Asp Glu Lys Glu Thr

465 470 475 480

Ile Asp Phe Lys Glu Leu Phe Pro His Leu Ala Lys Pro Leu Lys Leu

485 490 495

Val Pro Asn Phe Tyr Gly Asp Ser Lys Arg Glu Leu Tyr Lys Lys Tyr

500 505 510

Lys Asn Ala Ala Ile Tyr Thr Asp Ala Leu Trp Lys Ala Val Glu Lys

515 520 525

Ile Tyr Lys Ser Ala Phe Ser Ser Ser Leu Lys Asn Ser Phe Phe Asp

530 535 540

Thr Asp Phe Asp Lys Asp Phe Phe Ile Lys Arg Leu Gln Lys Ile Phe

545 550 555 560

Ser Val Tyr Arg Arg Phe Asn Thr Asp Lys Trp Lys Pro Ile Val Lys

565 570 575

Asn Ser Phe Ala Pro Tyr Cys Asp Ile Val Ser Leu Ala Glu Asn Glu

580 585 590

Val Leu Tyr Lys Pro Lys Gln Ser Arg Ser Arg Lys Ser Ala Ala Ile

595 600 605

Asp Lys Asn Arg Val Arg Leu Pro Ser Thr Glu Asn Ile Ala Lys Ala

610 615 620

Gly Ile Ala Leu Ala Arg Glu Leu Ser Val Ala Gly Phe Asp Trp Lys

625 630 635 640

Asp Leu Leu Lys Lys Glu Glu His Glu Glu Tyr Ile Asp Leu Ile Glu

645 650 655

Leu His Lys Thr Ala Leu Ala Leu Leu Leu Ala Val Thr Glu Thr Gln
660 665 670
Leu Asp Ile Ser Ala Leu Asp Phe Val Glu Asn Gly Thr Val Lys Asp
675 680 685
Phe Met Lys Thr Arg Asp Gly Asn Leu Val Leu Glu Gly Arg Phe Leu

690 695 700
Glu Met Phe Ser Gln Ser Ile Val Phe Ser Glu Leu Arg Gly Leu Ala
705 710 715 720
Gly Leu Met Ser Arg Lys Glu Phe Ile Thr Arg Ser Ala Ile Gln Thr
725 730 735
Met Asn Gly Lys Gln Ala Glu Leu Leu Tyr Ile Pro His Glu Phe Gln
740 745 750
Ser Ala Lys Ile Thr Thr Pro Lys Glu Met Ser Arg Ala Phe Leu Asp

755 760 765
Leu Ala Pro Ala Glu Phe Ala Thr Ser Leu Glu Pro Glu Ser Leu Ser
770 775 780
Glu Lys Ser Leu Leu Lys Leu Lys Gln Met Arg Tyr Tyr Pro His Tyr
785 790 795 800
Phe Gly Tyr Glu Leu Thr Arg Thr Gly Gln Gly Ile Asp Gly Gly Val
805 810 815
Ala Glu Asn Ala Leu Arg Leu Glu Lys Ser Pro Val Lys Lys Arg Glu

820 825 830
Ile Lys Cys Lys Gln Tyr Lys Thr Leu Gly Arg Gly Gln Asn Lys Ile
835 840 845
Val Leu Tyr Val Arg Ser Ser Tyr Tyr Gln Thr Gln Phe Leu Glu Trp
850 855 860
Phe Leu His Arg Pro Lys Asn Val Gln Thr Asp Val Ala Val Ser Gly
865 870 875 880
Ser Phe Leu Ile Asp Glu Lys Lys Val Lys Thr Arg Trp Asn Tyr Asp

885 890 895
Ala Leu Thr Val Ala Leu Glu Pro Val Ser Gly Ser Glu Arg Val Phe

900 905 910

Val Ser Gln Pro Phe Thr Ile Phe Pro Glu Lys Ser Ala Glu Glu Glu

915 920 925

Gly Gln Arg Tyr Leu Gly Ile Asp Ile Gly Glu Tyr Gly Ile Ala Tyr

930 935 940

Thr Ala Leu Glu Ile Thr Gly Asp Ser Ala Lys Ile Leu Asp Gln Asn

945 950 955 960

Phe Ile Ser Asp Pro Gln Leu Lys Thr Leu Arg Glu Glu Val Lys Gly

965 970 975

Leu Lys Leu Asp Gln Arg Arg Gly Thr Phe Ala Met Pro Ser Thr Lys

980 985 990

Ile Ala Arg Ile Arg Glu Ser Leu Val His Ser Leu Arg Asn Arg Ile

995 1000 1005

His His Leu Ala Leu Lys His Lys Ala Lys Ile Val Tyr Glu Leu

1010 1015 1020

Glu Val Ser Arg Phe Glu Glu Gly Lys Gln Lys Ile Lys Lys Val

1025 1030 1035

Tyr Ala Thr Leu Lys Lys Ala Asp Val Tyr Ser Glu Ile Asp Ala

1040 1045 1050

Asp Lys Asn Leu Gln Thr Thr Val Trp Gly Lys Leu Ala Val Ala

1055 1060 1065

Ser Glu Ile Ser Ala Ser Tyr Thr Ser Gln Phe Cys Gly Ala Cys

1070 1075 1080

Lys Lys Leu Trp Arg Ala Glu Met Gln Val Asp Glu Thr Ile Thr

1085 1090 1095

Thr Gln Glu Leu Ile Gly Thr Val Arg Val Ile Lys Gly Gly Thr

1100 1105 1110

Leu Ile Asp Ala Ile Lys Asp Phe Met Arg Pro Pro Ile Phe Asp

1115 1120 1125

Glu Asn Asp Thr Pro Phe Pro Lys Tyr Arg Asp Phe Cys Asp Lys

1130 1135 1140

His His Ile Ser Lys Lys Met Arg Gly Asn Ser Cys Leu Phe Ile

1145 1150 1155

Cys Pro Phe Cys Arg Ala Asn Ala Asp Ala Asp Ile Gln Ala Ser

1160 1165 1170

Gln Thr Ile Ala Leu Leu Arg Tyr Val Lys Glu Glu Lys Lys Val

1175 1180 1185

Glu Asp Tyr Phe Glu Arg Phe Arg Lys Leu Lys Asn Ile Lys Val

1190 1195 1200

Leu Gly Gln Met Lys Lys Ile

1205 1210

<210> 241

<211> 1082

<212> PRT

<213> Neisseria meningitidis

<400> 241

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp

1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Asp

20 25 30

Glu Asn Pro Ile Cys Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg

35 40 45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu

50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu

65 70 75 80

Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp

85 90 95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln

100 105 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser

115 120 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg

130 135 140

Lys Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys

145 150 155 160

Gly Val Ala Asp Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr

165 170 175

Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile

180 185 190

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu

195 200 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn

210 215 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met

225 230 235 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly

245 250 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr

260 265 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile

275 280 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr

290 295 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala

305 310 315 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg

325 330 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala

340 345 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys

355 360 365

Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr

370 375 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385 390 395 400
Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
405 410 415
Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val

420 425 430
Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
435 440 445
Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
450 455 460
Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465 470 475 480
Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly

485 490 495
Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
500 505 510
Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
515 520 525
Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
530 535 540
Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu

545 550 555 560
Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
565 570 575
Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
580 585 590
Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
595 600 605
Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn

610 615 620
Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu

625 630 635 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys

645 650 655

Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr

660 665 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr

675 680 685

Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn

690 695 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp

705 710 715 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala

725 730 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala

740 745 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln

755 760 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met

770 775 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala

785 790 795 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser

805 810 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg

820 825 830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys

835 840 845

Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu

850 855 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg

865 870 875 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys

885 890 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys

900 905 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val

915 920 925

Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn

930 935 940

Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr

945 950 955 960

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp

965 970 975

Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp

980 985 990

Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu

995 1000 1005

Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys

1010 1015 1020

His Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp

1025 1030 1035

His Lys Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys

1040 1045 1050

Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys

1055 1060 1065

Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg

1070 1075 1080

<210> 242

<211> 1082

<212> PRT

<213> Neisseria meningitidis

<400> 242

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp

1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu

20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg

35 40 45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu

50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu

65 70 75 80

Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp

85 90 95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln

100 105 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser

115 120 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg

130 135 140

Lys Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys

145 150 155 160

Gly Val Ala Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr

165 170 175

Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile

180 185 190

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu

195 200 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn

210 215 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met

225 230 235 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly

245 250 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
260 265 270
Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
275 280 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
290 295 300
Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305 310 315 320
Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
325 330 335
Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
340 345 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
355 360 365
Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr
370 375 380
Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385 390 395 400
Asp Arg Val Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
405 410 415

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
420 425 430
Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
435 440 445
Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
450 455 460
Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465 470 475 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
485 490 495
Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser

500 505 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys

515 520 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe

530 535 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu

545 550 555 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val

565 570 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe

580 585 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly

595 600 605

Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn

610 615 620

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu

625 630 635 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys

645 650 655

Phe Asp Glu Asp Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr

660 665 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp His Ile Leu Leu Thr

675 680 685

Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn

690 695 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp

705 710 715 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala

725 730 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala

740 745 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Lys Val Leu His Gln
755 760 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
770 775 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785 790 795 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
805 810 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
820 825 830

Ala Pro Asn Arg Lys Met Ser Gly Ala His Lys Asp Thr Leu Arg Ser
835 840 845

Ala Lys Arg Phe Val Lys His Asn Glu Lys Ile Ser Val Lys Arg Val
850 855 860

Trp Leu Thr Glu Ile Lys Leu Ala Asp Leu Glu Asn Met Val Asn Tyr
865 870 875 880

Lys Asn Gly Arg Glu Ile Glu Leu Tyr Glu Ala Leu Lys Ala Arg Leu
885 890 895

Glu Ala Tyr Gly Gly Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp Asn
900 905 910

Pro Phe Tyr Lys Lys Gly Gly Gln Leu Val Lys Ala Val Arg Val Glu
915 920 925

Lys Thr Gln Glu Ser Gly Val Leu Leu Asn Lys Lys Asn Ala Tyr Thr
930 935 940

Ile Ala Asp Asn Gly Asp Met Val Arg Val Asp Val Phe Cys Lys Val
945 950 955 960

Asp Lys Lys Gly Lys Asn Gln Tyr Phe Ile Val Pro Ile Tyr Ala Trp
965 970 975

Gln Val Ala Glu Asn Ile Leu Pro Asp Ile Asp Cys Lys Gly Tyr Arg
980 985 990

Ile Asp Asp Ser Tyr Thr Phe Cys Phe Ser Leu His Lys Tyr Asp Leu

995 1000 1005

Ile Ala Phe Gln Lys Asp Glu Lys Ser Lys Val Glu Phe Ala Tyr

1010 1015 1020

Tyr Ile Asn Cys Asp Ser Ser Asn Gly Arg Phe Tyr Leu Ala Trp

1025 1030 1035

His Asp Lys Gly Ser Lys Glu Gln Gln Phe Arg Ile Ser Thr Gln

1040 1045 1050

Asn Leu Val Leu Ile Gln Lys Tyr Gln Val Asn Glu Leu Gly Lys

1055 1060 1065

Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg

1070 1075 1080

<210> 243

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<220><223> Description of Combined DNA/RNA Molecule: Synthetic oligonucleotide

<220><221> modified_base

<222> (98)..(117)

<223> a, c, t, g, unknown or other

<400> 243

guucugtcuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacnnn nnnnnnnnnn nnnnnnn 117

<210> 244

<211> 108

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<220><221> modified_base

<222> (89)..(108)

<223> a, c, u, g, unknown or other

<400> 244

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggagcaccug aaaacaggug cuuggcacnn nnnnnnnnnn nnnnnnnn 108

<210> 245

<211> 119

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<220><221> modified_base

<222> (100)..(119)

<223> a, c, u, g, unknown or other

<400> 245

gucuaaagga cagaauuuuu caacgggugu gccaauggcc acuuuccagg uggcaaagcc 60

cguugaacuu cucaaaaaga acgaucugag aaguggcacn nnnnnnnnnn nnnnnnnn 119

<210> 246

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 246

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys

85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr

100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile

115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln

130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr

145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr

165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser

180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu

195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys

210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu

225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg

245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr

260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys

275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile

290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys

305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser

325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn

805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 247

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 247

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn

450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu

995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 248
<211> 1300
<212> PRT
<213> Francisella novicida

<400> 248
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15
Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45
Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80
Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys

85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn

805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290

Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 249

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 249

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys

210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn

805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu

915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile

930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290

Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 250

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 250

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys

20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys

35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu

50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser

65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys

85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met

340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys

355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln

370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr

385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala

405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn

420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala

435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn

450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys

485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp

515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His

530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His

545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val

565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu

915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile

930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu

995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300
<210> 251
<211> 1300

<212> PRT

<213> Francisella novicida

<400> 251

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys

485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp

515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His

530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His

545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val

565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300
<210> 252
<211> 1300
<212> PRT
<213> Francisella novicida
<400> 252
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15
Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45
Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80
Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met

340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu

915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile

930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu

995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn
1295 1300
<210> 253

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 253

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys

485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp

515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His

530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His

545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val

565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu

995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 254
<211> 1052
<212> PRT
<213> Staphylococcus aureus

<400> 254
Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1 5 10 15
Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val
20 25 30
Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser
35 40 45
Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln
50 55 60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser
65 70 75 80
Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser
85 90 95
Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala

100 105 110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly

115 120 125

Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu

130 135 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp

145 150 155 160

Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr Val

165 170 175

Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu

180 185 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg

195 200 205

Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp

210 215 220

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro

225 230 235 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn

245 250 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu

260 265 270

Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys

275 280 285

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val

290 295 300

Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro

305 310 315 320

Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala

325 330 335

Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys

340 345 350

Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
355 360 365
Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
370 375 380

Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn
385 390 395 400
Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile
405 410 415
Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln
420 425 430
Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val
435 440 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile
450 455 460
Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu
465 470 475 480
Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg
485 490 495
Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly
500 505 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met
515 520 525
Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp
530 535 540
Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg
545 550 555 560
Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln
565 570 575

Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser
580 585 590
Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu

595 600 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr

610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe

625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met

645 650 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val

660 665 670

Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys

675 680 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala

690 695 700

Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu

705 710 715 720

Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln

725 730 735

Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile

740 745 750

Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr

755 760 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn

770 775 780

Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile

785 790 795 800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys

805 810 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp

820 825 830

Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp

835 840 845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
850 855 860
Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865 870 875 880
Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
885 890 895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
900 905 910
Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
915 920 925
Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
930 935 940
Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945 950 955 960

Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu
965 970 975
Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
980 985 990
Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
995 1000 1005
Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr
1010 1015 1020

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
1025 1030 1035
Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 255
<211> 1052
<212> PRT
<213> Staphylococcus aureus
<400> 255
Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly

1 5 10 15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val

20 25 30

Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser

35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln

50 55 60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser

65 70 75 80

Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser

85 90 95

Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala

100 105 110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly

115 120 125

Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu

130 135 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp

145 150 155 160

Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr Val

165 170 175

Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu

180 185 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg

195 200 205

Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp

210 215 220

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro

225 230 235 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn

245 250 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
260 265 270
Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys

275 280 285
Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
290 295 300
Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305 310 315 320
Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
325 330 335
Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys

340 345 350
Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
355 360 365
Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
370 375 380
Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn
385 390 395 400
Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile

405 410 415
Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln
420 425 430
Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val
435 440 445
Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile
450 455 460
Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu

465 470 475 480
Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg
485 490 495
Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly

500 505 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met

515 520 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp

530 535 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg

545 550 555 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln

565 570 575

Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser

580 585 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu

595 600 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr

610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe

625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met

645 650 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val

660 665 670

Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys

675 680 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala

690 695 700

Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu

705 710 715 720

Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln

725 730 735

Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile

740 745 750

Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr

755 760 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn

770 775 780

Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile

785 790 795 800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys

805 810 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp

820 825 830

Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp

835 840 845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu

850 855 860

Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys

865 870 875 880

Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr

885 890 895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg

900 905 910

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys

915 920 925

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys

930 935 940

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu

945 950 955 960

Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu

965 970 975

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu

980 985 990

Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn

995 1000 1005

Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr

1010 1015 1020

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr

1025 1030 1035

Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly

1040 1045 1050

<210> 256

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 256

Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly

1 5 10 15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val

20 25 30

Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser

35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln

50 55 60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser

65 70 75 80

Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser

85 90 95

Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala

100 105 110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly

115 120 125

Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu

130 135 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp

145 150 155 160

Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr Val
165 170 175

Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu
180 185 190
Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg
195 200 205
Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp
210 215 220
Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225 230 235 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn
245 250 255
Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
260 265 270
Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
275 280 285
Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
290 295 300

Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305 310 315 320
Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
325 330 335
Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
340 345 350
Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
355 360 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
370 375 380
Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn
385 390 395 400
Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile

405 410 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln

420 425 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val

435 440 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile

450 455 460

Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu

465 470 475 480

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg

485 490 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly

500 505 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met

515 520 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp

530 535 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg

545 550 555 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln

565 570 575

Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser

580 585 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu

595 600 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr

610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe

625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met

645 650 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val
660 665 670
Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
675 680 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
690 695 700
Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705 710 715 720
Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
725 730 735
Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
740 745 750

Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
755 760 765
Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Lys Leu Ile Asn
770 775 780
Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile
785 790 795 800
Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys
805 810 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
820 825 830
Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
835 840 845
Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
850 855 860
Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865 870 875 880

Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
885 890 895
Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg

900 905 910

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys

915 920 925

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys

930 935 940

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu

945 950 955 960

Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile Asn Gly Glu

965 970 975

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu

980 985 990

Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn

995 1000 1005

Asp Lys Arg Pro Pro His Ile Ile Lys Thr Ile Ala Ser Lys Thr

1010 1015 1020

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr

1025 1030 1035

Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly

1040 1045 1050

<210> 257

<211> 1410

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 257

Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met

1 5 10 15

Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys

20 25 30

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

35 40 45

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln
50 55 60

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
65 70 75 80
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys
85 90 95
Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Met Gln Pro Thr Val
100 105 110
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
115 120 125

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg
130 135 140
Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr
145 150 155 160
Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
165 170 175
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Lys Phe
180 185 190

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
195 200 205
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp
210 215 220
Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp
225 230 235 240
Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala
245 250 255

Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
260 265 270
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
275 280 285
Thr Asn Leu Gly Ala Pro Arg Ala Phe Lys Tyr Phe Asp Thr Thr Ile

290 295 300

Ala Arg Lys Glu Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
305 310 315 320

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
325 330 335

Gln Leu Gly Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
340 345 350

Ser Gly Gly Ser Gly Gly Ser Gly Gly Met Asp Lys Lys Tyr Ser Ile
355 360 365

Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp
370 375 380

Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp
385 390 395 400

Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser
405 410 415

Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg
420 425 430

Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser
435 440 445

Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu
450 455 460

Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe
465 470 475 480

Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile
485 490 495

Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu
500 505 510

Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His
515 520 525

Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys
530 535 540

Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn
545 550 555 560
Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg
565 570 575

Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly
580 585 590
Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly
595 600 605
Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys
610 615 620
Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu
625 630 635 640

Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn
645 650 655
Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu
660 665 670
Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu
675 680 685
His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu
690 695 700

Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr
705 710 715 720
Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe
725 730 735
Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val
740 745 750
Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn
755 760 765

Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu
770 775 780
Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys

785 790 795 800

Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu

805 810 815

Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu

820 825 830

Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser

835 840 845

Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro

850 855 860

Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr

865 870 875 880

Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg

885 890 895

Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu

900 905 910

Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp

915 920 925

Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val

930 935 940

Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys

945 950 955 960

Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile

965 970 975

Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met

980 985 990

Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val

995 1000 1005

Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu

1010 1015 1020

Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys

1025 1030 1035

Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
1040 1045 1050
Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp
1055 1060 1065
Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu
1070 1075 1080

His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
1085 1090 1095
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
1100 1105 1110
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn
1115 1120 1125
Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
1130 1135 1140

Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
1145 1150 1155
Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr
1160 1165 1170
Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu
1175 1180 1185
Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val
1190 1195 1200

Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
1205 1210 1215
Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser
1220 1225 1230
Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu
1235 1240 1245
Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys
1250 1255 1260

Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile

1265 1270 1275

Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala

1280 1285 1290

Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp

1295 1300 1305

Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu

1310 1315 1320

Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

1325 1330 1335

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

1340 1345 1350

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu

1355 1360 1365

Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile

1370 1375 1380

Ala Lys Ser Glu Gln Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly

1385 1390 1395

Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1400 1405 1410

<210> 258

<211> 1129

<212> PRT

<213> Alicyclobacillus acidoterrestris

<400> 258

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Asp Asp Met Pro

1 5 10 15

Glu Ile Arg Ala Gly Leu Trp Lys Leu His Lys Glu Val Asn Ala Gly

20 25 30

Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu

35 40 45

Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Asp Lys Thr

50 55 60

Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln
65 70 75 80
Val Glu Asn Gly His Arg Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu

85 90 95
Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly
100 105 110
Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu
115 120 125
Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn
130 135 140
Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu

145 150 155 160
Glu Glu Lys Glu Lys Ala Glu Thr Arg Lys Ser Ala Asp Arg Thr Ala
165 170 175
Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg
180 185 190
Val Tyr Thr Asp Ser Glu Met Ser Ser Val Glu Trp Lys Pro Leu Arg
195 200 205
Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala

210 215 220
Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Gln
225 230 235 240
Glu Tyr Ala Lys Leu Val Glu Gln Lys Asn Arg Phe Glu Gln Lys Asn
245 250 255
Phe Val Gly Gln Glu His Leu Val His Leu Val Asn Gln Leu Gln Gln
260 265 270
Asp Met Lys Glu Ala Ser Pro Gly Leu Glu Ser Lys Glu Gln Thr Ala

275 280 285
His Tyr Val Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu
290 295 300
Lys Trp Gly Lys Leu Ala Pro Asp Ala Pro Phe Asp Leu Tyr Asp Ala

305 310 315 320

Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His

325 330 335

Asp Leu Phe Ala Lys Leu Ala Glu Pro Glu Tyr Gln Ala Leu Trp Arg

340 345 350

Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Leu

355 360 365

Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp

370 375 380

Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn

385 390 395 400

Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Arg Arg His

405 410 415

Ala Ile Arg Phe His Lys Leu Leu Lys Val Glu Asn Gly Val Ala Arg

420 425 430

Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Glu Gln Leu Asp

435 440 445

Asn Leu Leu Pro Arg Asp Pro Asn Glu Pro Ile Ala Leu Tyr Phe Arg

450 455 460

Asp Tyr Gly Ala Glu Gln His Phe Thr Gly Glu Phe Gly Gly Ala Lys

465 470 475 480

Ile Gln Cys Arg Arg Asp Gln Leu Ala His Met His Arg Arg Arg Gly

485 490 495

Ala Arg Asp Val Tyr Leu Asn Val Ser Val Arg Val Gln Ser Gln Ser

500 505 510

Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu

515 520 525

Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp

530 535 540

Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu

545 550 555 560

Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser
565 570 575
Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro
580 585 590
Asn Ser Lys Gly Arg Val Pro Phe Phe Phe Pro Ile Lys Gly Asn Asp

595 600 605
Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly
610 615 620
Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg
625 630 635 640
Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val
645 650 655
Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys

660 665 670
Leu Ile Glu Gln Pro Val Asp Ala Ala Asn His Met Thr Pro Asp Trp
675 680 685
Arg Glu Ala Phe Glu Asn Glu Leu Gln Lys Leu Lys Ser Leu His Gly
690 695 700
Ile Cys Ser Asp Lys Glu Trp Met Asp Ala Val Tyr Glu Ser Val Arg
705 710 715 720
Arg Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp

725 730 735
Val Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Ala Lys Asp Val
740 745 750
Val Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr
755 760 765
Lys Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val
770 775 780
Ile Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His

785 790 795 800
Ile Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile

805 810 815

Ile Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Glu Arg Gly Lys

820 825 830

Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu

835 840 845

Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn

850 855 860

Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Ile

865 870 875 880

Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala

885 890 895

Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys

900 905 910

Arg Arg Val Pro Ala Arg Cys Thr Gln Glu His Asn Pro Glu Pro Phe

915 920 925

Pro Trp Trp Leu Asn Lys Phe Val Val Glu His Thr Leu Asp Ala Cys

930 935 940

Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Ile Phe

945 950 955 960

Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala

965 970 975

Asp Leu Asn Ala Ala Gln Asn Leu Gln Gln Arg Leu Trp Ser Asp Phe

980 985 990

Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp Gly

995 1000 1005

Glu Leu Val Leu Ile Pro Arg Leu Thr Gly Lys Arg Thr Ala Asp

1010 1015 1020

Ser Tyr Ser Asn Lys Val Phe Tyr Thr Asn Thr Gly Val Thr Tyr

1025 1030 1035

Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala Gln

1040 1045 1050

Glu Lys Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala Asp
1055 1060 1065
Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser Gly
1070 1075 1080
Ile Ile Asn Arg Gly Asn Trp Thr Arg Gln Lys Glu Phe Trp Ser
1085 1090 1095
Met Val Asn Gln Arg Ile Glu Gly Tyr Leu Val Lys Gln Ile Arg

1100 1105 1110
Ser Arg Val Pro Leu Gln Asp Ser Ala Cys Glu Asn Thr Gly Asp
1115 1120 1125
Ile

<210> 259
<211> 1129
<212> PRT
<213> Alicyclobacillus acidiphilus
<400> 259
Met Ala Val Lys Ser Met Lys Val Lys Leu Arg Leu Asp Asn Met Pro
1 5 10 15
Glu Ile Arg Ala Gly Leu Trp Lys Leu His Thr Glu Val Asn Ala Gly
20 25 30

Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu
35 40 45
Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Tyr Lys Thr
50 55 60
Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln
65 70 75 80
Val Glu Asn Gly His Cys Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu
85 90 95

Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly
100 105 110
Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu

115 120 125

Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn

130 135 140

Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu

145 150 155 160

Glu Glu Lys Ala Lys Ala Glu Ala Arg Lys Ser Thr Asp Arg Thr Ala

165 170 175

Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg

180 185 190

Val Tyr Thr Asp Ser Asp Met Ser Ser Val Gln Trp Lys Pro Leu Arg

195 200 205

Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala

210 215 220

Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Glu

225 230 235 240

Ala Tyr Ala Lys Leu Val Glu Gln Lys Ser Arg Phe Glu Gln Lys Asn

245 250 255

Phe Val Gly Gln Glu His Leu Val Gln Leu Val Asn Gln Leu Gln Gln

260 265 270

Asp Met Lys Glu Ala Ser His Gly Leu Glu Ser Lys Glu Gln Thr Ala

275 280 285

His Tyr Leu Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu

290 295 300

Lys Trp Glu Lys Leu Asp Pro Asp Ala Pro Phe Asp Leu Tyr Asp Thr

305 310 315 320

Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His

325 330 335

Asp Leu Phe Ala Lys Leu Ala Glu Pro Lys Tyr Gln Ala Leu Trp Arg

340 345 350

Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Val

355 360 365

Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp
370 375 380
Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn
385 390 395 400
Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Gly Arg His
405 410 415

Ala Ile Arg Phe Gln Lys Leu Leu Thr Val Glu Asp Gly Val Ala Lys
420 425 430
Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Ala Gln Leu Asp
435 440 445
Asp Leu Leu Pro Arg Asp Pro His Glu Leu Val Ala Leu Tyr Phe Gln
450 455 460
Asp Tyr Gly Ala Glu Gln His Leu Ala Gly Glu Phe Gly Gly Ala Lys
465 470 475 480

Ile Gln Tyr Arg Arg Asp Gln Leu Asn His Leu His Ala Arg Arg Gly
485 490 495
Ala Arg Asp Val Tyr Leu Asn Leu Ser Val Arg Val Gln Ser Gln Ser
500 505 510
Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu
515 520 525
Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp
530 535 540

Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu
545 550 555 560
Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser
565 570 575
Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro
580 585 590
Asn Ser Glu Gly Arg Val Pro Phe Cys Phe Pro Ile Glu Gly Asn Glu
595 600 605

Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly

610 615 620

Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg

625 630 635 640

Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val

645 650 655

Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys

660 665 670

Leu Ile Glu Gln Pro Met Asp Ala Asn Gln Met Thr Pro Asp Trp Arg

675 680 685

Glu Ala Phe Glu Asp Glu Leu Gln Lys Leu Lys Ser Leu Tyr Gly Ile

690 695 700

Cys Gly Asp Arg Glu Trp Thr Glu Ala Val Tyr Glu Ser Val Arg Arg

705 710 715 720

Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp Val

725 730 735

Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Gln Lys Asp Val Val

740 745 750

Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr Lys

755 760 765

Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val Ile

770 775 780

Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His Ile

785 790 795 800

Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile Ile

805 810 815

Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Asp Glu Arg Gly Lys

820 825 830

Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu

835 840 845

Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn

850 855 860

Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Leu
865 870 875 880
Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala
885 890 895
Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys
900 905 910
Arg Arg Val Pro Ala Arg Cys Ala Arg Glu Gln Asn Pro Glu Pro Phe
915 920 925

Pro Trp Trp Leu Asn Lys Phe Val Ala Glu His Lys Leu Asp Gly Cys
930 935 940
Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Phe Phe
945 950 955 960
Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala
965 970 975
Asp Leu Asn Ala Ala Gln Asn Leu Gln Arg Arg Leu Trp Ser Asp Phe
980 985 990

Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp Gly
995 1000 1005
Glu Pro Val Leu Ile Pro Arg Thr Thr Gly Lys Arg Thr Ala Asp
1010 1015 1020
Ser Tyr Gly Asn Lys Val Phe Tyr Thr Lys Thr Gly Val Thr Tyr
1025 1030 1035
Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala Gln
1040 1045 1050

Glu Glu Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala Asp
1055 1060 1065
Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser Gly
1070 1075 1080
Ile Ile Asn Arg Gly Asp Trp Thr Arg Gln Lys Glu Phe Trp Ser
1085 1090 1095
Met Val Asn Gln Arg Ile Glu Gly Tyr Leu Val Lys Gln Ile Arg

1100 1105 1110

Ser Arg Val Arg Leu Gln Glu Ser Ala Cys Glu Asn Thr Gly Asp

1115 1120 1125

Ile

<210> 260

<211> 1140

<212> PRT

<213> Bacillus hisashii

<400> 260

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala

1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val

20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala

35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu

50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu

65 70 75 80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn

85 90 95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg

100 105 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu

115 120 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn

130 135 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp

145 150 155 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys

165 170 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly

180 185 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp

195 200 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg

210 215 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

225 230 235 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

245 250 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

260 265 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg

275 280 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu

290 295 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

305 310 315 320

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys

325 330 335

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr

340 345 350

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

355 360 365

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

370 375 380

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn

385 390 395 400

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn

405 410 415

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys

420 425 430

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

435 440 445

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg

450 455 460

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His

465 470 475 480

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr

485 490 495

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

500 505 510

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met

515 520 525

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys

530 535 540

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu

545 550 555 560

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

565 570 575

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly

580 585 590

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys

595 600 605

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu

610 615 620

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

625 630 635 640

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu

645 650 655

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

660 665 670

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

675 680 685

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

690 695 700

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

705 710 715 720

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

725 730 735

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

740 745 750

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

755 760 765

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

770 775 780

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu

785 790 795 800

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

805 810 815

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln

820 825 830

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn

835 840 845

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met

850 855 860

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu

865 870 875 880

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg

885 890 895

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr

900 905 910

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu

915 920 925

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu

930 935 940

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg

945 950 955 960

Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln

965 970 975

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys

980 985 990

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp

995 1000 1005

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile

1010 1015 1020

Leu Lys Asp Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys

1025 1030 1035

Ile Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp

1040 1045 1050

Ser Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys

1055 1060 1065

Gly Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe

1070 1075 1080

Pro Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu

1085 1090 1095

Glu Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser

1100 1105 1110

Thr Ile Glu Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala

1115 1120 1125

Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Lys

1130 1135 1140

<210> 261

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 261

gggaaataag agagaaaaga agagtaagaa gaaatataag agccacc 47

<210> 262

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 262

gctggagcct cggtggccat gcttcttgcc ccttgggcct ccccccagcc cctcctcccc 60

ttcctgcacc cgtacccccg tggtctttga ataaagtctg a 101

<210> 263

<211> 3420

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 263

atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc cgccaccaga 60

tccttcatcc tgaagatcga gcccaacgag gaagtgaaga aaggcctctg gaaaacccac 120

gaggtgctga accacggaat cgcctactac atgaatatcc tgaagctgat ccggcaagag 180

gccatctacg agcaccacga gcaggacccc aagaatccca agaaggtgtc caaggccgag 240

atccaggccg agctgtggga tttcgtgctg aagatgcaga agtgcaacag cttcacacac 300

gaggtggaca aggacgaggt gttcaacatc ctgagagagc tgtacgagga actggtgccc 360

agcagcgtgg aaaagaaggg cgaagccaac cagctgagca acaagtttct gtaccctctg 420

gtggacccca acagccagtc tggaaaggga acagccagca gcggcagaaa gcccagatgg 480

tacaacctga agattgccgg cgatccctcc tgggaagaag agaagaagaa gtgggaagaa 540

gataagaaaa aggacccgct ggccaagatc ctgggcaagc tggctgagta cggactgatc 600

cctctgttca tcccctacac cgacagcaac gagcccatcg tgaaagaaat caagtggatg 660
gaaaagtccc ggaaccagag cgtgcggcgg ctggataagg acatgttcat tcaggccctg 720
gaacggttcc tgagctggga gagctggaac ctgaaagtga aagaggaata cgagaaggtc 780
gagaaagagt acaagaccct ggaagagagg atcaaagagg acatccaggc tctgaaggct 840
ctggaacagt atgagaaaga gcggcaagaa cagctgctgc gggacaccct gaacaccaac 900
gagtaccggc tgagcaagag aggccttaga ggctggcggg aaatcatcca gaaatggctg 960
aaaatggacg agaacgagcc ctccgagaag tacctggaag tgttcaagga ctaccagcgg 1020

aagcacccta gagaggccgg cgattacagc gtgtacgagt tcctgtccaa gaaagagaac 1080
cacttcatct ggcggaatca ccctgagtac ccctacctgt acgccacctt ctgcgagatc 1140
gacaagaaaa agaaggacgc caagcagcag gccaccttca cactggccga tcctatcaat 1200
caccctctgt gggtccgatt cgaggaaaga agcggcagca acctgaacaa gtacagaatc 1260
ctgaccgagc agctgcacac cgagaagctg aagaaaaagc tgacagtgca gctggaccgg 1320
ctgatctacc ctacagaatc tggcggctgg gaagagaagg gcaaagtgga cattgtgctg 1380
ctgcccagcc ggcagttcta caaccagatc ttcctggaca tcgaggaaaa gggcaagcac 1440

gccttcacct acaaggatga gagcatcaag ttccctctga agggcacact cggcggagcc 1500
agagtgcagt tcgacagaga tcacctgaga agataccctc acaaggtgga aagcggcaac 1560
gtgggcagaa tctacttcaa catgaccgtg aacatcgagc ctacagagtc cccagtgtcc 1620
aagtctctga agatccaccg ggacgacttc cccaaggtgg tcaacttcaa gcccaaagaa 1680
ctgaccgagt ggatcaagga cagcaagggc aagaaactga agtccggcat cgagtccctg 1740
gaaatcggcc tgagagtgat gagcatcgac ctgggacaga gacaggccgc tgccgcctct 1800
attttcgagg tggtggatca gaagcccgac atcgaaggca agctgttttt cccaatcaag 1860

ggcaccgagc tgtatgccgt gcacagagcc agcttcaaca tcaagctgcc cggcgagaca 1920
ctggtcaaga gcagagaagt gctgcggaag gccagagagg acaatctgaa actgatgaac 1980
cagaagctca acttcctgcg gaacgtgctg cacttccagc agttcgagga catcaccgag 2040
agagagaagc gggtcaccaa gtggatcagc agacaagaga acagcgacgt gcccctggtg 2100
taccaggatg agctgatcca gatccgcgag ctgatgtaca agccttacaa ggactgggtc 2160
gccttcctga agcagctcca caagagactg gaagtcgaga tcggcaaaga agtgaagcac 2220
tggcggaagt ccctgagcga cggaagaaag ggcctgtacg gcatctccct gaagaacatc 2280

gacgagatcg atcggacccg gaagttcctg ctgagatggt ccctgaggcc taccgaacct 2340

ggcgaagtgc gtagactgga acccggccag agattcgcca tcgaccagct gaatcacctg 2400

aacgccctga aagaagatcg gctgaagaag atggccaaca ccatcatcat gcacgccctg 2460

ggctactgct acgacgtgcg gaagaagaaa tggcaggcta agaaccccgc ctgccagatc 2520

atcctgttcg aggatctgag caactacaac ccctacgagg aaaggtcccg cttcgagaac 2580

agcaagctca tgaagtggtc cagacgcgag atccccagac aggttgcact gcagggcgag 2640

atctatggcc tgcaagtggg agaagtgggc gctcagttca gcagcagatt ccacgccaag 2700

acaggcagcc ctggcatcag atgtagcgtc gtgaccaaag agaagctgca ggacaatcgg 2760

ttcttcaaga atctgcagag agagggcaga ctgaccctgg acaaaatcgc cgtgctgaaa 2820

gagggcgatc tgtacccaga caaaggcggc gagaagttca tcagcctgag caaggatcgg 2880

aagtgcgtga ccacacacgc cgacatcaac gccgctcaga acctgcagaa gcggttctgg 2940

acaagaaccc acggcttcta caaggtgtac tgcaaggcct accaggtgga cggccagacc 3000

gtgtacatcc ctgagagcaa ggaccagaag cagaagatca tcgaagagtt cggcgagggc 3060

tacttcattc tgaaggacgg ggtgtacgaa tgggtcaacg ccggcaagct gaaaatcaag 3120

aagggcagct ccaagcagag cagcagcgag ctggtggata gcgacatcct gaaagacagc 3180

ttcgacctgg cctccgagct gaaaggcgaa aagctgatgc tgtacaggga ccccagcggc 3240

aatgtgttcc ccagcgacaa atggatggcc gctggcgtgt tcttcggaaa gctggaacgc 3300

atcctgatca gcaagctgac caaccagtac tccatcagca ccatcgagga cgacagcagc 3360

aagcagtcta tgaaaaggcc ggcggccacg aaaaaggccg gccaggcaaa aaagaaaaag 3420

<210> 264

<211> 1112

<212> PRT

<213> Bacillus sp.

<400> 264

Met Ala Ile Arg Ser Ile Lys Leu Lys Met Lys Thr Asn Ser Gly Thr

1 5 10 15

Asp Ser Ile Tyr Leu Arg Lys Ala Leu Trp Arg Thr His Gln Leu Ile

20 25 30

Asn Glu Gly Ile Ala Tyr Tyr Met Asn Leu Leu Thr Leu Tyr Arg Gln

35 40 45

Glu Ala Ile Gly Asp Lys Thr Lys Glu Ala Tyr Gln Ala Glu Leu Ile

50 55 60

Asn Ile Ile Arg Asn Gln Gln Arg Asn Asn Gly Ser Ser Glu Glu His

65 70 75 80

Gly Ser Asp Gln Glu Ile Leu Ala Leu Leu Arg Gln Leu Tyr Glu Leu

85 90 95

Ile Ile Pro Ser Ser Ile Gly Glu Ser Gly Asp Ala Asn Gln Leu Gly

100 105 110

Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn Ser Gln Ser Gly Lys

115 120 125

Gly Thr Ser Asn Ala Gly Arg Lys Pro Arg Trp Lys Arg Leu Lys Glu

130 135 140

Glu Gly Asn Pro Asp Trp Glu Leu Glu Lys Lys Lys Asp Glu Glu Arg

145 150 155 160

Lys Ala Lys Asp Pro Thr Val Lys Ile Phe Asp Asn Leu Asn Lys Tyr

165 170 175

Gly Leu Leu Pro Leu Phe Pro Leu Phe Thr Asn Ile Gln Lys Asp Ile

180 185 190

Glu Trp Leu Pro Leu Gly Lys Arg Gln Ser Val Arg Lys Trp Asp Lys

195 200 205

Asp Met Phe Ile Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu Ser Trp

210 215 220

Asn Arg Arg Val Ala Asp Glu Tyr Lys Gln Leu Lys Glu Lys Thr Glu

225 230 235 240

Ser Tyr Tyr Lys Glu His Leu Thr Gly Gly Glu Glu Trp Ile Glu Lys

245 250 255

Ile Arg Lys Phe Glu Lys Glu Arg Asn Met Glu Leu Glu Lys Asn Ala

260 265 270

Phe Ala Pro Asn Asp Gly Tyr Phe Ile Thr Ser Arg Gln Ile Arg Gly

275 280 285

Trp Asp Arg Val Tyr Glu Lys Trp Ser Lys Leu Pro Glu Ser Ala Ser

290 295 300

Pro Glu Glu Leu Trp Lys Val Val Ala Glu Gln Gln Asn Lys Met Ser

305 310 315 320

Glu Gly Phe Gly Asp Pro Lys Val Phe Ser Phe Leu Ala Asn Arg Glu

325 330 335

Asn Arg Asp Ile Trp Arg Gly His Ser Glu Arg Ile Tyr His Ile Ala

340 345 350

Ala Tyr Asn Gly Leu Gln Lys Lys Leu Ser Arg Thr Lys Glu Gln Ala

355 360 365

Thr Phe Thr Leu Pro Asp Ala Ile Glu His Pro Leu Trp Ile Arg Tyr

370 375 380

Glu Ser Pro Gly Gly Thr Asn Leu Asn Leu Phe Lys Leu Glu Glu Lys

385 390 395 400

Gln Lys Lys Asn Tyr Tyr Val Thr Leu Ser Lys Ile Ile Trp Pro Ser

405 410 415

Glu Glu Lys Trp Ile Glu Lys Glu Asn Ile Glu Ile Pro Leu Ala Pro

420 425 430

Ser Ile Gln Phe Asn Arg Gln Ile Lys Leu Lys Gln His Val Lys Gly

435 440 445

Lys Gln Glu Ile Ser Phe Ser Asp Tyr Ser Ser Arg Ile Ser Leu Asp

450 455 460

Gly Val Leu Gly Gly Ser Arg Ile Gln Phe Asn Arg Lys Tyr Ile Lys

465 470 475 480

Asn His Lys Glu Leu Leu Gly Glu Gly Asp Ile Gly Pro Val Phe Phe

485 490 495

Asn Leu Val Val Asp Val Ala Pro Leu Gln Glu Thr Arg Asn Gly Arg

500 505 510

Leu Gln Ser Pro Ile Gly Lys Ala Leu Lys Val Ile Ser Ser Asp Phe

515 520 525

Ser Lys Val Ile Asp Tyr Lys Pro Lys Glu Leu Met Asp Trp Met Asn

530 535 540

Thr Gly Ser Ala Ser Asn Ser Phe Gly Val Ala Ser Leu Leu Glu Gly

545 550 555 560

Met Arg Val Met Ser Ile Asp Met Gly Gln Arg Thr Ser Ala Ser Val
565 570 575
Ser Ile Phe Glu Val Val Lys Glu Leu Pro Lys Asp Gln Glu Gln Lys

580 585 590
Leu Phe Tyr Ser Ile Asn Asp Thr Glu Leu Phe Ala Ile His Lys Arg
595 600 605
Ser Phe Leu Leu Asn Leu Pro Gly Glu Val Val Thr Lys Asn Asn Lys
610 615 620
Gln Gln Arg Gln Glu Arg Arg Lys Lys Arg Gln Phe Val Arg Ser Gln
625 630 635 640
Ile Arg Met Leu Ala Asn Val Leu Arg Leu Glu Thr Lys Lys Thr Pro

645 650 655
Asp Glu Arg Lys Lys Ala Ile His Lys Leu Met Glu Ile Val Gln Ser
660 665 670
Tyr Asp Ser Trp Thr Ala Ser Gln Lys Glu Val Trp Glu Lys Glu Leu
675 680 685
Asn Leu Leu Thr Asn Met Ala Ala Phe Asn Asp Glu Ile Trp Lys Glu
690 695 700
Ser Leu Val Glu Leu His His Arg Ile Glu Pro Tyr Val Gly Gln Ile

705 710 715 720
Val Ser Lys Trp Arg Lys Gly Leu Ser Glu Gly Arg Lys Asn Leu Ala
725 730 735
Gly Ile Ser Met Trp Asn Ile Asp Glu Leu Glu Asp Thr Arg Arg Leu
740 745 750
Leu Ile Ser Trp Ser Lys Arg Ser Arg Thr Pro Gly Glu Ala Asn Arg
755 760 765
Ile Glu Thr Asp Glu Pro Phe Gly Ser Ser Leu Leu Gln His Ile Gln

770 775 780
Asn Val Lys Asp Asp Arg Leu Lys Gln Met Ala Asn Leu Ile Ile Met
785 790 795 800
Thr Ala Leu Gly Phe Lys Tyr Asp Lys Glu Glu Lys Asp Arg Tyr Lys

805 810 815

Arg Trp Lys Glu Thr Tyr Pro Ala Cys Gln Ile Ile Leu Phe Glu Asn

820 825 830

Leu Asn Arg Tyr Leu Phe Asn Leu Asp Arg Ser Arg Arg Glu Asn Ser

835 840 845

Arg Leu Met Lys Trp Ala His Arg Ser Ile Pro Arg Thr Val Ser Met

850 855 860

Gln Gly Glu Met Phe Gly Leu Gln Val Gly Asp Val Arg Ser Glu Tyr

865 870 875 880

Ser Ser Arg Phe His Ala Lys Thr Gly Ala Pro Gly Ile Arg Cys His

885 890 895

Ala Leu Thr Glu Glu Asp Leu Lys Ala Gly Ser Asn Thr Leu Lys Arg

900 905 910

Leu Ile Glu Asp Gly Phe Ile Asn Glu Ser Glu Leu Ala Tyr Leu Lys

915 920 925

Lys Gly Asp Ile Ile Pro Ser Gln Gly Gly Glu Leu Phe Val Thr Leu

930 935 940

Ser Lys Arg Tyr Lys Lys Asp Ser Asp Asn Asn Glu Leu Thr Val Ile

945 950 955 960

His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln

965 970 975

Gln Asn Ser Glu Val Tyr Arg Val Pro Cys Gln Leu Ala Arg Met Gly

980 985 990

Glu Asp Lys Leu Tyr Ile Pro Lys Ser Gln Thr Glu Thr Ile Lys Lys

995 1000 1005

Tyr Phe Gly Lys Gly Ser Phe Val Lys Asn Asn Thr Glu Gln Glu

1010 1015 1020

Val Tyr Lys Trp Glu Lys Ser Glu Lys Met Lys Ile Lys Thr Asp

1025 1030 1035

Thr Thr Phe Asp Leu Gln Asp Leu Asp Gly Phe Glu Asp Ile Ser

1040 1045 1050

Lys Thr Ile Glu Leu Ala Gln Glu Gln Gln Lys Lys Tyr Leu Thr
1055 1060 1065
Met Phe Arg Asp Pro Ser Gly Tyr Phe Phe Asn Asn Glu Thr Trp
1070 1075 1080
Arg Pro Gln Lys Glu Tyr Trp Ser Ile Val Asn Asn Ile Ile Lys

1085 1090 1095
Ser Cys Leu Lys Lys Lys Ile Leu Ser Asn Lys Val Glu Leu
1100 1105 1110
<210> 265
<211> 1108
<212> PRT
<213> Bacillus thermoamylovorans
<400> 265
Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
1 5 10 15
Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
35 40 45
His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
50 55 60
Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
65 70 75 80
Ser Phe Thr His Glu Val Asp Lys Asp Val Val Phe Asn Ile Leu Arg
85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
100 105 110
Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
115 120 125
Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
130 135 140
Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys

145 150 155 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
165 170 175
Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Phe Thr Asp
180 185 190
Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
195 200 205
Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
210 215 220

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
225 230 235 240
Tyr Glu Lys Val Glu Lys Glu His Lys Thr Leu Glu Glu Arg Ile Lys
245 250 255
Glu Asp Ile Gln Ala Phe Lys Ser Leu Glu Gln Tyr Glu Lys Glu Arg
260 265 270
Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
275 280 285

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
290 295 300
Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
305 310 315 320
Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
325 330 335
Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
340 345 350

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
355 360 365
Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
370 375 380
His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
385 390 395 400

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
405 410 415

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
420 425 430

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
435 440 445

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
450 455 460

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
465 470 475 480

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
485 490 495

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
500 505 510

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
515 520 525

Ile His Arg Asp Asp Phe Pro Lys Phe Val Asn Phe Lys Pro Lys Glu
530 535 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
545 550 555 560

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
565 570 575

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
580 585 590

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
595 600 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
610 615 620

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625 630 635 640

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

645 650 655

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

660 665 670

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

675 680 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

690 695 700

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

705 710 715 720

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

725 730 735

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

740 745 750

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

755 760 765

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu

770 775 780

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

785 790 795 800

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln

805 810 815

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn

820 825 830

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met

835 840 845

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu

850 855 860

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg

865 870 875 880

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr

885 890 895

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
900 905 910
Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
915 920 925

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
930 935 940
Lys Leu Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
945 950 955 960
Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
965 970 975
Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp
980 985 990

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu
995 1000 1005
Lys Asp Gly Val Tyr Glu Trp Gly Asn Ala Gly Lys Leu Lys Ile
1010 1015 1020
Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser
1025 1030 1035
Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly
1040 1045 1050

Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro
1055 1060 1065
Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu
1070 1075 1080
Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr
1085 1090 1095
Ile Glu Asp Asp Ser Ser Lys Gln Ser Met
1100 1105
<210> 266
<211> 1302
<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12c1 sequence

<400> 266

Met Gln Thr Lys Lys Thr His Leu His Leu Ile Ser Ala Lys Ala Ser
1 5 10 15

Arg Lys Tyr Arg Arg Thr Ile Ala Cys Leu Ser Asp Thr Ala Lys Lys
20 25 30

Asp Leu Glu Arg Arg Lys Gln Ser Gly Ala Ala Asp Pro Ala Gln Glu
35 40 45

Leu Ser Cys Leu Lys Thr Ile Lys Phe Lys Leu Glu Val Pro Glu Gly

50 55 60

Ser Lys Leu Pro Ser Phe Asp Arg Ile Ser Gln Ile Tyr Asn Ala Leu
65 70 75 80

Glu Thr Ile Glu Lys Gly Ser Leu Ser Tyr Leu Leu Phe Ala Leu Ile
85 90 95

Leu Ser Gly Phe Arg Ile Phe Pro Asn Ser Ser Ala Ala Lys Thr Phe
100 105 110

Ala Ser Ser Ser Cys Tyr Lys Asn Asp Gln Phe Ala Ser Gln Ile Lys

115 120 125

Glu Ile Phe Gly Glu Met Val Lys Asn Phe Ile Pro Ser Glu Leu Glu
130 135 140

Ser Ile Leu Lys Lys Gly Arg Arg Lys Asn Asn Lys Asp Trp Thr Glu
145 150 155 160

Glu Asn Ile Lys Arg Val Leu Asn Ser Glu Phe Gly Arg Lys Asn Ser
165 170 175

Glu Gly Ser Ser Ala Leu Phe Asp Ser Phe Leu Ser Lys Phe Ser Gln

180 185 190

Glu Leu Phe Arg Lys Phe Asp Ser Trp Asn Glu Val Asn Lys Lys Tyr
195 200 205

Leu Glu Ala Ala Glu Leu Leu Asp Ser Met Leu Ala Ser Tyr Gly Pro
210 215 220

Phe Asp Ser Val Cys Lys Met Ile Gly Asp Ser Asp Ser Arg Asn Ser
225 230 235 240
Leu Pro Asp Lys Ser Thr Ile Ala Phe Thr Asn Asn Ala Glu Ile Thr

245 250 255
Val Asp Ile Glu Ser Ser Val Met Pro Tyr Met Ala Ile Ala Ala Leu
260 265 270
Leu Arg Glu Tyr Arg Gln Ser Lys Ser Lys Ala Ala Pro Val Ala Tyr
275 280 285
Val Gln Ser His Leu Thr Thr Thr Asn Gly Asn Gly Leu Ser Trp Phe
290 295 300
Phe Lys Phe Gly Leu Asp Leu Ile Arg Lys Ala Pro Val Ser Ser Lys

305 310 315 320
Gln Ser Thr Ser Asp Gly Ser Lys Ser Leu Gln Glu Leu Phe Ser Val
325 330 335
Pro Asp Asp Lys Leu Asp Gly Leu Lys Phe Ile Lys Glu Ala Cys Glu
340 345 350
Ala Leu Pro Glu Ala Ser Leu Leu Cys Gly Glu Lys Gly Glu Leu Leu
355 360 365
Gly Tyr Gln Asp Phe Arg Thr Ser Phe Ala Gly His Ile Asp Ser Trp

370 375 380
Val Ala Asn Tyr Val Asn Arg Leu Phe Glu Leu Ile Glu Leu Val Asn
385 390 395 400
Gln Leu Pro Glu Ser Ile Lys Leu Pro Ser Ile Leu Thr Gln Lys Asn
405 410 415
His Asn Leu Val Ala Ser Leu Gly Leu Gln Glu Ala Glu Val Ser His
420 425 430
Ser Leu Glu Leu Phe Glu Gly Leu Val Lys Asn Val Arg Gln Thr Leu

435 440 445
Lys Lys Leu Ala Gly Ile Asp Ile Ser Ser Ser Pro Asn Glu Gln Asp
450 455 460
Ile Lys Glu Phe Tyr Ala Phe Ser Asp Val Leu Asn Arg Leu Gly Ser

465 470 475 480

Ile Arg Asn Gln Ile Glu Asn Ala Val Gln Thr Ala Lys Lys Asp Lys

485 490 495

Ile Asp Leu Glu Ser Ala Ile Glu Trp Lys Glu Trp Lys Lys Leu Lys

500 505 510

Lys Leu Pro Lys Leu Asn Gly Leu Gly Gly Gly Val Pro Lys Gln Gln

515 520 525

Glu Leu Leu Asp Lys Ala Leu Glu Ser Val Lys Gln Ile Arg His Tyr

530 535 540

Gln Arg Ile Asp Phe Glu Arg Val Ile Gln Trp Ala Val Asn Glu His

545 550 555 560

Cys Leu Glu Thr Val Pro Lys Phe Leu Val Asp Ala Glu Lys Lys Lys

565 570 575

Ile Asn Lys Glu Ser Ser Thr Asp Phe Ala Ala Lys Glu Asn Ala Val

580 585 590

Arg Phe Leu Leu Glu Gly Ile Gly Ala Ala Ala Arg Gly Lys Thr Asp

595 600 605

Ser Val Ser Lys Ala Ala Tyr Asn Trp Phe Val Val Asn Asn Phe Leu

610 615 620

Ala Lys Lys Asp Leu Asn Arg Tyr Phe Ile Asn Cys Gln Gly Cys Ile

625 630 635 640

Tyr Lys Pro Pro Tyr Ser Lys Arg Arg Ser Leu Ala Phe Ala Leu Arg

645 650 655

Ser Asp Asn Lys Asp Thr Ile Glu Val Val Trp Glu Lys Phe Glu Thr

660 665 670

Phe Tyr Lys Glu Ile Ser Lys Glu Ile Glu Lys Phe Asn Ile Phe Ser

675 680 685

Gln Glu Phe Gln Thr Phe Leu His Leu Glu Asn Leu Arg Met Lys Leu

690 695 700

Leu Leu Arg Arg Ile Gln Lys Pro Ile Pro Ala Glu Ile Ala Phe Phe

705 710 715 720

Ser Leu Pro Gln Glu Tyr Tyr Asp Ser Leu Pro Pro Asn Val Ala Phe
725 730 735
Leu Ala Leu Asn Gln Glu Ile Thr Pro Ser Glu Tyr Ile Thr Gln Phe
740 745 750
Asn Leu Tyr Ser Ser Phe Leu Asn Gly Asn Leu Ile Leu Leu Arg Arg

755 760 765
Ser Arg Ser Tyr Leu Arg Ala Lys Phe Ser Trp Val Gly Asn Ser Lys
770 775 780
Leu Ile Tyr Ala Ala Lys Glu Ala Arg Leu Trp Lys Ile Pro Asn Ala
785 790 795 800
Tyr Trp Lys Ser Asp Glu Trp Lys Met Ile Leu Asp Ser Asn Val Leu
805 810 815
Val Phe Asp Lys Ala Gly Asn Val Leu Pro Ala Pro Thr Leu Lys Lys

820 825 830
Val Cys Glu Arg Glu Gly Asp Leu Arg Leu Phe Tyr Pro Leu Leu Arg
835 840 845
Gln Leu Pro His Asp Trp Cys Tyr Arg Asn Pro Phe Val Lys Ser Val
850 855 860
Gly Arg Glu Lys Asn Val Ile Glu Val Asn Lys Glu Gly Glu Pro Lys
865 870 875 880
Val Ala Ser Ala Leu Pro Gly Ser Leu Phe Arg Leu Ile Gly Pro Ala

885 890 895
Pro Phe Lys Ser Leu Leu Asp Asp Cys Phe Phe Asn Pro Leu Asp Lys
900 905 910
Asp Leu Arg Glu Cys Met Leu Ile Val Asp Gln Glu Ile Ser Gln Lys
915 920 925
Val Glu Ala Gln Lys Val Glu Ala Ser Leu Glu Ser Cys Thr Tyr Ser
930 935 940
Ile Ala Val Pro Ile Arg Tyr His Leu Glu Glu Pro Lys Val Ser Asn

945 950 955 960
Gln Phe Glu Asn Val Leu Ala Ile Asp Gln Gly Glu Ala Gly Leu Ala

965 970 975

Tyr Ala Val Phe Ser Leu Lys Ser Ile Gly Glu Ala Glu Thr Lys Pro

980 985 990

Ile Ala Val Gly Thr Ile Arg Ile Pro Ser Ile Arg Arg Leu Ile His

995 1000 1005

Ser Val Ser Thr Tyr Arg Lys Lys Lys Gln Arg Leu Gln Asn Phe

1010 1015 1020

Lys Gln Asn Tyr Asp Ser Thr Ala Phe Ile Met Arg Glu Asn Val

1025 1030 1035

Thr Gly Asp Val Cys Ala Lys Ile Val Gly Leu Met Lys Glu Phe

1040 1045 1050

Asn Ala Phe Pro Val Leu Glu Tyr Asp Val Lys Asn Leu Glu Ser

1055 1060 1065

Gly Ser Arg Gln Leu Ser Ala Val Tyr Lys Ala Val Asn Ser His

1070 1075 1080

Phe Leu Tyr Phe Lys Glu Pro Gly Arg Asp Ala Leu Arg Lys Gln

1085 1090 1095

Leu Trp Tyr Gly Gly Asp Ser Trp Thr Ile Asp Gly Ile Glu Ile

1100 1105 1110

Val Thr Arg Glu Arg Lys Glu Asp Gly Lys Glu Gly Val Glu Lys

1115 1120 1125

Ile Val Pro Leu Lys Val Phe Pro Gly Arg Ser Val Ser Ala Arg

1130 1135 1140

Phe Thr Ser Lys Thr Cys Ser Cys Cys Gly Arg Asn Val Phe Asp

1145 1150 1155

Trp Leu Phe Thr Glu Lys Lys Ala Lys Thr Asn Lys Lys Phe Asn

1160 1165 1170

Val Asn Ser Lys Gly Glu Leu Thr Thr Ala Asp Gly Val Ile Gln

1175 1180 1185

Leu Phe Glu Ala Asp Arg Ser Lys Gly Pro Lys Phe Tyr Ala Arg

1190 1195 1200

Arg Lys Glu Arg Thr Pro Leu Thr Lys Pro Ile Ala Lys Gly Ser
1205 1210 1215
Tyr Ser Leu Glu Glu Ile Glu Arg Arg Val Arg Thr Asn Leu Arg
1220 1225 1230
Arg Ala Pro Lys Ser Lys Gln Ser Arg Asp Thr Ser Gln Ser Gln
1235 1240 1245
Tyr Phe Cys Val Tyr Lys Asp Cys Ala Leu His Phe Ser Gly Met

1250 1255 1260
Gln Ala Asp Glu Asn Ala Ala Ile Asn Ile Gly Arg Arg Phe Leu
1265 1270 1275
Thr Ala Leu Arg Lys Asn Arg Arg Ser Asp Phe Pro Ser Asn Val
1280 1285 1290
Lys Ile Ser Asp Arg Leu Leu Asp Asn
1295 1300

<210> 267
<211> 1218
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
Cas12c2 sequence

<400> 267
Met Thr Lys His Ser Ile Pro Leu His Ala Phe Arg Asn Ser Gly Ala
1 5 10 15
Asp Ala Arg Lys Trp Lys Gly Arg Ile Ala Leu Leu Ala Lys Arg Gly
20 25 30
Lys Glu Thr Met Arg Thr Leu Gln Phe Pro Leu Glu Met Ser Glu Pro
35 40 45
Glu Ala Ala Ala Ile Asn Thr Thr Pro Phe Ala Val Ala Tyr Asn Ala
50 55 60

Ile Glu Gly Thr Gly Lys Gly Thr Leu Phe Asp Tyr Trp Ala Lys Leu
65 70 75 80
His Leu Ala Gly Phe Arg Phe Phe Pro Ser Gly Gly Ala Ala Thr Ile

85 90 95

Phe Arg Gln Gln Ala Val Phe Glu Asp Ala Ser Trp Asn Ala Ala Phe

100 105 110

Cys Gln Gln Ser Gly Lys Asp Trp Pro Trp Leu Val Pro Ser Lys Leu

115 120 125

Tyr Glu Arg Phe Thr Lys Ala Pro Arg Glu Val Ala Lys Lys Asp Gly

130 135 140

Ser Lys Lys Ser Ile Glu Phe Thr Gln Glu Asn Val Ala Asn Glu Ser

145 150 155 160

His Val Ser Leu Val Gly Ala Ser Ile Thr Asp Lys Thr Pro Glu Asp

165 170 175

Gln Lys Glu Phe Phe Leu Lys Met Ala Gly Ala Leu Ala Glu Lys Phe

180 185 190

Asp Ser Trp Lys Ser Ala Asn Glu Asp Arg Ile Val Ala Met Lys Val

195 200 205

Ile Asp Glu Phe Leu Lys Ser Glu Gly Leu His Leu Pro Ser Leu Glu

210 215 220

Asn Ile Ala Val Lys Cys Ser Val Glu Thr Lys Pro Asp Asn Ala Thr

225 230 235 240

Val Ala Trp His Asp Ala Pro Met Ser Gly Val Gln Asn Leu Ala Ile

245 250 255

Gly Val Phe Ala Thr Cys Ala Ser Arg Ile Asp Asn Ile Tyr Asp Leu

260 265 270

Asn Gly Gly Lys Leu Ser Lys Leu Ile Gln Glu Ser Ala Thr Thr Pro

275 280 285

Asn Val Thr Ala Leu Ser Trp Leu Phe Gly Lys Gly Leu Glu Tyr Phe

290 295 300

Arg Thr Thr Asp Ile Asp Thr Ile Met Gln Asp Phe Asn Ile Pro Ala

305 310 315 320

Ser Ala Lys Glu Ser Ile Lys Pro Leu Val Glu Ser Ala Gln Ala Ile

325 330 335

Pro Thr Met Thr Val Leu Gly Lys Lys Asn Tyr Ala Pro Phe Arg Pro
340 345 350

Asn Phe Gly Gly Lys Ile Asp Ser Trp Ile Ala Asn Tyr Ala Ser Arg
355 360 365

Leu Met Leu Leu Asn Asp Ile Leu Glu Gln Ile Glu Pro Gly Phe Glu
370 375 380

Leu Pro Gln Ala Leu Leu Asp Asn Glu Thr Leu Met Ser Gly Ile Asp
385 390 395 400

Met Thr Gly Asp Glu Leu Lys Glu Leu Ile Glu Ala Val Tyr Ala Trp
405 410 415

Val Asp Ala Ala Lys Gln Gly Leu Ala Thr Leu Leu Gly Arg Gly Gly
420 425 430

Asn Val Asp Asp Ala Val Gln Thr Phe Glu Gln Phe Ser Ala Met Met
435 440 445

Asp Thr Leu Asn Gly Thr Leu Asn Thr Ile Ser Ala Arg Tyr Val Arg
450 455 460

Ala Val Glu Met Ala Gly Lys Asp Glu Ala Arg Leu Glu Lys Leu Ile
465 470 475 480

Glu Cys Lys Phe Asp Ile Pro Lys Trp Cys Lys Ser Val Pro Lys Leu
485 490 495

Val Gly Ile Ser Gly Gly Leu Pro Lys Val Glu Glu Glu Ile Lys Val
500 505 510

Met Asn Ala Ala Phe Lys Asp Val Arg Ala Arg Met Phe Val Arg Phe
515 520 525

Glu Glu Ile Ala Ala Tyr Val Ala Ser Lys Gly Ala Gly Met Asp Val
530 535 540

Tyr Asp Ala Leu Glu Lys Arg Glu Leu Glu Gln Ile Lys Lys Leu Lys
545 550 555 560

Ser Ala Val Pro Glu Arg Ala His Ile Gln Ala Tyr Arg Ala Val Leu
565 570 575

His Arg Ile Gly Arg Ala Val Gln Asn Cys Ser Glu Lys Thr Lys Gln

580 585 590

Leu Phe Ser Ser Lys Val Ile Glu Met Gly Val Phe Lys Asn Pro Ser

595 600 605

His Leu Asn Asn Phe Ile Phe Asn Gln Lys Gly Ala Ile Tyr Arg Ser

610 615 620

Pro Phe Asp Arg Ser Arg His Ala Pro Tyr Gln Leu His Ala Asp Lys

625 630 635 640

Leu Leu Lys Asn Asp Trp Leu Glu Leu Leu Ala Glu Ile Ser Ala Thr

645 650 655

Leu Met Ala Ser Glu Ser Thr Glu Gln Met Glu Asp Ala Leu Arg Leu

660 665 670

Glu Arg Thr Arg Leu Gln Leu Gln Leu Ser Gly Leu Pro Asp Trp Glu

675 680 685

Tyr Pro Ala Ser Leu Ala Lys Pro Asp Ile Glu Val Glu Ile Gln Thr

690 695 700

Ala Leu Lys Met Gln Leu Ala Lys Asp Thr Val Thr Ser Asp Val Leu

705 710 715 720

Gln Arg Ala Phe Asn Leu Tyr Ser Ser Val Leu Ser Gly Leu Thr Phe

725 730 735

Lys Leu Leu Arg Arg Ser Phe Ser Leu Lys Met Arg Phe Ser Val Ala

740 745 750

Asp Thr Thr Gln Leu Ile Tyr Val Pro Lys Val Cys Asp Trp Ala Ile

755 760 765

Pro Lys Gln Tyr Leu Gln Ala Glu Gly Glu Ile Gly Ile Ala Ala Arg

770 775 780

Val Val Thr Glu Ser Ser Pro Ala Lys Met Val Thr Glu Val Glu Met

785 790 795 800

Lys Glu Pro Lys Ala Leu Gly His Phe Met Gln Gln Ala Pro His Asp

805 810 815

Trp Tyr Phe Asp Ala Ser Leu Gly Gly Thr Gln Val Ala Gly Arg Ile

820 825 830

Val Glu Lys Gly Lys Glu Val Gly Lys Glu Arg Lys Leu Val Gly Tyr
835 840 845

Arg Met Arg Gly Asn Ser Ala Tyr Lys Thr Val Leu Asp Lys Ser Leu
850 855 860

Val Gly Asn Thr Glu Leu Ser Gln Cys Ser Met Ile Ile Glu Ile Pro
865 870 875 880

Tyr Thr Gln Thr Val Asp Ala Asp Phe Arg Ala Gln Val Gln Ala Gly
885 890 895

Leu Pro Lys Val Ser Ile Asn Leu Pro Val Lys Glu Thr Ile Thr Ala
900 905 910

Ser Asn Lys Asp Glu Gln Met Leu Phe Asp Arg Phe Val Ala Ile Asp
915 920 925

Leu Gly Glu Arg Gly Leu Gly Tyr Ala Val Phe Asp Ala Lys Thr Leu
930 935 940

Glu Leu Gln Glu Ser Gly His Arg Pro Ile Lys Ala Ile Thr Asn Leu
945 950 955 960

Leu Asn Arg Thr His His Tyr Glu Gln Arg Pro Asn Gln Arg Gln Lys
965 970 975

Phe Gln Ala Lys Phe Asn Val Asn Leu Ser Glu Leu Arg Glu Asn Thr
980 985 990

Val Gly Asp Val Cys His Gln Ile Asn Arg Ile Cys Ala Tyr Tyr Asn
995 1000 1005

Ala Phe Pro Val Leu Glu Tyr Met Val Pro Asp Arg Leu Asp Lys
1010 1015 1020

Gln Leu Lys Ser Val Tyr Glu Ser Val Thr Asn Arg Tyr Ile Trp
1025 1030 1035

Ser Ser Thr Asp Ala His Lys Ser Ala Arg Val Gln Phe Trp Leu
1040 1045 1050

Gly Gly Glu Thr Trp Glu His Pro Tyr Leu Lys Ser Ala Lys Asp
1055 1060 1065

Lys Lys Pro Leu Val Leu Ser Pro Gly Arg Gly Ala Ser Gly Lys

1070 1075 1080

Gly Thr Ser Gln Thr Cys Ser Cys Cys Gly Arg Asn Pro Phe Asp
1085 1090 1095
Leu Ile Lys Asp Met Lys Pro Arg Ala Lys Ile Ala Val Val Asp
1100 1105 1110
Gly Lys Ala Lys Leu Glu Asn Ser Glu Leu Lys Leu Phe Glu Arg
1115 1120 1125
Asn Leu Glu Ser Lys Asp Asp Met Leu Ala Arg Arg His Arg Asn
1130 1135 1140

Glu Arg Ala Gly Met Glu Gln Pro Leu Thr Pro Gly Asn Tyr Thr
1145 1150 1155
Val Asp Glu Ile Lys Ala Leu Leu Arg Ala Asn Leu Arg Arg Ala
1160 1165 1170
Pro Lys Asn Arg Arg Thr Lys Asp Thr Thr Val Ser Glu Tyr His
1175 1180 1185
Cys Val Phe Ser Asp Cys Gly Lys Thr Met His Ala Asp Glu Asn
1190 1195 1200

Ala Ala Val Asn Ile Gly Gly Lys Phe Ile Ala Asp Ile Glu Lys
1205 1210 1215

<210> 268
<211> 1252
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
OspCas12c sequence
<400> 268
Met Thr Lys Leu Arg His Arg Gln Lys Lys Leu Thr His Asp Trp Ala
1 5 10 15
Gly Ser Lys Lys Arg Glu Val Leu Gly Ser Asn Gly Lys Leu Gln Asn
20 25 30

Pro Leu Leu Met Pro Val Lys Lys Gly Gln Val Thr Glu Phe Arg Lys

35 40 45

Ala Phe Ser Ala Tyr Ala Arg Ala Thr Lys Gly Glu Met Thr Asp Gly

50 55 60

Arg Lys Asn Met Phe Thr His Ser Phe Glu Pro Phe Lys Thr Lys Pro

65 70 75 80

Ser Leu His Gln Cys Glu Leu Ala Asp Lys Ala Tyr Gln Ser Leu His

85 90 95

Ser Tyr Leu Pro Gly Ser Leu Ala His Phe Leu Leu Ser Ala His Ala

100 105 110

Leu Gly Phe Arg Ile Phe Ser Lys Ser Gly Glu Ala Thr Ala Phe Gln

115 120 125

Ala Ser Ser Lys Ile Glu Ala Tyr Glu Ser Lys Leu Ala Ser Glu Leu

130 135 140

Ala Cys Val Asp Leu Ser Ile Gln Asn Leu Thr Ile Ser Thr Leu Phe

145 150 155 160

Asn Ala Leu Thr Thr Ser Val Arg Gly Lys Gly Glu Glu Thr Ser Ala

165 170 175

Asp Pro Leu Ile Ala Arg Phe Tyr Thr Leu Leu Thr Gly Lys Pro Leu

180 185 190

Ser Arg Asp Thr Gln Gly Pro Glu Arg Asp Leu Ala Glu Val Ile Ser

195 200 205

Arg Lys Ile Ala Ser Ser Phe Gly Thr Trp Lys Glu Met Thr Ala Asn

210 215 220

Pro Leu Gln Ser Leu Gln Phe Phe Glu Glu Glu Leu His Ala Leu Asp

225 230 235 240

Ala Asn Val Ser Leu Ser Pro Ala Phe Asp Val Leu Ile Lys Met Asn

245 250 255

Asp Leu Gln Gly Asp Leu Lys Asn Arg Thr Ile Val Phe Asp Pro Asp

260 265 270

Ala Pro Val Phe Glu Tyr Asn Ala Glu Asp Pro Ala Asp Ile Ile Ile

275 280 285

Lys Leu Thr Ala Arg Tyr Ala Lys Glu Ala Val Ile Lys Asn Gln Asn
290 295 300
Val Gly Asn Tyr Val Lys Asn Ala Ile Thr Thr Thr Asn Ala Asn Gly
305 310 315 320
Leu Gly Trp Leu Leu Asn Lys Gly Leu Ser Leu Leu Pro Val Ser Thr
325 330 335
Asp Asp Glu Leu Leu Glu Phe Ile Gly Val Glu Arg Ser His Pro Ser
340 345 350

Cys His Ala Leu Ile Glu Leu Ile Ala Gln Leu Glu Ala Pro Glu Leu
355 360 365
Phe Glu Lys Asn Val Phe Ser Asp Thr Arg Ser Glu Val Gln Gly Met
370 375 380
Ile Asp Ser Ala Val Ser Asn His Ile Ala Arg Leu Ser Ser Ser Arg
385 390 395 400
Asn Ser Leu Ser Met Asp Ser Glu Glu Leu Glu Arg Leu Ile Lys Ser
405 410 415

Phe Gln Ile His Thr Pro His Cys Ser Leu Phe Ile Gly Ala Gln Ser
420 425 430
Leu Ser Gln Gln Leu Glu Ser Leu Pro Glu Ala Leu Gln Ser Gly Val
435 440 445
Asn Ser Ala Asp Ile Leu Leu Gly Ser Thr Gln Tyr Met Leu Thr Asn
450 455 460
Ser Leu Val Glu Glu Ser Ile Ala Thr Tyr Gln Arg Thr Leu Asn Arg
465 470 475 480

Ile Asn Tyr Leu Ser Gly Val Ala Gly Gln Ile Asn Gly Ala Ile Lys
485 490 495
Arg Lys Ala Ile Asp Gly Glu Lys Ile His Leu Pro Ala Ala Trp Ser
500 505 510
Glu Leu Ile Ser Leu Pro Phe Ile Gly Gln Pro Val Ile Asp Val Glu
515 520 525
Ser Asp Leu Ala His Leu Lys Asn Gln Tyr Gln Thr Leu Ser Asn Glu

530 535 540

Phe Asp Thr Leu Ile Ser Ala Leu Gln Lys Asn Phe Asp Leu Asn Phe
545 550 555 560
Asn Lys Ala Leu Leu Asn Arg Thr Gln His Phe Glu Ala Met Cys Arg
565 570 575
Ser Thr Lys Lys Asn Ala Leu Ser Lys Pro Glu Ile Val Ser Tyr Arg
580 585 590
Asp Leu Leu Ala Arg Leu Thr Ser Cys Leu Tyr Arg Gly Ser Leu Val
595 600 605

Leu Arg Arg Ala Gly Ile Glu Val Leu Lys Lys His Lys Ile Phe Glu
610 615 620
Ser Asn Ser Glu Leu Arg Glu His Val His Glu Arg Lys His Phe Val
625 630 635 640
Phe Val Ser Pro Leu Asp Arg Lys Ala Lys Lys Leu Leu Arg Leu Thr
645 650 655
Asp Ser Arg Pro Asp Leu Leu His Val Ile Asp Glu Ile Leu Gln His
660 665 670

Asp Asn Leu Glu Asn Lys Asp Arg Glu Ser Leu Trp Leu Val Arg Ser
675 680 685
Gly Tyr Leu Leu Ala Gly Leu Pro Asp Gln Leu Ser Ser Ser Phe Ile
690 695 700
Asn Leu Pro Ile Ile Thr Gln Lys Gly Asp Arg Arg Leu Ile Asp Leu
705 710 715 720
Ile Gln Tyr Asp Gln Ile Asn Arg Asp Ala Phe Val Met Leu Val Thr
725 730 735

Ser Ala Phe Lys Ser Asn Leu Ser Gly Leu Gln Tyr Arg Ala Asn Lys
740 745 750
Gln Ser Phe Val Val Thr Arg Thr Leu Ser Pro Tyr Leu Gly Ser Lys
755 760 765
Leu Val Tyr Val Pro Lys Asp Lys Asp Trp Leu Val Pro Ser Gln Met
770 775 780

Phe Glu Gly Arg Phe Ala Asp Ile Leu Gln Ser Asp Tyr Met Val Trp
785 790 795 800

Lys Asp Ala Gly Arg Leu Cys Val Ile Asp Thr Ala Lys His Leu Ser
805 810 815
Asn Ile Lys Lys Ser Val Phe Ser Ser Glu Glu Val Leu Ala Phe Leu
820 825 830
Arg Glu Leu Pro His Arg Thr Phe Ile Gln Thr Glu Val Arg Gly Leu
835 840 845
Gly Val Asn Val Asp Gly Ile Ala Phe Asn Asn Gly Asp Ile Pro Ser
850 855 860

Leu Lys Thr Phe Ser Asn Cys Val Gln Val Lys Val Ser Arg Thr Asn
865 870 875 880
Thr Ser Leu Val Gln Thr Leu Asn Arg Trp Phe Glu Gly Gly Lys Val
885 890 895
Ser Pro Pro Ser Ile Gln Phe Glu Arg Ala Tyr Tyr Lys Lys Asp Asp
900 905 910
Gln Ile His Glu Asp Ala Ala Lys Arg Lys Ile Arg Phe Gln Met Pro
915 920 925

Ala Thr Glu Leu Val His Ala Ser Asp Asp Ala Gly Trp Thr Pro Ser
930 935 940
Tyr Leu Leu Gly Ile Asp Pro Gly Glu Tyr Gly Met Gly Leu Ser Leu
945 950 955 960
Val Ser Ile Asn Asn Gly Glu Val Leu Asp Ser Gly Phe Ile His Ile
965 970 975
Asn Ser Leu Ile Asn Phe Ala Ser Lys Lys Ser Asn His Gln Thr Lys
980 985 990

Val Val Pro Arg Gln Gln Tyr Lys Ser Pro Tyr Ala Asn Tyr Leu Glu
995 1000 1005
Gln Ser Lys Asp Ser Ala Ala Gly Asp Ile Ala His Ile Leu Asp
1010 1015 1020
Arg Leu Ile Tyr Lys Leu Asn Ala Leu Pro Val Phe Glu Ala Leu

1025 1030 1035

Ser Gly Asn Ser Gln Ser Ala Ala Asp Gln Val Trp Thr Lys Val

1040 1045 1050

Leu Ser Phe Tyr Thr Trp Gly Asp Asn Asp Ala Gln Asn Ser Ile

1055 1060 1065

Arg Lys Gln His Trp Phe Gly Ala Ser His Trp Asp Ile Lys Gly

1070 1075 1080

Met Leu Arg Gln Pro Pro Thr Glu Lys Lys Pro Lys Pro Tyr Ile

1085 1090 1095

Ala Phe Pro Gly Ser Gln Val Ser Ser Tyr Gly Asn Ser Gln Arg

1100 1105 1110

Cys Ser Cys Cys Gly Arg Asn Pro Ile Glu Gln Leu Arg Glu Met

1115 1120 1125

Ala Lys Asp Thr Ser Ile Lys Glu Leu Lys Ile Arg Asn Ser Glu

1130 1135 1140

Ile Gln Leu Phe Asp Gly Thr Ile Lys Leu Phe Asn Pro Asp Pro

1145 1150 1155

Ser Thr Val Ile Glu Arg Arg Arg His Asn Leu Gly Pro Ser Arg

1160 1165 1170

Ile Pro Val Ala Asp Arg Thr Phe Lys Asn Ile Ser Pro Ser Ser

1175 1180 1185

Leu Glu Phe Lys Glu Leu Ile Thr Ile Val Ser Arg Ser Ile Arg

1190 1195 1200

His Ser Pro Glu Phe Ile Ala Lys Lys Arg Gly Ile Gly Ser Glu

1205 1210 1215

Tyr Phe Cys Ala Tyr Ser Asp Cys Asn Ser Ser Leu Asn Ser Glu

1220 1225 1230

Ala Asn Ala Ala Ala Asn Val Ala Gln Lys Phe Gln Lys Gln Leu

1235 1240 1245

Phe Phe Glu Leu

1250

<210> 269

<211> 767

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12g1 sequence

<400> 269

Met Ala Gln Ala Ser Ser Thr Pro Ala Val Ser Pro Arg Pro Arg Pro
1 5 10 15

Arg Tyr Arg Glu Glu Arg Thr Leu Val Arg Lys Leu Leu Pro Arg Pro
20 25 30

Gly Gln Ser Lys Gln Glu Phe Arg Glu Asn Val Lys Lys Leu Arg Lys
35 40 45

Ala Phe Leu Gln Phe Asn Ala Asp Val Ser Gly Val Cys Gln Trp Ala
50 55 60

Ile Gln Phe Arg Pro Arg Tyr Gly Lys Pro Ala Glu Pro Thr Glu Thr
65 70 75 80

Phe Trp Lys Phe Phe Leu Glu Pro Glu Thr Ser Leu Pro Pro Asn Asp
85 90 95

Ser Arg Ser Pro Glu Phe Arg Arg Leu Gln Ala Phe Glu Ala Ala Ala
100 105 110

Gly Ile Asn Gly Ala Ala Ala Leu Asp Asp Pro Ala Phe Thr Asn Glu
115 120 125

Leu Arg Asp Ser Ile Leu Ala Val Ala Ser Arg Pro Lys Thr Lys Glu
130 135 140

Ala Gln Arg Leu Phe Ser Arg Leu Lys Asp Tyr Gln Pro Ala His Arg
145 150 155 160

Met Ile Leu Ala Lys Val Ala Ala Glu Trp Ile Glu Ser Arg Tyr Arg
165 170 175

Arg Ala His Gln Asn Trp Glu Arg Asn Tyr Glu Glu Trp Lys Lys Glu
180 185 190

Lys Gln Glu Trp Glu Gln Asn His Pro Glu Leu Thr Pro Glu Ile Arg

195 200 205

Glu Ala Phe Asn Gln Ile Phe Gln Gln Leu Glu Val Lys Glu Lys Arg

210 215 220

Val Arg Ile Cys Pro Ala Ala Arg Leu Leu Gln Asn Lys Asp Asn Cys

225 230 235 240

Gln Tyr Ala Gly Lys Asn Lys His Ser Val Leu Cys Asn Gln Phe Asn

245 250 255

Glu Phe Lys Lys Asn His Leu Gln Gly Lys Ala Ile Lys Phe Phe Tyr

260 265 270

Lys Asp Ala Glu Lys Tyr Leu Arg Cys Gly Leu Gln Ser Leu Lys Pro

275 280 285

Asn Val Gln Gly Pro Phe Arg Glu Asp Trp Asn Lys Tyr Leu Arg Tyr

290 295 300

Met Asn Leu Lys Glu Glu Thr Leu Arg Gly Lys Asn Gly Gly Arg Leu

305 310 315 320

Pro His Cys Lys Asn Leu Gly Gln Glu Cys Glu Phe Asn Pro His Thr

325 330 335

Ala Leu Cys Lys Gln Tyr Gln Gln Gln Leu Ser Ser Arg Pro Asp Leu

340 345 350

Val Gln His Asp Glu Leu Tyr Arg Lys Trp Arg Arg Glu Tyr Trp Arg

355 360 365

Glu Pro Arg Lys Pro Val Phe Arg Tyr Pro Ser Val Lys Arg His Ser

370 375 380

Ile Ala Lys Ile Phe Gly Glu Asn Tyr Phe Gln Ala Asp Phe Lys Asn

385 390 395 400

Ser Val Val Gly Leu Arg Leu Asp Ser Met Pro Ala Gly Gln Tyr Leu

405 410 415

Glu Phe Ala Phe Ala Pro Trp Pro Arg Asn Tyr Arg Pro Gln Pro Gly

420 425 430

Glu Thr Glu Ile Ser Ser Val His Leu His Phe Val Gly Thr Arg Pro

435 440 445

Arg Ile Gly Phe Arg Phe Arg Val Pro His Lys Arg Ser Arg Phe Asp
450 455 460
Cys Thr Gln Glu Glu Leu Asp Glu Leu Arg Ser Arg Thr Phe Pro Arg
465 470 475 480

Lys Ala Gln Asp Gln Lys Phe Leu Glu Ala Ala Arg Lys Arg Leu Leu
485 490 495
Glu Thr Phe Pro Gly Asn Ala Glu Gln Glu Leu Arg Leu Leu Ala Val
500 505 510
Asp Leu Gly Thr Asp Ser Ala Arg Ala Ala Phe Phe Ile Gly Lys Thr
515 520 525
Phe Gln Gln Ala Phe Pro Leu Lys Ile Val Lys Ile Glu Lys Leu Tyr
530 535 540

Glu Gln Trp Pro Asn Gln Lys Gln Ala Gly Asp Arg Arg Asp Ala Ser
545 550 555 560
Ser Lys Gln Pro Arg Pro Gly Leu Ser Arg Asp His Val Gly Arg His
565 570 575
Leu Gln Lys Met Arg Ala Gln Ala Ser Glu Ile Ala Gln Lys Arg Gln
580 585 590
Glu Leu Thr Gly Thr Pro Ala Pro Glu Thr Thr Thr Asp Gln Ala Ala
595 600 605

Lys Lys Ala Thr Leu Gln Pro Phe Asp Leu Arg Gly Leu Thr Val His
610 615 620
Thr Ala Arg Met Ile Arg Asp Trp Ala Arg Leu Asn Ala Arg Gln Ile
625 630 635 640
Ile Gln Leu Ala Glu Glu Asn Gln Val Asp Leu Ile Val Leu Glu Ser
645 650 655
Leu Arg Gly Phe Arg Pro Pro Gly Tyr Glu Asn Leu Asp Gln Glu Lys
660 665 670

Lys Arg Arg Val Ala Phe Phe Ala His Gly Arg Ile Arg Arg Lys Val
675 680 685
Thr Glu Lys Ala Val Glu Arg Gly Met Arg Val Val Thr Val Pro Tyr

690 695 700

Leu Ala Ser Ser Lys Val Cys Ala Glu Cys Arg Lys Lys Gln Lys Asp

705 710 715 720

Asn Lys Gln Trp Glu Lys Asn Lys Lys Arg Gly Leu Phe Lys Cys Glu

725 730 735

Gly Cys Gly Ser Gln Ala Gln Val Asp Glu Asn Ala Ala Arg Val Leu

740 745 750

Gly Arg Val Phe Trp Gly Glu Ile Glu Leu Pro Thr Ala Ile Pro

755 760 765

<210> 270

<211> 870

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12h1 sequence

<400> 270

Met Lys Val His Glu Ile Pro Arg Ser Gln Leu Leu Lys Ile Lys Gln

1 5 10 15

Tyr Glu Gly Ser Phe Val Glu Trp Tyr Arg Asp Leu Gln Glu Asp Arg

20 25 30

Lys Lys Phe Ala Ser Leu Leu Phe Arg Trp Ala Ala Phe Gly Tyr Ala

35 40 45

Ala Arg Glu Asp Asp Gly Ala Thr Tyr Ile Ser Pro Ser Gln Ala Leu

50 55 60

Leu Glu Arg Arg Leu Leu Leu Gly Asp Ala Glu Asp Val Ala Ile Lys

65 70 75 80

Phe Leu Asp Val Leu Phe Lys Gly Gly Ala Pro Ser Ser Ser Cys Tyr

85 90 95

Ser Leu Phe Tyr Glu Asp Phe Ala Leu Arg Asp Lys Ala Lys Tyr Ser

100 105 110

Gly Ala Lys Arg Glu Phe Ile Glu Gly Leu Ala Thr Met Pro Leu Asp

115 120 125

Lys Ile Ile Glu Arg Ile Arg Gln Asp Glu Gln Leu Ser Lys Ile Pro
130 135 140

Ala Glu Glu Trp Leu Ile Leu Gly Ala Glu Tyr Ser Pro Glu Glu Ile
145 150 155 160
Trp Glu Gln Val Ala Pro Arg Ile Val Asn Val Asp Arg Ser Leu Gly
165 170 175
Lys Gln Leu Arg Glu Arg Leu Gly Ile Lys Cys Arg Arg Pro His Asp
180 185 190
Ala Gly Tyr Cys Lys Ile Leu Met Glu Val Val Ala Arg Gln Leu Arg
195 200 205

Ser His Asn Glu Thr Tyr His Glu Tyr Leu Asn Gln Thr His Glu Met
210 215 220
Lys Thr Lys Val Ala Asn Asn Leu Thr Asn Glu Phe Asp Leu Val Cys
225 230 235 240
Glu Phe Ala Glu Val Leu Glu Glu Lys Asn Tyr Gly Leu Gly Trp Tyr
245 250 255
Val Leu Trp Gln Gly Val Lys Gln Ala Leu Lys Glu Gln Lys Lys Pro
260 265 270

Thr Lys Ile Gln Ile Ala Val Asp Gln Leu Arg Gln Pro Lys Phe Ala
275 280 285
Gly Leu Leu Thr Ala Lys Trp Arg Ala Leu Lys Gly Ala Tyr Asp Thr
290 295 300
Trp Lys Leu Lys Lys Arg Leu Glu Lys Arg Lys Ala Phe Pro Tyr Met
305 310 315 320
Pro Asn Trp Asp Asn Asp Tyr Gln Ile Pro Val Gly Leu Thr Gly Leu
325 330 335

Gly Val Phe Thr Leu Glu Val Lys Arg Thr Glu Val Val Val Asp Leu
340 345 350
Lys Glu His Gly Lys Leu Phe Cys Ser His Ser His Tyr Phe Gly Asp
355 360 365
Leu Thr Ala Glu Lys His Pro Ser Arg Tyr His Leu Lys Phe Arg His

370 375 380

Lys Leu Lys Leu Arg Lys Arg Asp Ser Arg Val Glu Pro Thr Ile Gly

385 390 395 400

Pro Trp Ile Glu Ala Ala Leu Arg Glu Ile Thr Ile Gln Lys Lys Pro

405 410 415

Asn Gly Val Phe Tyr Leu Gly Leu Pro Tyr Ala Leu Ser His Gly Ile

420 425 430

Asp Asn Phe Gln Ile Ala Lys Arg Phe Phe Ser Ala Ala Lys Pro Asp

435 440 445

Lys Glu Val Ile Asn Gly Leu Pro Ser Glu Met Val Val Gly Ala Ala

450 455 460

Asp Leu Asn Leu Ser Asn Ile Val Ala Pro Val Lys Ala Arg Ile Gly

465 470 475 480

Lys Gly Leu Glu Gly Pro Leu His Ala Leu Asp Tyr Gly Tyr Gly Glu

485 490 495

Leu Ile Asp Gly Pro Lys Ile Leu Thr Pro Asp Gly Pro Arg Cys Gly

500 505 510

Glu Leu Ile Ser Leu Lys Arg Asp Ile Val Glu Ile Lys Ser Ala Ile

515 520 525

Lys Glu Phe Lys Ala Cys Gln Arg Glu Gly Leu Thr Met Ser Glu Glu

530 535 540

Thr Thr Thr Trp Leu Ser Glu Val Glu Ser Pro Ser Asp Ser Pro Arg

545 550 555 560

Cys Met Ile Gln Ser Arg Ile Ala Asp Thr Ser Arg Arg Leu Asn Ser

565 570 575

Phe Lys Tyr Gln Met Asn Lys Glu Gly Tyr Gln Asp Leu Ala Glu Ala

580 585 590

Leu Arg Leu Leu Asp Ala Met Asp Ser Tyr Asn Ser Leu Leu Glu Ser

595 600 605

Tyr Gln Arg Met His Leu Ser Pro Gly Glu Gln Ser Pro Lys Glu Ala

610 615 620

Lys Phe Asp Thr Lys Arg Ala Ser Phe Arg Asp Leu Leu Arg Arg Arg
625 630 635 640
Val Ala His Thr Ile Val Glu Tyr Phe Asp Asp Cys Asp Ile Val Phe
645 650 655

Phe Glu Asp Leu Asp Gly Pro Ser Asp Ser Asp Ser Arg Asn Asn Ala
660 665 670
Leu Val Lys Leu Leu Ser Pro Arg Thr Leu Leu Leu Tyr Ile Arg Gln
675 680 685
Ala Leu Glu Lys Arg Gly Ile Gly Met Val Glu Val Ala Lys Asp Gly
690 695 700
Thr Ser Gln Asn Asn Pro Ile Ser Gly His Val Gly Trp Arg Asn Lys
705 710 715 720

Gln Asn Lys Ser Glu Ile Tyr Phe Tyr Glu Asp Lys Glu Leu Leu Val
725 730 735
Met Asp Ala Asp Glu Val Gly Ala Met Asn Ile Leu Cys Arg Gly Leu
740 745 750
Asn His Ser Val Cys Pro Tyr Ser Phe Val Thr Lys Ala Pro Glu Lys
755 760 765
Lys Asn Asp Glu Lys Lys Glu Gly Asp Tyr Gly Lys Arg Val Lys Arg
770 775 780

Phe Leu Lys Asp Arg Tyr Gly Ser Ser Asn Val Arg Phe Leu Val Ala
785 790 795 800
Ser Met Gly Phe Val Thr Val Thr Thr Lys Arg Pro Lys Asp Ala Leu
805 810 815
Val Gly Lys Arg Leu Tyr Tyr His Gly Gly Glu Leu Val Thr His Asp
820 825 830
Leu His Asn Arg Met Lys Asp Glu Ile Lys Tyr Leu Val Glu Lys Glu
835 840 845

Val Leu Ala Arg Arg Val Ser Leu Ser Asp Ser Thr Ile Lys Ser Tyr
850 855 860
Lys Ser Phe Ala His Val

865 870

<210> 271

<211> 1093

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12i1 sequence

<400> 271

Met Ser Asn Lys Glu Lys Asn Ala Ser Glu Thr Arg Lys Ala Tyr Thr
1 5 10 15

Thr Lys Met Ile Pro Arg Ser His Asp Arg Met Lys Leu Leu Gly Asn
20 25 30

Phe Met Asp Tyr Leu Met Asp Gly Thr Pro Ile Phe Phe Glu Leu Trp
35 40 45

Asn Gln Phe Gly Gly Gly Ile Asp Arg Asp Ile Ile Ser Gly Thr Ala
50 55 60

Asn Lys Asp Lys Ile Ser Asp Asp Leu Leu Leu Ala Val Asn Trp Phe
65 70 75 80

Lys Val Met Pro Ile Asn Ser Lys Pro Gln Gly Val Ser Pro Ser Asn
85 90 95

Leu Ala Asn Leu Phe Gln Gln Tyr Ser Gly Ser Glu Pro Asp Ile Gln
100 105 110

Ala Gln Glu Tyr Phe Ala Ser Asn Phe Asp Thr Glu Lys His Gln Trp
115 120 125

Lys Asp Met Arg Val Glu Tyr Glu Arg Leu Leu Ala Glu Leu Gln Leu
130 135 140

Ser Arg Ser Asp Met His His Asp Leu Lys Leu Met Tyr Lys Glu Lys
145 150 155 160

Cys Ile Gly Leu Ser Leu Ser Thr Ala His Tyr Ile Thr Ser Val Met
165 170 175

Phe Gly Thr Gly Ala Lys Asn Asn Arg Gln Thr Lys His Gln Phe Tyr
180 185 190

Ser Lys Val Ile Gln Leu Leu Glu Glu Ser Thr Gln Ile Asn Ser Val
195 200 205

Glu Gln Leu Ala Ser Ile Ile Leu Lys Ala Gly Asp Cys Asp Ser Tyr
210 215 220

Arg Lys Leu Arg Ile Arg Cys Ser Arg Lys Gly Ala Thr Pro Ser Ile
225 230 235 240

Leu Lys Ile Val Gln Asp Tyr Glu Leu Gly Thr Asn His Asp Asp Glu
245 250 255

Val Asn Val Pro Ser Leu Ile Ala Asn Leu Lys Glu Lys Leu Gly Arg
260 265 270

Phe Glu Tyr Glu Cys Glu Trp Lys Cys Met Glu Lys Ile Lys Ala Phe
275 280 285

Leu Ala Ser Lys Val Gly Pro Tyr Tyr Leu Gly Ser Tyr Ser Ala Met
290 295 300

Leu Glu Asn Ala Leu Ser Pro Ile Lys Gly Met Thr Thr Lys Asn Cys
305 310 315 320

Lys Phe Val Leu Lys Gln Ile Asp Ala Lys Asn Asp Ile Lys Tyr Glu
325 330 335

Asn Glu Pro Phe Gly Lys Ile Val Glu Gly Phe Phe Asp Ser Pro Tyr
340 345 350

Phe Glu Ser Asp Thr Asn Val Lys Trp Val Leu His Pro His His Ile
355 360 365

Gly Glu Ser Asn Ile Lys Thr Leu Trp Glu Asp Leu Asn Ala Ile His
370 375 380

Ser Lys Tyr Glu Glu Asp Ile Ala Ser Leu Ser Glu Asp Lys Lys Glu
385 390 395 400

Lys Arg Ile Lys Val Tyr Gln Gly Asp Val Cys Gln Thr Ile Asn Thr
405 410 415

Tyr Cys Glu Glu Val Gly Lys Glu Ala Lys Thr Pro Leu Val Gln Leu
420 425 430

Leu Arg Tyr Leu Tyr Ser Arg Lys Asp Asp Ile Ala Val Asp Lys Ile

435 440 445

Ile Asp Gly Ile Thr Phe Leu Ser Lys Lys His Lys Val Glu Lys Gln

450 455 460

Lys Ile Asn Pro Val Ile Gln Lys Tyr Pro Ser Phe Asn Phe Gly Asn

465 470 475 480

Asn Ser Lys Leu Leu Gly Lys Ile Ile Ser Pro Lys Asp Lys Leu Lys

485 490 495

His Asn Leu Lys Cys Asn Arg Asn Gln Val Asp Asn Tyr Ile Trp Ile

500 505 510

Glu Ile Lys Val Leu Asn Thr Lys Thr Met Arg Trp Glu Lys His His

515 520 525

Tyr Ala Leu Ser Ser Thr Arg Phe Leu Glu Glu Val Tyr Tyr Pro Ala

530 535 540

Thr Ser Glu Asn Pro Pro Asp Ala Leu Ala Ala Arg Phe Arg Thr Lys

545 550 555 560

Thr Asn Gly Tyr Glu Gly Lys Pro Ala Leu Ser Ala Glu Gln Ile Glu

565 570 575

Gln Ile Arg Ser Ala Pro Val Gly Leu Arg Lys Val Lys Lys Arg Gln

580 585 590

Met Arg Leu Glu Ala Ala Arg Gln Gln Asn Leu Leu Pro Arg Tyr Thr

595 600 605

Trp Gly Lys Asp Phe Asn Ile Asn Ile Cys Lys Arg Gly Asn Asn Phe

610 615 620

Glu Val Thr Leu Ala Thr Lys Val Lys Lys Lys Lys Glu Lys Asn Tyr

625 630 635 640

Lys Val Val Leu Gly Tyr Asp Ala Asn Ile Val Arg Lys Asn Thr Tyr

645 650 655

Ala Ala Ile Glu Ala His Ala Asn Gly Asp Gly Val Ile Asp Tyr Asn

660 665 670

Asp Leu Pro Val Lys Pro Ile Glu Ser Gly Phe Val Thr Val Glu Ser

675 680 685

Gln Val Arg Asp Lys Ser Tyr Asp Gln Leu Ser Tyr Asn Gly Val Lys
690 695 700
Leu Leu Tyr Cys Lys Pro His Val Glu Ser Arg Arg Ser Phe Leu Glu
705 710 715 720
Lys Tyr Arg Asn Gly Thr Met Lys Asp Asn Arg Gly Asn Asn Ile Gln
725 730 735

Ile Asp Phe Met Lys Asp Phe Glu Ala Ile Ala Asp Asp Glu Thr Ser
740 745 750
Leu Tyr Tyr Phe Asn Met Lys Tyr Cys Lys Leu Leu Gln Ser Ser Ile
755 760 765
Arg Asn His Ser Ser Gln Ala Lys Glu Tyr Arg Glu Glu Ile Phe Glu
770 775 780
Leu Leu Arg Asp Gly Lys Leu Ser Val Leu Lys Leu Ser Ser Leu Ser
785 790 795 800

Asn Leu Ser Phe Val Met Phe Lys Val Ala Lys Ser Leu Ile Gly Thr
805 810 815
Tyr Phe Gly His Leu Leu Lys Lys Pro Lys Asn Ser Lys Ser Asp Val
820 825 830
Lys Ala Pro Pro Ile Thr Asp Glu Asp Lys Gln Lys Ala Asp Pro Glu
835 840 845
Met Phe Ala Leu Arg Leu Ala Leu Glu Glu Lys Arg Leu Asn Lys Val
850 855 860

Lys Ser Lys Lys Glu Val Ile Ala Asn Lys Ile Val Ala Lys Ala Leu
865 870 875 880
Glu Leu Arg Asp Lys Tyr Gly Pro Val Leu Ile Lys Gly Glu Asn Ile
885 890 895
Ser Asp Thr Thr Lys Lys Gly Lys Lys Ser Ser Thr Asn Ser Phe Leu
900 905 910
Met Asp Trp Leu Ala Arg Gly Val Ala Asn Lys Val Lys Glu Met Val
915 920 925

Met Met His Gln Gly Leu Glu Phe Val Glu Val Asn Pro Asn Phe Thr

930 935 940

Ser His Gln Asp Pro Phe Val His Lys Asn Pro Glu Asn Thr Phe Arg

945 950 955 960

Ala Arg Tyr Ser Arg Cys Thr Pro Ser Glu Leu Thr Glu Lys Asn Arg

965 970 975

Lys Glu Ile Leu Ser Phe Leu Ser Asp Lys Pro Ser Lys Arg Pro Thr

980 985 990

Asn Ala Tyr Tyr Asn Glu Gly Ala Met Ala Phe Leu Ala Thr Tyr Gly

995 1000 1005

Leu Lys Lys Asn Asp Val Leu Gly Val Ser Leu Glu Lys Phe Lys

1010 1015 1020

Gln Ile Met Ala Asn Ile Leu His Gln Arg Ser Glu Asp Gln Leu

1025 1030 1035

Leu Phe Pro Ser Arg Gly Gly Met Phe Tyr Leu Ala Thr Tyr Lys

1040 1045 1050

Leu Asp Ala Asp Ala Thr Ser Val Asn Trp Asn Gly Lys Gln Phe

1055 1060 1065

Trp Val Cys Asn Ala Asp Leu Val Ala Ala Tyr Asn Val Gly Leu

1070 1075 1080

Val Asp Ile Gln Lys Asp Phe Lys Lys Lys

1085 1090

<210> 272

<211> 1054

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12i2 sequence

<400> 272

Met Ser Ser Ala Ile Lys Ser Tyr Lys Ser Val Leu Arg Pro Asn Glu

1 5 10 15

Arg Lys Asn Gln Leu Leu Lys Ser Thr Ile Gln Cys Leu Glu Asp Gly

20 25 30

Ser Ala Phe Phe Phe Lys Met Leu Gln Gly Leu Phe Gly Gly Ile Thr
35 40 45
Pro Glu Ile Val Arg Phe Ser Thr Glu Gln Glu Lys Gln Gln Gln Asp
50 55 60
Ile Ala Leu Trp Cys Ala Val Asn Trp Phe Arg Pro Val Ser Gln Asp

65 70 75 80
Ser Leu Thr His Thr Ile Ala Ser Asp Asn Leu Val Glu Lys Phe Glu
85 90 95
Glu Tyr Tyr Gly Gly Thr Ala Ser Asp Ala Ile Lys Gln Tyr Phe Ser
100 105 110
Ala Ser Ile Gly Glu Ser Tyr Tyr Trp Asn Asp Cys Arg Gln Gln Tyr
115 120 125
Tyr Asp Leu Cys Arg Glu Leu Gly Val Glu Val Ser Asp Leu Thr His

130 135 140
Asp Leu Glu Ile Leu Cys Arg Glu Lys Cys Leu Ala Val Ala Thr Glu
145 150 155 160
Ser Asn Gln Asn Asn Ser Ile Ile Ser Val Leu Phe Gly Thr Gly Glu
165 170 175
Lys Glu Asp Arg Ser Val Lys Leu Arg Ile Thr Lys Lys Ile Leu Glu
180 185 190
Ala Ile Ser Asn Leu Lys Glu Ile Pro Lys Asn Val Ala Pro Ile Gln

195 200 205
Glu Ile Ile Leu Asn Val Ala Lys Ala Thr Lys Glu Thr Phe Arg Gln
210 215 220
Val Tyr Ala Gly Asn Leu Gly Ala Pro Ser Thr Leu Glu Lys Phe Ile
225 230 235 240
Ala Lys Asp Gly Gln Lys Glu Phe Asp Leu Lys Lys Leu Gln Thr Asp
245 250 255
Leu Lys Lys Val Ile Arg Gly Lys Ser Lys Glu Arg Asp Trp Cys Cys

260 265 270
Gln Glu Glu Leu Arg Ser Tyr Val Glu Gln Asn Thr Ile Gln Tyr Asp

275 280 285

Leu Trp Ala Trp Gly Glu Met Phe Asn Lys Ala His Thr Ala Leu Lys

290 295 300

Ile Lys Ser Thr Arg Asn Tyr Asn Phe Ala Lys Gln Arg Leu Glu Gln

305 310 315 320

Phe Lys Glu Ile Gln Ser Leu Asn Asn Leu Leu Val Val Lys Lys Leu

325 330 335

Asn Asp Phe Phe Asp Ser Glu Phe Phe Ser Gly Glu Glu Thr Tyr Thr

340 345 350

Ile Cys Val His His Leu Gly Gly Lys Asp Leu Ser Lys Leu Tyr Lys

355 360 365

Ala Trp Glu Asp Asp Pro Ala Asp Pro Glu Asn Ala Ile Val Val Leu

370 375 380

Cys Asp Asp Leu Lys Asn Asn Phe Lys Lys Glu Pro Ile Arg Asn Ile

385 390 395 400

Leu Arg Tyr Ile Phe Thr Ile Arg Gln Glu Cys Ser Ala Gln Asp Ile

405 410 415

Leu Ala Ala Ala Lys Tyr Asn Gln Gln Leu Asp Arg Tyr Lys Ser Gln

420 425 430

Lys Ala Asn Pro Ser Val Leu Gly Asn Gln Gly Phe Thr Trp Thr Asn

435 440 445

Ala Val Ile Leu Pro Glu Lys Ala Gln Arg Asn Asp Arg Pro Asn Ser

450 455 460

Leu Asp Leu Arg Ile Trp Leu Tyr Leu Lys Leu Arg His Pro Asp Gly

465 470 475 480

Arg Trp Lys Lys His His Ile Pro Phe Tyr Asp Thr Arg Phe Phe Gln

485 490 495

Glu Ile Tyr Ala Ala Gly Asn Ser Pro Val Asp Thr Cys Gln Phe Arg

500 505 510

Thr Pro Arg Phe Gly Tyr His Leu Pro Lys Leu Thr Asp Gln Thr Ala

515 520 525

Ile Arg Val Asn Lys Lys His Val Lys Ala Ala Lys Thr Glu Ala Arg
530 535 540
Ile Arg Leu Ala Ile Gln Gln Gly Thr Leu Pro Val Ser Asn Leu Lys
545 550 555 560
Ile Thr Glu Ile Ser Ala Thr Ile Asn Ser Lys Gly Gln Val Arg Ile
565 570 575
Pro Val Lys Phe Asp Val Gly Arg Gln Lys Gly Thr Leu Gln Ile Gly

580 585 590
Asp Arg Phe Cys Gly Tyr Asp Gln Asn Gln Thr Ala Ser His Ala Tyr
595 600 605
Ser Leu Trp Glu Val Val Lys Glu Gly Gln Tyr His Lys Glu Leu Gly
610 615 620
Cys Phe Val Arg Phe Ile Ser Ser Gly Asp Ile Val Ser Ile Thr Glu
625 630 635 640
Asn Arg Gly Asn Gln Phe Asp Gln Leu Ser Tyr Glu Gly Leu Ala Tyr

645 650 655
Pro Gln Tyr Ala Asp Trp Arg Lys Lys Ala Ser Lys Phe Val Ser Leu
660 665 670
Trp Gln Ile Thr Lys Lys Asn Lys Lys Lys Glu Ile Val Thr Val Glu
675 680 685
Ala Lys Glu Lys Phe Asp Ala Ile Cys Lys Tyr Gln Pro Arg Leu Tyr
690 695 700
Lys Phe Asn Lys Glu Tyr Ala Tyr Leu Leu Arg Asp Ile Val Arg Gly

705 710 715 720
Lys Ser Leu Val Glu Leu Gln Gln Ile Arg Gln Glu Ile Phe Arg Phe
725 730 735
Ile Glu Gln Asp Cys Gly Val Thr Arg Leu Gly Ser Leu Ser Leu Ser
740 745 750
Thr Leu Glu Thr Val Lys Ala Val Lys Gly Ile Ile Tyr Ser Tyr Phe
755 760 765
Ser Thr Ala Leu Asn Ala Ser Lys Asn Asn Pro Ile Ser Asp Glu Gln

770 775 780

Arg Lys Glu Phe Asp Pro Glu Leu Phe Ala Leu Leu Glu Lys Leu Glu

785 790 795 800

Leu Ile Arg Thr Arg Lys Lys Lys Gln Lys Val Glu Arg Ile Ala Asn

805 810 815

Ser Leu Ile Gln Thr Cys Leu Glu Asn Asn Ile Lys Phe Ile Arg Gly

820 825 830

Glu Gly Asp Leu Ser Thr Thr Asn Asn Ala Thr Lys Lys Lys Ala Asn

835 840 845

Ser Arg Ser Met Asp Trp Leu Ala Arg Gly Val Phe Asn Lys Ile Arg

850 855 860

Gln Leu Ala Pro Met His Asn Ile Thr Leu Phe Gly Cys Gly Ser Leu

865 870 875 880

Tyr Thr Ser His Gln Asp Pro Leu Val His Arg Asn Pro Asp Lys Ala

885 890 895

Met Lys Cys Arg Trp Ala Ala Ile Pro Val Lys Asp Ile Gly Asp Trp

900 905 910

Val Leu Arg Lys Leu Ser Gln Asn Leu Arg Ala Lys Asn Ile Gly Thr

915 920 925

Gly Glu Tyr Tyr His Gln Gly Val Lys Glu Phe Leu Ser His Tyr Glu

930 935 940

Leu Gln Asp Leu Glu Glu Glu Leu Leu Lys Trp Arg Ser Asp Arg Lys

945 950 955 960

Ser Asn Ile Pro Cys Trp Val Leu Gln Asn Arg Leu Ala Glu Lys Leu

965 970 975

Gly Asn Lys Glu Ala Val Val Tyr Ile Pro Val Arg Gly Gly Arg Ile

980 985 990

Tyr Phe Ala Thr His Lys Val Ala Thr Gly Ala Val Ser Ile Val Phe

995 1000 1005

Asp Gln Lys Gln Val Trp Val Cys Asn Ala Asp His Val Ala Ala

1010 1015 1020

Ala Asn Ile Ala Leu Thr Val Lys Gly Ile Gly Glu Gln Ser Ser

1025 1030 1035

Asp Glu Glu Asn Pro Asp Gly Ser Arg Ile Lys Leu Gln Leu Thr

1040 1045 1050

Ser

<210> 273

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 273

Gly Gly Ser Gly Gly Ser

1 5

<210> 274

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 274

gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat cccggaggct ctggaggaag ctccgaagtc 540

gagttttccc atgagtactg gatgagacac gcattgactc tcgcaaagag ggctcgagat 600

gaacgcgagg tgcccgtggg ggcagtactc gtgctcaaca atcgcgtaat cggcgaaggt 660

tggaataggg caatcggact ccacgacccc actgcacatg cggaaatcat ggcccttcga 720

cagggagggc ttgtgatgca gaattatcga ctttatgatg cgacgctgta cgtcacgttt 780

gaaccttgcg taatgtgcgc gggagctatg attcactccc gcattggacg agttgtattc 840

ggtgttcgca acgccaagac gggtgccgca ggttcactga tggacgtgct gcatcatcca 900

ggcatgaacc accgggtaga aatcacagaa ggcatattgg cggacgaatg tgcggcgctg 960

ttgtgtcgtt tttttcgcat gcccaggcgg gtctttaacg cccagaaaaa agcacaatcc 1020

tctactgacg gctcttctgg atctgaaaca cctggcacaa gcgagagcgc caccctgag 1080

agctctggct cctgggaaga agagaagaag aagtgggaag aagataagaa aaaggacccg 1140

ctggccaaga tcctgggcaa gctggctgag tacggactga tccctctgtt catcccctac 1200

accgacagca acgagcccat cgtgaaagaa atcaagtgga tggaaaagtc ccggaaccag 1260

agcgtgcggc ggctggataa ggacatgttc attcaggccc tggaacggtt cctgagctgg 1320

gagagctgga acctgaaagt gaaagaggaa tacgagaagg tcgagaaaga gtacaagacc 1380

ctggaagaga ggatcaaaga ggacatccag gctctgaagg ctctggaaca gtatgagaaa 1440

gagcggcaag aacagctgct gcgggacacc ctgaacacca acgagtaccg gctgagcaag 1500

agaggcctta gaggctggcg ggaaatcatc cagaaatggc tgaaaatgga cgagaacgag 1560

ccctccgaga agtacctgga agtgttcaag gactaccagc ggaagcaccc tagagaggcc 1620

ggcgattaca gcgtgtacga gttcctgtcc aagaaagaga accacttcat ctggcggaat 1680

caccctgagt acccctacct gtacgccacc ttctgcgaga tcgacaagaa aaagaaggac 1740

gccaagcagc aggccacctt cacactggcc gatcctatca atcaccctct gtgggtccga 1800

ttcgaggaaa gaagcggcag caacctgaac aagtacagaa tcctgaccga gcagctgcac 1860

accgagaagc tgaagaaaaa gctgacagtg cagctggacc ggctgatcta ccctacagaa 1920

tctggcggct gggaagagaa gggcaaagtg gacattgtgc tgctgcccag ccggcagttc 1980

tacaaccaga tcttcctgga catcgaggaa aagggcaagc acgccttcac ctacaaggat 2040

gagagcatca agttccctct gaagggcaca ctcggcggag ccagagtgca gttcgacaga 2100

gatcacctga gaagataccc tcacaaggtg gaaagcggca acgtgggcag aatctacttc 2160

aacatgaccg tgaacatcga gcctacagag tccccagtgt ccaagtctct gaagatccac 2220

cgggacgact tccccaaggt ggtcaacttc aagcccaaag aactgaccga gtggatcaag 2280

gacagcaagg gcaagaaact gaagtccggc atcgagtccc tggaaatcgg cctgagagtg 2340

atgagcatcg acctgggaca gagacaggcc gctgccgcct ctattttcga ggtggtggat 2400

cagaagcccg acatcgaagg caagctgttt ttcccaatca agggcaccga gctgtatgcc 2460

gtgcacagag ccagcttcaa catcaagctg cccggcgaga cactggtcaa gagcagagaa 2520

gtgctgcgga aggccagaga ggacaatctg aaactgatga accagaagct caacttcctg 2580
cggaacgtgc tgcacttcca gcagttcgag gacatcaccg agagagagaa gcgggtcacc 2640
aagtggatca gcagacaaga gaacagcgac gtgcccctgg tgtaccagga tgagctgatc 2700
cagatccgcg agctgatgta caagccttac aaggactggg tcgccttcct gaagcagctc 2760
cacaagagac tggaagtcga gatcggcaaa gaagtgaagc actggcggaa gtccctgagc 2820
gacggaagaa agggcctgta cggcatctcc ctgaagaaca tcgacgagat cgatcggacc 2880
cggaagttcc tgctgagatg gtccctgagg cctaccgaac ctggcgaagt gcgtagactg 2940

gaacccggcc agagattcgc catcgaccag ctgaatcacc tgaacgccct gaaagaagat 3000
cggctgaaga agatggccaa caccatcatc atgcacgccc tgggctactg ctacgacgtg 3060
cggaagaaga aatggcaggc taagaacccc gcctgccaga tcatcctgtt cgaggatctg 3120
agcaactaca acccctacga ggaaaggtcc cgcttcgaga acagcaagct catgaagtgg 3180
tccagacgcg agatccccag acaggttgca ctgcagggcg agatctatgg cctgcaagtg 3240
ggagaagtgg gcgctcagtt cagcagcaga ttccacgcca agacaggcag ccctggcatc 3300
agatgtagcg tcgtgaccaa agagaagctg caggacaatc ggttcttcaa gaatctgcag 3360

agagagggca gactgaccct ggacaaaatc gccgtgctga aagagggcga tctgtaccca 3420
gacaaaggcg gcgagaagtt catcagcctg agcaaggatc ggaagtgcgt gaccacacac 3480
gccgacatca acgccgctca gaacctgcag aagcggttct ggacaagaac ccacggcttc 3540
tacaaggtgt actgcaaggc ctaccaggtg gacggccaga ccgtgtacat ccctgagagc 3600
aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660
ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720
agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840
aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900
accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960
ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020
gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080
gactatgcct aa 4092

<210> 275
<211> 1361
<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 275

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
115 120 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
130 135 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145 150 155 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Gly Gly Ser Gly Gly Ser Ser
165 170 175

Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu
180 185 190

Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
195 200 205

Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly

210 215 220

Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly

225 230 235 240

Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val

245 250 255

Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg

260 265 270

Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala

275 280 285

Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg Val

290 295 300

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys

305 310 315 320

Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala

325 330 335

Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser

340 345 350

Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Trp Glu Glu Glu Lys Lys

355 360 365

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly

370 375 380

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp

385 390 395 400

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg

405 410 415

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

420 425 430

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

435 440 445

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

450 455 460

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
465 470 475 480
Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
485 490 495
Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

500 505 510
Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
515 520 525
Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
530 535 540
Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
545 550 555 560
Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

565 570 575
Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
580 585 590
His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
595 600 605
Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
610 615 620
Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

625 630 635 640
Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
645 650 655
Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
660 665 670
Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
675 680 685
Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

690 695 700
Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met

705 710 715 720

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys

725 730 735

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu

740 745 750

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

755 760 765

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly

770 775 780

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys

785 790 795 800

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu

805 810 815

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

820 825 830

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu

835 840 845

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

850 855 860

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

865 870 875 880

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

885 890 895

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

900 905 910

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

915 920 925

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

930 935 940

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

945 950 955 960

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
965 970 975
Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
980 985 990
Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
995 1000 1005
Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp

1010 1015 1020
Gln Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu
1025 1030 1035
Ser Asn Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser
1040 1045 1050
Lys Leu Met Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala
1055 1060 1065
Leu Gln Gly Glu Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala

1070 1075 1080
Gln Phe Ser Ser Arg Phe His Ala Lys Thr Gly Ser Pro Gly Ile
1085 1090 1095
Arg Cys Ser Val Val Thr Lys Glu Lys Leu Gln Asp Asn Arg Phe
1100 1105 1110
Phe Lys Asn Leu Gln Arg Glu Gly Arg Leu Thr Leu Asp Lys Ile
1115 1120 1125
Ala Val Leu Lys Glu Gly Asp Leu Tyr Pro Asp Lys Gly Gly Glu

1130 1135 1140
Lys Phe Ile Ser Leu Ser Lys Asp Arg Lys Cys Val Thr Thr His
1145 1150 1155
Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Thr
1160 1165 1170
Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys Ala Tyr Gln Val
1175 1180 1185
Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp Gln Lys Gln

1190 1195 1200

Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu Lys Asp

1205 1210 1215

Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys Ile Lys Lys

1220 1225 1230

Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser Asp Ile

1235 1240 1245

Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly Glu Lys

1250 1255 1260

Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro Ser Asp

1265 1270 1275

Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu Arg Ile

1280 1285 1290

Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr Ile Glu

1295 1300 1305

Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala Ala Thr Lys

1310 1315 1320

Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp

1325 1330 1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr

1340 1345 1350

Pro Tyr Asp Val Pro Asp Tyr Ala

1355 1360

<210> 276

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 276

gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aaggaggctc tggaggaagc 840

tccgaagtcg agttttccca tgagtactgg atgagacacg cattgactct cgcaaagagg 900

gctcgagatg aacgcgaggt gcccgtgggg gcagtactcg tgctcaacaa tcgcgtaatc 960

ggcgaaggtt ggaatagggc aatcggactc cacgacccca ctgcacatgc ggaaatcatg 1020

gcccttcgac agggagggct tgtgatgcag aattatcgac tttatgatgc gacgctgtac 1080

gtcacgtttg aaccttgcgt aatgtgcgcg ggagctatga ttcactcccg cattggacga 1140

gttgtattcg gtgttcgcaa cgccaagacg ggtgccgcag gttcactgat ggacgtgctg 1200

catcatccag gcatgaacca ccgggtagaa atcacagaag gcatattggc ggacgaatgt 1260

gcggcgctgt tgtgtcgttt ttttcgcatg cccaggcggg tctttaacgc ccagaaaaaa 1320

gcacaatcct ctactgacgg ctcttctgga tctgaaacac ctggcacaag cgagagcgcc 1380

acccctgaga gctctggcga ggacatccag gctctgaagg ctctggaaca gtatgagaaa 1440

gagcggcaag aacagctgct gcgggacacc ctgaacacca acgagtaccg gctgagcaag 1500

agaggcctta gaggctggcg ggaaatcatc cagaaatggc tgaaaatgga cgagaacgag 1560

ccctccgaga agtacctgga agtgttcaag gactaccagc ggaagcaccc tagagaggcc 1620

ggcgattaca gcgtgtacga gttcctgtcc aagaaagaga accacttcat ctggcggaat 1680

caccctgagt acccctacct gtacgccacc ttctgcgaga tcgacaagaa aaagaaggac 1740

gccaagcagc aggccacctt cacactggcc gatcctatca atcaccctct gtgggtccga 1800

ttcgaggaaa gaagcggcag caacctgaac aagtacagaa tcctgaccga gcagctgcac 1860

accgagaagc tgaagaaaaa gctgacagtg cagctggacc ggctgatcta ccctacagaa 1920

tctggcggct gggaagagaa gggcaaagtg gacattgtgc tgctgcccag ccggcagttc 1980

tacaaccaga tcttcctgga catcgaggaa aagggcaagc acgccttcac ctacaaggat 2040

gagagcatca agttccctct gaagggcaca ctcggcggag ccagagtgca gttcgacaga 2100

gatcacctga gaagataccc tcacaaggtg gaaagcggca acgtgggcag aatctacttc 2160

aacatgaccg tgaacatcga gcctacagag tccccagtgt ccaagtctct gaagatccac 2220

cgggacgact tccccaaggt ggtcaacttc aagcccaaag aactgaccga gtggatcaag 2280

gacagcaagg gcaagaaact gaagtccggc atcgagtccc tggaaatcgg cctgagagtg 2340

atgagcatcg acctgggaca gagacaggcc gctgccgcct ctattttcga ggtggtggat 2400

cagaagcccg acatcgaagg caagctgttt ttcccaatca agggcaccga gctgtatgcc 2460

gtgcacagag ccagcttcaa catcaagctg cccggcgaga cactggtcaa gagcagagaa 2520

gtgctgcgga aggccagaga ggacaatctg aaactgatga accagaagct caacttcctg 2580

cggaacgtgc tgcacttcca gcagttcgag gacatcaccg agagagagaa gcgggtcacc 2640

aagtggatca gcagacaaga gaacagcgac gtgcccctgg tgtaccagga tgagctgatc 2700

cagatccgcg agctgatgta caagccttac aaggactggg tcgccttcct gaagcagctc 2760

cacaagagac tggaagtcga gatcggcaaa gaagtgaagc actggcggaa gtccctgagc 2820

gacggaagaa agggcctgta cggcatctcc ctgaagaaca tcgacgagat cgatcggacc 2880

cggaagttcc tgctgagatg gtccctgagg cctaccgaac ctggcgaagt gcgtagactg 2940

gaacccggcc agagattcgc catcgaccag ctgaatcacc tgaacgccct gaaagaagat 3000

cggctgaaga agatggccaa caccatcatc atgcacgccc tgggctactg ctacgacgtg 3060

cggaagaaga aatggcaggc taagaacccc gcctgccaga tcatcctgtt cgaggatctg 3120

agcaactaca acccctacga ggaaaggtcc cgcttcgaga acagcaagct catgaagtgg 3180

tccagacgcg agatccccag acaggttgca ctgcagggcg agatctatgg cctgcaagtg 3240

ggagaagtgg gcgctcagtt cagcagcaga ttccacgcca agacaggcag ccctggcatc 3300

agatgtagcg tcgtgaccaa agagaagctg caggacaatc ggttcttcaa gaatctgcag 3360

agagagggca gactgaccct ggacaaaatc gccgtgctga aagagggcga tctgtaccca 3420

gacaaaggcg gcgagaagtt catcagcctg agcaaggatc ggaagtgcgt gaccacacac 3480

gccgacatca acgccgctca gaacctgcag aagcggttct ggacaagaac ccacggcttc 3540

tacaaggtgt actgcaaggc ctaccaggtg gacggccaga ccgtgtacat ccctgagagc 3600

aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660

ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840
aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900
accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960
ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020
gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080
gactatgcct aa 4092

<210> 277
<211> 1361
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 277

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15
Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30
Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45
Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60
His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80
Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95
Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110
Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
115 120 125
Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn

130 135 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp

145 150 155 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys

165 170 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly

180 185 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp

195 200 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg

210 215 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

225 230 235 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

245 250 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

260 265 270

Gly Gly Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr Trp

275 280 285

Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu

290 295 300

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu

305 310 315 320

Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu

325 330 335

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu

340 345 350

Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala

355 360 365

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg

370 375 380

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His His
385 390 395 400
Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
405 410 415
Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val
420 425 430
Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly

435 440 445
Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly
450 455 460
Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
465 470 475 480
Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
485 490 495
Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

500 505 510
Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
515 520 525
Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
530 535 540
Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
545 550 555 560
Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

565 570 575
Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
580 585 590
His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
595 600 605
Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
610 615 620
Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

625 630 635 640

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg

645 650 655

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His

660 665 670

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr

675 680 685

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

690 695 700

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met

705 710 715 720

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys

725 730 735

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu

740 745 750

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

755 760 765

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly

770 775 780

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys

785 790 795 800

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu

805 810 815

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

820 825 830

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu

835 840 845

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

850 855 860

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

865 870 875 880

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

885 890 895

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

900 905 910

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

915 920 925

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

930 935 940

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

945 950 955 960

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

965 970 975

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu

980 985 990

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

995 1000 1005

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp

1010 1015 1020

Gln Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu

1025 1030 1035

Ser Asn Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser

1040 1045 1050

Lys Leu Met Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala

1055 1060 1065

Leu Gln Gly Glu Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala

1070 1075 1080

Gln Phe Ser Ser Arg Phe His Ala Lys Thr Gly Ser Pro Gly Ile

1085 1090 1095

Arg Cys Ser Val Val Thr Lys Glu Lys Leu Gln Asp Asn Arg Phe

1100 1105 1110

Phe Lys Asn Leu Gln Arg Glu Gly Arg Leu Thr Leu Asp Lys Ile

1115 1120 1125

Ala Val Leu Lys Glu Gly Asp Leu Tyr Pro Asp Lys Gly Gly Glu

1130 1135 1140

Lys Phe Ile Ser Leu Ser Lys Asp Arg Lys Cys Val Thr Thr His

1145 1150 1155

Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Thr

1160 1165 1170

Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys Ala Tyr Gln Val

1175 1180 1185

Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp Gln Lys Gln

1190 1195 1200

Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu Lys Asp

1205 1210 1215

Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys Ile Lys Lys

1220 1225 1230

Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser Asp Ile

1235 1240 1245

Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly Glu Lys

1250 1255 1260

Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro Ser Asp

1265 1270 1275

Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu Arg Ile

1280 1285 1290

Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr Ile Glu

1295 1300 1305

Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala Ala Thr Lys

1310 1315 1320

Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp

1325 1330 1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr

1340 1345 1350

Pro Tyr Asp Val Pro Asp Tyr Ala

1355 1360

<210> 278

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 278

gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aagaggacat ccaggctctg 840

aaggctctgg aacagtatga gaaagagcgg caagaacagc tgctgcggga caccctgaac 900

accaacgagt accggctgag caagagaggc cttagaggct ggcgggaaat catccagaaa 960

tggctgaaaa tggacggagg ctctggagga agctccgaag tcgagttttc ccatgagtac 1020

tggatgagac acgcattgac tctcgcaaag agggctcgag atgaacgcga ggtgcccgtg 1080

ggggcagtac tcgtgctcaa caatcgcgta atcggcgaag gttggaatag ggcaatcgga 1140

ctccacgacc ccactgcaca tgcggaaatc atggcccttc gacagggagg gcttgtgatg 1200

cagaattatc gactttatga tgcgacgctg tacgtcacgt ttgaaccttg cgtaatgtgc 1260

gcgggagcta tgattcactc ccgcattgga cgagttgtat tcggtgttcg caacgccaag 1320

acgggtgccg caggttcact gatggacgtg ctgcatcatc caggcatgaa ccaccgggta 1380

gaaatcacag aaggcatatt ggcggacgaa tgtgcggcgc tgttgtgtcg ttttttttcgc 1440

atgcccaggc gggtctttaa cgcccagaaa aaagcacaat cctctactga cggctcttct 1500

ggatctgaaa cacctggcac aagcgagagc gccacccctg agagctctgg cgagaacgag 1560

ccctccgaga agtacctgga agtgttcaag gactaccagc ggaagcaccc tagagaggcc 1620

ggcgattaca gcgtgtacga gttcctgtcc aagaaagaga accacttcat ctggcggaat 1680

caccctgagt acccctacct gtacgccacc ttctgcgaga tcgacaagaa aaagaaggac 1740

gccaagcagc aggccacctt cacactggcc gatcctatca atcaccctct gtgggtccga 1800

ttcgaggaaa gaagcggcag caacctgaac aagtacagaa tcctgaccga gcagctgcac 1860

accgagaagc tgaagaaaaa gctgacagtg cagctggacc ggctgatcta ccctacagaa 1920

tctggcggct gggaagagaa gggcaaagtg gacattgtgc tgctgcccag ccggcagttc 1980

tacaaccaga tcttcctgga catcgaggaa aagggcaagc acgccttcac ctacaaggat 2040

gagagcatca agttccctct gaagggcaca ctcggcggag ccagagtgca gttcgacaga 2100

gatcacctga gaagataccc tcacaaggtg gaaagcggca acgtgggcag aatctacttc 2160

aacatgaccg tgaacatcga gcctacagag tccccagtgt ccaagtctct gaagatccac 2220

cgggacgact tccccaaggt ggtcaacttc aagcccaaag aactgaccga gtggatcaag 2280

gacagcaagg gcaagaaact gaagtccggc atcgagtccc tggaaatcgg cctgagagtg 2340

atgagcatcg acctgggaca gagacaggcc gctgccgcct ctattttcga ggtggtggat 2400

cagaagcccg acatcgaagg caagctgttt ttcccaatca agggcaccga gctgtatgcc 2460

gtgcacagag ccagcttcaa catcaagctg cccggcgaga cactggtcaa gagcagagaa 2520

gtgctgcgga aggccagaga ggacaatctg aaactgatga accagaagct caacttcctg 2580

cggaacgtgc tgcacttcca gcagttcgag gacatcaccg agagagagaa gcgggtcacc 2640

aagtggatca gcagacaaga gaacagcgac gtgcccctgg tgtaccagga tgagctgatc 2700

cagatccgcg agctgatgta caagccttac aaggactggg tcgccttcct gaagcagctc 2760

cacaagagac tggaagtcga gatcggcaaa gaagtgaagc actggcggaa gtccctgagc 2820

gacggaagaa agggcctgta cggcatctcc ctgaagaaca tcgacgagat cgatcggacc 2880

cggaagttcc tgctgagatg gtccctgagg cctaccgaac ctggcgaagt gcgtagactg 2940

gaacccggcc agagattcgc catcgaccag ctgaatcacc tgaacgccct gaaagaagat 3000

cggctgaaga agatggccaa caccatcatc atgcacgccc tgggctactg ctacgacgtg 3060

cggaagaaga aatggcaggc taagaacccc gcctgccaga tcatcctgtt cgaggatctg 3120

agcaactaca acccctacga ggaaaggtcc cgcttcgaga acagcaagct catgaagtgg 3180

tccagacgcg agatccccag acaggttgca ctgcagggcg agatctatgg cctgcaagtg 3240

ggagaagtgg gcgctcagtt cagcagcaga ttccacgcca agacaggcag ccctggcatc 3300

agatgtagcg tcgtgaccaa agagaagctg caggacaatc ggttcttcaa gaatctgcag 3360

agagagggca gactgaccct ggacaaaatc gccgtgctga aagagggcga tctgtaccca 3420

gacaaaggcg gcgagaagtt catcagcctg agcaaggatc ggaagtgcgt gaccacacac 3480

gccgacatca acgccgctca gaacctgcag aagcggttct ggacaagaac ccacggcttc 3540

tacaaggtgt actgcaaggc ctaccaggtg gacggccaga ccgtgtacat ccctgagagc 3600

aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660

ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 279

<211> 1361

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 279

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80
Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95
Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110
Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu

115 120 125
Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
130 135 140
Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145 150 155 160
Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
165 170 175
Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly

180 185 190
Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
195 200 205
Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
210 215 220
Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
225 230 235 240
Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

245 250 255
Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
260 265 270
Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
275 280 285
Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
290 295 300
Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

305 310 315 320

Lys Met Asp Gly Gly Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His

325 330 335

Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp

340 345 350

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val

355 360 365

Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala

370 375 380

His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn

385 390 395 400

Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val

405 410 415

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

420 425 430

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val

435 440 445

Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile

450 455 460

Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro

465 470 475 480

Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly

485 490 495

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu

500 505 510

Ser Ser Gly Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys

515 520 525

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr

530 535 540

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

545 550 555 560

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

565 570 575

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn

580 585 590

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn

595 600 605

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys

610 615 620

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

625 630 635 640

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg

645 650 655

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His

660 665 670

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr

675 680 685

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

690 695 700

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met

705 710 715 720

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys

725 730 735

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu

740 745 750

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

755 760 765

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly

770 775 780

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys

785 790 795 800

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu

805 810 815

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

820 825 830

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu

835 840 845

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

850 855 860

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

865 870 875 880

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

885 890 895

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

900 905 910

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

915 920 925

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

930 935 940

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

945 950 955 960

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

965 970 975

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu

980 985 990

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

995 1000 1005

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp

1010 1015 1020

Gln Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu

1025 1030 1035

Ser Asn Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser

1040 1045 1050

Lys Leu Met Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala
1055 1060 1065
Leu Gln Gly Glu Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala

1070 1075 1080
Gln Phe Ser Ser Arg Phe His Ala Lys Thr Gly Ser Pro Gly Ile
1085 1090 1095
Arg Cys Ser Val Val Thr Lys Glu Lys Leu Gln Asp Asn Arg Phe
1100 1105 1110
Phe Lys Asn Leu Gln Arg Glu Gly Arg Leu Thr Leu Asp Lys Ile
1115 1120 1125
Ala Val Leu Lys Glu Gly Asp Leu Tyr Pro Asp Lys Gly Gly Glu

1130 1135 1140
Lys Phe Ile Ser Leu Ser Lys Asp Arg Lys Cys Val Thr Thr His
1145 1150 1155
Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Thr
1160 1165 1170
Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys Ala Tyr Gln Val
1175 1180 1185
Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp Gln Lys Gln

1190 1195 1200
Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu Lys Asp
1205 1210 1215
Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys Ile Lys Lys
1220 1225 1230
Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser Asp Ile
1235 1240 1245
Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly Glu Lys

1250 1255 1260
Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro Ser Asp
1265 1270 1275
Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu Arg Ile

1280 1285 1290

Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr Ile Glu

1295 1300 1305

Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala Ala Thr Lys

1310 1315 1320

Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp

1325 1330 1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr

1340 1345 1350

Pro Tyr Asp Val Pro Asp Tyr Ala

1355 1360

<210> 280

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 280

gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aagaggacat ccaggctctg 840

aaggctctgg aacagtatga gaaagagcgg caagaacagc tgctgcggga caccctgaac 900
accaacgagt accggctgag caagagaggc cttagaggct ggcgggaaat catccagaaa 960
tggctgaaaa tggacgagaa cgagccctcc gagaagtacc tggaagtgtt caaggactac 1020
cagcggaagc accctagaga ggccggcgat tacagcgtgt acgagttcct gtccaagaaa 1080
gagaaccact tcatctggcg gaatcaccct gagtaccct acctgtacgc caccttctgc 1140
gagatcgaca agaaaaagaa ggacgccaag cagcaggcca ccttcacact ggccgatcct 1200
atcaatcacc ctctgtgggt ccgattcgag gaaagaagcg gcagcaacct gaacaagtac 1260

agaatcctga ccgagcagct gcacaccgag aagctgaaga aaaagctgac agtgcagctg 1320
gaccggctga tctaccctac agaatctggc ggctgggaag agaagggcaa agtggacatt 1380
gtgctgctgc ccagccggca gttctacaac cagatcttcc tggacatcga ggaaaagggc 1440
aagcacgcct tcacctacaa ggatgagagc atcaagttcc ctctgaaggg cacactcggc 1500
ggagccagag tgcagttcga cagagatcac ctgagaagat accctcacaa ggtggaaagc 1560
ggcaacgtgg gcagaatcta cttcaacatg accgtgaaca tcgagcctac agagtcccca 1620
gtgtccaagt ctctgaagat ccaccgggac gacttcccca aggtggtcaa cttcaagccc 1680

aaagaactga ccgagtggat caaggacagc aagggcaaga aactgaagtc cggcatcgag 1740
tccctggaaa tcggcctgag agtgatgagc atcgacctgg gacagagaca ggccgctgcc 1800
gcctctattt tcgaggtggt ggatcagaag cccgacatcg aaggcaagct gtttttccca 1860
atcaagggca ccgagctgta tgccgtgcac agagccagct tcaacatcaa gctgcccggc 1920
gagacactgg tcaagagcag agaagtgctg cggaaggcca gagaggacaa tctgaaactg 1980
atgaaccaga agctcaactt cctgcggaac gtgctgcact tccagcagtt cgaggacatc 2040
accgagagag agaagcgggt caccaagtgg atcagcagac aagagaacag cgacgtgccc 2100

ctggtgtacc aggatgagct gatccagatc cgcgagctga tgtacaagcc ttacaaggac 2160
tgggtcgcct tcctgaagca gctccacaag agactggaag tcgagatcgg caaagaagtg 2220
aagcactggc ggaagtccct gagcgacgga agaaagggcc tgtacggcat ctccctgaag 2280
aacatcgacg agatcgatcg gacccggaag ttcctgctga gatggtccct gaggcctacc 2340
gaacctggcg aagtgcgtag actggaaccc ggccagagat tcgccatcga ccagctgaat 2400
cacctgaacg ccctgaaaga agatcggctg aagaagatgg ccaacaccat catcatgcac 2460
gccctgggct actgctacga cgtgcggaag aagaaatggc aggctaagaa ccccgcctgc 2520

cagatcatcc tgttcgagga tctgagcaac tacaacccct acgaggaaag gtcccgcttc 2580

gagaacagca agctcatgaa gtggtccaga cgcgagatcc ccagacaggt tgcactgcag 2640

ggcgagatct atggcctgca agtgggagaa gtgggcgctc agttcagcag cagattccac 2700

gccaagacag gcagccctgg catcagatgt agcgtcgtga ccaaagagaa gctgcaggac 2760

aatcggttct tcaagaatct gcagagagag ggcagactga ccctggacaa aatcgccgtg 2820

ctgaaagagg gcgatctgta cccagacaaa ggcggcgaga agttcatcag cctgagcaag 2880

gatcggaagt gcgtgaccac acacgccgac atcaacgccg ctcagaacct gcagaagcgg 2940

ttctggacaa gaacccacgg cttctacaag gtgtactgca aggcctacca ggtggacgga 3000

ggctctggag gaagctccga agtcgagttt tcccatgagt actggatgag acacgcattg 3060

actctcgcaa agagggctcg agatgaacgc gaggtgcccg tgggggcagt actcgtgctc 3120

aacaatcgcg taatcggcga aggttggaat agggcaatcg gactccacga ccccactgca 3180

catgcggaaa tcatggccct tcgacaggga gggcttgtga tgcagaatta tcgactttat 3240

gatgcgacgc tgtacgtcac gtttgaacct tgcgtaatgt gcgcgggagc tatgattcac 3300

tcccgcattg gacgagttgt attcggtgtt cgcaacgcca agacgggtgc cgcaggttca 3360

ctgatggacg tgctgcatca tccaggcatg aaccaccggg tagaaatcac agaaggcata 3420

ttggcggacg aatgtgcggc gctgttgtgt cgtttttttc gcatgcccag gcgggtcttt 3480

aacgcccaga aaaaagcaca atcctctact gacggctctt ctggatctga aacacctggc 3540

acaagcgaga gcgccacccc tgagagctct ggcggccaga ccgtgtacat ccctgagagc 3600

aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660

ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 281

<211> 1361

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 281

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
115 120 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
130 135 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145 150 155 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
165 170 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
180 185 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
195 200 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
210 215 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

225 230 235 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

245 250 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

260 265 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg

275 280 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu

290 295 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

305 310 315 320

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys

325 330 335

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr

340 345 350

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

355 360 365

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

370 375 380

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn

385 390 395 400

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn

405 410 415

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys

420 425 430

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

435 440 445

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg

450 455 460

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His

465 470 475 480

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
485 490 495
Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

500 505 510
Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
515 520 525
Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
530 535 540
Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
545 550 555 560
Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

565 570 575
Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
580 585 590
Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
595 600 605
Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
610 615 620
Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

625 630 635 640
Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
645 650 655
Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
660 665 670
Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
675 680 685
Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

690 695 700
Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
705 710 715 720
Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

725 730 735

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

740 745 750

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

755 760 765

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

770 775 780

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu

785 790 795 800

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

805 810 815

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln

820 825 830

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn

835 840 845

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met

850 855 860

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu

865 870 875 880

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg

885 890 895

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr

900 905 910

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu

915 920 925

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu

930 935 940

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg

945 950 955 960

Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln

965 970 975

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
980 985 990
Ala Tyr Gln Val Asp Gly Gly Ser Gly Gly Ser Ser Glu Val Glu Phe
995 1000 1005
Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg

1010 1015 1020
Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu
1025 1030 1035
Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu
1040 1045 1050
His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly
1055 1060 1065
Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr

1070 1075 1080
Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
1085 1090 1095
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr
1100 1105 1110
Gly Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met
1115 1120 1125
Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys

1130 1135 1140
Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe
1145 1150 1155
Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly
1160 1165 1170
Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
1175 1180 1185
Gly Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp Gln Lys Gln

1190 1195 1200
Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu Lys Asp

1205 1210 1215

Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys Ile Lys Lys

1220 1225 1230

Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser Asp Ile

1235 1240 1245

Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly Glu Lys

1250 1255 1260

Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro Ser Asp

1265 1270 1275

Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu Arg Ile

1280 1285 1290

Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr Ile Glu

1295 1300 1305

Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala Ala Thr Lys

1310 1315 1320

Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp

1325 1330 1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr

1340 1345 1350

Pro Tyr Asp Val Pro Asp Tyr Ala

1355 1360

<210> 282

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 282

gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aagaggacat ccaggctctg 840

aaggctctgg aacagtatga gaaagagcgg caagaacagc tgctgcggga caccctgaac 900

accaacgagt accggctgag caagagaggc cttagaggct ggcgggaaat catccagaaa 960

tggctgaaaa tggacgagaa cgagccctcc gagaagtacc tggaagtgtt caaggactac 1020

cagcggaagc accctagaga ggccggcgat tacagcgtgt acgagttcct gtccaagaaa 1080

gagaaccact tcatctggcg gaatcaccct gagtacccct acctgtacgc caccttctgc 1140

gagatcgaca agaaaaagaa ggacgccaag cagcaggcca ccttcacact ggccgatcct 1200

atcaatcacc ctctgtgggt ccgattcgag gaaagaagcg gcagcaacct gaacaagtac 1260

agaatcctga ccgagcagct gcacaccgag aagctgaaga aaaagctgac agtgcagctg 1320

gaccggctga tctaccctac agaatctggc ggctgggaag agaagggcaa agtggacatt 1380

gtgctgctgc ccagccggca gttctacaac cagatcttcc tggacatcga ggaaaagggc 1440

aagcacgcct tcacctacaa ggatgagagc atcaagttcc ctctgaaggg cacactcggc 1500

ggagccagag tgcagttcga cagagatcac ctgagaagat accctcacaa ggtggaaagc 1560

ggcaacgtgg gcagaatcta cttcaacatg accgtgaaca tcgagcctac agagtcccca 1620

gtgtccaagt ctctgaagat ccaccgggac gacttcccca aggtggtcaa cttcaagccc 1680

aaagaactga ccgagtggat caaggacagc aagggcaaga aactgaagtc cggcatcgag 1740

tccctggaaa tcggcctgag agtgatgagc atcgacctgg gacagagaca ggccgctgcc 1800

gcctctattt tcgaggtggt ggatcagaag cccgacatcg aaggcaagct gtttttccca 1860

atcaagggca ccgagctgta tgccgtgcac agagccagct tcaacatcaa gctgcccggc 1920

gagacactgg tcaagagcag agaagtgctg cggaaggcca gagaggacaa tctgaaactg 1980

atgaaccaga agctcaactt cctgcggaac gtgctgcact tccagcagtt cgaggacatc 2040

accgagagag agaagcgggt caccaagtgg atcagcagac aagagaacag cgacgtgccc 2100

ctggtgtacc aggatgagct gatccagatc cgcgagctga tgtacaagcc ttacaaggac 2160
tgggtcgcct tcctgaagca gctccacaag agactggaag tcgagatcgg caaagaagtg 2220
aagcactggc ggaagtccct gagcgacgga agaaagggcc tgtacggcat ctccctgaag 2280
aacatcgacg agatcgatcg gacccggaag ttcctgctga gatggtccct gaggcctacc 2340
gaacctggcg aagtgcgtag actggaaccc ggccagagat tcgccatcga ccagctgaat 2400
cacctgaacg ccctgaaaga agatcggctg aagaagatgg ccaacaccat catcatgcac 2460
gccctgggct actgctacga cgtgcggaag aagaaatggc aggctaagaa ccccgcctgc 2520

cagatcatcc tgttcgagga tctgagcaac tacaacccct acgaggaaag gtcccgcttc 2580
gagaacagca agctcatgaa gtggtccaga cgcgagatcc ccagacaggt tgcactgcag 2640
ggcgagatct atggcctgca agtgggagaa gtgggcgctc agttcagcag cagattccac 2700
gccaagacag gcagccctgg catcagatgt agcgtcgtga ccaaagagaa gctgcaggac 2760
aatcggttct tcaagaatct gcagagagag ggcagactga ccctggacaa aatcgccgtg 2820
ctgaaagagg gcgatctgta cccagacaaa ggcggcgaga agttcatcag cctgagcaag 2880
gatcggaagt gcgtgaccac acacgccgac atcaacgccg ctcagaacct gcagaagcgg 2940

ttctggacaa gaacccacgg cttctacaag gtgtactgca aggcctacca ggtggacggc 3000
cagaccgtgt acatccctga gagcaaggac cagaagcaga agatcatcga agagttcggc 3060
gagggctact tcattctgaa ggacggggtg tacgaatggg tcaacgccgg caagggaggc 3120
tctggaggaa gctccgaagt cgagttttcc catgagtact ggatgagaca cgcattgact 3180
ctcgcaaaga gggctcgaga tgaacgcgag gtgcccgtgg gggcagtact cgtgctcaac 3240
aatcgcgtaa tcggcgaagg ttggaatagg gcaatcggac tccacgaccc cactgcacat 3300
gcggaaatca tggcccttcg acagggaggg cttgtgatgc agaattatcg actttatgat 3360

gcgacgctgt acgtcacgtt tgaaccttgc gtaatgtgcg cgggagctat gattcactcc 3420
cgcattggac gagttgtatt cggtgttcgc aacgccaaga cgggtgccgc aggttcactg 3480
atggacgtgc tgcatcatcc aggcatgaac caccgggtag aaatcacaga aggcatattg 3540
gcggacgaat gtgcggcgct gttgtgtcgt ttttttcgca tgcccaggcg ggtctttaac 3600
gcccagaaaa aagcacaatc ctctactgac ggctcttctg gatctgaaac acctggcaca 3660
agcgagagcg ccacccctga gagctctggc ctgaaaatca agaagggcag ctccaagcag 3720
agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 283

<211> 1361

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 283

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
115 120 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
130 135 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp

145 150 155 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys

165 170 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly

180 185 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp

195 200 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg

210 215 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

225 230 235 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

245 250 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

260 265 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg

275 280 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu

290 295 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

305 310 315 320

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys

325 330 335

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr

340 345 350

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

355 360 365

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

370 375 380

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn

385 390 395 400

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
405 410 415
Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
420 425 430
Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

435 440 445
Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
450 455 460
Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
465 470 475 480
Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
485 490 495
Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

500 505 510
Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
515 520 525
Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
530 535 540
Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
545 550 555 560
Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

565 570 575
Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
580 585 590
Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
595 600 605
Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
610 615 620
Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

625 630 635 640
Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu

645 650 655

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

660 665 670

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

675 680 685

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

690 695 700

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

705 710 715 720

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

725 730 735

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

740 745 750

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

755 760 765

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

770 775 780

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu

785 790 795 800

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

805 810 815

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln

820 825 830

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn

835 840 845

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met

850 855 860

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu

865 870 875 880

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg

885 890 895

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr

900 905 910

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu

915 920 925

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu

930 935 940

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg

945 950 955 960

Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln

965 970 975

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys

980 985 990

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp

995 1000 1005

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile

1010 1015 1020

Leu Lys Asp Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Gly Gly

1025 1030 1035

Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met

1040 1045 1050

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu

1055 1060 1065

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

1070 1075 1080

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His

1085 1090 1095

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn

1100 1105 1110

Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys

1115 1120 1125

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

1130 1135 1140

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu

1145 1150 1155

Met Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile

1160 1165 1170

Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg

1175 1180 1185

Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala

1190 1195 1200

Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1205 1210 1215

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Leu Lys Ile Lys Lys

1220 1225 1230

Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser Asp Ile

1235 1240 1245

Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly Glu Lys

1250 1255 1260

Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro Ser Asp

1265 1270 1275

Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu Arg Ile

1280 1285 1290

Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr Ile Glu

1295 1300 1305

Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala Ala Thr Lys

1310 1315 1320

Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp

1325 1330 1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr

1340 1345 1350

Pro Tyr Asp Val Pro Asp Tyr Ala

1355 1360

<210> 284

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 284

Gly Gly Gly Ser Gly Gly Ser

1 5

<210> 285

<211> 707

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas-phi-1 sequence

<400> 285

Met Ala Asp Thr Pro Thr Leu Phe Thr Gln Phe Leu Arg His His Leu

1 5 10 15

Pro Gly Gln Arg Phe Arg Lys Asp Ile Leu Lys Gln Ala Gly Arg Ile

20 25 30

Leu Ala Asn Lys Gly Glu Asp Ala Thr Ile Ala Phe Leu Arg Gly Lys

35 40 45

Ser Glu Glu Ser Pro Pro Asp Phe Gln Pro Pro Val Lys Cys Pro Ile

50 55 60

Ile Ala Cys Ser Arg Pro Leu Thr Glu Trp Pro Ile Tyr Gln Ala Ser

65 70 75 80

Val Ala Ile Gln Gly Tyr Val Tyr Gly Gln Ser Leu Ala Glu Phe Glu

85 90 95

Ala Ser Asp Pro Gly Cys Ser Lys Asp Gly Leu Leu Gly Trp Phe Asp

100 105 110

Lys Thr Gly Val Cys Thr Asp Tyr Phe Ser Val Gln Gly Leu Asn Leu

115 120 125

Ile Phe Gln Asn Ala Arg Lys Arg Tyr Ile Gly Val Gln Thr Lys Val

130 135 140

Thr Asn Arg Asn Glu Lys Arg His Lys Lys Leu Lys Arg Ile Asn Ala

145 150 155 160

Lys Arg Ile Ala Glu Gly Leu Pro Glu Leu Thr Ser Asp Glu Pro Glu

165 170 175

Ser Ala Leu Asp Glu Thr Gly His Leu Ile Asp Pro Pro Gly Leu Asn

180 185 190

Thr Asn Ile Tyr Cys Tyr Gln Gln Val Ser Pro Lys Pro Leu Ala Leu

195 200 205

Ser Glu Val Asn Gln Leu Pro Thr Ala Tyr Ala Gly Tyr Ser Thr Ser

210 215 220

Gly Asp Asp Pro Ile Gln Pro Met Val Thr Lys Asp Arg Leu Ser Ile

225 230 235 240

Ser Lys Gly Gln Pro Gly Tyr Ile Pro Glu His Gln Arg Ala Leu Leu

245 250 255

Ser Gln Lys Lys His Arg Arg Met Arg Gly Tyr Gly Leu Lys Ala Arg

260 265 270

Ala Leu Leu Val Ile Val Arg Ile Gln Asp Asp Trp Ala Val Ile Asp

275 280 285

Leu Arg Ser Leu Leu Arg Asn Ala Tyr Trp Arg Arg Ile Val Gln Thr

290 295 300

Lys Glu Pro Ser Thr Ile Thr Lys Leu Leu Lys Leu Val Thr Gly Asp

305 310 315 320

Pro Val Leu Asp Ala Thr Arg Met Val Ala Thr Phe Thr Tyr Lys Pro

325 330 335

Gly Ile Val Gln Val Arg Ser Ala Lys Cys Leu Lys Asn Lys Gln Gly

340 345 350

Ser Lys Leu Phe Ser Glu Arg Tyr Leu Asn Glu Thr Val Ser Val Thr

355 360 365

Ser Ile Asp Leu Gly Ser Asn Asn Leu Val Ala Val Ala Thr Tyr Arg

370 375 380

Leu Val Asn Gly Asn Thr Pro Glu Leu Leu Gln Arg Phe Thr Leu Pro
385 390 395 400
Ser His Leu Val Lys Asp Phe Glu Arg Tyr Lys Gln Ala His Asp Thr
405 410 415
Leu Glu Asp Ser Ile Gln Lys Thr Ala Val Ala Ser Leu Pro Gln Gly

420 425 430
Gln Gln Thr Glu Ile Arg Met Trp Ser Met Tyr Gly Phe Arg Glu Ala
435 440 445
Gln Glu Arg Val Cys Gln Glu Leu Gly Leu Ala Asp Gly Ser Ile Pro
450 455 460
Trp Asn Val Met Thr Ala Thr Ser Thr Ile Leu Thr Asp Leu Phe Leu
465 470 475 480
Ala Arg Gly Gly Asp Pro Lys Lys Cys Met Phe Thr Ser Glu Pro Lys

485 490 495
Lys Lys Lys Asn Ser Lys Gln Val Leu Tyr Lys Ile Arg Asp Arg Ala
500 505 510
Trp Ala Lys Met Tyr Arg Thr Leu Leu Ser Lys Glu Thr Arg Glu Ala
515 520 525
Trp Asn Lys Ala Leu Trp Gly Leu Lys Arg Gly Ser Pro Asp Tyr Ala
530 535 540
Arg Leu Ser Lys Arg Lys Glu Glu Leu Ala Arg Arg Cys Val Asn Tyr

545 550 555 560
Thr Ile Ser Thr Ala Glu Lys Arg Ala Gln Cys Gly Arg Thr Ile Val
565 570 575
Ala Leu Glu Asp Leu Asn Ile Gly Phe Phe His Gly Arg Gly Lys Gln
580 585 590
Glu Pro Gly Trp Val Gly Leu Phe Thr Arg Lys Lys Glu Asn Arg Trp
595 600 605
Leu Met Gln Ala Leu His Lys Ala Phe Leu Glu Leu Ala His His Arg

610 615 620
Gly Tyr His Val Ile Glu Val Asn Pro Ala Tyr Thr Ser Gln Thr Cys

625 630 635 640

Pro Val Cys Arg His Cys Asp Pro Asp Asn Arg Asp Gln His Asn Arg

645 650 655

Glu Ala Phe His Cys Ile Gly Cys Gly Phe Arg Gly Asn Ala Asp Leu

660 665 670

Asp Val Ala Thr His Asn Ile Ala Met Val Ala Ile Thr Gly Glu Ser

675 680 685

Leu Lys Arg Ala Arg Gly Ser Val Ala Ser Lys Thr Pro Gln Pro Leu

690 695 700

Ala Ala Glu

705

<210> 286

<211> 757

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas-phi-2 sequence

<400> 286

Met Pro Lys Pro Ala Val Glu Ser Glu Phe Ser Lys Val Leu Lys Lys

1 5 10 15

His Phe Pro Gly Glu Arg Phe Arg Ser Ser Tyr Met Lys Arg Gly Gly

20 25 30

Lys Ile Leu Ala Ala Gln Gly Glu Glu Ala Val Val Ala Tyr Leu Gln

35 40 45

Gly Lys Ser Glu Glu Glu Pro Pro Asn Phe Gln Pro Pro Ala Lys Cys

50 55 60

His Val Val Thr Lys Ser Arg Asp Phe Ala Glu Trp Pro Ile Met Lys

65 70 75 80

Ala Ser Glu Ala Ile Gln Arg Tyr Ile Tyr Ala Leu Ser Thr Thr Glu

85 90 95

Arg Ala Ala Cys Lys Pro Gly Lys Ser Ser Glu Ser His Ala Ala Trp

100 105 110

Phe Ala Ala Thr Gly Val Ser Asn His Gly Tyr Ser His Val Gln Gly
115 120 125
Leu Asn Leu Ile Phe Asp His Thr Leu Gly Arg Tyr Asp Gly Val Leu
130 135 140
Lys Lys Val Gln Leu Arg Asn Glu Lys Ala Arg Ala Arg Leu Glu Ser

145 150 155 160
Ile Asn Ala Ser Arg Ala Asp Glu Gly Leu Pro Glu Ile Lys Ala Glu
165 170 175
Glu Glu Glu Val Ala Thr Asn Glu Thr Gly His Leu Leu Gln Pro Pro
180 185 190
Gly Ile Asn Pro Ser Phe Tyr Val Tyr Gln Thr Ile Ser Pro Gln Ala
195 200 205
Tyr Arg Pro Arg Asp Glu Ile Val Leu Pro Pro Glu Tyr Ala Gly Tyr

210 215 220
Val Arg Asp Pro Asn Ala Pro Ile Pro Leu Gly Val Val Arg Asn Arg
225 230 235 240
Cys Asp Ile Gln Lys Gly Cys Pro Gly Tyr Ile Pro Glu Trp Gln Arg
245 250 255
Glu Ala Gly Thr Ala Ile Ser Pro Lys Thr Gly Lys Ala Val Thr Val
260 265 270
Pro Gly Leu Ser Pro Lys Lys Asn Lys Arg Met Arg Arg Tyr Trp Arg

275 280 285
Ser Glu Lys Glu Lys Ala Gln Asp Ala Leu Leu Val Thr Val Arg Ile
290 295 300
Gly Thr Asp Trp Val Val Ile Asp Val Arg Gly Leu Leu Arg Asn Ala
305 310 315 320
Arg Trp Arg Thr Ile Ala Pro Lys Asp Ile Ser Leu Asn Ala Leu Leu
325 330 335
Asp Leu Phe Thr Gly Asp Pro Val Ile Asp Val Arg Arg Asn Ile Val

340 345 350
Thr Phe Thr Tyr Thr Leu Asp Ala Cys Gly Thr Tyr Ala Arg Lys Trp

355 360 365

Thr Leu Lys Gly Lys Gln Thr Lys Ala Thr Leu Asp Lys Leu Thr Ala

370 375 380

Thr Gln Thr Val Ala Leu Val Ala Ile Asp Leu Gly Gln Thr Asn Pro

385 390 395 400

Ile Ser Ala Gly Ile Ser Arg Val Thr Gln Glu Asn Gly Ala Leu Gln

405 410 415

Cys Glu Pro Leu Asp Arg Phe Thr Leu Pro Asp Asp Leu Leu Lys Asp

420 425 430

Ile Ser Ala Tyr Arg Ile Ala Trp Asp Arg Asn Glu Glu Glu Leu Arg

435 440 445

Ala Arg Ser Val Glu Ala Leu Pro Glu Ala Gln Gln Ala Glu Val Arg

450 455 460

Ala Leu Asp Gly Val Ser Lys Glu Thr Ala Arg Thr Gln Leu Cys Ala

465 470 475 480

Asp Phe Gly Leu Asp Pro Lys Arg Leu Pro Trp Asp Lys Met Ser Ser

485 490 495

Asn Thr Thr Phe Ile Ser Glu Ala Leu Leu Ser Asn Ser Val Ser Arg

500 505 510

Asp Gln Val Phe Phe Thr Pro Ala Pro Lys Lys Gly Ala Lys Lys Lys

515 520 525

Ala Pro Val Glu Val Met Arg Lys Asp Arg Thr Trp Ala Arg Ala Tyr

530 535 540

Lys Pro Arg Leu Ser Val Glu Ala Gln Lys Leu Lys Asn Glu Ala Leu

545 550 555 560

Trp Ala Leu Lys Arg Thr Ser Pro Glu Tyr Leu Lys Leu Ser Arg Arg

565 570 575

Lys Glu Glu Leu Cys Arg Arg Ser Ile Asn Tyr Val Ile Glu Lys Thr

580 585 590

Arg Arg Arg Thr Gln Cys Gln Ile Val Ile Pro Val Ile Glu Asp Leu

595 600 605

Asn Val Arg Phe Phe His Gly Ser Gly Lys Arg Leu Pro Gly Trp Asp
610 615 620
Asn Phe Phe Thr Ala Lys Lys Glu Asn Arg Trp Phe Ile Gln Gly Leu
625 630 635 640
His Lys Ala Phe Ser Asp Leu Arg Thr His Arg Ser Phe Tyr Val Phe
645 650 655
Glu Val Arg Pro Glu Arg Thr Ser Ile Thr Cys Pro Lys Cys Gly His

660 665 670
Cys Glu Val Gly Asn Arg Asp Gly Glu Ala Phe Gln Cys Leu Ser Cys
675 680 685
Gly Lys Thr Cys Asn Ala Asp Leu Asp Val Ala Thr His Asn Leu Thr
690 695 700
Gln Val Ala Leu Thr Gly Lys Thr Met Pro Lys Arg Glu Glu Pro Arg
705 710 715 720
Asp Ala Gln Gly Thr Ala Pro Ala Arg Lys Thr Lys Lys Ala Ser Lys

725 730 735
Ser Lys Ala Pro Pro Ala Glu Arg Glu Asp Gln Thr Pro Ala Gln Glu
740 745 750
Pro Ser Gln Thr Ser
755

<210> 287
<211> 766
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
Cas-phi-3 sequence
<400> 287
Met Glu Lys Glu Ile Thr Glu Leu Thr Lys Ile Arg Arg Glu Phe Pro
1 5 10 15
Asn Lys Lys Phe Ser Ser Thr Asp Met Lys Lys Ala Gly Lys Leu Leu

20 25 30
Lys Ala Glu Gly Pro Asp Ala Val Arg Asp Phe Leu Asn Ser Cys Gln

35 40 45

Glu Ile Ile Gly Asp Phe Lys Pro Pro Val Lys Thr Asn Ile Val Ser

50 55 60

Ile Ser Arg Pro Phe Glu Glu Trp Pro Val Ser Met Val Gly Arg Ala

65 70 75 80

Ile Gln Glu Tyr Tyr Phe Ser Leu Thr Lys Glu Glu Leu Glu Ser Val

85 90 95

His Pro Gly Thr Ser Ser Glu Asp His Lys Ser Phe Phe Asn Ile Thr

100 105 110

Gly Leu Ser Asn Tyr Asn Tyr Thr Ser Val Gln Gly Leu Asn Leu Ile

115 120 125

Phe Lys Asn Ala Lys Ala Ile Tyr Asp Gly Thr Leu Val Lys Ala Asn

130 135 140

Asn Lys Asn Lys Lys Leu Glu Lys Lys Phe Asn Glu Ile Asn His Lys

145 150 155 160

Arg Ser Leu Glu Gly Leu Pro Ile Ile Thr Pro Asp Phe Glu Glu Pro

165 170 175

Phe Asp Glu Asn Gly His Leu Asn Asn Pro Pro Gly Ile Asn Arg Asn

180 185 190

Ile Tyr Gly Tyr Gln Gly Cys Ala Ala Lys Val Phe Val Pro Ser Lys

195 200 205

His Lys Met Val Ser Leu Pro Lys Glu Tyr Glu Gly Tyr Asn Arg Asp

210 215 220

Pro Asn Leu Ser Leu Ala Gly Phe Arg Asn Arg Leu Glu Ile Pro Glu

225 230 235 240

Gly Glu Pro Gly His Val Pro Trp Phe Gln Arg Met Asp Ile Pro Glu

245 250 255

Gly Gln Ile Gly His Val Asn Lys Ile Gln Arg Phe Asn Phe Val His

260 265 270

Gly Lys Asn Ser Gly Lys Val Lys Phe Ser Asp Lys Thr Gly Arg Val

275 280 285

Lys Arg Tyr His His Ser Lys Tyr Lys Asp Ala Thr Lys Pro Tyr Lys
290 295 300
Phe Leu Glu Glu Ser Lys Lys Val Ser Ala Leu Asp Ser Ile Leu Ala
305 310 315 320
Ile Ile Thr Ile Gly Asp Asp Trp Val Val Phe Asp Ile Arg Gly Leu
325 330 335
Tyr Arg Asn Val Phe Tyr Arg Glu Leu Ala Gln Lys Gly Leu Thr Ala

340 345 350
Val Gln Leu Leu Asp Leu Phe Thr Gly Asp Pro Val Ile Asp Pro Lys
355 360 365
Lys Gly Val Val Thr Phe Ser Tyr Lys Glu Gly Val Val Pro Val Phe
370 375 380
Ser Gln Lys Ile Val Pro Arg Phe Lys Ser Arg Asp Thr Leu Glu Lys
385 390 395 400
Leu Thr Ser Gln Gly Pro Val Ala Leu Leu Ser Val Asp Leu Gly Gln

405 410 415
Asn Glu Pro Val Ala Ala Arg Val Cys Ser Leu Lys Asn Ile Asn Asp
420 425 430
Lys Ile Thr Leu Asp Asn Ser Cys Arg Ile Ser Phe Leu Asp Asp Tyr
435 440 445
Lys Lys Gln Ile Lys Asp Tyr Arg Asp Ser Leu Asp Glu Leu Glu Ile
450 455 460
Lys Ile Arg Leu Glu Ala Ile Asn Ser Leu Glu Thr Asn Gln Gln Val

465 470 475 480
Glu Ile Arg Asp Leu Asp Val Phe Ser Ala Asp Arg Ala Lys Ala Asn
485 490 495
Thr Val Asp Met Phe Asp Ile Asp Pro Asn Leu Ile Ser Trp Asp Ser
500 505 510
Met Ser Asp Ala Arg Val Ser Thr Gln Ile Ser Asp Leu Tyr Leu Lys
515 520 525
Asn Gly Gly Asp Glu Ser Arg Val Tyr Phe Glu Ile Asn Asn Lys Arg

530 535 540

Ile Lys Arg Ser Asp Tyr Asn Ile Ser Gln Leu Val Arg Pro Lys Leu

545 550 555 560

Ser Asp Ser Thr Arg Lys Asn Leu Asn Asp Ser Ile Trp Lys Leu Lys

565 570 575

Arg Thr Ser Glu Glu Tyr Leu Lys Leu Ser Lys Arg Lys Leu Glu Leu

580 585 590

Ser Arg Ala Val Val Asn Tyr Thr Ile Arg Gln Ser Lys Leu Leu Ser

595 600 605

Gly Ile Asn Asp Ile Val Ile Ile Leu Glu Asp Leu Asp Val Lys Lys

610 615 620

Lys Phe Asn Gly Arg Gly Ile Arg Asp Ile Gly Trp Asp Asn Phe Phe

625 630 635 640

Ser Ser Arg Lys Glu Asn Arg Trp Phe Ile Pro Ala Phe His Lys Ala

645 650 655

Phe Ser Glu Leu Ser Ser Asn Arg Gly Leu Cys Val Ile Glu Val Asn

660 665 670

Pro Ala Trp Thr Ser Ala Thr Cys Pro Asp Cys Gly Phe Cys Ser Lys

675 680 685

Glu Asn Arg Asp Gly Ile Asn Phe Thr Cys Arg Lys Cys Gly Val Ser

690 695 700

Tyr His Ala Asp Ile Asp Val Ala Thr Leu Asn Ile Ala Arg Val Ala

705 710 715 720

Val Leu Gly Lys Pro Met Ser Gly Pro Ala Asp Arg Glu Arg Leu Gly

725 730 735

Asp Thr Lys Lys Pro Arg Val Ala Arg Ser Arg Lys Thr Met Lys Arg

740 745 750

Lys Asp Ile Ser Asn Ser Thr Val Glu Ala Met Val Thr Ala

755 760 765

<210> 288

<211> 765

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas-phi-4 sequence

<400> 288

Met Tyr Ser Leu Glu Met Ala Asp Leu Lys Ser Glu Pro Ser Leu Leu
1 5 10 15

Ala Lys Leu Leu Arg Asp Arg Phe Pro Gly Lys Tyr Trp Leu Pro Lys
20 25 30

Tyr Trp Lys Leu Ala Glu Lys Lys Arg Leu Thr Gly Gly Glu Glu Ala
35 40 45

Ala Cys Glu Tyr Met Ala Asp Lys Gln Leu Asp Ser Pro Pro Pro Asn
50 55 60

Phe Arg Pro Pro Ala Arg Cys Val Ile Leu Ala Lys Ser Arg Pro Phe
65 70 75 80

Glu Asp Trp Pro Val His Arg Val Ala Ser Lys Ala Gln Ser Phe Val
85 90 95

Ile Gly Leu Ser Glu Gln Gly Phe Ala Ala Leu Arg Ala Ala Pro Pro
100 105 110

Ser Thr Ala Asp Ala Arg Arg Asp Trp Leu Arg Ser His Gly Ala Ser
115 120 125

Glu Asp Asp Leu Met Ala Leu Glu Ala Gln Leu Leu Glu Thr Ile Met
130 135 140

Gly Asn Ala Ile Ser Leu His Gly Gly Val Leu Lys Lys Ile Asp Asn
145 150 155 160

Ala Asn Val Lys Ala Ala Lys Arg Leu Ser Gly Arg Asn Glu Ala Arg
165 170 175

Leu Asn Lys Gly Leu Gln Glu Leu Pro Pro Glu Gln Glu Gly Ser Ala
180 185 190

Tyr Gly Ala Asp Gly Leu Leu Val Asn Pro Pro Gly Leu Asn Leu Asn
195 200 205

Ile Tyr Cys Arg Lys Ser Cys Cys Pro Lys Pro Val Lys Asn Thr Ala

210 215 220

Arg Phe Val Gly His Tyr Pro Gly Tyr Leu Arg Asp Ser Asp Ser Ile

225 230 235 240

Leu Ile Ser Gly Thr Met Asp Arg Leu Thr Ile Ile Glu Gly Met Pro

245 250 255

Gly His Ile Pro Ala Trp Gln Arg Glu Gln Gly Leu Val Lys Pro Gly

260 265 270

Gly Arg Arg Arg Arg Leu Ser Gly Ser Glu Ser Asn Met Arg Gln Lys

275 280 285

Val Asp Pro Ser Thr Gly Pro Arg Arg Ser Thr Arg Ser Gly Thr Val

290 295 300

Asn Arg Ser Asn Gln Arg Thr Gly Arg Asn Gly Asp Pro Leu Leu Val

305 310 315 320

Glu Ile Arg Met Lys Glu Asp Trp Val Leu Leu Asp Ala Arg Gly Leu

325 330 335

Leu Arg Asn Leu Arg Trp Arg Glu Ser Lys Arg Gly Leu Ser Cys Asp

340 345 350

His Glu Asp Leu Ser Leu Ser Gly Leu Leu Ala Leu Phe Ser Gly Asp

355 360 365

Pro Val Ile Asp Pro Val Arg Asn Glu Val Val Phe Leu Tyr Gly Glu

370 375 380

Gly Ile Ile Pro Val Arg Ser Thr Lys Pro Val Gly Thr Arg Gln Ser

385 390 395 400

Lys Lys Leu Leu Glu Arg Gln Ala Ser Met Gly Pro Leu Thr Leu Ile

405 410 415

Ser Cys Asp Leu Gly Gln Thr Asn Leu Ile Ala Gly Arg Ala Ser Ala

420 425 430

Ile Ser Leu Thr His Gly Ser Leu Gly Val Arg Ser Ser Val Arg Ile

435 440 445

Glu Leu Asp Pro Glu Ile Ile Lys Ser Phe Glu Arg Leu Arg Lys Asp

450 455 460

Ala Asp Arg Leu Glu Thr Glu Ile Leu Thr Ala Ala Lys Glu Thr Leu
465 470 475 480
Ser Asp Glu Gln Arg Gly Glu Val Asn Ser His Glu Lys Asp Ser Pro
485 490 495
Gln Thr Ala Lys Ala Ser Leu Cys Arg Glu Leu Gly Leu His Pro Pro
500 505 510
Ser Leu Pro Trp Gly Gln Met Gly Pro Ser Thr Thr Phe Ile Ala Asp
515 520 525

Met Leu Ile Ser His Gly Arg Asp Asp Asp Ala Phe Leu Ser His Gly
530 535 540
Glu Phe Pro Thr Leu Glu Lys Arg Lys Lys Phe Asp Lys Arg Phe Cys
545 550 555 560
Leu Glu Ser Arg Pro Leu Leu Ser Ser Glu Thr Arg Lys Ala Leu Asn
565 570 575
Glu Ser Leu Trp Glu Val Lys Arg Thr Ser Ser Glu Tyr Ala Arg Leu
580 585 590

Ser Gln Arg Lys Lys Glu Met Ala Arg Arg Ala Val Asn Phe Val Val
595 600 605
Glu Ile Ser Arg Arg Lys Thr Gly Leu Ser Asn Val Ile Val Asn Ile
610 615 620
Glu Asp Leu Asn Val Arg Ile Phe His Gly Gly Gly Lys Gln Ala Pro
625 630 635 640
Gly Trp Asp Gly Phe Phe Arg Pro Lys Ser Glu Asn Arg Trp Phe Ile
645 650 655

Gln Ala Ile His Lys Ala Phe Ser Asp Leu Ala Ala His His Gly Ile
660 665 670
Pro Val Ile Glu Ser Asp Pro Gln Arg Thr Ser Met Thr Cys Pro Glu
675 680 685
Cys Gly His Cys Asp Ser Lys Asn Arg Asn Gly Val Arg Phe Leu Cys
690 695 700
Lys Gly Cys Gly Ala Ser Met Asp Ala Asp Phe Asp Ala Ala Cys Arg

705 710 715 720

Asn Leu Glu Arg Val Ala Leu Thr Gly Lys Pro Met Pro Lys Pro Ser
725 730 735
Thr Ser Cys Glu Arg Leu Leu Ser Ala Thr Thr Gly Lys Val Cys Ser
740 745 750
Asp His Ser Leu Ser His Asp Ala Ile Glu Lys Ala Ser
755 760 765

<210> 289
<211> 793
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
Cas-phi-5 sequence
<400> 289

Met Ser Ser Leu Pro Thr Pro Leu Glu Leu Leu Lys Gln Lys His Ala

1 5 10 15
Asp Leu Phe Lys Gly Leu Gln Phe Ser Ser Lys Asp Asn Lys Met Ala
20 25 30
Gly Lys Val Leu Lys Lys Asp Gly Glu Glu Ala Ala Leu Ala Phe Leu
35 40 45
Ser Glu Arg Gly Val Ser Arg Gly Glu Leu Pro Asn Phe Arg Pro Pro
50 55 60
Ala Lys Thr Leu Val Val Ala Gln Ser Arg Pro Phe Glu Glu Phe Pro

65 70 75 80
Ile Tyr Arg Val Ser Glu Ala Ile Gln Leu Tyr Val Tyr Ser Leu Ser
85 90 95
Val Lys Glu Leu Glu Thr Val Pro Ser Gly Ser Ser Thr Lys Lys Glu
100 105 110
His Gln Arg Phe Phe Gln Asp Ser Ser Val Pro Asp Phe Gly Tyr Thr
115 120 125
Ser Val Gln Gly Leu Asn Lys Ile Phe Gly Leu Ala Arg Gly Ile Tyr

130 135 140

Leu Gly Val Ile Thr Arg Gly Glu Asn Gln Leu Gln Lys Ala Lys Ser

145 150 155 160

Lys His Glu Ala Leu Asn Lys Lys Arg Arg Ala Ser Gly Glu Ala Glu

165 170 175

Thr Glu Phe Asp Pro Thr Pro Tyr Glu Tyr Met Thr Pro Glu Arg Lys

180 185 190

Leu Ala Lys Pro Pro Gly Val Asn His Ser Ile Met Cys Tyr Val Asp

195 200 205

Ile Ser Val Asp Glu Phe Asp Phe Arg Asn Pro Asp Gly Ile Val Leu

210 215 220

Pro Ser Glu Tyr Ala Gly Tyr Cys Arg Glu Ile Asn Thr Ala Ile Glu

225 230 235 240

Lys Gly Thr Val Asp Arg Leu Gly His Leu Lys Gly Gly Pro Gly Tyr

245 250 255

Ile Pro Gly His Gln Arg Lys Glu Ser Thr Thr Glu Gly Pro Lys Ile

260 265 270

Asn Phe Arg Lys Gly Arg Ile Arg Arg Ser Tyr Thr Ala Leu Tyr Ala

275 280 285

Lys Arg Asp Ser Arg Arg Val Arg Gln Gly Lys Leu Ala Leu Pro Ser

290 295 300

Tyr Arg His His Met Met Arg Leu Asn Ser Asn Ala Glu Ser Ala Ile

305 310 315 320

Leu Ala Val Ile Phe Phe Gly Lys Asp Trp Val Val Phe Asp Leu Arg

325 330 335

Gly Leu Leu Arg Asn Val Arg Trp Arg Asn Leu Phe Val Asp Gly Ser

340 345 350

Thr Pro Ser Thr Leu Leu Gly Met Phe Gly Asp Pro Val Ile Asp Pro

355 360 365

Lys Arg Gly Val Val Ala Phe Cys Tyr Lys Glu Gln Ile Val Pro Val

370 375 380

Val Ser Lys Ser Ile Thr Lys Met Val Lys Ala Pro Glu Leu Leu Asn

385 390 395 400

Lys Leu Tyr Leu Lys Ser Glu Asp Pro Leu Val Leu Val Ala Ile Asp

405 410 415

Leu Gly Gln Thr Asn Pro Val Gly Val Gly Val Tyr Arg Val Met Asn

420 425 430

Ala Ser Leu Asp Tyr Glu Val Val Thr Arg Phe Ala Leu Glu Ser Glu

435 440 445

Leu Leu Arg Glu Ile Glu Ser Tyr Arg Gln Arg Thr Asn Ala Phe Glu

450 455 460

Ala Gln Ile Arg Ala Glu Thr Phe Asp Ala Met Thr Ser Glu Glu Gln

465 470 475 480

Glu Glu Ile Thr Arg Val Arg Ala Phe Ser Ala Ser Lys Ala Lys Glu

485 490 495

Asn Val Cys His Arg Phe Gly Met Pro Val Asp Ala Val Asp Trp Ala

500 505 510

Thr Met Gly Ser Asn Thr Ile His Ile Ala Lys Trp Val Met Arg His

515 520 525

Gly Asp Pro Ser Leu Val Glu Val Leu Glu Tyr Arg Lys Asp Asn Glu

530 535 540

Ile Lys Leu Asp Lys Asn Gly Val Pro Lys Lys Val Lys Leu Thr Asp

545 550 555 560

Lys Arg Ile Ala Asn Leu Thr Ser Ile Arg Leu Arg Phe Ser Gln Glu

565 570 575

Thr Ser Lys His Tyr Asn Asp Thr Met Trp Glu Leu Arg Arg Lys His

580 585 590

Pro Val Tyr Gln Lys Leu Ser Lys Ser Lys Ala Asp Phe Ser Arg Arg

595 600 605

Val Val Asn Ser Ile Ile Arg Arg Val Asn His Leu Val Pro Arg Ala

610 615 620

Arg Ile Val Phe Ile Ile Glu Asp Leu Lys Asn Leu Gly Lys Val Phe

625 630 635 640

His Gly Ser Gly Lys Arg Glu Leu Gly Trp Asp Ser Tyr Phe Glu Pro

645 650 655

Lys Ser Glu Asn Arg Trp Phe Ile Gln Val Leu His Lys Ala Phe Ser

660 665 670

Glu Thr Gly Lys His Lys Gly Tyr Tyr Ile Ile Glu Cys Trp Pro Asn

675 680 685

Trp Thr Ser Cys Thr Cys Pro Lys Cys Ser Cys Cys Asp Ser Glu Asn

690 695 700

Arg His Gly Glu Val Phe Arg Cys Leu Ala Cys Gly Tyr Thr Cys Asn

705 710 715 720

Thr Asp Phe Gly Thr Ala Pro Asp Asn Leu Val Lys Ile Ala Thr Thr

725 730 735

Gly Lys Gly Leu Pro Gly Pro Lys Lys Arg Cys Lys Gly Ser Ser Lys

740 745 750

Gly Lys Asn Pro Lys Ile Ala Arg Ser Ser Glu Thr Gly Val Ser Val

755 760 765

Thr Glu Ser Gly Ala Pro Lys Val Lys Lys Ser Ser Pro Thr Gln Thr

770 775 780

Ser Gln Ser Ser Ser Gln Ser Ala Pro

785 790

<210> 290

<211> 441

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas-phi-6 sequence

<400> 290

Met Asn Lys Ile Glu Lys Glu Lys Thr Pro Leu Ala Lys Leu Met Asn

1 5 10 15

Glu Asn Phe Ala Gly Leu Arg Phe Pro Phe Ala Ile Ile Lys Gln Ala

20 25 30

Gly Lys Lys Leu Leu Lys Glu Gly Glu Leu Lys Thr Ile Glu Tyr Met
35 40 45
Thr Gly Lys Gly Ser Ile Glu Pro Leu Pro Asn Phe Lys Pro Pro Val
50 55 60
Lys Cys Leu Ile Val Ala Lys Arg Arg Asp Leu Lys Tyr Phe Pro Ile
65 70 75 80
Cys Lys Ala Ser Cys Glu Ile Gln Ser Tyr Val Tyr Ser Leu Asn Tyr
85 90 95

Lys Asp Phe Met Asp Tyr Phe Ser Thr Pro Met Thr Ser Gln Lys Gln
100 105 110
His Glu Glu Phe Phe Lys Lys Ser Gly Leu Asn Ile Glu Tyr Gln Asn
115 120 125
Val Ala Gly Leu Asn Leu Ile Phe Asn Asn Val Lys Asn Thr Tyr Asn
130 135 140
Gly Val Ile Leu Lys Val Lys Asn Arg Asn Glu Lys Leu Lys Lys Lys
145 150 155 160

Ala Ile Lys Asn Asn Tyr Glu Phe Glu Glu Ile Lys Thr Phe Asn Asp
165 170 175
Asp Gly Cys Leu Ile Asn Lys Pro Gly Ile Asn Asn Val Ile Tyr Cys
180 185 190
Phe Gln Ser Ile Ser Pro Lys Ile Leu Lys Asn Ile Thr His Leu Pro
195 200 205
Lys Glu Tyr Asn Asp Tyr Asp Cys Ser Val Asp Arg Asn Ile Ile Gln
210 215 220

Lys Tyr Val Ser Arg Leu Asp Ile Pro Glu Ser Gln Pro Gly His Val
225 230 235 240
Pro Glu Trp Gln Arg Lys Leu Pro Glu Phe Asn Asn Thr Asn Asn Pro
245 250 255
Arg Arg Arg Arg Lys Trp Tyr Ser Asn Gly Arg Asn Ile Ser Lys Gly
260 265 270

Tyr Ser Val Asp Gln Val Asn Gln Ala Lys Ile Glu Asp Ser Leu Leu
275 280 285

Ala Gln Ile Lys Ile Gly Glu Asp Trp Ile Ile Leu Asp Ile Arg Gly
290 295 300

Leu Leu Arg Asp Leu Asn Arg Arg Glu Leu Ile Ser Tyr Lys Asn Lys
305 310 315 320

Leu Thr Ile Lys Asp Val Leu Gly Phe Phe Ser Asp Tyr Pro Ile Ile
325 330 335

Asp Ile Lys Lys Asn Leu Val Thr Phe Cys Tyr Lys Glu Gly Val Ile
340 345 350

Gln Val Val Ser Gln Lys Ser Ile Gly Asn Lys Lys Ser Lys Gln Leu
355 360 365

Leu Glu Lys Leu Ile Glu Asn Lys Pro Ile Ala Leu Val Ser Ile Asp
370 375 380

Leu Gly Gln Thr Asn Pro Val Ser Val Lys Ile Ser Lys Leu Asn Lys
385 390 395 400

Ile Asn Asn Lys Ile Ser Ile Glu Ser Phe Thr Tyr Arg Phe Leu Asn
405 410 415

Glu Glu Ile Leu Lys Glu Ile Glu Lys Tyr Arg Lys Asp Tyr Asp Lys
420 425 430

Leu Glu Leu Lys Leu Ile Asn Glu Ala
435 440

<210> 291

<211> 772

<212> PRT

<213> Unknown

<220><223> Description of Unknown:
Cas-phi-7 sequence

<400> 291

Met Ser Asn Thr Ala Val Ser Thr Arg Glu His Met Ser Asn Lys Thr
1 5 10 15

Thr Pro Pro Ser Pro Leu Ser Leu Leu Leu Arg Ala His Phe Pro Gly

20 25 30

Leu Lys Phe Glu Ser Gln Asp Tyr Lys Ile Ala Gly Lys Lys Leu Arg

35 40 45

Asp Gly Gly Pro Glu Ala Val Ile Ser Tyr Leu Thr Gly Lys Gly Gln

50 55 60

Ala Lys Leu Lys Asp Val Lys Pro Pro Ala Lys Ala Phe Val Ile Ala

65 70 75 80

Gln Ser Arg Pro Phe Ile Glu Trp Asp Leu Val Arg Val Ser Arg Gln

85 90 95

Ile Gln Glu Lys Ile Phe Gly Ile Pro Ala Thr Lys Gly Arg Pro Lys

100 105 110

Gln Asp Gly Leu Ser Glu Thr Ala Phe Asn Glu Ala Val Ala Ser Leu

115 120 125

Glu Val Asp Gly Lys Ser Lys Leu Asn Glu Glu Thr Arg Ala Ala Phe

130 135 140

Tyr Glu Val Leu Gly Leu Asp Ala Pro Ser Leu His Ala Gln Ala Gln

145 150 155 160

Asn Ala Leu Ile Lys Ser Ala Ile Ser Ile Arg Glu Gly Val Leu Lys

165 170 175

Lys Val Glu Asn Arg Asn Glu Lys Asn Leu Ser Lys Thr Lys Arg Arg

180 185 190

Lys Glu Ala Gly Glu Glu Ala Thr Phe Val Glu Glu Lys Ala His Asp

195 200 205

Glu Arg Gly Tyr Leu Ile His Pro Pro Gly Val Asn Gln Thr Ile Pro

210 215 220

Gly Tyr Gln Ala Val Val Ile Lys Ser Cys Pro Ser Asp Phe Ile Gly

225 230 235 240

Leu Pro Ser Gly Cys Leu Ala Lys Glu Ser Ala Glu Ala Leu Thr Asp

245 250 255

Tyr Leu Pro His Asp Arg Met Thr Ile Pro Lys Gly Gln Pro Gly Tyr

260 265 270

Val Pro Glu Trp Gln His Pro Leu Leu Asn Arg Arg Lys Asn Arg Arg

275 280 285

Arg Arg Asp Trp Tyr Ser Ala Ser Leu Asn Lys Pro Lys Ala Thr Cys

290 295 300

Ser Lys Arg Ser Gly Thr Pro Asn Arg Lys Asn Ser Arg Thr Asp Gln

305 310 315 320

Ile Gln Ser Gly Arg Phe Lys Gly Ala Ile Pro Val Leu Met Arg Phe

325 330 335

Gln Asp Glu Trp Val Ile Ile Asp Ile Arg Gly Leu Leu Arg Asn Ala

340 345 350

Arg Tyr Arg Lys Leu Leu Lys Glu Lys Ser Thr Ile Pro Asp Leu Leu

355 360 365

Ser Leu Phe Thr Gly Asp Pro Ser Ile Asp Met Arg Gln Gly Val Cys

370 375 380

Thr Phe Ile Tyr Lys Ala Gly Gln Ala Cys Ser Ala Lys Met Val Lys

385 390 395 400

Thr Lys Asn Ala Pro Glu Ile Leu Ser Glu Leu Thr Lys Ser Gly Pro

405 410 415

Val Val Leu Val Ser Ile Asp Leu Gly Gln Thr Asn Pro Ile Ala Ala

420 425 430

Lys Val Ser Arg Val Thr Gln Leu Ser Asp Gly Gln Leu Ser His Glu

435 440 445

Thr Leu Leu Arg Glu Leu Leu Ser Asn Asp Ser Ser Asp Gly Lys Glu

450 455 460

Ile Ala Arg Tyr Arg Val Ala Ser Asp Arg Leu Arg Asp Lys Leu Ala

465 470 475 480

Asn Leu Ala Val Glu Arg Leu Ser Pro Glu His Lys Ser Glu Ile Leu

485 490 495

Arg Ala Lys Asn Asp Thr Pro Ala Leu Cys Lys Ala Arg Val Cys Ala

500 505 510

Ala Leu Gly Leu Asn Pro Glu Met Ile Ala Trp Asp Lys Met Thr Pro
515 520 525
Tyr Thr Glu Phe Leu Ala Thr Ala Tyr Leu Glu Lys Gly Gly Asp Arg

530 535 540
Lys Val Ala Thr Leu Lys Pro Lys Asn Arg Pro Glu Met Leu Arg Arg
545 550 555 560
Asp Ile Lys Phe Lys Gly Thr Glu Gly Val Arg Ile Glu Val Ser Pro
565 570 575
Glu Ala Ala Glu Ala Tyr Arg Glu Ala Gln Trp Asp Leu Gln Arg Thr
580 585 590
Ser Pro Glu Tyr Leu Arg Leu Ser Thr Trp Lys Gln Glu Leu Thr Lys

595 600 605
Arg Ile Leu Asn Gln Leu Arg His Lys Ala Ala Lys Ser Ser Gln Cys
610 615 620
Glu Val Val Val Met Ala Phe Glu Asp Leu Asn Ile Lys Met Met His
625 630 635 640
Gly Asn Gly Lys Trp Ala Asp Gly Gly Trp Asp Ala Phe Phe Ile Lys
645 650 655
Lys Arg Glu Asn Arg Trp Phe Met Gln Ala Phe His Lys Ser Leu Thr

660 665 670
Glu Leu Gly Ala His Lys Gly Val Pro Thr Ile Glu Val Thr Pro His
675 680 685
Arg Thr Ser Ile Thr Cys Thr Lys Cys Gly His Cys Asp Lys Ala Asn
690 695 700
Arg Asp Gly Glu Arg Phe Ala Cys Gln Lys Cys Gly Phe Val Ala His
705 710 715 720
Ala Asp Leu Glu Ile Ala Thr Asp Asn Ile Glu Arg Val Ala Leu Thr

725 730 735
Gly Lys Pro Met Pro Lys Pro Glu Ser Glu Arg Ser Gly Asp Ala Lys
740 745 750
Lys Ser Val Gly Ala Arg Lys Ala Ala Phe Lys Pro Glu Glu Asp Ala

755 760 765

Glu Ala Ala Glu

770

<210> 292

<211> 717

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas-phi-8 sequence

<400> 292

Met Ile Lys Pro Thr Val Ser Gln Phe Leu Thr Pro Gly Phe Lys Leu

1 5 10 15

Ile Arg Asn His Ser Arg Thr Ala Gly Leu Lys Leu Lys Asn Glu Gly

20 25 30

Glu Glu Ala Cys Lys Lys Phe Val Arg Glu Asn Glu Ile Pro Lys Asp

35 40 45

Glu Cys Pro Asn Phe Gln Gly Gly Pro Ala Ile Ala Asn Ile Ile Ala

50 55 60

Lys Ser Arg Glu Phe Thr Glu Trp Glu Ile Tyr Gln Ser Ser Leu Ala

65 70 75 80

Ile Gln Glu Val Ile Phe Thr Leu Pro Lys Asp Lys Leu Pro Glu Pro

85 90 95

Ile Leu Lys Glu Glu Trp Arg Ala Gln Trp Leu Ser Glu His Gly Leu

100 105 110

Asp Thr Val Pro Tyr Lys Glu Ala Ala Gly Leu Asn Leu Ile Ile Lys

115 120 125

Asn Ala Val Asn Thr Tyr Lys Gly Val Gln Val Lys Val Asp Asn Lys

130 135 140

Asn Lys Asn Asn Leu Ala Lys Ile Asn Arg Lys Asn Glu Ile Ala Lys

145 150 155 160

Leu Asn Gly Glu Gln Glu Ile Ser Phe Glu Glu Ile Lys Ala Phe Asp

165 170 175

Asp Lys Gly Tyr Leu Leu Gln Lys Pro Ser Pro Asn Lys Ser Ile Tyr
180 185 190
Cys Tyr Gln Ser Val Ser Pro Lys Pro Phe Ile Thr Ser Lys Tyr His

195 200 205
Asn Val Asn Leu Pro Glu Glu Tyr Ile Gly Tyr Tyr Arg Lys Ser Asn
210 215 220
Glu Pro Ile Val Ser Pro Tyr Gln Phe Asp Arg Leu Arg Ile Pro Ile
225 230 235 240
Gly Glu Pro Gly Tyr Val Pro Lys Trp Gln Tyr Thr Phe Leu Ser Lys
245 250 255
Lys Glu Asn Lys Arg Arg Lys Leu Ser Lys Arg Ile Lys Asn Val Ser

260 265 270
Pro Ile Leu Gly Ile Ile Cys Ile Lys Lys Asp Trp Cys Val Phe Asp
275 280 285
Met Arg Gly Leu Leu Arg Thr Asn His Trp Lys Lys Tyr His Lys Pro
290 295 300
Thr Asp Ser Ile Asn Asp Leu Phe Asp Tyr Phe Thr Gly Asp Pro Val
305 310 315 320
Ile Asp Thr Lys Ala Asn Val Val Arg Phe Arg Tyr Lys Met Glu Asn

325 330 335
Gly Ile Val Asn Tyr Lys Pro Val Arg Glu Lys Lys Gly Lys Glu Leu
340 345 350
Leu Glu Asn Ile Cys Asp Gln Asn Gly Ser Cys Lys Leu Ala Thr Val
355 360 365
Asp Val Gly Gln Asn Asn Pro Val Ala Ile Gly Leu Phe Glu Leu Lys
370 375 380
Lys Val Asn Gly Glu Leu Thr Lys Thr Leu Ile Ser Arg His Pro Thr

385 390 395 400
Pro Ile Asp Phe Cys Asn Lys Ile Thr Ala Tyr Arg Glu Arg Tyr Asp
405 410 415
Lys Leu Glu Ser Ser Ile Lys Leu Asp Ala Ile Lys Gln Leu Thr Ser

420 425 430

Glu Gln Lys Ile Glu Val Asp Asn Tyr Asn Asn Asn Phe Thr Pro Gln

435 440 445

Asn Thr Lys Gln Ile Val Cys Ser Lys Leu Asn Ile Asn Pro Asn Asp

450 455 460

Leu Pro Trp Asp Lys Met Ile Ser Gly Thr His Phe Ile Ser Glu Lys

465 470 475 480

Ala Gln Val Ser Asn Lys Ser Glu Ile Tyr Phe Thr Ser Thr Asp Lys

485 490 495

Gly Lys Thr Lys Asp Val Met Lys Ser Asp Tyr Lys Trp Phe Gln Asp

500 505 510

Tyr Lys Pro Lys Leu Ser Lys Glu Val Arg Asp Ala Leu Ser Asp Ile

515 520 525

Glu Trp Arg Leu Arg Arg Glu Ser Leu Glu Phe Asn Lys Leu Ser Lys

530 535 540

Ser Arg Glu Gln Asp Ala Arg Gln Leu Ala Asn Trp Ile Ser Ser Met

545 550 555 560

Cys Asp Val Ile Gly Ile Glu Asn Leu Val Lys Lys Asn Asn Phe Phe

565 570 575

Gly Gly Ser Gly Lys Arg Glu Pro Gly Trp Asp Asn Phe Tyr Lys Pro

580 585 590

Lys Lys Glu Asn Arg Trp Trp Ile Asn Ala Ile His Lys Ala Leu Thr

595 600 605

Glu Leu Ser Gln Asn Lys Gly Lys Arg Val Ile Leu Leu Pro Ala Met

610 615 620

Arg Thr Ser Ile Thr Cys Pro Lys Cys Lys Tyr Cys Asp Ser Lys Asn

625 630 635 640

Arg Asn Gly Glu Lys Phe Asn Cys Leu Lys Cys Gly Ile Glu Leu Asn

645 650 655

Ala Asp Ile Asp Val Ala Thr Glu Asn Leu Ala Thr Val Ala Ile Thr

660 665 670

Ala Gln Ser Met Pro Lys Pro Thr Cys Glu Arg Ser Gly Asp Ala Lys
675 680 685
Lys Pro Val Arg Ala Arg Lys Ala Lys Ala Pro Glu Phe His Asp Lys
690 695 700
Leu Ala Pro Ser Tyr Thr Val Val Leu Arg Glu Ala Val

705 710 715
<210> 293
<211> 793
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
Cas-phi-9 sequence
<400> 293
Met Arg Ser Ser Arg Glu Ile Gly Asp Lys Ile Leu Met Arg Gln Pro
1 5 10 15
Ala Glu Lys Thr Ala Phe Gln Val Phe Arg Gln Glu Val Ile Gly Thr
20 25 30
Gln Lys Leu Ser Gly Gly Asp Ala Lys Thr Ala Gly Arg Leu Tyr Lys

35 40 45
Gln Gly Lys Met Glu Ala Ala Arg Glu Trp Leu Leu Lys Gly Ala Arg
50 55 60
Asp Asp Val Pro Pro Asn Phe Gln Pro Pro Ala Lys Cys Leu Val Val
65 70 75 80
Ala Val Ser His Pro Phe Glu Glu Trp Asp Ile Ser Lys Thr Asn His
85 90 95
Asp Val Gln Ala Tyr Ile Tyr Ala Gln Pro Leu Gln Ala Glu Gly His

100 105 110
Leu Asn Gly Leu Ser Glu Lys Trp Glu Asp Thr Ser Ala Asp Gln His
115 120 125
Lys Leu Trp Phe Glu Lys Thr Gly Val Pro Asp Arg Gly Leu Pro Val
130 135 140
Gln Ala Ile Asn Lys Ile Ala Lys Ala Ala Val Asn Arg Ala Phe Gly

145 150 155 160

Val Val Arg Lys Val Glu Asn Arg Asn Glu Lys Arg Arg Ser Arg Asp

165 170 175

Asn Arg Ile Ala Glu His Asn Arg Glu Asn Gly Leu Thr Glu Val Val

180 185 190

Arg Glu Ala Pro Glu Val Ala Thr Asn Ala Asp Gly Phe Leu Leu His

195 200 205

Pro Pro Gly Ile Asp Pro Ser Ile Leu Ser Tyr Ala Ser Val Ser Pro

210 215 220

Val Pro Tyr Asn Ser Ser Lys His Ser Phe Val Arg Leu Pro Glu Glu

225 230 235 240

Tyr Gln Ala Tyr Asn Val Glu Pro Asp Ala Pro Ile Pro Gln Phe Val

245 250 255

Val Glu Asp Arg Phe Ala Ile Pro Pro Gly Gln Pro Gly Tyr Val Pro

260 265 270

Glu Trp Gln Arg Leu Lys Cys Ser Thr Asn Lys His Arg Arg Met Arg

275 280 285

Gln Trp Ser Asn Gln Asp Tyr Lys Pro Lys Ala Gly Arg Arg Ala Lys

290 295 300

Pro Leu Glu Phe Gln Ala His Leu Thr Arg Glu Arg Ala Lys Gly Ala

305 310 315 320

Leu Leu Val Val Met Arg Ile Lys Glu Asp Trp Val Val Phe Asp Val

325 330 335

Arg Gly Leu Leu Arg Asn Val Glu Trp Arg Lys Val Leu Ser Glu Glu

340 345 350

Ala Arg Glu Lys Leu Thr Leu Lys Gly Leu Leu Asp Leu Phe Thr Gly

355 360 365

Asp Pro Val Ile Asp Thr Lys Arg Gly Ile Val Thr Phe Leu Tyr Lys

370 375 380

Ala Glu Ile Thr Lys Ile Leu Ser Lys Arg Thr Val Lys Thr Lys Asn

385 390 395 400

Ala Arg Asp Leu Leu Leu Arg Leu Thr Glu Pro Gly Glu Asp Gly Leu
405 410 415
Arg Arg Glu Val Gly Leu Val Ala Val Asp Leu Gly Gln Thr His Pro

420 425 430
Ile Ala Ala Ala Ile Tyr Arg Ile Gly Arg Thr Ser Ala Gly Ala Leu
435 440 445
Glu Ser Thr Val Leu His Arg Gln Gly Leu Arg Glu Asp Gln Lys Glu
450 455 460
Lys Leu Lys Glu Tyr Arg Lys Arg His Thr Ala Leu Asp Ser Arg Leu
465 470 475 480
Arg Lys Glu Ala Phe Glu Thr Leu Ser Val Glu Gln Gln Lys Glu Ile

485 490 495
Val Thr Val Ser Gly Ser Gly Ala Gln Ile Thr Lys Asp Lys Val Cys
500 505 510
Asn Tyr Leu Gly Val Asp Pro Ser Thr Leu Pro Trp Glu Lys Met Gly
515 520 525
Ser Tyr Thr His Phe Ile Ser Asp Asp Phe Leu Arg Arg Gly Gly Asp
530 535 540
Pro Asn Ile Val His Phe Asp Arg Gln Pro Lys Lys Gly Lys Val Ser

545 550 555 560
Lys Lys Ser Gln Arg Ile Lys Arg Ser Asp Ser Gln Trp Val Gly Arg
565 570 575
Met Arg Pro Arg Leu Ser Gln Glu Thr Ala Lys Ala Arg Met Glu Ala
580 585 590
Asp Trp Ala Ala Gln Asn Glu Asn Glu Glu Tyr Lys Arg Leu Ala Arg
595 600 605
Ser Lys Gln Glu Leu Ala Arg Trp Cys Val Asn Thr Leu Leu Gln Asn

610 615 620
Thr Arg Cys Ile Thr Gln Cys Asp Glu Ile Val Val Val Ile Glu Asp
625 630 635 640
Leu Asn Val Lys Ser Leu His Gly Lys Gly Ala Arg Glu Pro Gly Trp

645 650 655

Asp Asn Phe Phe Thr Pro Lys Thr Glu Asn Arg Trp Phe Ile Gln Ile

660 665 670

Leu His Lys Thr Phe Ser Glu Leu Pro Lys His Arg Gly Glu His Val

675 680 685

Ile Glu Gly Cys Pro Leu Arg Thr Ser Ile Thr Cys Pro Ala Cys Ser

690 695 700

Tyr Cys Asp Lys Asn Ser Arg Asn Gly Glu Lys Phe Val Cys Val Ala

705 710 715 720

Cys Gly Ala Thr Phe His Ala Asp Phe Glu Val Ala Thr Tyr Asn Leu

725 730 735

Val Arg Leu Ala Thr Thr Gly Met Pro Met Pro Lys Ser Leu Glu Arg

740 745 750

Gln Gly Gly Gly Glu Lys Ala Gly Gly Ala Arg Lys Ala Arg Lys Lys

755 760 765

Ala Lys Gln Val Glu Lys Ile Val Val Gln Ala Asn Ala Asn Val Thr

770 775 780

Met Asn Gly Ala Ser Leu His Ser Pro

785 790

<210> 294

<211> 812

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas-phi-10 sequence

<400> 294

Met Asp Met Leu Asp Thr Glu Thr Asn Tyr Ala Thr Glu Thr Pro Ala

1 5 10 15

Gln Gln Gln Asp Tyr Ser Pro Lys Pro Pro Lys Lys Ala Gln Arg Ala

20 25 30

Pro Lys Gly Phe Ser Lys Lys Ala Arg Pro Glu Lys Lys Pro Pro Lys

35 40 45

Pro Ile Thr Leu Phe Thr Gln Lys His Phe Ser Gly Val Arg Phe Leu
50 55 60

Lys Arg Val Ile Arg Asp Ala Ser Lys Ile Leu Lys Leu Ser Glu Ser
65 70 75 80
Arg Thr Ile Thr Phe Leu Glu Gln Ala Ile Glu Arg Asp Gly Ser Ala
85 90 95
Pro Pro Asp Val Thr Pro Pro Val His Asn Thr Ile Met Ala Val Thr
100 105 110
Arg Pro Phe Glu Glu Trp Pro Glu Val Ile Leu Ser Lys Ala Leu Gln
115 120 125

Lys His Cys Tyr Ala Leu Thr Lys Lys Ile Lys Ile Lys Thr Trp Pro
130 135 140
Lys Lys Gly Pro Gly Lys Lys Cys Leu Ala Ala Trp Ser Ala Arg Thr
145 150 155 160
Lys Ile Pro Leu Ile Pro Gly Gln Val Gln Ala Thr Asn Gly Leu Phe
165 170 175
Asp Arg Ile Gly Ser Ile Tyr Asp Gly Val Glu Lys Lys Val Thr Asn
180 185 190

Arg Asn Ala Asn Lys Lys Leu Glu Tyr Asp Glu Ala Ile Lys Glu Gly
195 200 205
Arg Asn Pro Ala Val Pro Glu Tyr Glu Thr Ala Tyr Asn Ile Asp Gly
210 215 220
Thr Leu Ile Asn Lys Pro Gly Tyr Asn Pro Asn Leu Tyr Ile Thr Gln
225 230 235 240
Ser Arg Thr Pro Arg Leu Ile Thr Glu Ala Asp Arg Pro Leu Val Glu
245 250 255

Lys Ile Leu Trp Gln Met Val Glu Lys Lys Thr Gln Ser Arg Asn Gln
260 265 270
Ala Arg Arg Ala Arg Leu Glu Lys Ala Ala His Leu Gln Gly Leu Pro
275 280 285
Val Pro Lys Phe Val Pro Glu Lys Val Asp Arg Ser Gln Lys Ile Glu

290 295 300

Ile Arg Ile Ile Asp Pro Leu Asp Lys Ile Glu Pro Tyr Met Pro Gln
305 310 315 320

Asp Arg Met Ala Ile Lys Ala Ser Gln Asp Gly His Val Pro Tyr Trp
325 330 335

Gln Arg Pro Phe Leu Ser Lys Arg Arg Asn Arg Arg Val Arg Ala Gly
340 345 350

Trp Gly Lys Gln Val Ser Ser Ile Gln Ala Trp Leu Thr Gly Ala Leu
355 360 365

Leu Val Ile Val Arg Leu Gly Asn Glu Ala Phe Leu Ala Asp Ile Arg
370 375 380

Gly Ala Leu Arg Asn Ala Gln Trp Arg Lys Leu Leu Lys Pro Asp Ala
385 390 395 400

Thr Tyr Gln Ser Leu Phe Asn Leu Phe Thr Gly Asp Pro Val Val Asn
405 410 415

Thr Arg Thr Asn His Leu Thr Met Ala Tyr Arg Glu Gly Val Val Asn
420 425 430

Ile Val Lys Ser Arg Ser Phe Lys Gly Arg Gln Thr Arg Glu His Leu
435 440 445

Leu Thr Leu Leu Gly Gln Gly Lys Thr Val Ala Gly Val Ser Phe Asp
450 455 460

Leu Gly Gln Lys His Ala Ala Gly Leu Leu Ala Ala His Phe Gly Leu
465 470 475 480

Gly Glu Asp Gly Asn Pro Val Phe Thr Pro Ile Gln Ala Cys Phe Leu
485 490 495

Pro Gln Arg Tyr Leu Asp Ser Leu Thr Asn Tyr Arg Asn Arg Tyr Asp
500 505 510

Ala Leu Thr Leu Asp Met Arg Arg Gln Ser Leu Leu Ala Leu Thr Pro
515 520 525

Ala Gln Gln Gln Glu Phe Ala Asp Ala Gln Arg Asp Pro Gly Gly Gln
530 535 540

Ala Lys Arg Ala Cys Cys Leu Lys Leu Asn Leu Asn Pro Asp Glu Ile
545 550 555 560
Arg Trp Asp Leu Val Ser Gly Ile Ser Thr Met Ile Ser Asp Leu Tyr
565 570 575

Ile Glu Arg Gly Gly Asp Pro Arg Asp Val His Gln Gln Val Glu Thr
580 585 590
Lys Pro Lys Gly Lys Arg Lys Ser Glu Ile Arg Ile Leu Lys Ile Arg
595 600 605
Asp Gly Lys Trp Ala Tyr Asp Phe Arg Pro Lys Ile Ala Asp Glu Thr
610 615 620
Arg Lys Ala Gln Arg Glu Gln Leu Trp Lys Leu Gln Lys Ala Ser Ser
625 630 635 640

Glu Phe Glu Arg Leu Ser Arg Tyr Lys Ile Asn Ile Ala Arg Ala Ile
645 650 655
Ala Asn Trp Ala Leu Gln Trp Gly Arg Glu Leu Ser Gly Cys Asp Ile
660 665 670
Val Ile Pro Val Leu Glu Asp Leu Asn Val Gly Ser Lys Phe Phe Asp
675 680 685
Gly Lys Gly Lys Trp Leu Leu Gly Trp Asp Asn Arg Phe Thr Pro Lys
690 695 700

Lys Glu Asn Arg Trp Phe Ile Lys Val Leu His Lys Ala Val Ala Glu
705 710 715 720
Leu Ala Pro His Arg Gly Val Pro Val Tyr Glu Val Met Pro His Arg
725 730 735
Thr Ser Met Thr Cys Pro Ala Cys His Tyr Cys His Pro Thr Asn Arg
740 745 750
Glu Gly Asp Arg Phe Glu Cys Gln Ser Cys His Val Val Lys Asn Thr
755 760 765

Asp Arg Asp Val Ala Pro Tyr Asn Ile Leu Arg Val Ala Val Glu Gly
770 775 780
Lys Thr Leu Asp Arg Trp Gln Ala Glu Lys Lys Pro Gln Ala Glu Pro

785 790 795 800

Asp Arg Pro Met Ile Leu Ile Asp Asn Gln Glu Ser

805 810

<210> 295

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 295

ccttacctag ggacgcagcc 20

<210> 296

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 296

ggatccccag cgccagctgc 20

<210> 297

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 297

agcgccagct gccggcctcc 20

<210> 298

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 298

gcgccagctg ccggcctcca 20

<210> 299

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 299

cctggctcag cgccacgggc 20

<210> 300

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 300

gctgcaggtg cagacaggtg 20

<210> 301

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 301

gcggtaccac gtcttgtaga 20

<210> 302

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 302

tgcctgtggg aacaaacaga 20

<210> 303

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 303

cttactttca ctatcctggg 20

<210> 304

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 304

tcccttactt tcactatcct 20

<210> 305

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 305

ctcccaggat agtgaaagta 20

<210> 306

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 306

tgatgtctga aagggcagag 20

<210> 307

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 307

tgcccaggag ctggtgcgga 20

<210> 308

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 308

ttgctgccac agaaccagaa 20

<210> 309

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 309

attcaaggga ttgaaaaccc 20

<210> 310

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 310

acctgcccag gggatacccg 20

<210> 311

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 311

cagcggcagc cttctgagtc 20

<210> 312

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 312

gctacaaaca agctcatctt 20

<210> 313

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 313

ccagccaagt acgtaagtag 20

<210> 314

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 314

ctggatatct gtgggacaag 20

<210> 315

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 315

cttacctggg ctggggaaga 20

<210> 316

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 316

ttcgtatctg taaaaccaag 20

<210> 317

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 317

tttcaaaacc tgtcagtgat 20

<210> 318

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 318

ttcaaaacct gtcagtgatt 20

<210> 319

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 319

ggaccctgca tagagagaga 20

<210> 320

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 320

tgccccagca gacgggcacg 20

<210> 321

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 321

gttacctggg ccatgtcccc 20

<210> 322

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 322

cttactgtta gatttatatc 20

<210> 323

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 323

ttactgttag atttatatca 20

<210> 324

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 324

tttgctatgg aaacacaaac 20

<210> 325

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 325

tccatagcaa atatccacat 20

<210> 326

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 326

gcagcaaccc tcacaacctt 20

<210> 327

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 327

cagcaaccct cacaaccttt 20

<210> 328

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 328

aggcaggctc ccctcgcctc 20

<210> 329

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 329

ggagcagcag gaccagcttc 20

<210> 330

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 330

caggaatacc tgagctttct 20

<210> 331

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 331

aggttccaga ttccattgct 20

<210> 332

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 332

ctgggcccag gggctgaggc 20

<210> 333

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 333

gatcgagtgg ccccaggtcc 20

<210> 334

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 334

tcacccgact tctgaacgtg 20

<210> 335

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 335

ttacctgccc actgttagcc 20

<210> 336

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 336

gaagccacag gaagtctgtg 20

<210> 337

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 337

actccagttc ctgacggctg 20

<210> 338

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 338

acctacagga gtacctgacg 20

<210> 339

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 339

ttcccagagc accagagcca 20

<210> 340

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 340

acgttcagaa gtcgggtgag 20

<210> 341

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 341

ttcagagcag tttgagacag 20

<210> 342

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 342

cctgctggga aacagacaac 20

<210> 343

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 343

cactcacagt ctagggtggc 20

<210> 344

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 344

caaccggcag cgagggaacg 20

<210> 345

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 345

acagggctgg ggcaggacag 20

<210> 346

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 346

catccaccag ctcgcagcac 20

<210> 347

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 347

atccaccagc tcgcagcaca 20

<210> 348

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 348

tccaccagct cgcagcacag 20

<210> 349

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 349

ccaccagctc gcagcacagg 20

<210> 350

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 350

tgtcacctgg caggaggagg 20

<210> 351

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 351

gtcacctggc aggaggaggc 20

<210> 352

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 352

ggcaccctgc agggagaaga 20

<210> 353

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 353

gcaccctgca gggagaagaa 20

<210> 354

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 354

attctgtcgt gggtaggggc 20

<210> 355

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 355

ctccattctg tcgtgggtag 20

<210> 356

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 356

cctcgggctg caaaggagag 20

<210> 357

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 357

cgggctgcaa aggagagggg 20

<210> 358

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 358

gctgaccttt ccggtatccc 20

<210> 359

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 359

ctgacctttc cggtatccca 20

<210> 360

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 360

tgttgcactg agatggggca 20

<210> 361

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 361

ctgttgcact gagatggggc 20

<210> 362

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 362

gccctgcagc ccccagaacc 20

<210> 363

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 363

ctcacccgca gtgacacaca 20

<210> 364

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 364

gcagcaccca gggcaggacc 20

<210> 365

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 365

cagcacccag ggcaggacca 20

<210> 366

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 366

gtccctgtgg aacagcagca 20

<210> 367

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 367

gtgggttcaa gttcggatgg 20

<210> 368

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 368

gccccacctg ggtctgagct 20

<210> 369

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 369

ggccccacct gggtctgagc 20

<210> 370

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 370

ccacctgggt ctgagctggg 20

<210> 371

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 371

aggcctccca ggagccctca 20

<210> 372

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 372

ctcccaggag ccctcaggga 20

<210> 373

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 373

cccccagctc acagtcacca 20

<210> 374

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 374

ggtctcaccg gtgcttatgt 20

<210> 375

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 375

tgctgcgccg acataagcac 20

<210> 376

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 376

ggcagggctg ggagagagca 20

<210> 377

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 377

cgagagcacg agcatgggtg 20

<210> 378

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 378

gagcacgagc atgggtgagg 20

<210> 379

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 379

agcacgagca tgggtgagga 20

<210> 380

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 380

gcacgagcat gggtgaggag 20

<210> 381

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 381

ctcacccatg ctcgtgctct 20

<210> 382

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 382

ggtgcctgcg cgggggaagg 20

<210> 383

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 383

gtgcctgcgc gggggaagga 20

<210> 384

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 384

cttggtgcct gcgcggggga 20

<210> 385

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 385

ggccccaggg ccctgccgcc 20

<210> 386

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 386

tctacgctca ggcaggggag 20

<210> 387

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 387

ccaccaggac ggcccccat 20

<210> 388

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 388

aggcccaggc ggcagggccc 20

<210> 389

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 389

ggcccaggcg gcagggccct 20

<210> 390

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 390

acgactggcc agggcgcctg 20

<210> 391

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 391

caccgcccag acgactggcc 20

<210> 392

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 392

ctacaactgg gctggcggcc 20

<210> 393

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 393

ggagtctgag agatggagag 20

<210> 394

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 394

cagcaaccag acggacaagc 20

<210> 395

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 395

gtgtcacaca actgcccaac 20

<210> 396

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 396

agccggccag ttccaaaccc 20

<210> 397

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 397

cagttccaaa ccctggtggt 20

<210> 398

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 398

cggccagttc caaaccctgg 20

<210> 399

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 399

ggacccagac tagcagcacc 20

<210> 400

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 400

gacgttacct cgtgcggccc 20

<210> 401

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 401

tccctgcaga gaaacacact 20

<210> 402

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 402

gagactcacc aggggctggc 20

<210> 403

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 403

cctccttctt tgaggagaaa 20

<210> 404

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 404

ttctctctgg aagggcacaa 20

<210> 405

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 405

ccagtggcga gagaagaccc 20

<210> 406

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 406

tgcccagcca ctgaggcctg 20

<210> 407

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 407

cgactggcca gggcgcctgt 20

<210> 408

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 408

accgcccaga cgactggcca 20

<210> 409

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 409

gtttctgcag ccgctttggg 20

<210> 410

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 410

ttacctggag ccacccaaag 20

<210> 411

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 411

tcactaggtg agcaaaagag 20

<210> 412

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 412

gcctctccag ccaggggctg 20

<210> 413

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 413

cttggcagca tcagccagac 20

<210> 414

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 414

ccactgggcg ggaaagagaa 20

<210> 415

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 415

acatactcga ggcctggccc 20

<210> 416

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 416

cctgcagccc cgcgtccagc 20

<210> 417

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 417

cgcgtccagc tggatgagcg 20

<210> 418

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 418

tgggccaggc ctcgagtatg 20

<210> 419

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 419

gggagttacc cagaacagtg 20

<210> 420

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 420

cctgccccaa gtcagcccca 20

<210> 421

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 421

gccagggccg agtccctgtc 20

<210> 422

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 422

ccagggccga gtccctgtcc 20

<210> 423

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 423

gccccagggc ccagagtcca 20

<210> 424

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 424

atgtgagcca ggcccaggct 20

<210> 425

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 425

gaggagtcca cttggcagtg 20

<210> 426

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 426

gagtcactga aaagagtaga 20

<210> 427

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 427

ctggacaaga acgctttgtg 20

<210> 428

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 428

ccatcccaga ggagtttctc 20

<210> 429

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 429

tggcaatgcc agctgtacca 20

<210> 430

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 430

taccaggggg agaggcttct 20

<210> 431

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 431

ggcattgcca aggctgggaa 20

<210> 432

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 432

ggcacctatg gagaaagtac 20

<210> 433

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 433

agacaggtga gccagggaca 20

<210> 434

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 434

ggctcacctg tcttctccaa 20

<210> 435

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 435

gtcgccactg tgagaagaga 20

<210> 436

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 436

gcaggctcag agcaagatag 20

<210> 437

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 437

gctggagcaa gaaccggagc 20

<210> 438

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 438

actcaccttt gcagaagaca 20

<210> 439

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 439

cactcacctt tgcagaagac 20

<210> 440

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 440

agggccaggt cctggtagcc 20

<210> 441

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 441

ggcccagcct gctgtggtac 20

<210> 442

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 442

gcttcggcag gctgacagcc 20

<210> 443

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 443

tatccaagga ctgagggcca 20

<210> 444

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 444

ggaacccaga tttatgtaat 20

<210> 445

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 445

gctcaccaat tacataaatc 20

<210> 446

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 446

ctcaccaatt acataaatct 20

<210> 447

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 447

ctcagctgaa cctggctacc 20

<210> 448

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 448

ggcgtctcaa acaggtatct 20

<210> 449

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 449

caaagcctga agagaaatcc 20

<210> 450

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 450

agagcctgaa ggagaagtct 20

<210> 451

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 451

tcccagtacc gggaaggcga 20

<210> 452

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 452

ggttcctgtg gagcacaggg 20

<210> 453

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 453

tggggttcct gtggagcaca 20

<210> 454

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 454

tcatttccta gatgggagac 20

<210> 455

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 455

ttcctagatg ggagacaggc 20

<210> 456

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 456

ccaggtgtct gaggaggaag 20

<210> 457

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 457

gtgtctgagg aggaagggga 20

<210> 458

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 458

tagtcctggg tggggtaggg 20

<210> 459

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 459

agtcctgggt ggggtagggt 20

<210> 460

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 460

cattacctca tagtaggcaa 20

<210> 461

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 461

ccattacctc atagtaggca 20

<210> 462

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 462

acagatctga gcagataaca 20

<210> 463

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 463

tcctacccac tggttgccct 20

<210> 464

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 464

aggtgcagta cctgaaggac 20

<210> 465

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 465

caggcagctg gcgggcgcca 20

<210> 466

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 466

gacttccagg gctccttctg 20

<210> 467

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 467

catccagata cattttgtca 20

<210> 468

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 468

acctgccaaa cacagacagt 20

<210> 469

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 469

cgtgcagatg caatggtcca 20

<210> 470

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 470

tgtaactgta acaaaacata 20

<210> 471

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 471

acttaccacc gaccatgcat 20

<210> 472

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 472

cttaccaaaa accttgactg 20

<210> 473

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 473

ccagtccaaa tcttcgatga 20

<210> 474

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 474

cagtccaaat cttcgatgat 20

<210> 475

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 475

aaaccatgtg atatttgctt 20

<210> 476

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 476

atgttccaca ctttatttcc 20

<210> 477

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 477

tcacgttgag gagtggtgtt 20

<210> 478

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 478

cattgtctag ggatataaag 20

<210> 479

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 479

acattgtcta gggatataaa 20

<210> 480

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 480

gacattgtct agggatataa 20

<210> 481

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 481

tggccaggac attccatctt 20

<210> 482

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 482

ccattctagg aacaaaatat 20

<210> 483

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 483

gagcttccaa gtcttctcca 20

<210> 484

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 484

tccccgaaaa ggtcgaattt 20

<210> 485

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 485

atgaagaaca gtcacaggac 20

<210> 486

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 486

gatttcgtct gtaggcacaa 20

<210> 487

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 487

taaacttacc tgaaacagcc 20

<210> 488

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 488

attcagacag tgccttcatg 20

<210> 489

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 489

gttgcactcg attgggacag 20

<210> 490

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 490

ttatttcaag ccctgattga 20

<210> 491

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 491

ttacctgtgt aacttttata 20

<210> 492

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 492

attgttcctg gaatttgggg 20

<210> 493

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 493

attatacctg ccatgccgta 20

<210> 494

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 494

gttcctggaa tttggggagg 20

<210> 495

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 495

ctgccatgcc gtaaggcaag 20

<210> 496

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 496

tctacctttg gtgaacccgt 20

<210> 497

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 497

cttaccttag ctccttctaa 20

<210> 498

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 498

gggatgtcga ctcctagggg 20

<210> 499

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 499

cgagggcacc atgcttcaag 20

<210> 500

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 500

aaactcactt tatgctaggg 20

<210> 501

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 501

ctcactttat gctagggagg 20

<210> 502

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 502

cttcacctgc atttaaagaa 20

<210> 503

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 503

ccaccagatt agctctggcc 20

<210> 504

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 504

ctacctttgg tgaacccgtt 20

<210> 505

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 505

atgttccaga tgtccagata 20

<210> 506

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 506

tgttccagat gtccagatat 20

<210> 507

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 507

agatagacaa acaagttgga 20

<210> 508

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 508

agcttgcacc aagtcacatg 20

<210> 509

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 509

acccaccttg gtgccttctc 20

<210> 510

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 510

actcacgctg gatagcctcc 20

<210> 511

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 511

tggagtacct gaggaatatc 20

<210> 512

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 512

ttaccccact taactatctt 20

<210> 513

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 513

tcgatctatg aaaaagacag 20

<210> 514

<211> 18

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 514

taccccactt aactatct 18

<210> 515

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 515

ctcaggtact ccaaagattc 20

<210> 516

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 516

cttaccccac ttaactatct 20

<210> 517

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 517

catttgccag acagaacctc 20

<210> 518

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 518

aaacaagacc aaaaggctaa 20

<210> 519

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 519

gtaagccaag aaagaaatcc 20

<210> 520

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 520

gcttcagatt ctgaatgagc 20

<210> 521

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 521

ttaaaacaaa atgaaatgaa 20

<210> 522

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 522

gttcctcagc ttccttcaga 20

<210> 523

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 523

caaagagcaa gagattctga 20

<210> 524

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 524

aaagagcaag agattctgaa 20

<210> 525

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 525

acacagcact atctgaaacc 20

<210> 526

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 526

cacccagaaa acaacacagc 20

<210> 527

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 527

taccaagttg aaatgaatca 20

<210> 528

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 528

atgaatcaag ggcagtccca 20

<210> 529

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 529

agggcagtcc caaggtacag 20

<210> 530

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 530

gttccaaaaa ccctcacacc 20

<210> 531

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 531

gaaacagcac ttgaatcaac 20

<210> 532

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 532

attacaaata aagaataaag 20

<210> 533

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 533

taatgtccaa atgggactgg 20

<210> 534

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 534

caaagcaaga tcttcttcac 20

<210> 535

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 535

acaacaagct tcagttctac 20

<210> 536

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 536

ctgcgcaact tgctcagcaa 20

<210> 537

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 537

cactcagacc cctccccaga 20

<210> 538

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 538

tttttccatg ttttgttttc 20

<210> 539

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 539

ttacctacac atctgcaaga 20

<210> 540

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 540

acacttaccc acttagcaat 20

<210> 541

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 541

catgcagaat ggcagcacat 20

<210> 542

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 542

aagctcagga ggagaaaaaa 20

<210> 543

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 543

cgcaagccag gctaaacagt 20

<210> 544

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 544

ttctccccag tctcagccga 20

<210> 545

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 545

tggtcaggaa aagcagccat 20

<210> 546

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 546

ctagtccagg gtgtggcttc 20

<210> 547

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 547

ccccaaaatc aacatggcag 20

<210> 548

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 548

tgtgatccag cagccttctt 20

<210> 549

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 549

gaccagatca aggccataag 20

<210> 550

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 550

caagacaaga aaagcatgaa 20

<210> 551

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 551

ctgaacaggg actgttagga 20

<210> 552

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 552

ccagagctca gagtggcaac 20

<210> 553

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 553

ttgctgcaga cgccccgtga 20

<210> 554

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 554

ctgcagacgc cccgtgacgg 20

<210> 555

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 555

cgacaagcag tgcctttgct 20

<210> 556

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 556

gcccagaacg tgattgacta 20

<210> 557

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 557

tgtgccaggg gtgttatgac 20

<210> 558

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 558

cagatccagt ctgatggcag 20

<210> 559

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 559

tgtacacgat ggtcagccat 20

<210> 560

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 560

ttttaccttg gggctctgac 20

<210> 561

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 561

agccccaagg taaaaaggcc 20

<210> 562

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 562

gagccccaag gtaaaaaggc 20

<210> 563

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 563

cagctcacag tgtgccacca 20

<210> 564

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 564

tatgaccaga tggacctggc 20

<210> 565

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 565

actggaccag tatgtcttcc 20

<210> 566

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 566

tgtcttccag gactcccagc 20

<210> 567

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 567

ttcaaccagg agccagcctc 20

<210> 568

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 568

gaccagattc ccagtatgtt 20

<210> 569

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 569

taacatactg ggaatctggt 20

<210> 570

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 570

aaaggcactg caagagacaa 20

<210> 571

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 571

ctctggcaaa tctctgaggc 20

<210> 572

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 572

agccaagtac cccctcccag 20

<210> 573

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 573

acctcccgag caaacatgac 20

<210> 574

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 574

ccttacctgt catgtttgct 20

<210> 575

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 575

tgctctggag atggagaagc 20

<210> 576

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 576

cccacccaat gcccggcagc 20

<210> 577

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 577

aggccatttt ggaagcttgt 20

<210> 578

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 578

accggctctg caaaggccag 20

<210> 579

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 579

tggtgcaggc caggctggag 20

<210> 580

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 580

gaacggcagc tggcccaagg 20

<210> 581

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 581

ggcccaagga ggcctggctg 20

<210> 582

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 582

gacacgagtg attgctgtgc 20

<210> 583

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 583

ctggtcaggg caagagctat 20

<210> 584

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 584

gggcccacag ccactcgtgg 20

<210> 585

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 585

ttccagaaga agctgctccg 20

<210> 586

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 586

cctggtccag agcctgagca 20

<210> 587

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 587

cagacatcaa agtaccctac 20

<210> 588

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 588

acatcaaagt accctacagg 20

<210> 589

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 589

cgcccaggtc ctcacgtctg 20

<210> 590

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 590

cttagtccaa cacccaccgc 20

<210> 591

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 591

cctcctgcaa tgcttcctgg 20

<210> 592

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 592

gagccagcca cagggccccc 20

<210> 593

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 593

ggaagcagaa ggtgcttgcg 20

<210> 594

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 594

ggctgcagcc ggggacactg 20

<210> 595

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 595

ctgccaaatt ccagcctcct 20

<210> 596

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 596

ggcgggccaa gacttctccc 20

<210> 597

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 597

agactcagag gtgagaggag 20

<210> 598

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 598

agcctaggag gcaaagagca 20

<210> 599

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 599

cccccaggct ttccccaaac 20

<210> 600

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 600

tcactccaga tgctgcaggg 20

<210> 601

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 601

aggctgcagg tggaatcaga 20

<210> 602

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 602

cttcccccag ctgaagtcct 20

<210> 603

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 603

cactcacttg agggtttcca 20

<210> 604

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 604

cagactgcgg ggacacagtg 20

<210> 605

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 605

ccactcacct tagcctgagc 20

<210> 606

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 606

cactcacctt agcctgagca 20

<210> 607

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 607

gtacaagctg tcggaaacag 20

<210> 608

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 608

acactcactc catcacccgg 20

<210> 609

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 609

cactcactcc atcacccgga 20

<210> 610

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 610

cgtccagtac aacaagttca 20

<210> 611

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 611

ccacatcctg caaggggga 20

<210> 612

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 612

cacatcctgc aaggggggat 20

<210> 613

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 613

tgggcgtcca catcctgcaa 20

<210> 614

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 614

gggcgtccac atcctgcaag 20

<210> 615

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 615

ggcgtccaca tcctgcaagg 20

<210> 616

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 616

gcgtccacat cctgcaaggg 20

<210> 617

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 617

gcccaaggta agagcttccc 20

<210> 618

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 618

gctcttacct tgggcagcca 20

<210> 619

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 619

agctcttacc ttgggcagcc 20

<210> 620

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 620

tgcacctctg gggaggacct 20

<210> 621

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 621

ctgcacctct ggggaggacc 20

<210> 622

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 622

cgcctgcagc tgtcggacac 20

<210> 623

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 623

cacctgccag gccatcacgg 20

<210> 624

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 624

ccctacctgt caccaggacc 20

<210> 625

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 625

cctacctgtc accaggacca 20

<210> 626

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 626

cctctgagaa ggaaaaaaga 20

<210> 627

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 627

cagaggaaca gtcccaagga 20

<210> 628

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 628

cactcacctg cccacagcag 20

<210> 629

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 629

ccactcacct gcccacagca 20

<210> 630

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 630

gccactcacc tgcccacagc 20

<210> 631

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 631

agggcccctg tggggaaacg 20

<210> 632

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 632

gggcccctgt ggggaaacga 20

<210> 633

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 633

gcaagtgcag gagtcagtga 20

<210> 634

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 634

cggaaccagt aaccatgaac 20

<210> 635

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 635

ggaaccagta accatgaact 20

<210> 636

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 636

gaaccagtaa ccatgaactg 20

<210> 637

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 637

gctagatcaa gaagtacagg 20

<210> 638

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 638

agaagtacag gaggagactc 20

<210> 639

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 639

caagtctagt gaggagaaag 20

<210> 640

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 640

aagtctagtg aggagaaaga 20

<210> 641

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 641

agtctagtga ggagaaagag 20

<210> 642

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 642

acaggcccag gacacagagc 20

<210> 643

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 643

acctgtcagg tgaagttcgc 20

<210> 644

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 644

acttacaggt gacgttgagc 20

<210> 645

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 645

aacatctagg agaggaagag 20

<210> 646

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 646

gttccacaga acccaacaac 20

<210> 647

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 647

ttcctacctg agccatctcc 20

<210> 648

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 648

atgctcacat gaagaagatg 20

<210> 649

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 649

gggaaacaag agaccagagc 20

<210> 650

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 650

tcactctgat gggagacacc 20

<210> 651

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 651

cactctgatg ggagacacca 20

<210> 652

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 652

tttcttatgg agaggaaaga 20

<210> 653

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 653

gtacaggtaa gagcaacgcc 20

<210> 654

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 654

ctcttacctg taccataacc 20

<210> 655

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 655

ttacctgtac cataaccagg 20

<210> 656

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 656

tgtatctgta ggaggagaag 20

<210> 657

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 657

gtatctgtag gaggagaagt 20

<210> 658

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 658

cagatacaaa ctggactctc 20

<210> 659

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 659

tcttaccttc cttcgtttgc 20

<210> 660

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 660

tttggatcta aaacggtgac 20

<210> 661

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 661

gatctaaaac ggtgacaggt 20

<210> 662

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 662

aaaggctcag cagttgacct 20

<210> 663

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 663

atttaccggc atagaatagt 20

<210> 664

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 664

tcactgccta agagagacat 20

<210> 665

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 665

aggatccacg tggagaatgg 20

<210> 666

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 666

ggatccacgt ggagaatggt 20

<210> 667

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 667

tctcactgtt ctcagggaag 20

<210> 668

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 668

cctgcccaag gcttcccacc 20

<210> 669

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 669

ctgcccaagg cttcccacct 20

<210> 670

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 670

gcctgctgcg gtaagcggta 20

<210> 671

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 671

gatgctcagg gaacacgtat 20

<210> 672

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 672

ttctcaaagt tcccacatcc 20

<210> 673

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 673

tcacagattg gtgagtagcc 20

<210> 674

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 674

ctcacctgtt ctgtgattac 20

<210> 675

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 675

tccttccagc tactcaatcc 20

<210> 676

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 676

cacagtacaa ataagagccc 20

<210> 677

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 677

ccccccagcg cttcggtgag 20

<210> 678

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 678

cccccagcgc ttcggtgagt 20

<210> 679

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 679

ccactcaccg aagcgctggg 20

<210> 680

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 680

tacctcggag gaaagagaaa 20

<210> 681

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 681

cagcttccaa aacgacaagc 20

<210> 682

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 682

aacataccag cttgtcgttt 20

<210> 683

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 683

agaccacctg cagagacgag 20

<210> 684

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 684

ccacctgcag agacgagagg 20

<210> 685

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 685

ccacacccaa aaggccacac 20

<210> 686

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 686

cccaccagct cagctccacg 20

<210> 687

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 687

cgctgtcaag tccagttcta 20

<210> 688

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 688

gctgtcaagt ccagttctac 20

<210> 689

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 689

cacccagatc gtcagcgccg 20

<210> 690

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 690

ccactcacct gctctacccc 20

<210> 691

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 691

ccacagtctg aaagaaagca 20

<210> 692

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 692

gacactgttg gcacggagga 20

<210> 693

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 693

ttaccatggc catcaacaca 20

<210> 694

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 694

ccacagtctg aaagaaaaca 20

<210> 695

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 695

cacagtctga aagaaaacag 20

<210> 696

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 696

ttaccatggc catcagcacg 20

<210> 697

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 697

tctgcgttca ggggtaagcg 20

<210> 698

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 698

gcgcttaccc ctgaacgcag 20

<210> 699

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 699

gactgggcag cggcccctgt 20

<210> 700

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 700

ggactgggca gcggcccctg 20

<210> 701

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 701

gtcatgcagc ccttgcctgc 20

<210> 702

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 702

tcatgcagcc cttgcctgct 20

<210> 703

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 703

catgcagccc ttgcctgctg 20

<210> 704

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 704

ctcaccagat tcccgaaggt 20

<210> 705

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 705

cagactcacc agattcccga 20

<210> 706

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 706

agcctaggca agtgcagagg 20

<210> 707

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 707

cagcctaggc aagtgcagag 20

<210> 708

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 708

accagcctag gcaagtgcag 20

<210> 709

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 709

tggaacccca ataccagcct 20

<210> 710

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 710

cacctgcaga agatgccaga 20

<210> 711

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 711

cgctcacctg caccagcctc 20

<210> 712

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 712

cttcctggca agacacaaga 20

<210> 713

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 713

aattcaggga gccattgaaa 20

<210> 714

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 714

ctactgacct gcctttcaa 20

<210> 715

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 715

gcctctcaaa gtcttatgtg 20

<210> 716

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 716

ttcccatgct gctttacttc 20

<210> 717

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 717

tttaggtcaa ccagatgtag 20

<210> 718

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 718

tgtgcctgaa agcaaaaatg 20

<210> 719

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 719

ttacctacta ttcttgccag 20

<210> 720

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 720

aaatgctgtt taaaaaaagg 20

<210> 721

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 721

ccccaaaaag tgaatatcaa 20

<210> 722

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 722

gtccagaatt tccagaagta 20

<210> 723

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 723

tccctggagg ggtgggggag 20

<210> 724

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 724

tcacttcaac aggactccta 20

<210> 725

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 725

ccttacactc aaaggcaaag 20

<210> 726

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 726

atggctgcag ggagaggaag 20

<210> 727

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 727

ggagcgccag gggatgctgg 20

<210> 728

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 728

tcacgtccca tgcagccaca 20

<210> 729

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 729

tggacgtctg gtggggagta 20

<210> 730

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 730

actcacattg atgaggaaga 20

<210> 731

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 731

cagcatctga gaaaggcaga 20

<210> 732

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 732

aatccaacac ctcagcgatg 20

<210> 733

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 733

atccaacacc tcagcgatgt 20

<210> 734

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 734

cgcggcgcag tacccgctgc 20

<210> 735

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 735

cactcaccgt ggtgggtcgg 20

<210> 736

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 736

actcaccgtg gtgggtcgga 20

<210> 737

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 737

tggctccctg tggggcaacg 20

<210> 738

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 738

gattccaggg ggaggcggtg 20

<210> 739

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 739

gctcctacct gtgctggacc 20

<210> 740

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 740

tcgccgccac agtggggtcg 20

<210> 741

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 741

cagactgaga gtggggagag 20

<210> 742

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 742

cctcctccag agtgtggttg 20

<210> 743

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 743

agccatccca gggagagagc 20

<210> 744

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 744

gccatcccag ggagagagct 20

<210> 745

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 745

ccatcccagg gagagagctg 20

<210> 746

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 746

ctcacaggcc acccaccagc 20

<210> 747

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 747

ccgcttccag aagtgcctgg 20

<210> 748

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 748

cttccagaag tgcctggcgg 20

<210> 749

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 749

acaactgcaa aggaatgggt 20

<210> 750

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 750

caactgcaaa ggaatgggta 20

<210> 751

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 751

gaactaggaa gacggtccag 20

<210> 752

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 752

ggctgaccag gacctgttgc 20

<210> 753

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 753

ctcacctgta cgccaggcgg 20

<210> 754

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 754

gctctcacct gtacgccagg 20

<210> 755

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 755

cttagacctg gcaggcagat 20

<210> 756

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 756

caatccagtc cccgaagcca 20

<210> 757

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 757

aatccagtcc ccgaagccac 20

<210> 758

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 758

actcaccggt gatgaggaca 20

<210> 759

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 759

ctcaccggtg atgaggacaa 20

<210> 760

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 760

ccggtctgcg ggaagggtac 20

<210> 761

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 761

tgggctgcag gagccgcggc 20

<210> 762

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 762

ttgtaccaaa tgcccctgtc 20

<210> 763

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 763

cggacagcag tcctccatta 20

<210> 764

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 764

aggtgcagca cagccccatg 20

<210> 765

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 765

ggtgcagcac agccccatgt 20

<210> 766

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 766

agttgccaga tgcgcttcga 20

<210> 767

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 767

gttgccagat gcgcttcgac 20

<210> 768

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 768

gtctcagctg ctcgacacgc 20

<210> 769

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 769

ttcttaccct ggaatagtcc 20

<210> 770

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 770

attacctgta tgctaatcga 20

<210> 771

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 771

ttgcaatgcg ttcgtggctt 20

<210> 772

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 772

actgacctgt gaccatagcc 20

<210> 773

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 773

tatctgcagg gacagagaaa 20

<210> 774

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 774

tgcggcgcag acatacagct 20

<210> 775

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 775

ccccgcaggc gggggcgtta 20

<210> 776

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 776

tttcagaagt gtctcagtgt 20

<210> 777

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 777

attacctgat ggaaagtctg 20

<210> 778

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 778

cttcagtgcc ttcgtggatt 20

<210> 779

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 779

tttctgcaga gggatagaga 20

<210> 780

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 780

agaccaccag agtaagggac 20

<210> 781

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 781

acttgcagcc ctcggccaaa 20

<210> 782

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 782

agggctcacc agaggtagga 20

<210> 783

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 783

ttcacctgcc caaggaaaaa 20

<210> 784

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 784

ctaagcctag aaaatcagtt 20

<210> 785

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 785

acatctagaa aaaaaaatac 20

<210> 786

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 786

attaccttcc tctttgcact 20

<210> 787

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 787

gatcctgtag gttggaacat 20

<210> 788

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 788

aagccacccc aagttagatc 20

<210> 789

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 789

aacttacccc aaacaattag 20

<210> 790

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 790

atacctacag gatttaaagt 20

<210> 791

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 791

gtttcctaga aagcaaaaaa 20

<210> 792

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 792

aagttcaact gcttcgtaat 20

<210> 793

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 793

aattcagact atcatcctca 20

<210> 794

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 794

gcttgcagct gtgggacacc 20

<210> 795

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 795

gaccactcaa ccagtgccca 20

<210> 796

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 796

ggaggcacag accactccac 20

<210> 797

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 797

ggcacagaca actcgactga 20

<210> 798

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 798

gcacagacca ctcaacccac 20

<210> 799

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 799

gaccactcaa cccacaggcc 20

<210> 800

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 800

gaccactcaa accacagcca 20

<210> 801

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 801

gaccactcaa cccacagcca 20

<210> 802

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 802

gaccactcaa ccagcagcca 20

<210> 803

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 803

ccuuaccuag ggacgcagcc 20

<210> 804

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 804

ggaucccag cgccagcugc 20

<210> 805

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 805

agcgccagcu gccggccucc 20

<210> 806

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 806

gcgccagcug ccggccucca 20

<210> 807

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 807

ccuggcucag cgccacgggc 20

<210> 808

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 808

gcugcaggug cagacaggug 20

<210> 809

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 809

gcgguaccac gucuuguaga 20

<210> 810

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 810

ugccuguggg aacaaacaga 20

<210> 811

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 811

cuuacuuuca cuauccuggg 20

<210> 812

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 812

ucccuuacuu ucacuauccu 20

<210> 813

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 813

cucccaggau agugaaagua 20

<210> 814

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 814

ugaugucuga aagggcagag 20

<210> 815

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 815

ugcccaggag cuggugcgga 20

<210> 816

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 816

uugcugccac agaaccagaa 20

<210> 817

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 817

auucaaggga uugaaaaccc 20

<210> 818

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 818

accugcccag gggauacccg 20

<210> 819

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 819

cagcggcagc cuucugaguc 20

<210> 820

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 820

gcuacaaaca agcucaucuu 20

<210> 821

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 821

ccagccaagu acguaaguag 20

<210> 822

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 822

cuggauaucu gugggacaag 20

<210> 823

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 823

cuuaccuggg cuggggaaga 20

<210> 824

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 824

uucguaucug uaaaaccaag 20

<210> 825

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 825

uuucaaaacc ugucagugau 20

<210> 826

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 826

uucaaaaccu gucagugauu 20

<210> 827

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 827

ggacccugca uagagagaga 20

<210> 828

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 828

ugccccagca gacgggcacg 20

<210> 829

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 829

guuaccuggg ccaugucccc 20

<210> 830

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 830

cuuacuguua gauuuauauc 20

<210> 831

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 831

uuacuguuag auuuauauca 20

<210> 832

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 832

uuugcuaugg aaacacaaac 20

<210> 833

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 833

uccauagcaa auauccacau 20

<210> 834

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 834

gcagcaaccc ucacaaccuu 20

<210> 835

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 835

cagcaacccu cacaaccuuu 20

<210> 836

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 836

aggcaggcuc cccucgccuc 20

<210> 837

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 837

ggagcagcag gaccagcuuc 20

<210> 838

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 838

caggaauacc ugagcuuucu 20

<210> 839

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 839

agguuccaga uuccauugcu 20

<210> 840

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 840

cugggcccag gggcugaggc 20

<210> 841

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 841

gaucgagugg ccccaggucc 20

<210> 842

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 842

ucacccgacu ucugaacgug 20

<210> 843

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 843

uuaccugccc acuguuagcc 20

<210> 844

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 844

gaagccacag gaagucugug 20

<210> 845

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 845

acuccaguuc cugacggcug 20

<210> 846

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 846

accuacagga guaccugacg 20

<210> 847

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 847

uucccagagc accagagcca 20

<210> 848

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 848

acguucagaa gucgggugag 20

<210> 849

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 849

uucagagcag uuugagacag 20

<210> 850

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 850

ccugcuggga aacagacaac 20

<210> 851

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 851

cacucacagu cuaggguggc 20

<210> 852

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 852

acagggcugg ggcaggacag 20

<210> 853

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 853

cauccaccag cucgcagcac 20

<210> 854

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 854

auccaccagc ucgcagcaca 20

<210> 855

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 855

uccaccagcu cgcagcacag 20

<210> 856

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 856

ccaccagcuc gcagcacagg 20

<210> 857

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 857

ugucaccugg caggaggagg 20

<210> 858

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 858

gucaccuggc aggaggaggc 20

<210> 859

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 859

ggcacccugc agggagaaga 20

<210> 860

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 860

gcacccugca gggagaagaa 20

<210> 861

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 861

auucugucgu ggguaggggc 20

<210> 862

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 862

cuccauucug ucguggguag 20

<210> 863

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 863

ccucgggcug caaaggagag 20

<210> 864

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 864

cgggcugcaa aggagagggg 20

<210> 865

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 865

gcugaccuuu ccgguauccc 20

<210> 866

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 866

cugaccuuuc cgguauccca 20

<210> 867

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 867

uguugcacug agaugggga 20

<210> 868

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 868

cuguugcacu gagaugggc 20

<210> 869

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 869

gcccugcagc ccccagaacc 20

<210> 870

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 870

cucacccgca gugacacaca 20

<210> 871

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 871

gucccugugg aacagcagca 20

<210> 872

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 872

guggguucaa guucggaugg 20

<210> 873

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 873

gccccaccug ggucugagcu 20

<210> 874

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 874

ggccccaccu gggucugagc 20

<210> 875

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 875

ccaccugggu cugagcuggg 20

<210> 876

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 876

aggccuccca ggagcccuca 20

<210> 877

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 877

cucccaggag cccucaggga 20

<210> 878

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 878

cccccagcuc acagucacca 20

<210> 879

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 879

ggucucaccg gugcuuaugu 20

<210> 880

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 880

ugcugcgccg acauaagcac 20

<210> 881

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 881

ggcagggcug ggagagagca 20

<210> 882

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 882

cgagagcacg agcaugggug 20

<210> 883

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 883

gagcacgagc augggugagg 20

<210> 884

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 884

agcacgagca ugggugagga 20

<210> 885

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 885

gcacgagcau gggugaggag 20

<210> 886

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 886

cucacccaug cucgugcucu 20

<210> 887

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 887

ggugccugcg cgggggaagg 20

<210> 888

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 888

gugccugcgc gggggaagga 20

<210> 889

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 889

cuuggugccu gcgcggggga 20

<210> 890

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 890

ggccccaggg cccugccgcc 20

<210> 891

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 891

ucuacgcuca ggcaggggag 20

<210> 892

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 892

ccaccaggac ggccccccau 20

<210> 893

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 893

ggcccaggcg gcagggcccu 20

<210> 894

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 894

acgacuggcc agggcgccug 20

<210> 895

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 895

caccgcccag acgacuggcc 20

<210> 896

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 896

cuacaacugg gcuggcggcc 20

<210> 897

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 897

caccuaccua agaaccaucc 20

<210> 898

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 898

ggagucugag agauggagag 20

<210> 899

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 899

gugucacaca acugcccaac 20

<210> 900

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 900

agccggccag uuccaaaccc 20

<210> 901

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 901

caguuccaaa cccugguggu 20

<210> 902

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 902

cggccaguuc caaacccugg 20

<210> 903

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 903

ggacccagac uagcagcacc 20

<210> 904

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 904

gacguuaccu cgugcggccc 20

<210> 905

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 905

ucccugcaga gaaacacacu 20

<210> 906

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 906

gagacucacc aggggcuggc 20

<210> 907

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 907

ccuccuucuu ugaggagaaa 20

<210> 908

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 908

gggguuccag ggccugucug 20

<210> 909

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 909

uucucucugg aagggcacaa 20

<210> 910

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 910

ccaguggcga gagaagaccc 20

<210> 911

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 911

ugcccagcca cugaggccug 20

<210> 912

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 912

cgacuggcca gggcgccugu 20

<210> 913

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 913

accgcccaga cgacuggcca 20

<210> 914

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 914

guuucugcag ccgcuuuggg 20

<210> 915

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 915

uuaccuggag ccacccaaag 20

<210> 916

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 916

ucacuaggug agcaaaagag 20

<210> 917

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 917

gccucuccag ccaggggcug 20

<210> 918

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 918

cuuggcagca ucagccagac 20

<210> 919

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 919

ccacugggcg ggaaagagaa 20

<210> 920

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 920

acauacucga ggccuggccc 20

<210> 921

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 921

ccugcagccc cgcguccagc 20

<210> 922

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 922

cgcguccagc uggaugagcg 20

<210> 923

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 923

ugggccaggc cucgaguaug 20

<210> 924

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 924

gggaguuacc cagaacagug 20

<210> 925

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 925

ccugccccaa gucagcccca 20

<210> 926

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 926

gccagggccg agucccuguc 20

<210> 927

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 927

ccagggccga gucccugucc 20

<210> 928

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 928

gccccagggc ccagagucca 20

<210> 929

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 929

augugagcca ggcccaggcu 20

<210> 930

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 930

gaggagucca cuuggcagug 20

<210> 931

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 931

gagucacuga aaagaguaga 20

<210> 932

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 932

cuggacaaga acgcuuugug 20

<210> 933

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 933

ccaucccaga ggaguuucuc 20

<210> 934

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 934

uggcaaugcc agcuguacca 20

<210> 935

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 935

uaccaggggg agaggcuucu 20

<210> 936

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 936

ggcauugcca aggcugggaa 20

<210> 937

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 937

ggcaccuaug gagaaaguac 20

<210> 938

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 938

agacagguga gccagggaca 20

<210> 939

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 939

ggcucaccug ucuucuccaa 20

<210> 940

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 940

gucgccacug ugagaagaga 20

<210> 941

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 941

gcaggcucag agcaagauag 20

<210> 942

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 942

gcuggagcaa gaaccggagc 20

<210> 943

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 943

acucaccuuu gcagaagaca 20

<210> 944

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 944

cacucaccuu ugcagaagac 20

<210> 945

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 945

agggccaggu ccugguagcc 20

<210> 946

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 946

ggcccagccu gcugugguac 20

<210> 947

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 947

gcuucggcag gcugacagcc 20

<210> 948

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 948

uauccaagga cugagggcca 20

<210> 949

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 949

ggaacccaga uuuauguaau 20

<210> 950

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 950

gcucaccaau uacauaaauc 20

<210> 951

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 951

cucaccaauu acauaaaucu 20

<210> 952

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 952

cucagcugaa ccuggcuacc 20

<210> 953

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 953

ggcgucucaa acagguaucu 20

<210> 954

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 954

caaagccuga agagaaaucc 20

<210> 955

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 955

agagccugaa ggagaagucu 20

<210> 956

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 956

ucccaguacc gggaaggcga 20

<210> 957

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 957

gguuccugug gagcacaggg 20

<210> 958

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 958

ugggguuccu guggagcaca 20

<210> 959

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 959

ucauuuccua gaugggagac 20

<210> 960

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 960

uuccuagaug ggagacaggc 20

<210> 961

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 961

ccaggugucu gaggaggaag 20

<210> 962

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 962

gugucugagg aggaagggga 20

<210> 963

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 963

uaguccuggg ugggguaggg 20

<210> 964

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 964

aguccugggu gggguagggu 20

<210> 965

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 965

cauuaccuca uaguaggcaa 20

<210> 966
<211> 20
<212> RNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 966
ccauuaccuc auaguaggca 20
<210> 967
<211> 20
<212> RNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 967
acagaucuga gcagauaaca 20
<210> 968
<211> 20
<212> RNA
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 968
uccuacccac ugguugcccu 20
<210> 969
<211> 20
<212> RNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 969
aggugcagua ccugaaggac 20

<210> 970

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 970

caggcagcug gcgggcgcca 20

<210> 971

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 971

gacuuccagg gcuccuucug 20

<210> 972

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 972

cauccagaua cauuuuguca 20

<210> 973

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 973

accugccaaa cacagacagu 20

<210> 974

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 974

cgugcagaug caauggucca 20

<210> 975

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 975

uguaacugua acaaaacaua 20

<210> 976

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 976

acuuaccacc gaccaugcau 20

<210> 977

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 977

cuuaccaaaa accuugacug 20

<210> 978

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 978

ccaguccaaa ucuucgauga 20

<210> 979

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 979

caguccaaau cuucgaugau 20

<210> 980

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 980

aaaccaugug auauuugcuu 20

<210> 981

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 981

auguuccaca cuuuauuucc 20

<210> 982

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 982

ucacguugag gagugguguu 20

<210> 983

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 983

cauugucuag ggauauaaag 20

<210> 984

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 984

acauugucua gggauauaaa 20

<210> 985

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 985

gacauugucu agggauauaa 20

<210> 986

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 986

uggccaggac auuccaucuu 20

<210> 987

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 987

ccauucuagg aacaaaauau 20

<210> 988

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 988

gagcuuccaa gucuucucca 20

<210> 989

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 989

uccccgaaaa ggucgaauuu 20

<210> 990

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 990

augaagaaca gucacaggac 20

<210> 991

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 991

gauuucgucu guaggcacaa 20

<210> 992

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 992

uaaacuuacc ugaaacagcc 20

<210> 993

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 993

auucagacag ugccuucaug 20

<210> 994

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 994

guugcacucg auugggacag 20

<210> 995

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 995

uuauuucaag cccugauuga 20

<210> 996

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 996

uuaccugugu aacuuuuaua 20

<210> 997

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 997

auuguuccug gaauuugggg 20

<210> 998

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 998

auuauaccug ccaugccgua 20

<210> 999

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 999

guuccuggaa uuuggggagg 20

<210> 1000

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1000

cugccaugcc guaaggcaag 20

<210> 1001

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1001

ucuaccuuug gugaacccgu 20

<210> 1002

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1002

cuuaccuuag cuccuucuaa 20

<210> 1003

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1003

gggaugucga cuccuagggg 20

<210> 1004

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1004

cgagggcacc augcuucaag 20

<210> 1005

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1005

aaacucacuu uaugcuaggg 20

<210> 1006

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1006

cucacuuuau gcuagggagg 20

<210> 1007

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1007

cuucaccugc auuuaaagaa 20

<210> 1008

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1008

ccaccagauu agcucuggcc 20

<210> 1009

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1009

cuaccuuugg ugaacccguu 20

<210> 1010

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1010

auguuccaga uguccagaua 20

<210> 1011

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1011

uguuccagau guccagauau 20

<210> 1012

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1012

agauagacaa acaaguugga 20

<210> 1013

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1013

agcuugcacc aagucacaug 20

<210> 1014

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1014

acccaccuug gugccuucuc 20

<210> 1015

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1015

acucacgcug gauagccucc 20

<210> 1016

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1016

uggaguaccu gaggaauauc 20

<210> 1017

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1017

uuaccccacu uaacuaucuu 20

<210> 1018

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1018

ucgaucuaug aaaaagacag 20

<210> 1019

<211> 18

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1019

uaccccacuu aacuaucu 18

<210> 1020

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1020

cucagguacu ccaaagauuc 20

<210> 1021

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1021

cuuaccccac uuaacuaucu 20

<210> 1022

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1022

cauuugccag acagaaccuc 20

<210> 1023

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1023

aaacaagacc aaaaggcuaa 20

<210> 1024

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1024

guaagccaag aaagaaaucc 20

<210> 1025

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1025

gcuucagauu cugaaugagc 20

<210> 1026

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1026

uuaaaacaaa augaaaugaa 20

<210> 1027

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1027

guuccucagc uuccuucaga 20

<210> 1028

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1028

caaagagcaa gagauucuga 20

<210> 1029

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1029

aaagagcaag agauucugaa 20

<210> 1030

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1030

acacagcacu aucugaaacc 20

<210> 1031

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1031

uaccaaguug aaaugaauca 20

<210> 1032

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1032

augaaucaag ggcaguccca 20

<210> 1033

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1033

agggcagucc caagguacag 20

<210> 1034

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1034

guuccaaaaa cccucacacc 20

<210> 1035

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1035

gaaacagcac uugaaucaac 20

<210> 1036

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1036

auuacaaaua aagaauaaag 20

<210> 1037

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1037

uaauguccaa augggacugg 20

<210> 1038

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1038

caaagcaaga ucuucuucac 20

<210> 1039

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1039

acaacaagcu ucaguucuac 20

<210> 1040

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1040

cugcgcaacu ugcucagcaa 20

<210> 1041

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1041

cacucagacc ccucccccaga 20

<210> 1042

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1042

uuuuuccaug uuuuguuuuc 20

<210> 1043

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1043

uuaccuacac aucugcaaga 20

<210> 1044

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1044

acacuuaccc acuuagcaau 20

<210> 1045

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1045

caugcagaau ggcagcacau 20

<210> 1046

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1046

aagcucagga ggagaaaaaa 20

<210> 1047

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1047

cgcaagccag gcuaaacagu 20

<210> 1048

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1048

uucucccccag ucucagccga 20

<210> 1049

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1049

uggucaggaa aagcagccau 20

<210> 1050

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1050

cuaguccagg guguggcuuc 20

<210> 1051

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1051

ccccaaaauc aacauggcag 20

<210> 1052

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1052

ugugauccag cagccuucuu 20

<210> 1053

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1053

gaccagauca aggccauaag 20

<210> 1054

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1054

caagacaaga aaagcaugaa 20

<210> 1055

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1055

cugaacaggg acuguuagga 20

<210> 1056

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1056

ccagagcuca gaguggcaac 20

<210> 1057

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1057

uugcugcaga cgccccguga 20

<210> 1058

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1058

cugcagacgc cccgugacgg 20

<210> 1059

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1059

cgacaagcag ugccuuugcu 20

<210> 1060

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1060

gcccagaacg ugauugacua 20

<210> 1061

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1061

ugugccaggg guguuaugac 20

<210> 1062

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1062

cagauccagu cugauggcag 20

<210> 1063

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1063

uguacacgau ggucagccau 20

<210> 1064

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1064

uuuuaccuug gggcucugac 20

<210> 1065

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1065

agccccaagg uaaaaaggcc 20

<210> 1066

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1066

gagccccaag guaaaaaggc 20

<210> 1067

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1067

cagcucacag ugugccacca 20

<210> 1068

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1068

uaugaccaga uggaccuggc 20

<210> 1069

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1069

acuggaccag uaugucuucc 20

<210> 1070

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1070

ugucuuccag gacucccagc 20

<210> 1071

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1071

uucaaccagg agccagccuc 20

<210> 1072

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1072

gaccagauuc ccaguauguu 20

<210> 1073

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1073

uaacauacug ggaaucuggu 20

<210> 1074

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1074

aaaggcacug caagagacaa 20

<210> 1075

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1075

cucuggcaaa ucucugaggc 20

<210> 1076

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1076

agccaaguac ccccucccag 20

<210> 1077

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1077

accucccgag caaacaugac 20

<210> 1078

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1078

ccuuaccugu cauguuugcu 20

<210> 1079

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1079

ugcucuggag auggagaagc 20

<210> 1080

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1080

cccacccaau gcccggcagc 20

<210> 1081

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1081

aggccauuuu ggaagcuugu 20

<210> 1082

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1082

accggcucug caaaggccag 20

<210> 1083

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1083

uggugcaggc caggcuggag 20

<210> 1084

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1084

gaacggcagc uggcccaagg 20

<210> 1085

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1085

ggcccaagga ggccuggcug 20

<210> 1086

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1086

gacacgagug auugcugugc 20

<210> 1087

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1087

cuggucaggg caagagcuau 20

<210> 1088

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1088

gggcccacag ccacucgugg 20

<210> 1089

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1089

uuccagaaga agcugcuccg 20

<210> 1090

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1090

ccugguccag agccugagca 20

<210> 1091

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1091

cagacaucaa aguacccuac 20

<210> 1092

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1092

acaucaaagu acccuacagg 20

<210> 1093

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1093

cgcccagguc cucacgucug 20

<210> 1094

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1094

cuuaguccaa cacccaccgc 20

<210> 1095

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1095

ccuccugcaa ugcuuccugg 20

<210> 1096

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1096

ggaagcagaa ggugcuugcg 20

<210> 1097

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1097

ggcugcagcc ggggacacug 20

<210> 1098

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1098

cugccaaauu ccagccuccu 20

<210> 1099

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1099

ggcgggccaa gacuucuccc 20

<210> 1100

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1100

agacucagag gugagaggag 20

<210> 1101

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1101

agccuaggag gcaaagagca 20

<210> 1102

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1102

ccccaggcu uuccccaaac 20

<210> 1103

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1103

ucacuccaga ugcugcaggg 20

<210> 1104

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1104

aggcugcagg uggaaucaga 20

<210> 1105

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1105

cuucccccag cugaaguccu 20

<210> 1106

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1106

cacucacuug aggguuucca 20

<210> 1107

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1107

cagacugcgg ggacacagug 20

<210> 1108

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1108

ccacucaccu uagccugagc 20

<210> 1109

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1109

cacucaccuu agccugagca 20

<210> 1110

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1110

guacaagcug ucggaaacag 20

<210> 1111

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1111

acacucacuc caucacccgg 20

<210> 1112

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1112

cacucacucc aucacccgga 20

<210> 1113

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1113

cguccaguac aacaaguuca 20

<210> 1114

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1114

ccacauccug caaggggga 20

<210> 1115

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1115

cacauccugc aaggggggau 20

<210> 1116

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1116

ugggcgucca cauccugcaa 20

<210> 1117

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1117

gggcguccac auccugcaag 20

<210> 1118

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1118

ggcguccaca uccugcaagg 20

<210> 1119

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1119

gcguccacau ccugcaaggg 20

<210> 1120

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1120

gcccaaggua agagcuuccc 20

<210> 1121

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1121

gcucuuaccu ugggcagcca 20

<210> 1122

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1122

agcucuuacc uugggcagcc 20

<210> 1123

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1123

ugcaccucug gggaggaccu 20

<210> 1124

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1124

cugcaccucu ggggaggacc 20

<210> 1125

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1125

cgccugcagc ugucggacac 20

<210> 1126

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1126

caccugccag gccaucacgg 20

<210> 1127

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1127

cccuaccugu caccaggacc 20

<210> 1128

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1128

ccuaccuguc accaggacca 20

<210> 1129

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1129

ccucugagaa ggaaaaaaga 20

<210> 1130

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1130

cagaggaaca gucccaagga 20

<210> 1131

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1131

cacucaccug cccacagcag 20

<210> 1132

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1132

ccacucaccu gcccacagca 20

<210> 1133

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1133

gccacucacc ugcccacagc 20

<210> 1134

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1134

agggccccug uggggaaacg 20

<210> 1135

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1135

gggccccugu ggggaaacga 20

<210> 1136

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1136

gcaagugcag gagucaguga 20

<210> 1137

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1137

cggaaccagu aaccaugaac 20

<210> 1138

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1138

ggaaccagua accaugaacu 20

<210> 1139

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1139

gaaccaguaa ccaugaacug 20

<210> 1140

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1140

gcuagaucaa gaaguacagg 20

<210> 1141

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1141

agaaguacag gaggagacuc 20

<210> 1142

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1142

caagucuagu gaggagaaag 20

<210> 1143

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1143

aagucuagug aggagaaaga 20

<210> 1144

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1144

agucuaguga ggagaaagag 20

<210> 1145

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1145

accugucagg ugaaguucgc 20

<210> 1146

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1146

acuuacaggu gacguugagc 20

<210> 1147

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1147

aacaucuagg agaggaagag 20

<210> 1148

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1148

guuccacaga acccaacaac 20

<210> 1149

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1149

uuccuaccug agccaucucc 20

<210> 1150

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1150

augcucacau gaagaagaug 20

<210> 1151

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1151

ucacucugau gggagacacc 20

<210> 1152

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1152

cacucugaug ggagacacca 20

<210> 1153

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1153

uuucuuaugg agaggaaaga 20

<210> 1154

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1154

guacagguaa gagcaacgcc 20

<210> 1155

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1155

cucuuaccug uaccauaacc 20

<210> 1156

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1156

uuaccuguac cauaaccagg 20

<210> 1157

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1157

uguaucugua ggaggagaag 20

<210> 1158

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1158

guaucuguag gaggagaagu 20

<210> 1159

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1159

cagauacaaa cuggacucuc 20

<210> 1160

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1160

ucuuaccuuc cuucguuugc 20

<210> 1161

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1161

uuuggaucua aaacggugac 20

<210> 1162

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1162

gaucuaaaac ggugacaggu 20

<210> 1163

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1163

aaaggcucag caguugaccu 20

<210> 1164

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1164

auuuaccggc auagaauagu 20

<210> 1165

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1165

ucacugccua agagagacau 20

<210> 1166

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1166

aggauccacg uggagaaugg 20

<210> 1167

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1167

ggauccacgu ggagaauggu 20

<210> 1168

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1168

ucucacuguu cucagggaag 20

<210> 1169

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1169

ccugcccaag gcuucccacc 20

<210> 1170

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1170

cugcccaagg cuucccaccu 20

<210> 1171

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1171

gccugcugcg guaagcggua 20

<210> 1172

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1172

gaugcucagg gaacacguau 20

<210> 1173

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1173

uucucaaagu ucccacaucc 20

<210> 1174

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1174

ucacagauug gugaguagcc 20

<210> 1175

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1175

cucaccuguu cugugauuac 20

<210> 1176

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1176

uccuuccagc uacucaaucc 20

<210> 1177

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1177

cacaguacaa auaagagccc 20

<210> 1178

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1178

ccccccagcg cuucggugag 20

<210> 1179

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1179

cccccagcgc uucggugagu 20

<210> 1180

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1180

ccacucaccg aagcgcuggg 20

<210> 1181

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1181

uaccucggag gaaagagaaa 20

<210> 1182

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1182

cagcuuccaa aacgacaagc 20

<210> 1183

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1183

aacauaccag cuugucguuu 20

<210> 1184

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1184

agaccaccug cagagacgag 20

<210> 1185

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1185

ccaccugcag agacgagagg 20

<210> 1186

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1186

cccaccagcu cagcuccacg 20

<210> 1187

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1187

cgcugucaag uccaguucua 20

<210> 1188

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1188

gcugucaagu ccaguucuac 20

<210> 1189

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1189

cacccagauc gucagcgccg 20

<210> 1190

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1190

ccacucaccu gcucuacccc 20

<210> 1191

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1191

ccacagucug aaagaaagca 20

<210> 1192

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1192

gacacuguug gcacggagga 20

<210> 1193

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1193

uuaccauggc caucaacaca 20

<210> 1194

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1194

ccacagucug aaagaaaaca 20

<210> 1195

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1195

cacagucuga aagaaaacag 20

<210> 1196

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1196

uuaccauggc caucagcacg 20

<210> 1197

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1197

ucugcguuca gggguaagcg 20

<210> 1198

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1198

gcgcuuaccc cugaacgcag 20

<210> 1199

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1199

gacugggcag cggccccugu 20

<210> 1200

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1200

ggacugggca gcggccccug 20

<210> 1201

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1201

gucaugcagc ccuugccugc 20

<210> 1202

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1202

ucaugcagcc cuugccugcu 20

<210> 1203

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1203

caugcagccc uugccugcug 20

<210> 1204

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1204

cucaccagau ucccgaaggu 20

<210> 1205

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1205

cagacucacc agauucccga 20

<210> 1206

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1206

agccuaggca agugcagagg 20

<210> 1207

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1207

cagccuaggc aagugcagag 20

<210> 1208

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1208

accagccuag gcaagugcag 20

<210> 1209

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1209

uggaacccca auaccagccu 20

<210> 1210

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1210

caccugcaga agaugccaga 20

<210> 1211

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1211

cgcucaccug caccagccuc 20

<210> 1212

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1212

cuuccuggca agacacaaga 20

<210> 1213

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1213

aauucaggga gccauugaaa 20

<210> 1214

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1214

cuacugaccu gccuuuucaa 20

<210> 1215

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1215

gccucucaaa gucuuaugug 20

<210> 1216

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1216

uucccaugcu gcuuuacuuc 20

<210> 1217

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1217

uuuaggucaa ccagauguag 20

<210> 1218

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1218

ugugccugaa agcaaaaaug 20

<210> 1219

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1219

uuaccuacua uucuugccag 20

<210> 1220

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1220

aaaugcuguu uaaaaaaagg 20

<210> 1221

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1221

ccccaaaaag ugaauaucaa 20

<210> 1222

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1222

guccagaauu uccagaagua 20

<210> 1223

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1223

ucccuggagg gguggggggag 20

<210> 1224

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1224

ucacuucaac aggacuccua 20

<210> 1225

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1225

ccuuacacuc aaaggcaaag 20

<210> 1226

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1226

auggcugcag ggagaggaag 20

<210> 1227

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1227

ggagcgccag gggaugcugg 20

<210> 1228

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1228

ucacguccca ugcagccaca 20

<210> 1229

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1229

uggacgucug guggggagua 20

<210> 1230

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1230

acucacauug augaggaaga 20

<210> 1231

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1231

cagcaucuga gaaaggcaga 20

<210> 1232

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1232

aauccaacac cucagcgaug 20

<210> 1233

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1233

auccaacacc ucagcgaugu 20

<210> 1234

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1234

cgcggcgcag uacccgcugc 20

<210> 1235

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1235

cacucaccgu ggugggucgg 20

<210> 1236

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1236

acucaccgug gugggucgga 20

<210> 1237

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1237

uggcucccug uggggcaacg 20

<210> 1238

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1238

gauuccaggg ggaggcggug 20

<210> 1239

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1239

gcuccuaccu gugcuggacc 20

<210> 1240

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1240

ucgccgccac agugggucg 20

<210> 1241

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1241

cagacugaga gugggagag 20

<210> 1242

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1242

ccuccuccag agugugguug 20

<210> 1243

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1243

agccauccca gggagagagc 20

<210> 1244

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1244

gccaucccag ggagagagcu 20

<210> 1245

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1245

ccaucccagg gagagagcug 20

<210> 1246

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1246

cucacaggcc acccaccagc 20

<210> 1247

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1247

ccgcuuccag aagugccugg 20

<210> 1248

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1248

cuuccagaag ugccuggcgg 20

<210> 1249

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1249

acaacugcaa aggaaugggu 20

<210> 1250

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1250

caacugcaaa ggaaugggua 20

<210> 1251

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1251

gaacuaggaa gacgguccag 20

<210> 1252

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1252

ggcugaccag gaccuguugc 20

<210> 1253

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1253

cucaccugua cgccaggcgg 20

<210> 1254

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1254

gcucucaccu guacgccagg 20

<210> 1255

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1255

cuuagaccug gcaggcagau 20

<210> 1256

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1256

caauccaguc cccgaagcca 20

<210> 1257

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1257

aauccagucc ccgaagccac 20

<210> 1258

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1258

acucaccggu gaugaggaca 20

<210> 1259

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1259

cucaccggug augaggacaa 20

<210> 1260

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1260

ccggucugcg ggaaggguac 20

<210> 1261

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1261

ugggcugcag gagccgcggc 20

<210> 1262

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1262

uuguaccaaa ugccccuguc 20

<210> 1263

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1263

cggacagcag uccuccauua 20

<210> 1264

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1264

aggugcagca cagccccaug 20

<210> 1265

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1265

ggugcagcac agccccaugu 20

<210> 1266

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1266

aguugccaga ugcgcuucga 20

<210> 1267

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1267

guugccagau gcgcuucgac 20

<210> 1268

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1268

gucucagcug cucgacacgc 20

<210> 1269

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1269

uucuuacccu ggaauagucc 20

<210> 1270

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1270

auuaccugua ugcuaaucga 20

<210> 1271

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1271

uugcaaugcg uucguggcuu 20

<210> 1272

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1272

acugaccugu gaccauagcc 20

<210> 1273

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1273

uaucugcagg gacagagaaa 20

<210> 1274

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1274

ugcggcgcag acauacagcu 20

<210> 1275

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1275

ccccgcaggc gggggcguua 20

<210> 1276

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1276

uuucagaagu gucucagugu 20

<210> 1277

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1277

auuaccugau ggaaagucug 20

<210> 1278

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1278

cuucagugcc uucguggauu 20

<210> 1279

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1279

uuucugcaga gggauagaga 20

<210> 1280

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1280

agaccaccag aguaagggac 20

<210> 1281

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1281

acuugcagcc cucggccaaa 20

<210> 1282

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1282

agggcucacc agagguagga 20

<210> 1283

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1283

uucaccugcc caaggaaaaa 20

<210> 1284

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1284

cuaagccuag aaaaucaguu 20

<210> 1285

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1285

acaucuagaa aaaaaaauac 20

<210> 1286

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1286

auuaccuucc ucuuugcacu 20

<210> 1287

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1287

gauccuguag guuggaacau 20

<210> 1288

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1288

aagccacccc aaguuagauc 20

<210> 1289

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1289

aacuuacccc aaacaauuag 20

<210> 1290

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1290

auaccuacag gauuuaaagu 20

<210> 1291

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1291

guuuccuaga aagcaaaaaa 20

<210> 1292

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1292

aaguucaacu gcuucguaau 20

<210> 1293

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1293

aauucagacu aucauccuca 20

<210> 1294

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1294

gcuugcagcu gugggacacc 20

<210> 1295

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1295

gaccacucaa ccagugccca 20

<210> 1296

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1296

ggaggcacag accacuccac 20

<210> 1297

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1297

ggcacagaca acucgacuga 20

<210> 1298

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1298

gcacagacca cucaacccac 20

<210> 1299

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1299

gaccacucaa cccacaggcc 20

<210> 1300

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1300

gaccacucaa accacagcca 20

<210> 1301

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1301

gaccacucaa cccacagcca 20

<210> 1302

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1302

gaccacucaa ccagcagcca 20

<210> 1303

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown: target sequence

<400> 1303

cctgcagaga aacacacttg 20

<210> 1304

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400

> 1304

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Val Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Glu Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 1305

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 1305

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 1306

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 1306

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Arg

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Glu Tyr Arg Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 1307

<211> 1359

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1307

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705 710 715 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095

Ile Gln Thr Val Gly Gln Asn Gly Gly Leu Phe Asp Asp Asn Pro
1100 1105 1110

Lys Ser Pro Leu Glu Val Thr Pro Ser Lys Leu Val Pro Leu Lys
1115 1120 1125

Lys Glu Leu Asn Pro Lys Lys Tyr Gly Gly Tyr Gln Lys Pro Thr
1130 1135 1140

Thr Ala Tyr Pro Val Leu Leu Ile Thr Asp Thr Lys Gln Leu Ile
1145 1150 1155

Pro Ile Ser Val Met Asn Lys Lys Gln Phe Glu Gln Asn Pro Val
1160 1165 1170

Lys Phe Leu Arg Asp Arg Gly Tyr Gln Gln Val Gly Lys Asn Asp
1175 1180 1185

Phe Ile Lys Leu Pro Lys Tyr Thr Leu Val Asp Ile Gly Asp Gly
1190 1195 1200

Ile Lys Arg Leu Trp Ala Ser Ser Lys Glu Ile His Lys Gly Asn
1205 1210 1215

Gln Leu Val Val Ser Lys Lys Ser Gln Ile Leu Leu Tyr His Ala

1220 1225 1230

His His Leu Asp Ser Asp Leu Ser Asn Asp Tyr Leu Gln Asn His

1235 1240 1245

Asn Gln Gln Phe Asp Val Leu Phe Asn Glu Ile Ile Ser Phe Ser

1250 1255 1260

Lys Lys Cys Lys Leu Gly Lys Glu His Ile Gln Lys Ile Glu Asn

1265 1270 1275

Val Tyr Ser Asn Lys Lys Asn Ser Ala Ser Ile Glu Glu Leu Ala

1280 1285 1290

Glu Ser Phe Ile Lys Leu Leu Gly Phe Thr Gln Leu Gly Ala Thr

1295 1300 1305

Ser Pro Phe Asn Phe Leu Gly Val Lys Leu Asn Gln Lys Gln Tyr

1310 1315 1320

Lys Gly Lys Lys Asp Tyr Ile Leu Pro Cys Thr Glu Gly Thr Leu

1325 1330 1335

Ile Arg Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Val Asp Leu

1340 1345 1350

Ser Lys Ile Gly Glu Asp

1355

<210> 1308

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 1308

Gly Gly Gly Ser

1

<210> 1309

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<

400> 1309

Glu Ala Ala Ala Lys

1 5

<210> 1310

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 1310

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1 5 10 15

Glu Ser Ser Gly

20

<210> 1311

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1311

ggaggctctg gaggaagc 18

<210> 1312

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1312

ggctcttctg gatctgaaac acctggcaca agcgagagcg ccacccctga gagctctggc 60

<210> 1313

<211> 17

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nuclear localization signal sequence

<400> 1313

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala

1 5 10 15

Ala

<210> 1314

<211> 51

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

nuclear localization signal sequence

<400> 1314

atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc c 51

<210> 1315

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 1315

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Gly Ser Ser Gly Ser Glu Thr Pro

1010 1015 1020

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val

1025 1030 1035

Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

1040 1045 1050

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
1055 1060 1065
Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
1070 1075 1080
Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

1085 1090 1095
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr
1100 1105 1110
Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met
1115 1120 1125
Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala
1130 1135 1140
Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro

1145 1150 1155
Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
1160 1165 1170
Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln
1175 1180 1185
Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Tyr Asp
1190 1195 1200
Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1205 1210 1215
Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1220 1225 1230
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260
Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 1316

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1316

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Gly
1010 1015 1020

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1025 1030 1035
Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met
1040 1045 1050
Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu
1055 1060 1065
Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
1070 1075 1080

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
1085 1090 1095
Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn
1100 1105 1110
Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys
1115 1120 1125
Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

1130 1135 1140

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
1145 1150 1155
Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile
1160 1165 1170
Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr
1175 1180 1185
Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala
1190 1195 1200

Gln Ser Ser Thr Asp Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala
1205 1210 1215
Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1220 1225 1230
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550
<210> 1317
<211> 1554
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 1317

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Ser Ser Gly Ser Glu Thr Pro Gly

1025 1030 1035

Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu

1040 1045 1050

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys

1055 1060 1065

Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val

1070 1075 1080

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly

1085 1090 1095

Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

1100 1105 1110

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

1115 1120 1125

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile

1130 1135 1140

His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys

1145 1150 1155

Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly

1160 1165 1170

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu

1175 1180 1185

Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val

1190 1195 1200

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1220 1225 1230
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260
Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1318

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1318

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

1040 1045 1050

Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu

1055 1060 1065

```
Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala Lys Arg  Ala Arg Asp
    1070                1075                1080
Glu Arg  Glu Val Pro Val Gly  Ala Val Leu Val Leu  Asn Asn Arg
    1085                1090                1095
Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile Gly Leu  His Asp Pro
    1100                1105                1110
Thr Ala  His Ala Glu Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val
    1115                1120                1125

Met Gln  Asn Tyr Arg Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe
    1130                1135                1140
Glu Pro  Cys Val Met Cys Ala  Gly Ala Met Ile His  Ser Arg Ile
    1145                1150                1155
Gly Arg  Val Val Phe Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala
    1160                1165                1170
Gly Ser  Leu Met Asp Val Leu  His Tyr Pro Gly Met  Asn His Arg
    1175                1180                1185

Val Glu  Ile Thr Glu Gly Ile  Leu Ala Asp Glu Cys  Ala Ala Leu
    1190                1195                1200
Leu Cys  Tyr Phe Phe Arg Met  Pro Arg Gln Val Phe  Asn Ala Gln
    1205                1210                1215
Lys Lys  Ala Gln Ser Ser Thr  Asp Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230
Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260
Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275
Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290
Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
```

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 1319

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1319

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

1070 1075 1080

Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp

1085 1090 1095

Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg

1100 1105 1110

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile

1115 1120 1125

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala

1130 1135 1140

```
His Ala  Glu Ile Met Ala Leu  Arg Gln Gly Gly Leu  Val Met Gln
    1145                1150                1155
Asn Tyr  Arg Leu Ile Asp Ala  Thr Leu Tyr Val Thr  Phe Glu Pro
    1160                1165                1170
Cys Val  Met Cys Ala Gly Ala  Met Ile His Ser Arg  Ile Gly Arg

    1175                1180                1185
Val Val  Phe Gly Val Arg Asn  Ala Lys Thr Gly Ala  Ala Gly Ser
    1190                1195                1200
Leu Met  Asp Val Leu His Tyr  Pro Gly Met Asn His  Arg Val Glu
    1205                1210                1215
Ile Thr  Glu Gly Ile Leu Ala  Asp Glu Cys Ala Ala  Leu Leu Cys
    1220                1225                1230
Tyr Phe  Phe Arg Met Pro Arg  Gln Val Phe Asn Ala  Gln Lys Lys

    1235                1240                1245
Ala Gln  Ser Ser Thr Asp Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260
Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275
Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290
Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg

    1295                1300                1305
Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320
Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335
Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350
Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu

    1355                1360                1365
Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
```

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1320

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1320

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Gly
1235 1240 1245
Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1250 1255 1260
Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met
1265 1270 1275
Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu
1280 1285 1290

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
1295 1300 1305
Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
1310 1315 1320
Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn
1325 1330 1335
Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys
1340 1345 1350

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
1355 1360 1365
Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
1370 1375 1380
Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile
1385 1390 1395
Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr
1400 1405 1410

Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala
1415 1420 1425
Gln Ser Ser Thr Asp Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550
<210> 1321
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1321
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala

1055 1060 1065

Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr

1070 1075 1080

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

1085 1090 1095

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val

1100 1105 1110

Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr

1115 1120 1125

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

1130 1135 1140

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

1145 1150 1155

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly

1160 1165 1170

Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly

1175 1180 1185

Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val

1190 1195 1200

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu

1205 1210 1215

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

1220 1225 1230

Lys Ala Gln Ser Ser Thr Asp Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 1322

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1322

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
1025 1030 1035

Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His
1040 1045 1050

Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg
1055 1060 1065

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn
1070 1075 1080

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp
1085 1090 1095

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu
1100 1105 1110

Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr
1115 1120 1125

Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg
1130 1135 1140

Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala

1145 1150 1155

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
1160 1165 1170
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala
1175 1180 1185
Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala
1190 1195 1200
Gln Lys Lys Ala Gln Ser Ser Thr Asp Gln Glu Ile Gly Lys Ala
1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1220 1225 1230
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260
Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550
<210> 1323
<211> 1540
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1323

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

770 775 780

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg

785 790 795 800

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

805 810 815

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp

820 825 830

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met

835 840 845

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

850 855 860

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

865 870 875 880

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

885 890 895

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly

900 905 910

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys

915 920 925

Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Thr Thr Gln Lys

930 935 940

Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile

945 950 955 960

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr

965 970 975

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg

980 985 990

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr

995 1000 1005

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser

1010 1015 1020

Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys

1025 1030 1035

Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

1040 1045 1050

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

1055 1060 1065

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu

1070 1075 1080

Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln

1085 1090 1095

Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr

1100 1105 1110

Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile

1115 1120 1125

Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln

1130 1135 1140

Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp

1145 1150 1155

Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr

1160 1165 1170

Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr

1175 1180 1185

Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys

1190 1195 1200

Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe

1205 1210 1215

Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro

1220 1225 1230

Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys

1235 1240 1245

Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln

1250 1255 1260

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser

1265 1270 1275

Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala

1280 1285 1290

Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

1295 1300 1305

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys

1310 1315 1320

Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile

1325 1330 1335

Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe

1340 1345 1350

Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile

1355 1360 1365

Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys

1370 1375 1380

Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu

1385 1390 1395

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His

1400 1405 1410

Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln

1415 1420 1425

Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu

1430 1435 1440

Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn

1445 1450 1455

Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro
1460 1465 1470
Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr
1475 1480 1485
Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
1490 1495 1500
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr

1505 1510 1515
Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1520 1525 1530
Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 1324

<211> 1540

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1324

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
1250 1255 1260
Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr
1265 1270 1275

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu
1280 1285 1290
Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val
1295 1300 1305
Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr
1310 1315 1320

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
1325 1330 1335

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu
1340 1345 1350
Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly
1355 1360 1365
Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly
1370 1375 1380
Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val
1385 1390 1395

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
1400 1405 1410
Cys Tyr Phe Phe Arg Met Pro Arg Glu Asp Asn Glu Gln Lys Gln
1415 1420 1425
Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu
1430 1435 1440
Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn
1445 1450 1455

Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro
1460 1465 1470
Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr
1475 1480 1485
Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
1490 1495 1500
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr
1505 1510 1515

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1520 1525 1530
Leu Ser Gln Leu Gly Gly Asp
1535 1540
<210> 1325

<211> 1543

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1325

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser
1250 1255 1260
Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser
1265 1270 1275
His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala

1280 1285 1290
Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn
1295 1300 1305
Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His
1310 1315 1320
Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly
1325 1330 1335
Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val

1340 1345 1350
Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
1355 1360 1365
Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
1370 1375 1380
Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn
1385 1390 1395
His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala

1400 1405 1410
Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Glu Asp Asn Glu

1415 1420 1425

Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu

1430 1435 1440

Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala

1445 1450 1455

Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg

1460 1465 1470

Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe

1475 1480 1485

Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

1490 1495 1500

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu

1505 1510 1515

Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr

1520 1525 1530

Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 1326

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1326

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser
1250 1255 1260

Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu Val
1265 1270 1275

Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

1280 1285 1290

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
1295 1300 1305
Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
1310 1315 1320
Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
1325 1330 1335
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr
1340 1345 1350

Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met
1355 1360 1365
Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala
1370 1375 1380
Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro
1385 1390 1395
Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
1400 1405 1410

Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Glu
1415 1420 1425
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1430 1435 1440
Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1445 1450 1455
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1460 1465 1470

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1475 1480 1485
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1490 1495 1500
Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1505 1510 1515

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540 1545

<210> 1327

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1327

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser
1250 1255 1260
Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu Val
1265 1270 1275
Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala
1280 1285 1290

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
1295 1300 1305
Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
1310 1315 1320
Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
1325 1330 1335
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr
1340 1345 1350

Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met
1355 1360 1365
Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

1370 1375 1380

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro

1385 1390 1395

Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp

1400 1405 1410

Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg Arg Glu

1415 1420 1425

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1445 1450 1455

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1460 1465 1470

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1475 1480 1485

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1505 1510 1515

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540 1545

<210> 1328

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1328

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

```
        755                 760                 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005
```

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu

1250 1255 1260

Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu Val Glu

1265 1270 1275

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys

1280 1285 1290

Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val

1295 1300 1305

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly

1310 1315 1320

Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

1325 1330 1335

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

1340 1345 1350

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile

1355 1360 1365

His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys

1370 1375 1380

Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly

1385 1390 1395

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu

1400 1405 1410

Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg Arg Pro Glu

1415 1420 1425

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1445 1450 1455

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1460 1465 1470

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1475 1480 1485
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1490 1495 1500
Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1505 1510 1515
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540 1545

<210> 1329

<211> 1551

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1329

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser
1250 1255 1260
Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Gly Ser Glu Val Glu
1265 1270 1275
Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys
1280 1285 1290

Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val
1295 1300 1305
Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala His Ala
1310 1315 1320
Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
1325 1330 1335

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val
1340 1345 1350

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val
1355 1360 1365
Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
1370 1375 1380
Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr
1385 1390 1395
Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe
1400 1405 1410

Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln
1415 1420 1425
Ser Ser Thr Asp Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1430 1435 1440
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1445 1450 1455
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1460 1465 1470

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1475 1480 1485
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1490 1495 1500
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1505 1510 1515
Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1520 1525 1530

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1535 1540 1545
Gly Gly Asp
1550
<210> 1330

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1330

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

```
    450                 455                 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465                 470                 475                 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

    530                 535                 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

        595                 600                 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

            660                 665                 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
```

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1250 1255 1260

Ser Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu

1265 1270 1275

Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu

1280 1285 1290

Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val

1295 1300 1305

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

1310 1315 1320

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu

1325 1330 1335

Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala

1340 1345 1350

Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

1355 1360 1365

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn

1370 1375 1380

Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr

1385 1390 1395

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

1400 1405 1410

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg Arg

1415 1420 1425

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1445 1450 1455

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1460 1465 1470

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1475 1480 1485

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1505 1510 1515

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540 1545

<210> 1331

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1331

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly

1250 1255 1260

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser

1265 1270 1275

Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1280 1285 1290

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

1295 1300 1305

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg

1310 1315 1320

Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala

1325 1330 1335

Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

1340 1345 1350

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly

1355 1360 1365

Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg

1370 1375 1380

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

1385 1390 1395

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

1400 1405 1410

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg

1415 1420 1425

Arg Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1445 1450 1455

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1460 1465 1470

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1475 1480 1485

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1505 1510 1515

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540 1545

<210> 1332

<211> 1533

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1332

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala

1250 1255 1260

Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr

1265 1270 1275

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

1280 1285 1290

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val

1295 1300 1305

Ile Gly Glu Gly Trp Asn Arg Ala His Ala Glu Ile Met Ala Leu

1310 1315 1320

Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala

1325 1330 1335

Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

1340 1345 1350

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn

1355 1360 1365

Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr

1370 1375 1380

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

1385 1390 1395

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg

1400 1405 1410

Gln Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1415 1420 1425

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1430 1435 1440

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1445 1450 1455

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1460 1465 1470

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1475 1480 1485

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1490 1495 1500

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1505 1510 1515

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1520 1525 1530

<210> 1333

<211> 1549

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1333

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

995 1000 1005

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

1010 1015 1020

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg

1025 1030 1035

Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala

1040 1045 1050

Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

1055 1060 1065

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly

1070 1075 1080

Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg

1085 1090 1095

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

1100 1105 1110

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

1115 1120 1125

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro

1130 1135 1140

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp

1145 1150 1155

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

1160 1165 1170

Pro Glu Ser Ser Gly Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu

1175 1180 1185

Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile

1190 1195 1200

Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe

1205 1210 1215

Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu

1220 1225 1230

Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly

1235 1240 1245

Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr

1250 1255 1260

Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys

1265 1270 1275

Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro

1280 1285 1290

Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp

1295 1300 1305

Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser

1310 1315 1320

Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu

1325 1330 1335

Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser

1340 1345 1350

Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr

1355 1360 1365

Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser

1370 1375 1380

Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala

1385 1390 1395

Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr

1400 1405 1410

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly

1415 1420 1425

Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His

1430 1435 1440

Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser

1445 1450 1455

Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser

1460 1465 1470

Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu

1475 1480 1485

Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala
1490 1495 1500
Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr
1505 1510 1515
Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile
1520 1525 1530
Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly

1535 1540 1545
Asp

<210> 1334
<211> 1555
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1334
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

995 1000 1005

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

1010 1015 1020

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg

1025 1030 1035

Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala

1040 1045 1050

Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

1055 1060 1065

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly

1070 1075 1080

Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg

1085 1090 1095

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

1100 1105 1110

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

1115 1120 1125

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro

1130 1135 1140

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp

1145 1150 1155

Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu

1160 1165 1170

Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Ile Lys Lys Tyr

1175 1180 1185

Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr

1190 1195 1200

Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys

1205 1210 1215

Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe

1220 1225 1230

Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro

1235 1240 1245

Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys

1250 1255 1260

Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln

1265 1270 1275

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser

1280 1285 1290

Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala

1295 1300 1305

Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

1310 1315 1320

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys

1325 1330 1335

Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile

1340 1345 1350

Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe

1355 1360 1365

Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile

1370 1375 1380

Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys

1385 1390 1395

Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu

1400 1405 1410

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His

1415 1420 1425

Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln

1430 1435 1440

Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu

1445 1450 1455

Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn

1460 1465 1470

Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro

1475 1480 1485

Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr

1490 1495 1500

Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile

1505 1510 1515

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr

1520 1525 1530

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp

1535 1540 1545

Leu Ser Gln Leu Gly Gly Asp

1550 1555

<210> 1335

<211> 1421

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1335

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu
785 790 795 800

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu
805 810 815
Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile
820 825 830
Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
835 840 845
Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
850 855 860

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met
865 870 875 880
Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
885 890 895
Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
900 905 910
His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
915 920 925

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
930 935 940
Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Gly

945 950 955 960

Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu

965 970 975

Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met

980 985 990

Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val

995 1000 1005

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1010 1015 1020

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1025 1030 1035

Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys

1040 1045 1050

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1055 1060 1065

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1070 1075 1080

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1085 1090 1095

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1100 1105 1110

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1115 1120 1125

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1130 1135 1140

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1145 1150 1155

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1160 1165 1170

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1175 1180 1185

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu
1190 1195 1200
Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
1205 1210 1215
Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp
1220 1225 1230

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile
1235 1240 1245
Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
1250 1255 1260
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu
1265 1270 1275
Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser
1280 1285 1290

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys
1295 1300 1305
Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
1310 1315 1320
Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
1325 1330 1335
Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys
1340 1345 1350

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1355 1360 1365
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1370 1375 1380
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1385 1390 1395
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1400 1405 1410

Asp Leu Ser Gln Leu Gly Gly Asp

1415 1420

<210> 1336

<211> 1552

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1336

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45
Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Tyr Phe Phe Arg Met Pro Arg Ser Gly Gly Ser Ser Gly Gly
145 150 155 160
Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu

165 170 175
Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly
180 185 190

Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu
195 200 205
Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg
210 215 220
His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly

225 230 235 240
Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr
245 250 255
Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn
260 265 270
Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser
275 280 285
Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly

290 295 300
Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr
305 310 315 320
His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg
325 330 335
Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe
340 345 350
Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu

355 360 365
Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro
370 375 380
Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu
385 390 395 400
Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu
405 410 415
Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu

420 425 430
Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu

435 440 445

Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala

450 455 460

Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu

465 470 475 480

Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile

485 490 495

Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His

500 505 510

His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro

515 520 525

Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala

530 535 540

Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile

545 550 555 560

Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys

565 570 575

Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly

580 585 590

Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg

595 600 605

Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile

610 615 620

Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala

625 630 635 640

Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr

645 650 655

Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala

660 665 670

Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn

675 680 685

Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val
690 695 700
Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys
705 710 715 720
Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu
725 730 735
Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr

740 745 750
Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu
755 760 765
Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile
770 775 780
Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu
785 790 795 800
Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile

805 810 815
Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met
820 825 830
Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg
835 840 845
Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu
850 855 860
Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu

865 870 875 880
Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln
885 890 895
Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala
900 905 910
Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val
915 920 925
Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val

930 935 940

Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn

945 950 955 960

Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly

965 970 975

Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn

980 985 990

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

995 1000 1005

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1010 1015 1020

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1025 1030 1035

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1040 1045 1050

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1055 1060 1065

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1070 1075 1080

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1085 1090 1095

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1100 1105 1110

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1115 1120 1125

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1130 1135 1140

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1145 1150 1155

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1160 1165 1170

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1175 1180 1185

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1190 1195 1200

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1205 1210 1215

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1220 1225 1230

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1235 1240 1245

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1250 1255 1260

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1265 1270 1275

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1280 1285 1290

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1295 1300 1305

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1310 1315 1320

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1325 1330 1335

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1340 1345 1350

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1355 1360 1365

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1370 1375 1380

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1385 1390 1395

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1400 1405 1410

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1415 1420 1425

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1430 1435 1440

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1445 1450 1455

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1460 1465 1470

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1475 1480 1485

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1490 1495 1500

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1505 1510 1515

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1520 1525 1530

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1535 1540 1545

Leu Gly Gly Asp

1550

<210> 1337

<211> 1545

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1337

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Thr
35 40 45
Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln
50 55 60
Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys
65 70 75 80
Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val
85 90 95

Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp
100 105 110
Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly
115 120 125
Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met
130 135 140
Pro Arg Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro
145 150 155 160

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly
165 170 175
Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser
180 185 190
Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys
195 200 205
Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu
210 215 220

Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg
225 230 235 240
Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile
245 250 255
Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp
260 265 270

Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys
275 280 285

Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala
290 295 300
Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val
305 310 315 320
Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala
325 330 335
His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn
340 345 350

Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr
355 360 365
Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp
370 375 380
Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu
385 390 395 400
Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly
405 410 415

Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn
420 425 430
Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr
435 440 445
Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala
450 455 460
Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser
465 470 475 480

Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala
485 490 495
Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu
500 505 510
Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe

515 520 525

Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala

530 535 540

Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met

545 550 555 560

Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu

565 570 575

Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His

580 585 590

Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro

595 600 605

Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg

610 615 620

Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala

625 630 635 640

Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu

645 650 655

Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met

660 665 670

Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His

675 680 685

Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val

690 695 700

Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu

705 710 715 720

Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val

725 730 735

Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe

740 745 750

Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu

755 760 765

Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu
770 775 780
Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu
785 790 795 800

Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr
805 810 815
Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg
820 825 830
Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg
835 840 845
Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly
850 855 860

Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr
865 870 875 880
Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser
885 890 895
Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys
900 905 910
Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met
915 920 925

Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn
930 935 940
Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg
945 950 955 960
Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His
965 970 975
Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr
980 985 990

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn
995 1000 1005
Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe

1010 1015 1020

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp

1025 1030 1035

Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val

1040 1045 1050

Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu

1055 1060 1065

Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly

1070 1075 1080

Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu

1085 1090 1095

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp

1100 1105 1110

Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg

1115 1120 1125

Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe

1130 1135 1140

Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr

1145 1150 1155

His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala

1160 1165 1170

Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly

1175 1180 1185

Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu

1190 1195 1200

Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn

1205 1210 1215

Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu

1220 1225 1230

Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu

1235 1240 1245

```
Ile Val  Trp Asp Lys Gly Arg  Asp Phe Ala Thr Val  Arg Lys Val
    1250                 1255                 1260
Leu Ser  Met Pro Gln Val Asn  Ile Val Lys Lys Thr  Glu Val Gln
    1265                 1270                 1275
Thr Gly  Gly Phe Ser Lys Glu  Ser Ile Leu Pro Lys  Arg Asn Ser
    1280                 1285                 1290

Asp Lys  Leu Ile Ala Arg Lys  Lys Asp Trp Asp Pro  Lys Lys Tyr
    1295                 1300                 1305
Gly Gly  Phe Asp Ser Pro Thr  Val Ala Tyr Ser Val  Leu Val Val
    1310                 1315                 1320
Ala Lys  Val Glu Lys Gly Lys  Ser Lys Lys Leu Lys  Ser Val Lys
    1325                 1330                 1335
Glu Leu  Leu Gly Ile Thr Ile  Met Glu Arg Ser Ser  Phe Glu Lys
    1340                 1345                 1350

Asn Pro  Ile Asp Phe Leu Glu  Ala Lys Gly Tyr Lys  Glu Val Lys
    1355                 1360                 1365
Lys Asp  Leu Ile Ile Lys Leu  Pro Lys Tyr Ser Leu  Phe Glu Leu
    1370                 1375                 1380
Glu Asn  Gly Arg Lys Arg Met  Leu Ala Ser Ala Gly  Glu Leu Gln
    1385                 1390                 1395
Lys Gly  Asn Glu Leu Ala Leu  Pro Ser Lys Tyr Val  Asn Phe Leu
    1400                 1405                 1410

Tyr Leu  Ala Ser His Tyr Glu  Lys Leu Lys Gly Ser  Pro Glu Asp
    1415                 1420                 1425
Asn Glu  Gln Lys Gln Leu Phe  Val Glu Gln His Lys  His Tyr Leu
    1430                 1435                 1440
Asp Glu  Ile Ile Glu Gln Ile  Ser Glu Phe Ser Lys  Arg Val Ile
    1445                 1450                 1455
Leu Ala  Asp Ala Asn Leu Asp  Lys Val Leu Ser Ala  Tyr Asn Lys
    1460                 1465                 1470

His Arg  Asp Lys Pro Ile Arg  Glu Gln Ala Glu Asn  Ile Ile His
```

1475 1480 1485

Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr

1490 1495 1500

Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu

1505 1510 1515

Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr

1520 1525 1530

Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540 1545

<210> 1338

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1338

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Gly Ser Ser Gly Ser

995 1000 1005

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1010 1015 1020

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1025 1030 1035

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly

1040 1045 1050

Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

1055 1060 1065

Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met

1070 1075 1080

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile

1085 1090 1095

Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala

1100 1105 1110

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
1115 1120 1125
Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
1130 1135 1140
His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile
1145 1150 1155
Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met
1160 1165 1170

Pro Arg Gln Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185
Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly
1190 1195 1200
Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe
1205 1210 1215
Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg
1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
1235 1240 1245
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro
1250 1255 1260
Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
1265 1270 1275
Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp
1295 1300 1305
Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu
1310 1315 1320
Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
1325 1330 1335
Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile
1355 1360 1365
Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
1370 1375 1380
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu
1385 1390 1395
Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser
1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys
1415 1420 1425
Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
1430 1435 1440
Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
1445 1450 1455
Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys
1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1475 1480 1485
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1490 1495 1500
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1505 1510 1515
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540
<210> 1339
<211> 1541
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1339

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Gly Ser Ser Gly

995 1000 1005

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

1010 1015 1020

Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala

1025 1030 1035

Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val

1040 1045 1050

Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp

1055 1060 1065

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile

1070 1075 1080

Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu

1085 1090 1095

Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys

1100 1105 1110

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly

1115 1120 1125

Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val

1130 1135 1140

Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly

1145 1150 1155

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg

1160 1165 1170

Met Pro Arg Gln Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 1340

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1340

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Gly Ser Ser

995 1000 1005

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1010 1015 1020

Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

1025 1030 1035

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

1040 1045 1050

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly

1055 1060 1065

Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu
1070 1075 1080
Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
1085 1090 1095
Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met
1100 1105 1110
Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe
1115 1120 1125

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp
1130 1135 1140
Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
1145 1150 1155
Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe
1160 1165 1170
Arg Met Pro Arg Gln Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly
1190 1195 1200
Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe
1205 1210 1215
Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg
1220 1225 1230
Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro
1250 1255 1260
Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
1265 1270 1275
Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1280 1285 1290
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu
1310 1315 1320
Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
1325 1330 1335
Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp
1340 1345 1350
Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile
1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
1370 1375 1380
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu
1385 1390 1395
Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser
1400 1405 1410
His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys
1415 1420 1425

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
1430 1435 1440
Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
1445 1450 1455
Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys
1460 1465 1470
Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1490 1495 1500
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1505 1510 1515
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 1341

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1341

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Gly Ser

995 1000 1005

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu

1010 1015 1020

Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg

1025 1030 1035

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val

1040 1045 1050

Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu

1055 1060 1065

Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala

1070 1075 1080

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr

1085 1090 1095

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val

1100 1105 1110

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val

1115 1120 1125

Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met

1130 1135 1140

Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr

1145 1150 1155

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe
1160 1165 1170
Phe Arg Met Pro Arg Gln Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185
Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200
Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe
1205 1210 1215
Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg
1220 1225 1230
Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
1235 1240 1245
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260
Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
1265 1270 1275
Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1280 1285 1290
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp
1295 1300 1305
Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320
Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
1325 1330 1335
Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp
1340 1345 1350
Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile
1355 1360 1365
Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 1342

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 1342

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Gly Ser Ser Gly Ser Glu Thr Pro

785 790 795 800

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu

805 810 815

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg

820 825 830

Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn

835 840 845

Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp

850 855 860

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val

865 870 875 880

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

885 890 895

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg

900 905 910

Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu

915 920 925

Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr

930 935 940

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe

945 950 955 960

Arg Met Pro Arg Gln Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn

965 970 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly

980 985 990

Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp

995 1000 1005

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp

1010 1015 1020

Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
1025 1030 1035
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
1040 1045 1050
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg
1055 1060 1065
Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
1070 1075 1080

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
1085 1090 1095
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
1100 1105 1110
Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val
1115 1120 1125
Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe
1130 1135 1140

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His
1145 1150 1155
Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys
1160 1165 1170
Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185
Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly
1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe
1205 1210 1215
Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg
1220 1225 1230
Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
1235 1240 1245
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
1265 1270 1275
Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1280 1285 1290
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp
1295 1300 1305
Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu
1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
1325 1330 1335
Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp
1340 1345 1350
Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile
1355 1360 1365
Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu
1385 1390 1395
Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser
1400 1405 1410
His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys
1415 1420 1425
Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
1445 1450 1455
Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys
1460 1465 1470
Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1505 1510 1515
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1520 1525 1530
Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 1343

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1343

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

```
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

    130                 135                 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

        195                 200                 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

            260                 265                 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

                325                 330                 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
```

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Ser Gly Ser Glu Thr Pro Gly
785 790 795 800
Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe
805 810 815
Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala
820 825 830
Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

835 840 845
Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
850 855 860
Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

865 870 875 880

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro

885 890 895

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

900 905 910

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met

915 920 925

Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

930 935 940

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg

945 950 955 960

Met Pro Arg Gln Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn

965 970 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly

980 985 990

Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp

995 1000 1005

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp

1010 1015 1020

Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

1025 1030 1035

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

1040 1045 1050

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg

1055 1060 1065

Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu

1070 1075 1080

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg

1085 1090 1095

Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn

1100 1105 1110

Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val

1115 1120 1125

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1130 1135 1140

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1145 1150 1155

Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys

1160 1165 1170

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 1344

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 1344

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
```

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

785 790 795 800

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser

805 810 815

His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg

820 825 830

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg

835 840 845

Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr

850 855 860

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln

865 870 875 880

Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys

885 890 895

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val

900 905 910

Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

915 920 925

Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly

930 935 940

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met

945 950 955 960

Pro Arg Gln Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn

965 970 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
980 985 990

Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp
995 1000 1005

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp
1010 1015 1020

Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
1025 1030 1035

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
1040 1045 1050

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg
1055 1060 1065

Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
1070 1075 1080

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
1085 1090 1095

Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
1100 1105 1110

Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val
1115 1120 1125

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe
1130 1135 1140

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His
1145 1150 1155

Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys
1160 1165 1170

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly
1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg
1220 1225 1230
Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
1235 1240 1245
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro
1250 1255 1260
Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1280 1285 1290
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp
1295 1300 1305
Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu
1310 1315 1320
Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp
1340 1345 1350
Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile
1355 1360 1365
Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
1370 1375 1380
Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu
1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser
1400 1405 1410
His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys
1415 1420 1425
Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys
1460 1465 1470
Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1475 1480 1485
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1490 1495 1500
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1520 1525 1530
Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 1345
<211> 1484
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1345
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Ser Ser Gly Ser Glu Val Glu

1010 1015 1020

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys

1025 1030 1035

Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val

1040 1045 1050

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly

1055 1060 1065

Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

1070 1075 1080

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

1085 1090 1095

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile

1100 1105 1110

His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys

1115 1120 1125

Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly

1130 1135 1140

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu

1145 1150 1155

Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val

1160 1165 1170

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Glu Ile

1175 1180 1185

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu

1190 1195 1200

Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr

1205 1210 1215

Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp

1220 1225 1230

Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly

1235 1240 1245

Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala

1250 1255 1260

Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu

1265 1270 1275

Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn

1280 1285 1290

Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys

1295 1300 1305

Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu

1310 1315 1320

Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys

1325 1330 1335

Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr

1340 1345 1350

Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn

1355 1360 1365

Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp

1370 1375 1380

Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu

1385 1390 1395

Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His

1400 1405 1410

Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

1415 1420 1425

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe

1430 1435 1440

Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val

1445 1450 1455

Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu

1460 1465 1470

Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1475 1480

<210> 1346

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1346

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Met Asn

1055 1060 1065

His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala

1070 1075 1080

Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn

1085 1090 1095

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser

1100 1105 1110

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1115 1120 1125

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1130 1135 1140

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly

1145 1150 1155

Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

1160 1165 1170

Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met

1175 1180 1185

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile

1190 1195 1200

Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala

1205 1210 1215

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val

1220 1225 1230

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu

1235 1240 1245

His Tyr Pro Gly Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1347

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1347

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Gly Ser Ser
690 695 700

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
705 710 715 720

Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
725 730 735

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
740 745 750

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
755 760 765

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
770 775 780

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
785 790 795 800

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
805 810 815

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

820 825 830

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

835 840 845

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

850 855 860

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

865 870 875 880

Lys Ala Gln Ser Ser Thr Asp Leu Thr Phe Lys Glu Asp Ile Gln Lys

885 890 895

Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn

900 905 910

Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys

915 920 925

Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn

930 935 940

Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln

945 950 955 960

Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu

965 970 975

Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu

980 985 990

Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met

995 1000 1005

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1010 1015 1020

Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

1025 1030 1035

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser

1040 1045 1050

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr

1055 1060 1065

Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
1070 1075 1080
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
1085 1090 1095

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile
1100 1105 1110
Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys
1115 1120 1125
Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr
1130 1135 1140
Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe
1145 1150 1155

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala
1160 1165 1170
Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro
1175 1180 1185
Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp
1190 1195 1200
Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala
1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1220 1225 1230
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260
Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 1348

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1348

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu

1250 1255 1260

Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val

1265 1270 1275

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser

1280 1285 1290

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1295 1300 1305

Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

1310 1315 1320

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

1325 1330 1335

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly

1340 1345 1350

Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu

1355 1360 1365

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

1370 1375 1380

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

1385 1390 1395

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

1400 1405 1410

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

1415 1420 1425

Val Leu His Tyr Pro Gly Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1349

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1349

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
```

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

```
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035
Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050
Met Asn  His Arg Val Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu
    1055                1060                1065
Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val
    1070                1075                1080

Phe Asn  Ala Gln Lys Lys Ala  Gln Ser Ser Thr Asp  Gly Ser Ser
    1085                1090                1095
Gly Ser  Glu Thr Pro Gly Thr  Ser Glu Ser Ala Thr  Pro Glu Ser
    1100                1105                1110
Ser Gly  Ser Glu Val Glu Phe  Ser His Glu Tyr Trp  Met Arg His
    1115                1120                1125
Ala Leu  Thr Leu Ala Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro
    1130                1135                1140

Val Gly  Ala Val Leu Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly
    1145                1150                1155
Trp Asn  Arg Ala Ile Gly Leu  His Asp Pro Thr Ala  His Ala Glu
    1160                1165                1170
Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg
    1175                1180                1185
Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met
    1190                1195                1200

Cys Ala  Gly Ala Met Ile His  Ser Arg Ile Gly Arg  Val Val Phe
    1205                1210                1215
Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp
```

1220 1225 1230

Val Leu His Tyr Pro Gly Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 1350
<211> 1554
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1350
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
1040 1045 1050

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
1055 1060 1065

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly

1070 1075 1080

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1085 1090 1095

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met

1100 1105 1110

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu

1115 1120 1125

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

1130 1135 1140

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His

1145 1150 1155

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn

1160 1165 1170

Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys

1175 1180 1185

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

1190 1195 1200

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu

1205 1210 1215

Met Asp Val Leu His Tyr Pro Gly Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1351

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1351

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Met Asn His Arg Val Glu Ile Thr Glu Gly

1295 1300 1305

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg

1310 1315 1320

Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser

1325 1330 1335

Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

1340 1345 1350

Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu

1355 1360 1365

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp

1370 1375 1380

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg
1385 1390 1395
Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
1400 1405 1410
Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
1415 1420 1425

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe
1430 1435 1440
Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile
1445 1450 1455
Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala
1460 1465 1470
Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Asp Lys Pro Ile
1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550
<210> 1352
<211> 1454
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1352
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu
785 790 795 800
Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

805 810 815
Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile
820 825 830
Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
835 840 845
Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
850 855 860
Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

865 870 875 880
Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
885 890 895
Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
900 905 910
His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
915 920 925
Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg

930 935 940
Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Glu Glu
945 950 955 960
Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys
965 970 975
Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly
980 985 990
Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val

995 1000 1005

Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser

1010 1015 1020

Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu

1025 1030 1035

Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg

1040 1045 1050

Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His

1055 1060 1065

His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu

1070 1075 1080

Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp

1085 1090 1095

Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln

1100 1105 1110

Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile

1115 1120 1125

Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile

1130 1135 1140

Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile

1145 1150 1155

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu

1160 1165 1170

Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr

1175 1180 1185

Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp

1190 1195 1200

Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly

1205 1210 1215

Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala

1220 1225 1230

Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu

1235 1240 1245

Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn

1250 1255 1260

Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys

1265 1270 1275

Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu

1280 1285 1290

Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys

1295 1300 1305

Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr

1310 1315 1320

Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn

1325 1330 1335

Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp

1340 1345 1350

Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu

1355 1360 1365

Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His

1370 1375 1380

Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

1385 1390 1395

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe

1400 1405 1410

Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val

1415 1420 1425

Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu

1430 1435 1440

Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1445 1450

<210> 1353

<211> 1420

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1353

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu

785 790 795 800

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

805 810 815

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile

820 825 830

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His

835 840 845

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr

850 855 860

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

865 870 875 880

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly

885 890 895

Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu

900 905 910

His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

915 920 925

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg

930 935 940

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Leu
945 950 955 960
Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr

965 970 975
Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
980 985 990
Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile
995 1000 1005
Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln
1010 1015 1020
Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp

1025 1030 1035
Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr
1040 1045 1050
Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr
1055 1060 1065
Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys
1070 1075 1080
Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe

1085 1090 1095
Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro
1100 1105 1110
Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys
1115 1120 1125
Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln
1130 1135 1140
Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser

1145 1150 1155
Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala
1160 1165 1170
Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

1175 1180 1185

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys

1190 1195 1200

Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile

1205 1210 1215

Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe

1220 1225 1230

Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile

1235 1240 1245

Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys

1250 1255 1260

Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu

1265 1270 1275

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His

1280 1285 1290

Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln

1295 1300 1305

Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu

1310 1315 1320

Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn

1325 1330 1335

Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro

1340 1345 1350

Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr

1355 1360 1365

Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile

1370 1375 1380

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr

1385 1390 1395

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp

1400 1405 1410

Leu Ser Gln Leu Gly Gly Asp

1415 1420

<210> 1354

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1354

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Thr Ala His Ala Glu
995 1000 1005
Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
1010 1015 1020
Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met
1025 1030 1035

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe
1040 1045 1050
Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp
1055 1060 1065
Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
1070 1075 1080
Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe
1085 1090 1095

Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser
1100 1105 1110
Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
1115 1120 1125
Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His
1130 1135 1140
Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg
1145 1150 1155

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

1160 1165 1170

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp

1175 1180 1185

Pro Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1190 1195 1200

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 1355

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1355

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Thr Ala His Ala Glu Ile Met Ala

1010 1015 1020

Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

1025 1030 1035

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly

1040 1045 1050

Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg

1055 1060 1065

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

1070 1075 1080

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

1085 1090 1095

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro

1100 1105 1110

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp

1115 1120 1125

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

1130 1135 1140

Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp

1145 1150 1155

Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg

1160 1165 1170

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile

1175 1180 1185

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Tyr Asp

1190 1195 1200

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1356

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1356

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

```
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
```

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Thr

1010 1015 1020

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

1025 1030 1035

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

1040 1045 1050

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly

1055 1060 1065

Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly

1070 1075 1080

Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val

1085 1090 1095

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
1100 1105 1110
Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
1115 1120 1125
Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro
1130 1135 1140

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val
1145 1150 1155
Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala
1160 1165 1170
Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
1175 1180 1185
Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
1190 1195 1200

Gly Leu His Asp Pro Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala
1205 1210 1215
Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1220 1225 1230
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1357

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 1357

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Thr Ala His Ala Glu Ile Met Ala Leu

1025 1030 1035

Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala

1040 1045 1050

Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

1055 1060 1065

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn

1070 1075 1080

Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr

1085 1090 1095

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

1100 1105 1110

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg

1115 1120 1125

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly

1130 1135 1140

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1145 1150 1155

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met

1160 1165 1170

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu

1175 1180 1185

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

1190 1195 1200

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 1358
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1358
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly

1040 1045 1050

Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val

1055 1060 1065

Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser

1070 1075 1080

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

1085 1090 1095

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn

1100 1105 1110

His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala

1115 1120 1125

Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn

1130 1135 1140

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser

1145 1150 1155

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1160 1165 1170

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1175 1180 1185

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly

1190 1195 1200

Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

1205 1210 1215

Arg Ala Ile Gly Leu His Asp Pro Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1359

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1359

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu

1055 1060 1065

Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr

1070 1075 1080

Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg

1085 1090 1095

Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala

1100 1105 1110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

1115 1120 1125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala

1130 1135 1140

Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala

1145 1150 1155

Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu

1160 1165 1170

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser

1175 1180 1185

Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1190 1195 1200

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

1205 1210 1215

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg

1220 1225 1230

Ala Ile Gly Leu His Asp Pro Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 1360

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1360

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Thr

1235 1240 1245

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

1250 1255 1260

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

1265 1270 1275

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly

1280 1285 1290

Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly

1295 1300 1305

Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val

1310 1315 1320

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu

1325 1330 1335

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

1340 1345 1350

Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro

1355 1360 1365

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val

1370 1375 1380

Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

1385 1390 1395

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu

1400 1405 1410

Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

1415 1420 1425

Gly Leu His Asp Pro Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 1361
<211> 167
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
TadA sequence
<400> 1361
Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45
Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp
165

<210> 1362

<211> 178

<212> PRT

<213> Escherichia coli

<400> 1362

Met Arg Arg Ala Phe Ile Thr Gly Val Phe Phe Leu Ser Glu Val Glu
1 5 10 15

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
20 25 30

Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn
35 40 45

Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp
50 55 60

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65 70 75 80

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Leu Glu
85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
100 105 110

Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala Ala Gly Ser Leu
115 120 125
Met Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr
130 135 140
Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Ser Asp Phe Phe
145 150 155 160
Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys Ala Gln Ser Ser
165 170 175

Thr Asp

<210> 1363
<211> 160
<212> PRT
<213> Staphylococcus aureus
<400> 1363
Met Gly Ser His Met Thr Asn Asp Ile Tyr Phe Met Thr Leu Ala Ile
1 5 10 15
Glu Glu Ala Lys Lys Ala Ala Gln Leu Gly Glu Val Pro Ile Gly Ala
20 25 30
Ile Ile Thr Lys Asp Asp Glu Val Ile Ala Arg Ala His Asn Leu Arg
35 40 45
Glu Thr Leu Gln Gln Pro Thr Ala His Ala Glu His Ile Ala Ile Glu

50 55 60
Arg Ala Ala Lys Val Leu Gly Ser Trp Arg Leu Glu Gly Cys Thr Leu
65 70 75 80
Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Thr Ile Val Met
85 90 95
Ser Arg Ile Pro Arg Val Val Tyr Gly Ala Asp Asp Pro Lys Gly Gly
100 105 110

Cys Ser Gly Ser Leu Met Asn Leu Leu Gln Gln Ser Asn Phe Asn His

115 120 125

Arg Ala Ile Val Asp Lys Gly Val Leu Lys Glu Ala Cys Ser Thr Leu

130 135 140

Leu Thr Thr Phe Phe Lys Asn Leu Arg Ala Asn Lys Lys Ser Thr Asn

145 150 155 160

<210> 1364

<211> 161

<212> PRT

<213> Bacillus subtilis

<400> 1364

Met Thr Gln Asp Glu Leu Tyr Met Lys Glu Ala Ile Lys Glu Ala Lys

1 5 10 15

Lys Ala Glu Glu Lys Gly Glu Val Pro Ile Gly Ala Val Leu Val Ile

20 25 30

Asn Gly Glu Ile Ile Ala Arg Ala His Asn Leu Arg Glu Thr Glu Gln

35 40 45

Arg Ser Ile Ala His Ala Glu Met Leu Val Ile Asp Glu Ala Cys Lys

50 55 60

Ala Leu Gly Thr Trp Arg Leu Glu Gly Ala Thr Leu Tyr Val Thr Leu

65 70 75 80

Glu Pro Cys Pro Met Cys Ala Gly Ala Val Val Leu Ser Arg Val Glu

85 90 95

Lys Val Val Phe Gly Ala Phe Asp Pro Lys Gly Gly Cys Ser Gly Thr

100 105 110

Leu Met Asn Leu Leu Gln Glu Glu Arg Phe Asn His Gln Ala Glu Val

115 120 125

Val Ser Gly Val Leu Glu Glu Glu Cys Gly Gly Met Leu Ser Ala Phe

130 135 140

Phe Arg Glu Leu Arg Lys Lys Lys Lys Ala Ala Arg Lys Asn Leu Ser

145 150 155 160

Glu

<210> 1365

<211> 183

<212> PRT

<213> Salmonella typhimurium

<400> 1365

Met Pro Pro Ala Phe Ile Thr Gly Val Thr Ser Leu Ser Asp Val Glu
1 5 10 15
Leu Asp His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
20 25 30
Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn
35 40 45
His Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp
50 55 60
Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65 70 75 80
Leu Gln Asn Tyr Arg Leu Leu Asp Thr Thr Leu Tyr Val Thr Leu Glu
85 90 95
Pro Cys Val Met Cys Ala Gly Ala Met Val His Ser Arg Ile Gly Arg
100 105 110
Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala Ala Gly Ser Leu
115 120 125
Ile Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Ile
130 135 140
Glu Gly Val Leu Arg Asp Glu Cys Ala Thr Leu Leu Ser Asp Phe Phe
145 150 155 160
Arg Met Arg Arg Gln Glu Ile Lys Ala Leu Lys Lys Ala Asp Arg Ala
165 170 175
Glu Gly Ala Gly Pro Ala Val
180

<210> 1366

<211> 164

<212> PRT

<213> Shewanella putrefaciens

<400> 1366

Met Asp Glu Tyr Trp Met Gln Val Ala Met Gln Met Ala Glu Lys Ala
1 5 10 15
Glu Ala Ala Gly Glu Val Pro Val Gly Ala Val Leu Val Lys Asp Gly
20 25 30
Gln Gln Ile Ala Thr Gly Tyr Asn Leu Ser Ile Ser Gln His Asp Pro

35 40 45
Thr Ala His Ala Glu Ile Leu Cys Leu Arg Ser Ala Gly Lys Lys Leu
50 55 60
Glu Asn Tyr Arg Leu Leu Asp Ala Thr Leu Tyr Ile Thr Leu Glu Pro
65 70 75 80
Cys Ala Met Cys Ala Gly Ala Met Val His Ser Arg Ile Ala Arg Val
85 90 95
Val Tyr Gly Ala Arg Asp Glu Lys Thr Gly Ala Ala Gly Thr Val Val

100 105 110
Asn Leu Leu Gln His Pro Ala Phe Asn His Gln Val Glu Val Thr Ser
115 120 125
Gly Val Leu Ala Glu Ala Cys Ser Ala Gln Leu Ser Arg Phe Phe Lys
130 135 140
Arg Arg Arg Asp Glu Lys Lys Ala Leu Lys Leu Ala Gln Arg Ala Gln
145 150 155 160
Gln Gly Ile Glu

<210> 1367

<211> 173

<212> PRT

<213> Haemophilus influenzae

<400> 1367

Met Asp Ala Ala Lys Val Arg Ser Glu Phe Asp Glu Lys Met Met Arg

1 5 10 15

Tyr Ala Leu Glu Leu Ala Asp Lys Ala Glu Ala Leu Gly Glu Ile Pro

20 25 30

Val Gly Ala Val Leu Val Asp Asp Ala Arg Asn Ile Ile Gly Glu Gly

35 40 45

Trp Asn Leu Ser Ile Val Gln Ser Asp Pro Thr Ala His Ala Glu Ile

50 55 60

Ile Ala Leu Arg Asn Gly Ala Lys Asn Ile Gln Asn Tyr Arg Leu Leu

65 70 75 80

Asn Ser Thr Leu Tyr Val Thr Leu Glu Pro Cys Thr Met Cys Ala Gly

85 90 95

Ala Ile Leu His Ser Arg Ile Lys Arg Leu Val Phe Gly Ala Ser Asp

100 105 110

Tyr Lys Thr Gly Ala Ile Gly Ser Arg Phe His Phe Phe Asp Asp Tyr

115 120 125

Lys Met Asn His Thr Leu Glu Ile Thr Ser Gly Val Leu Ala Glu Glu

130 135 140

Cys Ser Gln Lys Leu Ser Thr Phe Phe Gln Lys Arg Arg Glu Glu Lys

145 150 155 160

Lys Ile Glu Lys Ala Leu Leu Lys Ser Leu Ser Asp Lys

165 170

<210> 1368

<211> 161

<212> PRT

<213> Caulobacter crescentus

<400

> 1368

Met Arg Thr Asp Glu Ser Glu Asp Gln Asp His Arg Met Met Arg Leu

1 5 10 15

Ala Leu Asp Ala Ala Arg Ala Ala Ala Glu Ala Gly Glu Thr Pro Val

20 25 30

Gly Ala Val Ile Leu Asp Pro Ser Thr Gly Glu Val Ile Ala Thr Ala

35 40 45

Gly Asn Gly Pro Ile Ala Ala His Asp Pro Thr Ala His Ala Glu Ile
50 55 60

Ala Ala Met Arg Ala Ala Ala Ala Lys Leu Gly Asn Tyr Arg Leu Thr
65 70 75 80
Asp Leu Thr Leu Val Val Thr Leu Glu Pro Cys Ala Met Cys Ala Gly
85 90 95
Ala Ile Ser His Ala Arg Ile Gly Arg Val Val Phe Gly Ala Asp Asp
100 105 110
Pro Lys Gly Gly Ala Val Val His Gly Pro Lys Phe Phe Ala Gln Pro
115 120 125

Thr Cys His Trp Arg Pro Glu Val Thr Gly Gly Val Leu Ala Asp Glu
130 135 140
Ser Ala Asp Leu Leu Arg Gly Phe Phe Arg Ala Arg Arg Lys Ala Lys
145 150 155 160
Ile

<210> 1369
<211> 179
<212> PRT
<213> Geobacter sulfurreducens
<400> 1369
Met Ser Ser Leu Lys Lys Thr Pro Ile Arg Asp Asp Ala Tyr Trp Met
1 5 10 15
Gly Lys Ala Ile Arg Glu Ala Ala Lys Ala Ala Ala Arg Asp Glu Val

20 25 30
Pro Ile Gly Ala Val Ile Val Arg Asp Gly Ala Val Ile Gly Arg Gly
35 40 45
His Asn Leu Arg Glu Gly Ser Asn Asp Pro Ser Ala His Ala Glu Met
50 55 60
Ile Ala Ile Arg Gln Ala Ala Arg Arg Ser Ala Asn Trp Arg Leu Thr
65 70 75 80

Gly Ala Thr Leu Tyr Val Thr Leu Glu Pro Cys Leu Met Cys Met Gly

85 90 95

Ala Ile Ile Leu Ala Arg Leu Glu Arg Val Val Phe Gly Cys Tyr Asp

100 105 110

Pro Lys Gly Gly Ala Ala Gly Ser Leu Tyr Asp Leu Ser Ala Asp Pro

115 120 125

Arg Leu Asn His Gln Val Arg Leu Ser Pro Gly Val Cys Gln Glu Glu

130 135 140

Cys Gly Thr Met Leu Ser Asp Phe Phe Arg Asp Leu Arg Arg Arg Lys

145 150 155 160

Lys Ala Lys Ala Thr Pro Ala Leu Phe Ile Asp Glu Arg Lys Val Pro

165 170 175

Pro Glu Pro

<210> 1370

<211> 167

<212> PRT

<213> Escherichia coli

<400> 1370

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp
165

<210> 1371
<211> 20
<212> RNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide
<400> 1371
augggcuca gaggcuguaa 20

<210> 1372
<211> 198
<212> PRT
<213> Homo sapiens
<400> 1372
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val

20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45
Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu

195

<210> 1373

<211> 198

<212> PRT

<213> Mus musculus

<400> 1373

Met Asp Ser Leu Leu Met Lys Gln Lys Lys Phe Leu Tyr His Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Cys Ser Leu Asp Phe Gly His

35 40 45

Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Glu
85 90 95
Phe Leu Arg Trp Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110
Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125
Leu His Arg Ala Gly Val Gln Ile Gly Ile Met Thr Phe Lys Asp Tyr

130 135 140
Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Arg Thr Phe Lys
145 150 155 160
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Thr Arg Gln Leu
165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190
Phe Arg Met Leu Gly Phe
195

<210> 1374
<211> 198
<212> PRT
<213> Canis lupus familiaris
<400> 1374
Met Asp Ser Leu Leu Met Lys Gln Arg Lys Phe Leu Tyr His Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly His
35 40 45
Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Ala Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Lys Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu

195

<210> 1375

<211> 199

<212> PRT

<213> Bos taurus

<400> 1375

Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His

35 40 45

Leu Arg Asn Lys Ala Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Asp Lys Glu Arg Lys Ala Glu Pro Glu Gly Leu Arg

115 120 125

Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp

130 135 140

Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe

145 150 155 160

Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln

165 170 175

Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp

180 185 190

Ala Phe Arg Thr Leu Gly Leu

195

<210> 1376

<211> 239

<212> PRT

<213> Rattus norvegicus

<400> 1376

Met Ala Val Gly Ser Lys Pro Lys Ala Ala Leu Val Gly Pro His Trp

1 5 10 15

Glu Arg Glu Arg Ile Trp Cys Phe Leu Cys Ser Thr Gly Leu Gly Thr

20 25 30

Gln Gln Thr Gly Gln Thr Ser Arg Trp Leu Arg Pro Ala Ala Thr Gln

35 40 45

Asp Pro Val Ser Pro Pro Arg Ser Leu Leu Met Lys Gln Arg Lys Phe

50 55 60

Leu Tyr His Phe Lys Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr

65 70 75 80

Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser

85 90 95

Leu Asp Phe Gly Tyr Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu
100 105 110
Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys
115 120 125
Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala
130 135 140
Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg

145 150 155 160
Ile Phe Thr Ala Arg Leu Thr Gly Trp Gly Ala Leu Pro Ala Gly Leu
165 170 175
Met Ser Pro Ala Arg Pro Ser Asp Tyr Phe Tyr Cys Trp Asn Thr Phe
180 185 190
Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu
195 200 205
Asn Ser Val Arg Leu Ser Arg Arg Leu Arg Arg Ile Leu Leu Pro Leu

210 215 220
Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly Leu
225 230 235
<210> 1377
<211> 198
<212> PRT
<213> Mus musculus
<400> 1377
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45
Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu

195

<210> 1378

<211> 229

<212> PRT

<213> Rattus norvegicus

<400> 1378

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys
225

<210> 1379
<211> 229
<212> PRT
<213> Mesocricetus auratus
<400> 1379
Met Ser Ser Glu Thr Gly Pro Val Val Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Asp Ala Phe Phe Asp Gln Gly Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Arg Trp Gly Gly Arg His

35 40 45

Asn Ile Trp Arg His Thr Gly Gln Asn Thr Ser Arg His Val Glu Ile

50 55 60

Asn Phe Ile Glu Lys Phe Thr Ser Glu Arg Tyr Phe Tyr Pro Ser Thr

65 70 75 80

Arg Cys Ser Ile Val Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Lys Ala Ile Thr Glu Phe Leu Ser Gly His Pro Asn Val Thr Leu

100 105 110

Phe Ile Tyr Ala Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Arg Gly Val Thr Ile Arg Ile Met

130 135 140

Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160

Pro Ser Asn Glu Val Tyr Trp Pro Arg Tyr Pro Asn Leu Trp Met Arg

165 170 175

Leu Tyr Ala Leu Glu Leu Tyr Cys Ile His Leu Gly Leu Pro Pro Cys

180 185 190

Leu Lys Ile Lys Arg Arg His Gln Tyr Pro Leu Thr Phe Phe Arg Leu

195 200 205

Asn Leu Gln Ser Cys His Tyr Gln Arg Ile Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Phe Ile

225

<210> 1380

<211> 236

<212> PRT

<213> Pongo pygmaeus

<400> 1380

Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Ser Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
35 40 45

Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
50 55 60
Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Arg Phe His Ser Ser Ile
65 70 75 80
Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95
Ser Gln Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu
100 105 110

Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160
Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
165 170 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
180 185 190
Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Ala Phe Phe Arg Leu
195 200 205
His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu
210 215 220
Ala Thr Gly Leu Ile His Pro Ser Val Thr Trp Arg
225 230 235
<210> 1381

<211> 236
<212> PRT

<213> Oryctolagus cuniculus

<400> 1381

Met Ala Ser Glu Lys Gly Pro Ser Asn Lys Asp Tyr Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro Trp Glu Phe Glu Val Phe Phe Asp Pro Gln Glu Leu
20 25 30

Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Ala Ser Ser
35 40 45

Lys Thr Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
50 55 60

Asn Phe Leu Glu Lys Leu Thr Ser Glu Gly Arg Leu Gly Pro Ser Thr
65 70 75 80

Cys Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95

Ser Met Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu
100 105 110

Ile Ile Phe Val Ala Arg Leu Phe Gln His Met Asp Arg Arg Asn Arg
115 120 125

Gln Gly Leu Lys Asp Leu Val Thr Ser Gly Val Thr Val Arg Val Met
130 135 140

Ser Val Ser Glu Tyr Cys Tyr Cys Trp Glu Asn Phe Val Asn Tyr Pro
145 150 155 160

Pro Gly Lys Ala Ala Gln Trp Pro Arg Tyr Pro Pro Arg Trp Met Leu
165 170 175

Met Tyr Ala Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Lys Ile Ser Arg Arg His Gln Lys Gln Leu Thr Phe Phe Ser Leu
195 200 205

Thr Pro Gln Tyr Cys His Tyr Lys Met Ile Pro Pro Tyr Ile Leu Leu
210 215 220

Ala Thr Gly Leu Leu Gln Pro Ser Val Pro Trp Arg
225 230 235

<210> 1382

<211> 235

<212> PRT

<213> Monodelphis domestica

<400> 1382

Met Asn Ser Lys Thr Gly Pro Ser Val Gly Asp Ala Thr Leu Arg Arg
1 5 10 15
Arg Ile Lys Pro Trp Glu Phe Val Ala Phe Phe Asn Pro Gln Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Asn Gln Asn
35 40 45
Ile Trp Arg His Ser Asn Gln Asn Thr Ser Gln His Ala Glu Ile Asn
50 55 60

Phe Met Glu Lys Phe Thr Ala Glu Arg His Phe Asn Ser Ser Val Arg
65 70 75 80
Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser
85 90 95
Lys Ala Ile Arg Lys Phe Leu Asp His Tyr Pro Asn Val Thr Leu Ala
100 105 110
Ile Phe Ile Ser Arg Leu Tyr Trp His Met Asp Gln Gln His Arg Gln
115 120 125

Gly Leu Lys Glu Leu Val His Ser Gly Val Thr Ile Gln Ile Met Ser
130 135 140
Tyr Ser Glu Tyr His Tyr Cys Trp Arg Asn Phe Val Asp Tyr Pro Gln
145 150 155 160
Gly Glu Glu Asp Tyr Trp Pro Lys Tyr Pro Tyr Leu Trp Ile Met Leu
165 170 175
Tyr Val Leu Glu Leu His Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu
180 185 190

Lys Ile Ser Gly Ser His Ser Asn Gln Leu Ala Leu Phe Ser Leu Asp
195 200 205

Leu Gln Asp Cys His Tyr Gln Lys Ile Pro Tyr Asn Val Leu Val Ala
210 215 220

Thr Gly Leu Val Gln Pro Phe Val Thr Trp Arg
225 230 235

<210> 1383

<211> 224

<212> PRT

<213> Pongo pygmaeus

<400> 1383

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Thr Glu Ala Ala Ser Gln
1 5 10 15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly
65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110

Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Leu
145 150 155 160

Glu Ile Gln Asp Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val

180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile

195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 1384

<211> 224

<212> PRT

<213> Bos taurus

<400> 1384

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Ala Glu Pro Ala Ser Gln

1 5 10 15

Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro

35 40 45

Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly

50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly

65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Thr

85 90 95

Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met Pro Thr Phe Asp

100 105 110

Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser Ser Ser Pro Cys

115 120 125

Ala Ala Cys Ala Asp Arg Ile Val Lys Thr Leu Asn Lys Thr Lys Asn

130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro

145 150 155 160

Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala Gly Cys Arg Leu

165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val

180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile

195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 1385

<211> 440

<212> PRT

<213> Mus musculus

<400> 1385

Met Gln Pro Gln Arg Leu Gly Pro Arg Ala Gly Met Gly Pro Phe Cys

1 5 10 15

Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro Ile Arg Asn Leu Ile

20 25 30

Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn Leu Gly Tyr Ala Lys

35 40 45

Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val Thr Arg Lys Asp Cys

50 55 60

Asp Ser Pro Val Ser Leu His His Gly Val Phe Lys Asn Lys Asp Asn

65 70 75 80

Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe His Asp Lys Val Leu

85 90 95

Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile Thr Trp Tyr Met Ser

100 105 110

Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile Val Arg Phe Leu Ala

115 120 125

Thr His His Asn Leu Ser Leu Asp Ile Phe Ser Ser Arg Leu Tyr Asn

130 135 140

Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys Arg Leu Val Gln Glu

145 150 155 160

Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu Phe Lys Lys Cys Trp
165 170 175
Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe Arg Pro Trp Lys Arg
180 185 190

Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys Leu Gln Glu Ile Leu
195 200 205
Arg Pro Cys Tyr Ile Ser Val Pro Ser Ser Ser Ser Ser Thr Leu Ser
210 215 220
Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr Arg Phe Trp Val Glu
225 230 235 240
Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu Phe Tyr Ser Gln Phe
245 250 255

Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr His Arg Met Lys Pro
260 265 270
Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly Gln Ala Pro Leu Lys
275 280 285
Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His Ala Glu Ile Leu Phe
290 295 300
Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln Val Thr Ile Thr Cys
305 310 315 320

Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala Trp Gln Leu Ala Ala
325 330 335
Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His Ile Tyr Thr Ser Arg
340 345 350
Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys Gly Leu Cys Ser Leu
355 360 365
Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp Leu Pro Gln Phe Thr
370 375 380

Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg Pro Phe Trp Pro Trp
385 390 395 400
Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln Arg Arg Leu Arg Arg

405 410 415

Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val Asn Asp Phe Gly Asn

420 425 430

Leu Gln Leu Gly Pro Pro Met Ser

435 440

<210> 1386

<211> 429

<212> PRT

<213> Mus musculus

<400> 1386

Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro

1 5 10 15

Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn

20 25 30

Leu Gly Tyr Ala Lys Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val

35 40 45

Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe

50 55 60

Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe

65 70 75 80

His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile

85 90 95

Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile

100 105 110

Val Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser

115 120 125

Ser Arg Leu Tyr Asn Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys

130 135 140

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu

145 150 155 160

Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe

165 170 175

Arg Pro Trp Lys Arg Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys
180 185 190

Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser
195 200 205

Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr
210 215 220

Arg Phe Cys Val Glu Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu
225 230 235 240

Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr
245 250 255

His Arg Met Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly
260 265 270

Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His
275 280 285

Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln
290 295 300

Val Thr Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
305 310 315 320

Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His
325 330 335

Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys
340 345 350

Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp
355 360 365

Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg
370 375 380

Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln
385 390 395 400

Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val
405 410 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser

420 425

<210> 1387

<211> 430

<212> PRT

<213> Rattus norvegicus

<400> 1387

Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro

1 5 10 15

Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn

20 25 30

Arg Leu Arg Tyr Ala Ile Asp Arg Lys Asp Thr Phe Leu Cys Tyr Glu

35 40 45

Val Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val

50 55 60

Phe Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp

65 70 75 80

Phe His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys

85 90 95

Ile Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln

100 105 110

Val Leu Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe

115 120 125

Ser Ser Arg Leu Tyr Asn Ile Arg Asp Pro Glu Asn Gln Gln Asn Leu

130 135 140

Cys Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr

145 150 155 160

Glu Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg

165 170 175

Phe Arg Pro Trp Lys Lys Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser

180 185 190

Lys Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser

195 200 205

Ser Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu

210 215 220

Thr Arg Phe Cys Val Glu Arg Arg Arg Val His Leu Leu Ser Glu Glu

225 230 235 240

Glu Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr

245 250 255

Tyr His Gly Val Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn

260 265 270

Gly Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln

275 280 285

His Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser

290 295 300

Gln Val Ile Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys

305 310 315 320

Ala Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu

325 330 335

His Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln

340 345 350

Lys Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met

355 360 365

Asp Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys

370 375 380

Arg Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr

385 390 395 400

Gln Arg Arg Leu His Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu

405 410 415

Val Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser

420 425 430

<210> 1388

<211> 199

<212> PRT

<213> Homo sapiens

<400> 1388

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His
1 5 10 15

Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr
20 25 30

Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met
35 40 45

Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys
50 55 60

Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro
65 70 75 80

Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile
85 90 95

Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala
100 105 110

Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg
115 120 125

Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg
130 135 140

Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His
145 150 155 160

Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp
165 170 175

Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala
180 185 190

Ile Leu Gln Asn Gln Gly Asn
195

<210> 1389

<211> 373

<212> PRT

<213> Homo sapiens

<400> 1389

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15
Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
20 25 30
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Arg
35 40 45
Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Gln Pro Glu His
50 55 60

His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro
65 70 75 80
Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys
85 90 95
Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ala Glu His Pro Asn
100 105 110
Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg
115 120 125

Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val
130 135 140
Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val
145 150 155 160
Tyr Ser Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn
165 170 175
Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met
180 185 190

Glu Ala Met Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg
195 200 205
Lys Ala Tyr Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val
210 215 220
Val Lys His His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn

225 230 235 240

Gln Val Asp Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser

245 250 255

Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr

260 265 270

Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala

275 280 285

Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala

290 295 300

Arg Leu Tyr Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser

305 310 315 320

Leu Ser Gln Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe

325 330 335

Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys

340 345 350

Pro Trp Lys Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu

355 360 365

Gln Glu Ile Leu Glu

370

<210> 1390

<211> 370

<212> PRT

<213> Macaca mulatta

<400> 1390

Met Val Glu Pro Met Asp Pro Arg Thr Phe Val Ser Asn Phe Asn Asn

1 5 10 15

Arg Pro Ile Leu Ser Gly Leu Asn Thr Val Trp Leu Cys Cys Glu Val

20 25 30

Lys Thr Lys Asp Pro Ser Gly Pro Pro Leu Asp Ala Lys Ile Phe Gln

35 40 45

Gly Lys Val Tyr Ser Lys Ala Lys Tyr His Pro Glu Met Arg Phe Leu

50 55 60

Arg Trp Phe His Lys Trp Arg Gln Leu His His Asp Gln Glu Tyr Lys
65 70 75 80
Val Thr Trp Tyr Val Ser Trp Ser Pro Cys Thr Arg Cys Ala Asn Ser
85 90 95
Val Ala Thr Phe Leu Ala Lys Asp Pro Lys Val Thr Leu Thr Ile Phe
100 105 110
Val Ala Arg Leu Tyr Tyr Phe Trp Lys Pro Asp Tyr Gln Gln Ala Leu
115 120 125

Arg Ile Leu Cys Gln Lys Arg Gly Gly Pro His Ala Thr Met Lys Ile
130 135 140
Met Asn Tyr Asn Glu Phe Gln Asp Cys Trp Asn Lys Phe Val Asp Gly
145 150 155 160
Arg Gly Lys Pro Phe Lys Pro Arg Asn Asn Leu Pro Lys His Tyr Thr
165 170 175
Leu Leu Gln Ala Thr Leu Gly Glu Leu Leu Arg His Leu Met Asp Pro
180 185 190

Gly Thr Phe Thr Ser Asn Phe Asn Asn Lys Pro Trp Val Ser Gly Gln
195 200 205
His Glu Thr Tyr Leu Cys Tyr Lys Val Glu Arg Leu His Asn Asp Thr
210 215 220
Trp Val Pro Leu Asn Gln His Arg Gly Phe Leu Arg Asn Gln Ala Pro
225 230 235 240
Asn Ile His Gly Phe Pro Lys Gly Arg His Ala Glu Leu Cys Phe Leu
245 250 255

Asp Leu Ile Pro Phe Trp Lys Leu Asp Gly Gln Gln Tyr Arg Val Thr
260 265 270
Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala
275 280 285
Lys Phe Ile Ser Asn Asn Glu His Val Ser Leu Cys Ile Phe Ala Ala
290 295 300

Arg Ile Tyr Asp Asp Gln Gly Arg Tyr Gln Glu Gly Leu Arg Ala Leu
305 310 315 320

His Arg Asp Gly Ala Lys Ile Ala Met Met Asn Tyr Ser Glu Phe Glu
325 330 335
Tyr Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe Gln Pro
340 345 350
Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg
355 360 365
Ala Ile
370

<210> 1391
<211> 384
<212> PRT
<213> Pan troglodytes
<400> 1391
Met Lys Pro His Phe Arg Asn Pro Val Glu Arg Met Tyr Gln Asp Thr

1 5 10 15
Phe Ser Asp Asn Phe Tyr Asn Arg Pro Ile Leu Ser His Arg Asn Thr
20 25 30
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
35 40 45
Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Lys Leu Lys Tyr
50 55 60
His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu

65 70 75 80
His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95
Cys Thr Lys Cys Thr Arg Asp Val Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110
Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly

130 135 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys

145 150 155 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn

165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile

180 185 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Ser Asn Phe Asn Asn

195 200 205

Glu Leu Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val

210 215 220

Glu Arg Leu His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly

225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg

245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp

260 265 270

Leu His Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys

275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Asn Asn Lys His

290 295 300

Val Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg

305 310 315 320

Cys Gln Glu Gly Leu Arg Thr Leu Ala Lys Ala Gly Ala Lys Ile Ser

325 330 335

Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp

340 345 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser

355 360 365

Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn

370 375 380

<210> 1392

<211> 377

<212> PRT

<213> Chlorocebus sabaeus

<400> 1392

Met Asn Pro Gln Ile Arg Asn Met Val Glu Gln Met Glu Pro Asp Ile
1 5 10 15

Phe Val Tyr Tyr Phe Asn Asn Arg Pro Ile Leu Ser Gly Arg Asn Thr
20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Asp Pro Ser Gly Pro Pro
35 40 45

Leu Asp Ala Asn Ile Phe Gln Gly Lys Leu Tyr Pro Glu Ala Lys Asp
50 55 60

His Pro Glu Met Lys Phe Leu His Trp Phe Arg Lys Trp Arg Gln Leu
65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Val Ser Trp Ser Pro
85 90 95

Cys Thr Arg Cys Ala Asn Ser Val Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Lys
115 120 125

Pro Asp Tyr Gln Gln Ala Leu Arg Ile Leu Cys Gln Glu Arg Gly Gly
130 135 140

Pro His Ala Thr Met Lys Ile Met Asn Tyr Asn Glu Phe Gln His Cys
145 150 155 160

Trp Asn Glu Phe Val Asp Gly Gln Gly Lys Pro Phe Lys Pro Arg Lys
165 170 175

Asn Leu Pro Lys His Tyr Thr Leu Leu His Ala Thr Leu Gly Glu Leu
180 185 190

Leu Arg His Val Met Asp Pro Gly Thr Phe Thr Ser Asn Phe Asn Asn

195 200 205

Lys Pro Trp Val Ser Gly Gln Arg Glu Thr Tyr Leu Cys Tyr Lys Val

210 215 220

Glu Arg Ser His Asn Asp Thr Trp Val Leu Leu Asn Gln His Arg Gly

225 230 235 240

Phe Leu Arg Asn Gln Ala Pro Asp Arg His Gly Phe Pro Lys Gly Arg

245 250 255

His Ala Glu Leu Cys Phe Leu Asp Leu Ile Pro Phe Trp Lys Leu Asp

260 265 270

Asp Gln Gln Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe

275 280 285

Ser Cys Ala Gln Lys Met Ala Lys Phe Ile Ser Asn Asn Lys His Val

290 295 300

Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys

305 310 315 320

Gln Glu Gly Leu Arg Thr Leu His Arg Asp Gly Ala Lys Ile Ala Val

325 330 335

Met Asn Tyr Ser Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Asp Arg

340 345 350

Gln Gly Arg Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln

355 360 365

Ala Leu Ser Gly Arg Leu Arg Ala Ile

370 375

<210> 1393

<211> 384

<212> PRT

<213> Homo sapiens

<400> 1393

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr

20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
35 40 45
Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr
50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65 70 75 80
His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95
Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110
Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140
Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145 150 155 160
Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175
Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
180 185 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
195 200 205
Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
210 215 220
Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225 230 235 240
Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
260 265 270
Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys

275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His

290 295 300

Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg

305 310 315 320

Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser

325 330 335

Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp

340 345 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser

355 360 365

Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn

370 375 380

<210> 1394

<211> 382

<212> PRT

<213> Homo sapiens

<400> 1394

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr

20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu

35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln

50 55 60

Tyr His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu

65 70 75 80

Pro Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro

85 90 95

Cys Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ser Glu His Pro

100 105 110

Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu

115 120 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg

130 135 140

Val Thr Ile Met Asp Tyr Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Glu Gly Gln Gln Phe Met Pro Trp Tyr Lys Phe Asp Glu

165 170 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Tyr Leu

180 185 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val

195 200 205

Leu Arg Arg Arg Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp

210 215 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn

225 230 235 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu

245 250 255

Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile

260 265 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly

275 280 285

Cys Ala Gly Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg

290 295 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys

305 310 315 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met

325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln

340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala

355 360 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn

370 375 380

<210> 1395

<211> 395

<212> PRT

<213> Rattus norvegicus

<400> 1395

Met Gln Pro Gln Gly Leu Gly Pro Asn Ala Gly Met Gly Pro Val Cys
1 5 10 15

Leu Gly Cys Ser His Arg Arg Pro Tyr Ser Pro Ile Arg Asn Pro Leu
20 25 30

Lys Lys Leu Tyr Gln Gln Thr Phe Tyr Phe His Phe Lys Asn Val Arg
35 40 45

Tyr Ala Trp Gly Arg Lys Asn Asn Phe Leu Cys Tyr Glu Val Asn Gly
50 55 60

Met Asp Cys Ala Leu Pro Val Pro Leu Arg Gln Gly Val Phe Arg Lys
65 70 75 80

Gln Gly His Ile His Ala Glu Leu Cys Phe Ile Tyr Trp Phe His Asp
85 90 95

Lys Val Leu Arg Val Leu Ser Pro Met Glu Glu Phe Lys Val Thr Trp
100 105 110

Tyr Met Ser Trp Ser Pro Cys Ser Lys Cys Ala Glu Gln Val Ala Arg
115 120 125

Phe Leu Ala Ala His Arg Asn Leu Ser Leu Ala Ile Phe Ser Ser Arg
130 135 140

Leu Tyr Tyr Tyr Leu Arg Asn Pro Asn Tyr Gln Gln Lys Leu Cys Arg
145 150 155 160

Leu Ile Gln Glu Gly Val His Val Ala Ala Met Asp Leu Pro Glu Phe
165 170 175

Lys Lys Cys Trp Asn Lys Phe Val Asp Asn Asp Gly Gln Pro Phe Arg

180 185 190

Pro Trp Met Arg Leu Arg Ile Asn Phe Ser Phe Tyr Asp Cys Lys Leu

195 200 205

Gln Glu Ile Phe Ser Arg Met Asn Leu Leu Arg Glu Asp Val Phe Tyr

210 215 220

Leu Gln Phe Asn Asn Ser His Arg Val Lys Pro Val Gln Asn Arg Tyr

225 230 235 240

Tyr Arg Arg Lys Ser Tyr Leu Cys Tyr Gln Leu Glu Arg Ala Asn Gly

245 250 255

Gln Glu Pro Leu Lys Gly Tyr Leu Leu Tyr Lys Lys Gly Glu Gln His

260 265 270

Val Glu Ile Leu Phe Leu Glu Lys Met Arg Ser Met Glu Leu Ser Gln

275 280 285

Val Arg Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala

290 295 300

Arg Gln Leu Ala Ala Phe Lys Lys Asp His Pro Asp Leu Ile Leu Arg

305 310 315 320

Ile Tyr Thr Ser Arg Leu Tyr Phe Trp Arg Lys Lys Phe Gln Lys Gly

325 330 335

Leu Cys Thr Leu Trp Arg Ser Gly Ile His Val Asp Val Met Asp Leu

340 345 350

Pro Gln Phe Ala Asp Cys Trp Thr Asn Phe Val Asn Pro Gln Arg Pro

355 360 365

Phe Arg Pro Trp Asn Glu Leu Glu Lys Asn Ser Trp Arg Ile Gln Arg

370 375 380

Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu

385 390 395

<210> 1396

<211> 226

<212> PRT

<213> Bos taurus

<400> 1396

Asp Gly Trp Glu Val Ala Phe Arg Ser Gly Thr Val Leu Lys Ala Gly
1 5 10 15
Val Leu Gly Val Ser Met Thr Glu Gly Trp Ala Gly Ser Gly His Pro
20 25 30

Gly Gln Gly Ala Cys Val Trp Thr Pro Gly Thr Arg Asn Thr Met Asn
35 40 45
Leu Leu Arg Glu Val Leu Phe Lys Gln Gln Phe Gly Asn Gln Pro Arg
50 55 60
Val Pro Ala Pro Tyr Tyr Arg Arg Lys Thr Tyr Leu Cys Tyr Gln Leu
65 70 75 80
Lys Gln Arg Asn Asp Leu Thr Leu Asp Arg Gly Cys Phe Arg Asn Lys
85 90 95

Lys Gln Arg His Ala Glu Arg Phe Ile Asp Lys Ile Asn Ser Leu Asp
100 105 110
Leu Asn Pro Ser Gln Ser Tyr Lys Ile Ile Cys Tyr Ile Thr Trp Ser
115 120 125
Pro Cys Pro Asn Cys Ala Asn Glu Leu Val Asn Phe Ile Thr Arg Asn
130 135 140
Asn His Leu Lys Leu Glu Ile Phe Ala Ser Arg Leu Tyr Phe His Trp
145 150 155 160

Ile Lys Ser Phe Lys Met Gly Leu Gln Asp Leu Gln Asn Ala Gly Ile
165 170 175
Ser Val Ala Val Met Thr His Thr Glu Phe Glu Asp Cys Trp Glu Gln
180 185 190
Phe Val Asp Asn Gln Ser Arg Pro Phe Gln Pro Trp Asp Lys Leu Glu
195 200 205
Gln Tyr Ser Ala Ser Ile Arg Arg Arg Leu Gln Arg Ile Leu Thr Ala
210 215 220

Pro Ile
225

<210> 1397

<211> 490

<212> PRT

<213> Pan troglodytes

<400> 1397

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Trp Met Tyr Gln Arg Thr
1 5 10 15

Phe Tyr Tyr Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Arg Arg Gly His Ser Asn Leu
35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Met Tyr Ser Gln Pro Glu
50 55 60

His His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65 70 75 80

Ser Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
85 90 95

Cys Pro Asp Cys Val Ala Lys Leu Ala Lys Phe Leu Ala Glu His Pro
100 105 110

Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu
115 120 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg
130 135 140

Val Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe
145 150 155 160

Val Tyr Asn Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp
165 170 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Ile Arg His Leu
180 185 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val
195 200 205

Leu Arg Arg His Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp
210 215 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn
225 230 235 240
Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu

245 250 255
Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile
260 265 270
Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly
275 280 285
Cys Ala Gly Gln Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg
290 295 300
Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys

305 310 315 320
Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met
325 330 335
Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
340 345 350
Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
355 360 365
Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Val Arg Ala Ser Ser Leu

370 375 380
Cys Met Val Pro His Arg Pro Pro Pro Pro Pro Gln Ser Pro Gly Pro
385 390 395 400
Cys Leu Pro Leu Cys Ser Glu Pro Pro Leu Gly Ser Leu Leu Pro Thr
405 410 415
Gly Arg Pro Ala Pro Ser Leu Pro Phe Leu Leu Thr Ala Ser Phe Ser
420 425 430
Phe Pro Pro Pro Ala Ser Leu Pro Pro Leu Pro Ser Leu Ser Leu Ser

435 440 445
Pro Gly His Leu Pro Val Pro Ser Phe His Ser Leu Thr Ser Cys Ser
450 455 460
Ile Gln Pro Pro Cys Ser Ser Arg Ile Arg Glu Thr Glu Gly Trp Ala

465 470 475 480

Ser Val Ser Lys Glu Gly Arg Asp Leu Gly

485 490

<210> 1398

<211> 190

<212> PRT

<213> Homo sapiens

<400> 1398

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr

1 5 10 15

Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu

20 25 30

Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val

35 40 45

Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His

50 55 60

Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu

65 70 75 80

Ser Pro Asn Thr Lys Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro

85 90 95

Cys Pro Asp Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser

100 105 110

Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Tyr

115 120 125

Pro Cys Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala

130 135 140

Val Glu Ile Met Asp Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Asp Asn Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Thr

165 170 175

Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ser Leu Gln

180 185 190

<210> 1399

<211> 190

<212> PRT

<213> Gorilla gorilla

<400

> 1399

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
1 5 10 15

Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
20 25 30

Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
35 40 45

Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His
50 55 60

Cys His Ala Glu Arg Cys Phe Leu Ser Trp Glu Cys Asp Asp Ile Leu
65 70 75 80

Ser Pro Asn Thr Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
85 90 95

Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
100 105 110

Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Asp
115 120 125

Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
130 135 140

Val Lys Ile Met Asp Tyr Lys Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145 150 155 160

Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
165 170 175

Asn Phe Arg Phe Leu Lys Arg Arg Leu Gln Glu Ile Leu Glu
180 185 190

<210> 1400

<211> 202

<212> PRT

<213> Macaca mulatta

<400> 1400

Met Asp Gly Ser Pro Ala Ser Arg Pro Arg His Leu Met Asp Pro Asn
1 5 10 15

Thr Phe Thr Phe Asn Phe Asn Asn Asp Leu Ser Val Arg Gly Arg His
20 25 30

Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Trp
35 40 45

Val Pro Met Asp Glu Arg Arg Gly Phe Leu Cys Asn Lys Ala Lys Asn
50 55 60

Val Pro Cys Gly Asp Tyr Gly Cys His Val Glu Leu Arg Phe Leu Cys
65 70 75 80

Glu Val Pro Ser Trp Gln Leu Asp Pro Ala Gln Thr Tyr Arg Val Thr
85 90 95

Trp Phe Ile Ser Trp Ser Pro Cys Phe Arg Arg Gly Cys Ala Gly Gln
100 105 110

Val Arg Val Phe Leu Gln Glu Asn Lys His Val Arg Leu Arg Ile Phe
115 120 125

Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Gln Glu Ala Leu Arg
130 135 140

Thr Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Glu Glu
145 150 155 160

Phe Lys His Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe
165 170 175

Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg
180 185 190

Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
195 200

<210> 1401

<211> 185

<212> PRT

<213> Bos taurus

<400> 1401

Met Asp Glu Tyr Thr Phe Thr Glu Asn Phe Asn Asn Gln Gly Trp Pro
1 5 10 15

Ser Lys Thr Tyr Leu Cys Tyr Glu Met Glu Arg Leu Asp Gly Asp Ala
20 25 30

Thr Ile Pro Leu Asp Glu Tyr Lys Gly Phe Val Arg Asn Lys Gly Leu
35 40 45

Asp Gln Pro Glu Lys Pro Cys His Ala Glu Leu Tyr Phe Leu Gly Lys
50 55 60

Ile His Ser Trp Asn Leu Asp Arg Asn Gln His Tyr Arg Leu Thr Cys
65 70 75 80

Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Gln Lys Leu Thr Thr
85 90 95

Phe Leu Lys Glu Asn His His Ile Ser Leu His Ile Leu Ala Ser Arg
100 105 110

Ile Tyr Thr His Asn Arg Phe Gly Cys His Gln Ser Gly Leu Cys Glu
115 120 125

Leu Gln Ala Ala Gly Ala Arg Ile Thr Ile Met Thr Phe Glu Asp Phe
130 135 140

Lys His Cys Trp Glu Thr Phe Val Asp His Lys Gly Lys Pro Phe Gln
145 150 155 160

Pro Trp Glu Gly Leu Asn Val Lys Ser Gln Ala Leu Cys Thr Glu Leu
165 170 175

Gln Ala Ile Leu Lys Thr Gln Gln Asn
180 185

<210> 1402

<211> 200

<212> PRT

<213> Homo sapiens

<400> 1402

Met Ala Leu Leu Thr Ala Glu Thr Phe Arg Leu Gln Phe Asn Asn Lys

1 5 10 15

Arg Arg Leu Arg Arg Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr

20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Glu

35 40 45

Asn Lys Lys Lys Cys His Ala Glu Ile Cys Phe Ile Asn Glu Ile Lys

50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu

65 70 75 80

Thr Trp Ser Pro Cys Ser Ser Cys Ala Trp Glu Leu Val Asp Phe Ile

85 90 95

Lys Ala His Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr

100 105 110

Tyr His Trp Cys Lys Pro Gln Gln Lys Gly Leu Arg Leu Leu Cys Gly

115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Phe Pro Lys Phe Ala Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Glu Lys Pro Leu Ser Phe Asn Pro Tyr

145 150 155 160

Lys Met Leu Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg

165 170 175

Leu Glu Arg Ile Lys Ile Pro Gly Val Arg Ala Gln Gly Arg Tyr Met

180 185 190

Asp Ile Leu Cys Asp Ala Glu Val

195 200

<210> 1403

<211> 210

<212> PRT

<213> Macaca mulatta

<400> 1403

Met Ala Leu Leu Thr Ala Lys Thr Phe Ser Leu Gln Phe Asn Asn Lys

1 5 10 15

Arg Arg Val Asn Lys Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr

20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly His Leu Lys

35 40 45

Asn Lys Lys Lys Asp His Ala Glu Ile Arg Phe Ile Asn Lys Ile Lys

50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu

65 70 75 80

Thr Trp Ser Pro Cys Pro Ser Cys Ala Gly Glu Leu Val Asp Phe Ile

85 90 95

Lys Ala His Arg His Leu Asn Leu Arg Ile Phe Ala Ser Arg Leu Tyr

100 105 110

Tyr His Trp Arg Pro Asn Tyr Gln Glu Gly Leu Leu Leu Leu Cys Gly

115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Leu Pro Glu Phe Thr Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Lys Glu Pro Pro Ser Phe Asn Pro Ser

145 150 155 160

Glu Lys Leu Glu Glu Leu Asp Lys Asn Ser Gln Ala Ile Lys Arg Arg

165 170 175

Leu Glu Arg Ile Lys Ser Arg Ser Val Asp Val Leu Glu Asn Gly Leu

180 185 190

Arg Ser Leu Gln Leu Gly Pro Val Thr Pro Ser Ser Ser Ile Arg Asn

195 200 205

Ser Arg

210

<210> 1404

<211> 386

<212

> PRT

<213> Homo sapiens

<400> 1404

```
Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1               5                  10                  15
Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
            20                  25                  30
Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
        35                  40                  45
Leu Trp Asp Thr Gly Val Phe Arg Gly Pro Val Leu Pro Lys Arg Gln
    50                  55                  60

Ser Asn His Arg Gln Glu Val Tyr Phe Arg Phe Glu Asn His Ala Glu
65                  70                  75                  80
Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Arg Leu Pro Ala Asn Arg
                85                  90                  95
Arg Phe Gln Ile Thr Trp Phe Val Ser Trp Asn Pro Cys Leu Pro Cys
            100                 105                 110
Val Val Lys Val Thr Lys Phe Leu Ala Glu His Pro Asn Val Thr Leu
        115                 120                 125

Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Arg Asp Arg Asp Trp Arg
    130                 135                 140
Trp Val Leu Leu Arg Leu His Lys Ala Gly Ala Arg Val Lys Ile Met
145                 150                 155                 160
Asp Tyr Glu Asp Phe Ala Tyr Cys Trp Glu Asn Phe Val Cys Asn Glu
                165                 170                 175
Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn Tyr Ala Ser
            180                 185                 190

Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met Glu Ala Met
        195                 200                 205
Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Leu Lys Ala Cys
    210                 215                 220
Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val Thr Lys His
225                 230                 235                 240
His Ser Ala Val Phe Arg Lys Arg Gly Val Phe Arg Asn Gln Val Asp
```

245 250 255

Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys
260 265 270
Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr
275 280 285
Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu
290 295 300
Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Cys
305 310 315 320

Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Cys Ser Leu Ser Gln
325 330 335
Glu Gly Ala Ser Val Lys Ile Met Gly Tyr Lys Asp Phe Val Ser Cys
340 345 350
Trp Lys Asn Phe Val Tyr Ser Asp Asp Glu Pro Phe Lys Pro Trp Lys
355 360 365
Gly Leu Gln Thr Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ile
370 375 380

Leu Gln
385

<210> 1405
<211> 236
<212> PRT
<213> Homo sapiens
<400> 1405

Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
35 40 45
Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val

50 55 60

Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met

65 70 75 80

Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys

85 90 95

Ser Gln Ala Ile Arg Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu

100 105 110

Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160

Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met

165 170 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys

180 185 190

Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu

195 200 205

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu

210 215 220

Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg

225 230 235

<210> 1406

<211> 229

<212> PRT

<213> Mus musculus

<400> 1406

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Val Trp Arg His Thr Ser Gln Asn Thr Ser Asn His Val Glu Val
50 55 60
Asn Phe Leu Glu Lys Phe Thr Thr Glu Arg Tyr Phe Arg Pro Asn Thr

65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg His Pro Tyr Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140
Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160
Pro Ser Asn Glu Ala Tyr Trp Pro Arg Tyr Pro His Leu Trp Val Lys
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Lys Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205
Thr Leu Gln Thr Cys His Tyr Gln Arg Ile Pro Pro His Leu Leu Trp
210 215 220
Ala Thr Gly Leu Lys
225

<210> 1407
<211> 224
<212> PRT
<213> Homo sapiens
<400> 1407
Met Ala Gln Lys Glu Glu Ala Ala Val Ala Thr Glu Ala Ala Ser Gln

1 5 10 15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro

35 40 45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly

50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly

65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala

85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp

100 105 110

Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys

115 120 125

Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn

130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro

145 150 155 160

Glu Ile Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu

165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val

180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile

195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 1408

<211> 224

<212> PRT

<213> Mus musculus

<400> 1408

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln
1 5 10 15
Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
20 25 30
Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro
35 40 45

Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60
Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Val Gln Ser Lys Gly
65 70 75 80
Gly Gln Ala Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly
85 90 95
Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110

Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125
Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140
Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160
Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190
Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205
Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 1409

<211> 224

<212> PRT

<213> Rattus norvegicus

<400> 1409

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln

1 5 10 15

Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro

35 40 45

Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly

50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly

65 70 75 80

Gly Gln Val Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly

85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp

100 105 110

Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys

115 120 125

Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn

130 135 140

Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro

145 150 155 160

Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu

165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Leu Trp Gln Asn Phe Val

180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile

195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 1410

<211> 208

<212> PRT

<213> Petromyzon marinus

<400> 1410

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr
1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
20 25 30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65 70 75 80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
85 90 95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
130 135 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145 150 155 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
180 185 190

Phe Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205

<210> 1411

<211> 381

<212> PRT

<213> Homo sapiens

<400> 1411

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr

20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro

35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr

50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu

65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro

85 90 95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro

100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp

115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly

130 135 140

Pro Arg Ala Thr Met Lys Phe Asn Tyr Asp Glu Phe Gln His Cys Trp

145 150 155 160

Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn Asn

165 170 175

Leu Pro Lys Tyr Tyr Ile Leu Leu His Phe Met Leu Gly Glu Ile Leu

180 185 190

Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu

195 200 205

Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu

210 215 220

Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe
225 230 235 240
Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His
245 250 255
Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu

260 265 270
Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe
275 280 285
Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Lys His Val Ser
290 295 300
Leu Cys Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln
305 310 315 320
Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr

325 330 335
Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly
340 345 350
Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu
355 360 365
Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
370 375 380

<210> 1412
<211> 182
<212> PRT
<213> Homo sapiens
<400> 1412

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Trp
1 5 10 15
Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His Asn
20 25 30
Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn Gln
35 40 45
Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu Cys

50 55 60

Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp Tyr
65 70 75 80
Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln
85 90 95
Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys Ile
100 105 110
Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly Leu
115 120 125

Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr Tyr Ser Glu
130 135 140
Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe
145 150 155 160
Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser Gly Arg
165 170 175
Leu Arg Ala Ile Leu Gln
180

<210> 1413
<211> 184
<212> PRT
<213> Homo sapiens
<400
> 1413

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val
1 5 10 15
Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His
20 25 30
Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn
35 40 45
Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu
50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp

65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala

85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys

100 105 110

Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln Glu Gly

115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Met Thr Tyr

130 135 140

Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys

145 150 155 160

Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser

165 170 175

Gly Arg Leu Arg Ala Ile Leu Gln

180

<210> 1414

<211> 367

<212> PRT

<213> Homo sapiens

<400> 1414

Met Glu Pro Ile Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val

1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro

20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe

35 40 45

Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Phe Pro Gln Thr Lys

50 55 60

His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln

65 70 75 80

Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser

85 90 95

Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn
100 105 110
Asp Ser Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asn
115 120 125

Glu Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Ile Thr
130 135 140
Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160
Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
165 170 175
Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly
180 185 190

Ile Trp His Ser Val Leu His Ser Phe Ile Ser Gly Val Ser Gly Ser
195 200 205
His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Ala Asp Arg His
210 215 220
Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Tyr Phe Thr
225 230 235 240
Asp Val Leu Leu Gln Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Glu
245 250 255

Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Phe Phe
260 265 270
Gln Met Pro Ser Gly Gln Leu Gln Pro Asn Leu Pro Pro Asp Leu Arg
275 280 285
Ala Pro Val Val Phe Val Leu Val Pro Leu Arg Asp Leu Pro Pro Met
290 295 300
His Met Gly Gln Asn Pro Asn Lys Pro Arg Asn Ile Val Arg His Leu
305 310 315 320

Asn Met Pro Gln Met Ser Phe Gln Glu Thr Lys Asp Leu Gly Arg Leu
325 330 335
Pro Thr Gly Arg Ser Val Glu Ile Val Glu Ile Thr Glu Gln Phe Ala

340 345 350

Ser Ser Lys Glu Ala Asp Glu Lys Lys Lys Lys Lys Gly Lys Lys

355 360 365

<210> 1415

<211> 388

<212> PRT

<213> Rattus norvegicus

<400> 1415

Met Glu Pro Leu Tyr Glu Glu Tyr Leu Thr His Ser Gly Thr Ile Val

1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Val Ser Leu Asn Cys Thr Asn Cys Pro

20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Pro Tyr Thr Glu Phe

35 40 45

His Gln Thr Phe Gly Phe Pro Trp Ser Thr Tyr Pro Gln Thr Lys His

50 55 60

Leu Thr Phe Tyr Glu Leu Arg Ser Ser Ser Gly Asn Leu Ile Gln Lys

65 70 75 80

Gly Leu Ala Ser Asn Cys Thr Gly Ser His Thr His Pro Glu Ser Met

85 90 95

Leu Phe Glu Arg Asp Gly Tyr Leu Asp Ser Leu Ile Phe His Asp Ser

100 105 110

Asn Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asp Glu

115 120 125

Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Met Asn Tyr

130 135 140

Pro Glu Val Thr Leu Ser Val Phe Phe Ser Gln Leu Tyr His Thr Glu

145 150 155 160

Asn Gln Phe Pro Thr Ser Ala Trp Asn Arg Glu Ala Leu Arg Gly Leu

165 170 175

Ala Ser Leu Trp Pro Gln Val Thr Leu Ser Ala Ile Ser Gly Gly Ile

180 185 190

Trp Gln Ser Ile Leu Glu Thr Phe Val Ser Gly Ile Ser Glu Gly Leu

195 200 205

Thr Ala Val Arg Pro Phe Thr Ala Gly Arg Thr Leu Thr Asp Arg Tyr

210 215 220

Asn Ala Tyr Glu Ile Asn Cys Ile Thr Glu Val Lys Pro Tyr Phe Thr

225 230 235 240

Asp Ala Leu His Ser Trp Gln Lys Glu Asn Gln Asp Gln Lys Val Trp

245 250 255

Ala Ala Ser Glu Asn Gln Pro Leu His Asn Thr Thr Pro Ala Gln Trp

260 265 270

Gln Pro Asp Met Ser Gln Asp Cys Arg Thr Pro Ala Val Phe Met Leu

275 280 285

Val Pro Tyr Arg Asp Leu Pro Pro Ile His Val Asn Pro Ser Pro Gln

290 295 300

Lys Pro Arg Thr Val Val Arg His Leu Asn Thr Leu Gln Leu Ser Ala

305 310 315 320

Ser Lys Val Lys Ala Leu Arg Lys Ser Pro Ser Gly Arg Pro Val Lys

325 330 335

Lys Glu Glu Ala Arg Lys Gly Ser Thr Arg Ser Gln Glu Ala Asn Glu

340 345 350

Thr Asn Lys Ser Lys Trp Lys Lys Gln Thr Leu Phe Ile Lys Ser Asn

355 360 365

Ile Cys His Leu Leu Glu Arg Glu Gln Lys Lys Ile Gly Ile Leu Ser

370 375 380

Ser Trp Ser Val

385

<210> 1416

<211> 363

<212> PRT

<213> Macaca fascicularis

<400> 1416

Met Glu Pro Thr Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val
1 5 10 15
Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro
20 25 30
Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe
35 40 45
Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Tyr Pro Gln Thr Lys

50 55 60
His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln
65 70 75 80
Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser
85 90 95
Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn
100 105 110
Asp Ser Ile Arg His Ile Ile Leu Tyr Cys Asn Asn Ser Pro Cys Asn

115 120 125
Glu Ala Asn His Cys Cys Ile Ser Lys Val Tyr Asn Phe Leu Ile Thr
130 135 140
Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160
Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
165 170 175
Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly

180 185 190
Ile Trp His Ser Val Leu His Ser Phe Val Ser Gly Val Ser Gly Ser
195 200 205
His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Thr Asp Arg Tyr
210 215 220
Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Phe Phe Thr
225 230 235 240
Asp Val Leu Leu His Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Met

245 250 255

Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Ser Phe

260 265 270

Gln Met Thr Ser Gly Ile Pro Pro Asp Leu Arg Ala Pro Val Val Phe

275 280 285

Val Leu Leu Pro Leu Arg Asp Leu Pro Pro Met His Met Gly Gln Asp

290 295 300

Pro Asn Lys Pro Arg Asn Ile Ile Arg His Leu Asn Met Pro Gln Met

305 310 315 320

Ser Phe Gln Glu Thr Lys Asp Leu Glu Arg Leu Pro Thr Arg Arg Ser

325 330 335

Val Glu Thr Val Glu Ile Thr Glu Arg Phe Ala Ser Ser Lys Gln Ala

340 345 350

Glu Glu Lys Thr Lys Lys Lys Lys Gly Lys Lys

355 360

<210> 1417

<211> 224

<212> PRT

<213> Petromyzon marinus

<400> 1417

Met Ala Gly Tyr Glu Cys Val Arg Val Ser Glu Lys Leu Asp Phe Asp

1 5 10 15

Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His

20 25 30

Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg

35 40 45

Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His

50 55 60

Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn

65 70 75 80

Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala

85 90 95

Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu
100 105 110
Gln Gly His Thr Leu Thr Met His Phe Ser Arg Ile Tyr Asp Arg Asp
115 120 125
Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn

130 135 140
Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu
145 150 155 160
Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr
165 170 175
Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu
180 185 190
Val Arg Cys Ala Gly Met Arg Glu Ser Gly Ile Pro Leu His Leu Phe

195 200 205
Thr Leu Gln Thr Pro Leu Leu Ser Gly Arg Val Val Trp Trp Arg Val
210 215 220
<210> 1418
<211> 331
<212> PRT
<213> Petromyzon marinus
<400> 1418
Met Glu Leu Arg Glu Val Val Asp Cys Ala Leu Ala Ser Cys Val Arg
1 5 10 15
His Glu Pro Leu Ser Arg Val Ala Phe Leu Arg Cys Phe Ala Ala Pro
20 25 30

Ser Gln Lys Pro Arg Gly Thr Val Ile Leu Phe Tyr Val Glu Gly Ala
35 40 45
Gly Arg Gly Val Thr Gly Gly His Ala Val Asn Tyr Asn Lys Gln Gly
50 55 60
Thr Ser Ile His Ala Glu Val Leu Leu Leu Ser Ala Val Arg Ala Ala
65 70 75 80
Leu Leu Arg Arg Arg Arg Cys Glu Asp Gly Glu Glu Ala Thr Arg Gly

85 90 95

Cys Thr Leu His Cys Tyr Ser Thr Tyr Ser Pro Cys Arg Asp Cys Val
100 105 110
Glu Tyr Ile Gln Glu Phe Gly Ala Ser Thr Gly Val Arg Val Val Ile
115 120 125
His Cys Cys Arg Leu Tyr Glu Leu Asp Val Asn Arg Arg Arg Ser Glu
130 135 140
Ala Glu Gly Val Leu Arg Ser Leu Ser Arg Leu Gly Arg Asp Phe Arg
145 150 155 160

Leu Met Gly Pro Arg Asp Ala Ile Ala Leu Leu Leu Gly Gly Arg Leu
165 170 175
Ala Asn Thr Ala Asp Gly Glu Ser Gly Ala Ser Gly Asn Ala Trp Val
180 185 190
Thr Glu Thr Asn Val Val Glu Pro Leu Val Asp Met Thr Gly Phe Gly
195 200 205
Asp Glu Asp Leu His Ala Gln Val Gln Arg Asn Lys Gln Ile Arg Glu
210 215 220

Ala Tyr Ala Asn Tyr Ala Ser Ala Val Ser Leu Met Leu Gly Glu Leu
225 230 235 240
His Val Asp Pro Asp Lys Phe Pro Phe Leu Ala Glu Phe Leu Ala Gln
245 250 255
Thr Ser Val Glu Pro Ser Gly Thr Pro Arg Glu Thr Arg Gly Arg Pro
260 265 270
Arg Gly Ala Ser Ser Arg Gly Pro Glu Ile Gly Arg Gln Arg Pro Ala
275 280 285

Asp Phe Glu Arg Ala Leu Gly Ala Tyr Gly Leu Phe Leu His Pro Arg
290 295 300
Ile Val Ser Arg Glu Ala Asp Arg Glu Glu Ile Lys Arg Asp Leu Ile
305 310 315 320
Val Val Met Arg Lys His Asn Tyr Gln Gly Pro
325 330

<210> 1419

<211> 209

<212> PRT

<213> Petromyzon marinus

<400> 1419

Met Ala Gly Asp Glu Asn Val Arg Val Ser Glu Lys Leu Asp Phe Asp
1 5 10 15

Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His
20 25 30

Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg
35 40 45

Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His
50 55 60

Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn
65 70 75 80

Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala
85 90 95

Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu
100 105 110

Gln Gly His Thr Leu Met Met His Phe Ser Arg Ile Tyr Asp Arg Asp
115 120 125

Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn
130 135 140

Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu
145 150 155 160

Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr
165 170 175

Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu
180 185 190

Val Arg Cys Ala Gly Met Arg Glu Ser Gly Met Pro Leu His Leu Phe
195 200 205

Thr

<210> 1420

<211> 158

<212> PRT

<213> Saccharomyces cerevisiae

<400> 1420

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1 5 10 15
Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro
20 25 30
Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
35 40 45
Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60
Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65 70 75 80
Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
85 90 95
Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val
100 105 110
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125
Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe
130 135 140
Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
145 150 155

<210> 1421

<211> 218

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 1421

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu

180 185 190

Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro

195 200 205

Pro His Ile Leu Trp Ala Thr Gly Leu Lys

210 215

<210> 1422

<211> 218

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1422

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln His Tyr Gln Arg Leu Pro
195 200 205

Pro His Ile Leu Trp Ala Thr Gly Leu Lys

210 215

<210> 1423

<211> 1367

<212

> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 1423

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Ala
485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Ala Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Ala Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Ala Ile Thr Lys
915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 1424

<211> 23

<212> PRT

<213> Unknown

<220><223> Description of Unknown:
nuclear localization sequence
<400> 1424
Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys
1 5 10 15
Pro Lys Lys Lys Arg Lys Val
20
<210> 1425
<211> 8
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 1425
Ser Gly Gly Ser Ser Gly Gly Ser

1 5
<210> 1426
<211> 1609
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1426
Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45
Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Thr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160
Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Glu Ile Gly Lys Ala Thr Ala Lys Tyr
195 200 205
Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
210 215 220
Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

225 230 235 240
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
245 250 255
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val
260 265 270
Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser
275 280 285
Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly

290 295 300
Gly Phe Met Gln Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys
305 310 315 320
Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu

325 330 335

Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

340 345 350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile

355 360 365

Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

370 375 380

Met Leu Ala Ser Ala Lys Phe Leu Gln Lys Gly Asn Glu Leu Ala Leu

385 390 395 400

Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

405 410 415

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

420 425 430

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe

435 440 445

Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser

450 455 460

Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

465 470 475 480

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Arg Ala Phe

485 490 495

Lys Tyr Phe Asp Thr Thr Ile Ala Arg Lys Glu Tyr Arg Ser Thr Lys

500 505 510

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr

515 520 525

Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Gly Gly Ser Gly

530 535 540

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

545 550 555 560

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

565 570 575

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
580 585 590
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
595 600 605
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

610 615 620
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
625 630 635 640
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
645 650 655
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
660 665 670
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

675 680 685
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
690 695 700
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
705 710 715 720
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
725 730 735
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

740 745 750
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
755 760 765
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
770 775 780
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
785 790 795 800
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

805 810 815
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

820 825 830

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

835 840 845

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

850 855 860

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

865 870 875 880

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

885 890 895

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

900 905 910

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

915 920 925

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

930 935 940

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

945 950 955 960

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

965 970 975

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

980 985 990

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

995 1000 1005

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala

1010 1015 1020

Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe

1025 1030 1035

Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu

1040 1045 1050

Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

1055 1060 1065

Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu
1070 1075 1080
Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
1085 1090 1095
Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe
1100 1105 1110
Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr

1115 1120 1125
Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val
1130 1135 1140
Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu
1145 1150 1155
Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu
1160 1165 1170
Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp

1175 1180 1185
Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe
1190 1195 1200
Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly
1205 1210 1215
Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
1220 1225 1230
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

1235 1240 1245
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr
1250 1255 1260
Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp
1265 1270 1275
Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile
1280 1285 1290
Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

1295 1300 1305

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met

1310 1315 1320

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

1325 1330 1335

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser

1340 1345 1350

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn

1355 1360 1365

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr

1370 1375 1380

Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val

1385 1390 1395

Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp

1400 1405 1410

Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp

1415 1420 1425

Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp

1430 1435 1440

Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp

1445 1450 1455

Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

1460 1465 1470

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

1475 1480 1485

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr

1490 1495 1500

Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu

1505 1510 1515

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr

1520 1525 1530

Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr

1535 1540 1545

Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys

1550 1555 1560

Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val

1565 1570 1575

Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Gly Ala Asp Lys Arg

1580 1585 1590

Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys

1595 1600 1605

Val

<210> 1427

<211> 1807

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1427

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro

35 40 45

Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Gly Gly Ser Ser Gly Gly Ser Ser Gly

165 170 175

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

180 185 190

Gly Ser Ser Gly Gly Ser Met Ser Glu Val Glu Phe Ser His Glu Tyr

195 200 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg

210 215 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

225 230 235 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala

245 250 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

260 265 270

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys

275 280 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val

290 295 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

305 310 315 320

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

325 330 335

Asp Glu Cys Ala Ala Leu Leu Cys Thr Phe Phe Arg Met Pro Arg Gln

340 345 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly

355 360 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

370 375 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Glu Ile Gly

385 390 395 400

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe

405 410 415

Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

420 425 430

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg

435 440 445

Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

450 455 460

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

465 470 475 480

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp

485 490 495

Asp Pro Lys Lys Tyr Gly Gly Phe Met Gln Pro Thr Val Ala Tyr Ser

500 505 510

Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

515 520 525

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

530 535 540

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

545 550 555 560

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu

565 570 575

Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Lys Phe Leu Gln Lys

580 585 590

Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu
595 600 605
Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln

610 615 620
Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
625 630 635 640
Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn
645 650 655
Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
660 665 670
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu

675 680 685
Gly Ala Pro Arg Ala Phe Lys Tyr Phe Asp Thr Thr Ile Ala Arg Lys
690 695 700
Glu Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
705 710 715 720
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly
725 730 735
Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

740 745 750
Ser Gly Gly Ser Gly Gly Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala
755 760 765
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
770 775 780
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
785 790 795 800
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

805 810 815
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
820 825 830
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

835 840 845

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

850 855 860

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

865 870 875 880

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

885 890 895

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

900 905 910

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

915 920 925

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

930 935 940

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

945 950 955 960

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

965 970 975

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

980 985 990

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

995 1000 1005

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln

1010 1015 1020

Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala

1025 1030 1035

Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn

1040 1045 1050

Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr

1055 1060 1065

Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr

1070 1075 1080

Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
1085 1090 1095
Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser
1100 1105 1110
Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu

1115 1120 1125
Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
1130 1135 1140
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
1145 1150 1155
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His
1160 1165 1170
Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr

1175 1180 1185
Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr
1190 1195 1200
Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
1205 1210 1215
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro
1220 1225 1230
Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

1235 1240 1245
Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
1250 1255 1260
Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val
1265 1270 1275
Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg
1280 1285 1290
Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp

1295 1300 1305
Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys

1310 1315 1320

Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile

1325 1330 1335

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His

1340 1345 1350

Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu

1355 1360 1365

Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

1370 1375 1380

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

1385 1390 1395

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg

1400 1405 1410

Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile

1415 1420 1425

Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser

1430 1435 1440

Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

1445 1450 1455

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly

1460 1465 1470

Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

1475 1480 1485

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp

1490 1495 1500

Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val

1505 1510 1515

Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys

1520 1525 1530

Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu

1535 1540 1545

Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln

1550 1555 1560

Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg

1565 1570 1575

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp

1580 1585 1590

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp

1595 1600 1605

Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

1610 1615 1620

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

1625 1630 1635

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg

1640 1645 1650

Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu

1655 1660 1665

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg

1670 1675 1680

Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn

1685 1690 1695

Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val

1700 1705 1710

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1715 1720 1725

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1730 1735 1740

Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys

1745 1750 1755

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1760 1765 1770

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Gly Ala

1775 1780 1785

Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys

1790 1795 1800

Lys Arg Lys Val

1805

<210> 1428

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1428

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300
Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365
Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro

565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln

610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670
Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685
Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700
Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735
Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750
Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800
Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815
Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880
Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940
Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945 950 955 960
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
980 985 990
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
995 1000 1005
Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1010 1015 1020
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His
1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1040 1045 1050
Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110
Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125
Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170
Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1280 1285 1290
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305
Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala
1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1520 1525 1530

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1429

<211> 1786

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1429

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45

Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190

Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
195 200 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg

210 215 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225 230 235 240
Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
245 250 255
Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
260 265 270
Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
275 280 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
290 295 300
Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305 310 315 320
His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
325 330 335
Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg
340 345 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
355 360 365
Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
370 375 380
Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys
385 390 395 400
Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
405 410 415

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
420 425 430
Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
435 440 445
Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
450 455 460

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
465 470 475 480

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
485 490 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
500 505 510

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
515 520 525

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
530 535 540

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
545 550 555 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
565 570 575

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
580 585 590

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
595 600 605

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
610 615 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
625 630 635 640

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
645 650 655

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
660 665 670

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
675 680 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
690 695 700

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg

705 710 715 720

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg

725 730 735

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys

740 745 750

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe

755 760 765

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu

770 775 780

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr

785 790 795 800

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His

805 810 815

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn

820 825 830

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val

835 840 845

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys

850 855 860

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys

865 870 875 880

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys

885 890 895

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu

900 905 910

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu

915 920 925

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile

930 935 940

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu

945 950 955 960

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
965 970 975
Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
980 985 990

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
995 1000 1005
Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu
1010 1015 1020
Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu
1025 1030 1035
Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
1040 1045 1050

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
1055 1060 1065
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly
1070 1075 1080
Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu
1085 1090 1095
Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
1100 1105 1110

Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
1115 1120 1125
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu
1130 1135 1140
Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
1145 1150 1155
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
1160 1165 1170

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly
1175 1180 1185
Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln

1190 1195 1200

Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met

1205 1210 1215

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1220 1225 1230

Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

1235 1240 1245

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser

1250 1255 1260

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr

1265 1270 1275

Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

1280 1285 1290

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

1295 1300 1305

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile

1310 1315 1320

Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

1325 1330 1335

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr

1340 1345 1350

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe

1355 1360 1365

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala

1370 1375 1380

Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro

1385 1390 1395

Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1400 1405 1410

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1415 1420 1425

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1430 1435 1440
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1445 1450 1455
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1460 1465 1470

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1475 1480 1485
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1490 1495 1500
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1505 1510 1515
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1520 1525 1530

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1535 1540 1545
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1550 1555 1560
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1565 1570 1575
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1580 1585 1590

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1595 1600 1605
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1610 1615 1620
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1625 1630 1635
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1640 1645 1650

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1655 1660 1665

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1670 1675 1680

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1685 1690 1695

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1700 1705 1710

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1715 1720 1725

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1730 1735 1740

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1745 1750 1755

Ser Gln Leu Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp

1760 1765 1770

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1775 1780 1785

<210> 1430

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1430

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300
Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620
Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685
Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700
Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750
Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1085 1090 1095
Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125
Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1130 1135 1140
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1145 1150 1155
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305
Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala
1325 1330 1335
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1415 1420 1425
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1520 1525 1530

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1431

<211> 1786

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1431

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45

Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145 150 155 160
Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
195 200 205
Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
210 215 220
Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225 230 235 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
245 250 255
Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
260 265 270
Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
275 280 285
Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
290 295 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305 310 315 320
His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
325 330 335
Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg
340 345 350
Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly

355 360 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
370 375 380
Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys
385 390 395 400
Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
405 410 415
Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
420 425 430

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
435 440 445
Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
450 455 460
Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
465 470 475 480
Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
485 490 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
500 505 510
Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
515 520 525
Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
530 535 540
Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
545 550 555 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
565 570 575
Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
580 585 590
Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
595 600 605

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
610 615 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
625 630 635 640
Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
645 650 655
Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
660 665 670
Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
675 680 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
690 695 700
Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
705 710 715 720
Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
725 730 735
Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys
740 745 750

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
755 760 765
Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
770 775 780
Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
785 790 795 800
Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
805 810 815

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
820 825 830
Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
835 840 845
Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys

```
    850                 855                 860
Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
865                 870                 875                 880

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                885                 890                 895
Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            900                 905                 910
Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        915                 920                 925
Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    930                 935                 940

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
945                 950                 955                 960
Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                965                 970                 975
Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            980                 985                 990
Leu Leu Lys Ile Ile Lys Asp Lys  Asp Phe Leu Asp Asn  Glu Glu Asn
        995                 1000                1005

Glu Asp  Ile Leu Glu Asp Ile  Val Leu Thr Leu Thr  Leu Phe Glu
    1010                1015                1020
Asp Arg  Glu Met Ile Glu Glu  Arg Leu Lys Thr Tyr  Ala His Leu
    1025                1030                1035
Phe Asp  Asp Lys Val Met Lys  Gln Leu Lys Arg Arg  Arg Tyr Thr
    1040                1045                1050
Gly Trp  Gly Arg Leu Ser Arg  Lys Leu Ile Asn Gly  Ile Arg Asp
    1055                1060                1065

Lys Gln  Ser Gly Lys Thr Ile  Leu Asp Phe Leu Lys  Ser Asp Gly
    1070                1075                1080
Phe Ala  Asn Arg Asn Phe Met  Gln Leu Ile His Asp  Asp Ser Leu
    1085                1090                1095
```

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
1100 1105 1110
Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
1115 1120 1125

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu
1130 1135 1140
Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
1145 1150 1155
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
1160 1165 1170
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly
1175 1180 1185

Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln
1190 1195 1200
Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met
1205 1210 1215
Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp
1220 1225 1230
Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
1235 1240 1245

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser
1250 1255 1260
Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr
1265 1270 1275
Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
1280 1285 1290
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
1295 1300 1305

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile
1310 1315 1320
Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

1325 1330 1335

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr

1340 1345 1350

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe

1355 1360 1365

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala

1370 1375 1380

Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro

1385 1390 1395

Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1400 1405 1410

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1415 1420 1425

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1430 1435 1440

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1445 1450 1455

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1460 1465 1470

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1475 1480 1485

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1490 1495 1500

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1505 1510 1515

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1520 1525 1530

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1535 1540 1545

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1550 1555 1560

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1565 1570 1575
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1580 1585 1590
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1595 1600 1605

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1610 1615 1620
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1625 1630 1635
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1640 1645 1650
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1655 1660 1665

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1670 1675 1680
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1685 1690 1695
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1700 1705 1710
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1715 1720 1725

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1730 1735 1740
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1745 1750 1755
Ser Gln Leu Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp
1760 1765 1770
Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1775 1780 1785

<210>

1432

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1432

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Ser Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415
Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735
Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800
Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
835 840 845
Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
900 905 910
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1280 1285 1290
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala
1325 1330 1335
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410
Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1520 1525 1530
Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1433
<211> 1786
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1433
Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro

35 40 45

Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr

195 200 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg

210 215 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

225 230 235 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala

245 250 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

260 265 270

Leu Ile Asp Ala Thr Leu Tyr Ser Thr Phe Glu Pro Cys Val Met Cys

275 280 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val

290 295 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

305 310 315 320

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

325 330 335

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Arg

340 345 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly

355 360 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

370 375 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys

385 390 395 400

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val

405 410 415

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly

420 425 430

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu

435 440 445

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala

450 455 460

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu

465 470 475 480

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg

485 490 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His

500 505 510

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr

515 520 525

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys

530 535 540

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe

545 550 555 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp

565 570 575

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe

580 585 590

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu

595 600 605

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln

610 615 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu

625 630 635 640

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu

645 650 655

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp

660 665 670

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala

675 680 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val

690 695 700

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg

705 710 715 720

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg

725 730 735

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys

740 745 750

```
Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        755                 760                 765
Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu


    770                 775                 780
Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
785                 790                 795                 800
Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                805                 810                 815
Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            820                 825                 830
Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val


        835                 840                 845
Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    850                 855                 860
Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
865                 870                 875                 880
Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                885                 890                 895
Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu


            900                 905                 910
Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        915                 920                 925
Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    930                 935                 940
Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
945                 950                 955                 960
Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile


                965                 970                 975
Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            980                 985                 990
Leu Leu Lys Ile Ile Lys Asp Lys  Asp Phe Leu Asp Asn  Glu Glu Asn
```

995 1000 1005

Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu

1010 1015 1020

Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu

1025 1030 1035

Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

1040 1045 1050

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

1055 1060 1065

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly

1070 1075 1080

Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu

1085 1090 1095

Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly

1100 1105 1110

Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

1115 1120 1125

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu

1130 1135 1140

Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

1145 1150 1155

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser

1160 1165 1170

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly

1175 1180 1185

Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln

1190 1195 1200

Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met

1205 1210 1215

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1220 1225 1230

Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

1235 1240 1245

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser

1250 1255 1260

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr

1265 1270 1275

Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

1280 1285 1290

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

1295 1300 1305

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile

1310 1315 1320

Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

1325 1330 1335

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr

1340 1345 1350

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe

1355 1360 1365

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala

1370 1375 1380

Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro

1385 1390 1395

Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1400 1405 1410

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1415 1420 1425

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1430 1435 1440

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1445 1450 1455

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1460 1465 1470

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1475 1480 1485

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1490 1495 1500

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1505 1510 1515

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1520 1525 1530

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1535 1540 1545

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1550 1555 1560

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1565 1570 1575

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1580 1585 1590

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1595 1600 1605

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1610 1615 1620

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1625 1630 1635

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1640 1645 1650

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1655 1660 1665

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1670 1675 1680

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1685 1690 1695

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1700 1705 1710

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1715 1720 1725

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1730 1735 1740

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1745 1750 1755

Ser Gln Leu Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp

1760 1765 1770

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1775 1780 1785

<210> 1434

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1434

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu

65 70 75 80

Tyr Ser Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala

195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys

210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser

225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro

565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln

610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880
Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940
Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945 950 955 960
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
980 985 990
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
995 1000 1005
Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1010 1015 1020
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035
Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1040 1045 1050
Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1250 1255 1260

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1520 1525 1530

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1435

<211> 1786

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1435

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45

Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
195 200 205
Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
210 215 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225 230 235 240
Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
245 250 255
Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
260 265 270
Leu Tyr Asp Ala Thr Leu Tyr Ser Thr Phe Glu Pro Cys Val Met Cys
275 280 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
290 295 300
Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305 310 315 320
His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
325 330 335
Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg
340 345 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
355 360 365
Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
370 375 380
Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys
385 390 395 400

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
405 410 415

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
420 425 430

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
435 440 445

Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
450 455 460

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
465 470 475 480

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
485 490 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
500 505 510

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
515 520 525

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
530 535 540

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
545 550 555 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
565 570 575

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
580 585 590

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
595 600 605

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
610 615 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
625 630 635 640

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu

645 650 655

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp

660 665 670

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala

675 680 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val

690 695 700

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg

705 710 715 720

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg

725 730 735

Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys

740 745 750

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe

755 760 765

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu

770 775 780

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr

785 790 795 800

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His

805 810 815

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn

820 825 830

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val

835 840 845

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys

850 855 860

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys

865 870 875 880

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys

885 890 895

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
900 905 910
Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
915 920 925

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
930 935 940
Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
945 950 955 960
Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
965 970 975
Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
980 985 990

Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn
995 1000 1005
Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu
1010 1015 1020
Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu
1025 1030 1035
Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
1040 1045 1050

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
1055 1060 1065
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly
1070 1075 1080
Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu
1085 1090 1095
Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
1100 1105 1110

Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
1115 1120 1125
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu

1130 1135 1140

Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

1145 1150 1155

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser

1160 1165 1170

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly

1175 1180 1185

Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln

1190 1195 1200

Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met

1205 1210 1215

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1220 1225 1230

Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

1235 1240 1245

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser

1250 1255 1260

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr

1265 1270 1275

Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

1280 1285 1290

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

1295 1300 1305

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile

1310 1315 1320

Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

1325 1330 1335

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr

1340 1345 1350

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe

1355 1360 1365

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala
1370 1375 1380
Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro
1385 1390 1395
Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp
1400 1405 1410

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala
1415 1420 1425
Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1430 1435 1440
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1445 1450 1455
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1460 1465 1470

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1475 1480 1485
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1490 1495 1500
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1505 1510 1515
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1520 1525 1530

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1535 1540 1545
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1550 1555 1560
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1565 1570 1575
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1580 1585 1590

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1595 1600 1605

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1610 1615 1620

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1625 1630 1635

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1640 1645 1650

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1655 1660 1665

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1670 1675 1680

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1685 1690 1695

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1700 1705 1710

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1715 1720 1725

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1730 1735 1740

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1745 1750 1755

Ser Gln Leu Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp

1760 1765 1770

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1775 1780 1785

<210> 1436

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1436

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45
Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Thr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His
1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305
Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala
1325 1330 1335
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1415 1420 1425
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1437

<211> 1588

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 1437

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Arg Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160
Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415
Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
450 455 460
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480
Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525
Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605
Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670
Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700
Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735
Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1040 1045 1050
Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1085 1090 1095
Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110
Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1130 1135 1140
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1145 1150 1155
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170
Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530
Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1438
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 1438

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45
Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Ser Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala

195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys

210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser

225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365
Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415
Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430
Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480
Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495
Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525
Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540
Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560
Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln

610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880
Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940
Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945 950 955 960
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
980 985 990
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
995 1000 1005
Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1010 1015 1020
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035
Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1040 1045 1050
Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1250 1255 1260

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln

1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1439

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1439

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300
Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365
Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
900 905 910
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940
Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945 950 955 960
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
965 970 975
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
995 1000 1005
Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1010 1015 1020
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His
1025 1030 1035
Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1085 1090 1095
Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125
Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1250 1255 1260

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1415 1420 1425
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585
<210> 1440
<211> 1588

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 1440

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Ser Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415
Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735
Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800
Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
835 840 845
Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
900 905 910
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1280 1285 1290
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala
1325 1330 1335
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410
Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530
Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1441
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 1441
Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Arg Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala

195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys

210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser

225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro

565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln

610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800
Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815
Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845
Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880
Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940
Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945 950 955 960
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
980 985 990
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1250 1255 1260

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln

1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1442

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1442

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Tyr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300
Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620
Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685
Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700
Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750
Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1085 1090 1095
Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125
Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1130 1135 1140
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1145 1150 1155
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245
Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305
Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala
1325 1330 1335
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1415 1420 1425
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1443

<211> 1588

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1443

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

```
Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
        115                 120                 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
    130                 135                 140
Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145                 150                 155                 160
Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
                245                 250                 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
```

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415
Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
450 455 460
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480
Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525
Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540
Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670
Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700
Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735
Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800
Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
835 840 845
Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110
Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1130 1135 1140
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1145 1150 1155
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170
Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1280 1285 1290
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln

1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1444

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1444

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala

195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys

210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser

225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430
Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480
Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495
Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525
Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540
Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560
Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605
Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620
Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1250 1255 1260

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln

1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1445

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1445

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Arg Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys

210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300
Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365
Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430
Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495
Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525
Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560
Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620
Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685
Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700
Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
965 970 975
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
995 1000 1005
Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1010 1015 1020
Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His
1025 1030 1035
Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1055 1060 1065
Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1070 1075 1080
Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1085 1090 1095
Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
1115 1120 1125
Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
1130 1135 1140
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
1145 1150 1155
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185
Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1250 1255 1260

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1446

<211> 1588

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 1446

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

```
            20                  25                  30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
        35                  40                  45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
    50                  55                  60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65                  70                  75                  80
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
        115                 120                 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
    130                 135                 140
Leu Cys Thr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145                 150                 155                 160
Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
                245                 250                 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270
```

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415
Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
450 455 460
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480
Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro

565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln

610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800
Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
835 840 845
Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
900 905 910
Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
915 920 925
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945 950 955 960
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
965 970 975
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
980 985 990
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1280 1285 1290
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala
1325 1330 1335
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410
Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530
Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575
Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1447
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 1447

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45
Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Thr Phe Phe Arg Met Pro Arg Ser Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala

195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys

210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser

225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

305 310 315 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

325 330 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro

565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln

610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly

645 650 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val

1160 1165 1170

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn

1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu

1190 1195 1200

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys

1205 1210 1215

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys

1220 1225 1230

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile

1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr

1250 1255 1260

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe

1265 1270 1275

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val

1280 1285 1290

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp

1310 1315 1320

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala

1325 1330 1335

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu

1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys

1370 1375 1380

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln

1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1448

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1448

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140
Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300
Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365
Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

370 375 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

385 390 395 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys

405 410 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro

435 440 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu

450 455 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile

465 470 475 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp

485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln

515 520 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys

530 535 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr

545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro

565 570 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn

580 585 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln

610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685
Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700
Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750
Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815
Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
835 840 845
Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser

900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

```
Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125
Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140
Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val Ile Thr  Leu Lys Ser
    1145                1150                1155
Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

Arg Glu  Ile Asn Asn Tyr His  His Ala His Asp Ala  Tyr Leu Asn
    1175                1180                1185
Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200
Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215
Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245
Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260
Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275
Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305
Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320
Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335
Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
```

1340 1345 1350

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1385 1390 1395
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
1415 1420 1425
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560
Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 1449

<211> 1588

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1449

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Ser Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
180 185 190
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
195 200 205
Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
210 215 220
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225 230 235 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
245 250 255
Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
260 265 270
Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
275 280 285
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
290 295 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305 310 315 320
Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
325 330 335
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
340 345 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
355 360 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
370 375 380
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385 390 395 400
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
405 410 415
Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys

420 425 430

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
435 440 445
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
450 455 460
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465 470 475 480
Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
485 490 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
500 505 510
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
515 520 525
Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
530 535 540
Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545 550 555 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
565 570 575
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
580 585 590
Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
595 600 605
Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
610 615 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625 630 635 640
Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
645 650 655
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
660 665 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
675 680 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
690 695 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
725 730 735

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
740 745 750

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
755 760 765

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
770 775 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785 790 795 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
805 810 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
820 825 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
835 840 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
850 855 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865 870 875 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
885 890 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
900 905 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser

915 920 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu

930 935 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu

945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg

965 970 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

980 985 990

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys

995 1000 1005

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp

1010 1015 1020

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His

1025 1030 1035

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys

1040 1045 1050

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val

1055 1060 1065

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1070 1075 1080

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu

1085 1090 1095

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly

1100 1105 1110

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His

1115 1120 1125

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

1130 1135 1140

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1145 1150 1155

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
1160 1165 1170
Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
1175 1180 1185

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
1190 1195 1200
Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
1205 1210 1215
Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
1220 1225 1230
Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
1235 1240 1245

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
1250 1255 1260
Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
1265 1270 1275
Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
1280 1285 1290
Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1295 1300 1305

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
1310 1315 1320
Trp Asp Pro Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala
1325 1330 1335
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
1340 1345 1350
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
1355 1360 1365

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1370 1375 1380
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1385 1390 1395

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala

1400 1405 1410

Ser Ala Arg Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu

1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu

1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val

1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln

1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln

1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu

1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1580 1585

<210> 1450

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<220><221> MISC_FEATURE

<222> (1)..(21)

<223> This sequence may encompass 1, 3, or 7 "Gly Gly Ser"
repeating units

<400> 1450

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1 5 10 15

Gly Ser Gly Gly Ser
20

<210> 1451

<211> 8961

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 1451

atatgccaag tacgcccccct attgacgtca atgacggtaa atggcccgcc tggcattatg 60

cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg 120

ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact 180

cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa 240

atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta 300

ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct 360

agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat gaaacggaca 420

gccgacggaa gcgagttcga gtcaccaaag aagaagcgga aagtctcctc agagactggg 480

cctgtcgccg tcgatccaac cctgcgccgc cggattgaac ctcacgagtt tgaagtgttc 540

tttgaccccc gggagctgag aaaggagaca tgcctgctgt acgagatcaa ctggggaggc 600

aggcactcca tctggaggca cacctctcag aacacaaata agcacgtgga ggtgaacttc 660

atcgagaagt ttaccacaga gcggtacttc tgccccaata ccagatgtag catcacatgg 720

tttctgagct ggtccccttg cggagagtgt agcagggcca tcaccgagtt cctgtccaga 780

tatccacacg tgacactgtt tatctacatc gccaggctgt atcaccacgc agacccaagg 840

aataggcagg gcctgcgcga tctgatcagc tccggcgtga ccatccagat catgacagag 900

caggagtccg gctactgctg gcggaacttc gtgaattatt ctcctagcaa cgaggcccac 960
tggcctaggt acccacacct gtgggtgcgc ctgtacgtgc tggagctgta ttgcatcatc 1020
ctgggcctgc ccccttgtct gaatatcctg cggagaaagc agccccagct gaccttcttt 1080
acaatcgccc tgcagtcttg tcactatcag aggctgccac cccacatcct gtgggccaca 1140
ggcctgaagt ctggaggatc tagcggagga tcctctggca gcgagacacc aggaacaagc 1200
gagtcagcaa caccagagag cagtggcggc agcagcggcg gcagcgacaa gaagtacagc 1260
atcggcctgg ccatcggcac caactctgtg ggctgggccg tgatcaccga cgagtacaag 1320

gtgcccagca agaaattcaa ggtgctgggc aacaccgacc ggcacagcat caagaagaac 1380
ctgatcggag ccctgctgtt cgacagcggc gaaacagccg aggccacccg gctgaagaga 1440
accgccagaa gaagatacac cagacggaag aaccggatct gctatctgca agagatcttc 1500
agcaacgaga tggccaaggt ggacgacagc ttcttccaca gactggaaga gtccttcctg 1560
gtggaagagg ataagaagca cgagcggcac cccatcttcg gcaacatcgt ggacgaggtg 1620
gcctaccacg agaagtaccc caccatctac cacctgagaa agaaactggt ggacagcacc 1680
gacaaggccg acctgcggct gatctatctg gccctggccc acatgatcaa gttccggggc 1740

cacttcctga tcgagggcga cctgaacccc gacaacagcg acgtggacaa gctgttcatc 1800
cagctggtgc agacctacaa ccagctgttc gaggaaaacc ccatcaacgc cagcggcgtg 1860
gacgccaagg ccatcctgtc tgccagactg agcaagagca gacggctgga aaatctgatc 1920
gcccagctgc ccggcgagaa gaagaatggc ctgttcggaa acctgattgc cctgagcctg 1980
ggcctgaccc ccaacttcaa gagcaacttc gacctggccg aggatgccaa actgcagctg 2040
agcaaggaca cctacgacga cgacctggac aacctgctgg cccagatcgg cgaccagtac 2100
gccgacctgt ttctggccgc caagaacctg tccgacgcca tcctgctgag cgacatcctg 2160

agagtgaaca ccgagatcac caaggccccc ctgagcgcct ctatgatcaa gagatacgac 2220
gagcaccacc aggacctgac cctgctgaaa gctctcgtgc ggcagcagct gcctgagaag 2280
tacaaagaga ttttcttcga ccagagcaag aacggctacg ccggctacat tgacggcgga 2340
gccagccagg aagagttcta caagttcatc aagcccatcc tggaaaagat ggacggcacc 2400
gaggaactgc tcgtgaagct gaacagagag gacctgctgc ggaagcagcg gaccttcgac 2460
aacggcagca tcccccacca gatccacctg ggagagctgc acgccattct gcggcggcag 2520
gaagattttt acccattcct gaaggacaac cgggaaaaga tcgagaagat cctgaccttc 2580

cgcatcccct actacgtggg ccctctggcc aggggaaaca gcagattcgc ctggatgacc 2640

agaaagagcg aggaaaccat caccccctgg aacttcgagg aagtggtgga caagggcgct 2700

tccgcccaga gcttcatcga gcggatgacc aacttcgata agaacctgcc caacgagaag 2760

gtgctgccca agcacagcct gctgtacgag tacttcaccg tgtataacga gctgaccaaa 2820

gtgaaatacg tgaccgaggg aatgagaaag cccgccttcc tgagcggcga gcagaaaaag 2880

gccatcgtgg acctgctgtt caagaccaac cggaaagtga ccgtgaagca gctgaaagag 2940

gactacttca agaaaatcga gtgcttcgac tccgtggaaa tctccggcgt ggaagatcgg 3000

ttcaacgcct ccctgggcac ataccacgat ctgctgaaaa ttatcaagga caaggacttc 3060

ctggacaatg aggaaaacga ggacattctg gaagatatcg tgctgaccct gacactgttt 3120

gaggacagag agatgatcga ggaacggctg aaaacctatg cccacctgtt cgacgacaaa 3180

gtgatgaagc agctgaagcg gcggagatac accggctggg gcaggctgag ccggaagctg 3240

atcaacggca tccgggacaa gcagtccggc aagacaatcc tggatttcct gaagtccgac 3300

ggcttcgcca acagaaactt catgcagctg atccacgacg acagcctgac ctttaaagag 3360

gacatccaga aagcccaggt gtccggccag ggcgatagcc tgcacgagca cattgccaat 3420

ctggccggca gccccgccat taagaagggc atcctgcaga cagtgaaggt ggtggacgag 3480

ctcgtgaaag tgatgggccg gcacaagccc gagaacatcg tgatcgaaat ggccagagag 3540

aaccagacca cccagaaggg acagaagaac agccgcgaga gaatgaagcg gatcgaagag 3600

ggcatcaaag agctgggcag ccagatcctg aaagaacacc ccgtggaaaa cacccagctg 3660

cagaacgaga agctgtacct gtactacctg cagaatgggc gggatatgta cgtggaccag 3720

gaactggaca tcaaccggct gtccgactac gatgtggacc atatcgtgcc tcagagcttt 3780

ctgaaggacg actccatcga caacaaggtg ctgaccagaa gcgacaagaa ccggggcaag 3840

agcgacaacg tgccctccga agaggtcgtg aagaagatga agaactactg gcggcagctg 3900

ctgaacgcca agctgattac ccagagaaag ttcgacaatc tgaccaaggc cgagaggagc 3960

ggcctgagcg aactggataa ggccggcttc atcaagagac agctggtgga aacccggcag 4020

atcacaaagc acgtggcaca gatcctggac tcccggatga acactaagta cgacgagaat 4080

gacaagctga tccgggaagt gaaagtgatc accctgaagt ccaagctggt gtccgatttc 4140

cggaaggatt tccagtttta caaagtgcgc gagatcaaca actaccacca cgcccacgac 4200

gcctacctga acgccgtcgt gggaaccgcc ctgatcaaaa agtaccctaa gctggaaagc 4260

gagttcgtgt acggcgacta caaggtgtac gacgtgcgga agatgatcgc caagagcgag 4320

caggaaatcg gcaaggctac cgccaagtac ttcttctaca gcaacatcat gaacttttc 4380

aagaccgaga ttaccctggc caacggcgag atccggaagc ggcctctgat cgagacaaac 4440

ggcgaaaccg gggagatcgt gtgggataag ggccgggatt ttgccaccgt gcggaaagtg 4500

ctgagcatgc cccaagtgaa tatcgtgaaa aagaccgagg tgcagacagg cggcttcagc 4560

aaagagtcta tcctgcccaa gaggaacagc gataagctga tcgccagaaa gaaggactgg 4620

gaccctaaga agtacggcgg cttcgacagc cccaccgtgg cctattctgt gctggtggtg 4680

gccaaagtgg aaaagggcaa gtccaagaaa ctgaagagtg tgaaagagct gctggggatc 4740

accatcatgg aaagaagcag cttcgagaag aatcccatcg actttctgga agccaagggc 4800

tacaaagaag tgaaaaagga cctgatcatc aagctgccta agtactccct gttcgagctg 4860

gaaaacggcc ggaagagaat gctggcctct gccggcgaac tgcagaaggg aaacgaactg 4920

gccctgccct ccaaatatgt gaacttcctg tacctggcca gccactatga gaagctgaag 4980

ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa gcactacctg 5040

gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc cgacgctaat 5100

ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag agagcaggcc 5160

gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc cttcaagtac 5220

tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct ggacgccacc 5280

ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc tcagctggga 5340

ggtgacagcg gcgggagcgg cgggagcggg gggagcacta atctgagcga catcattgag 5400

aaggagactg ggaaacagct ggtcattcag gagtccatcc tgatgctgcc tgaggaggtg 5460

gaggaagtga tcggcaacaa gccagagtct gacatcctgg tgcacaccgc ctacgacgag 5520

tccacagatg agaatgtgat gctgctgacc tctgacgccc ccgagtataa gccttgggcc 5580

ctggtcatcc aggattctaa cggcgagaat aagatcaaga tgctgagcgg aggatccgga 5640

ggatctggag gcagcaccaa cctgtctgac atcatcgaga aggagacagg caagcagctg 5700

gtcatccagg agagcatcct gatgctgccc gaagaagtcg aagaagtgat cggaaacaag 5760

cctgagagcg atatcctggt ccataccgcc tacgacgaga gtaccgacga aaatgtgatg 5820

ctgctgacat ccgacgcccc agagtataag ccctgggctc tggtcatcca ggattccaac 5880

ggagagaaca aaatcaaaat gctgtctggc ggctcaaaaa gaaccgccga cggcagcgaa 5940

ttcgagccca agaagaagag gaaagtctaa ccggtcatca tcaccatcac cattgagttt 6000

aaacccgctg atcagcctcg actgtgcctt ctagttgcca gccatctgtt gtttgccccc 6060

cccccgtgcc ttccttgacc ctggaaggtg ccactcccac tgtcctttcc taataaaatg 6120

aggaaattgc atcgcattgt ctgagtaggt gtcattctat tctgggggggt ggggtggggc 6180

aggacagcaa gggggaggat tgggaagaca atagcaggca tgctggggat gcggtgggct 6240

ctatggcttc tgaggcggaa agaaccagct ggggctcgat accgtcgacc tctagctaga 6300

gcttggcgta atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc 6360

cacacaacat acgagccgga agcataaagt gtaaagccta gggtgcctaa tgagtgagct 6420
aactcacatt aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc 6480
agctgcatta atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt 6540
ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag 6600
ctcactcaaa ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca 6660
tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt 6720
tccataggct ccgcccccct gacgagcatc acaaaaatcg acgctcaagt cagaggtggc 6780

gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc ctcgtgcgct 6840
ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg 6900
tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca 6960
agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta tccggtaact 7020
atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta 7080
acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tggtggccta 7140
actacggcta cactagaaga acagtatttg gtatctgcgc tctgctgaag ccagttacct 7200

tcggaaaaag agttggtagc tcttgatccg gcaaacaaac caccgctggt agcggtggtt 7260
tttttgtttg caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga 7320
tcttttctac ggggtctgac actcagtgga acgaaaactc acgttaaggg attttggtca 7380
tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga agttttaaat 7440
caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg 7500
cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc cccgtcgtgt 7560
agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 7620

acccacgctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 7680
gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 7740
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca 7800
tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa 7860
ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctccttc ggtcctccga 7920
tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca gcactgcata 7980
attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 8040

agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 8100

ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa cgttcttcgg 8160

ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 8220

cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 8280

gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 8340

tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 8400

tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 8460

tgccacctga cgtcgacgga tcgggagatc gatctcccga tcccctaggg tcgactctca 8520

gtacaatctg ctctgatgcc gcatagttaa gccagtatct gctccctgct tgtgtgttgg 8580

aggtcgctga gtagtgcgcg agcaaaattt aagctacaac aaggcaaggc ttgaccgaca 8640

attgcatgaa gaatctgctt agggttaggc gttttgcgct gcttcgcgat gtacgggcca 8700

gatatacgcg ttgacattga ttattgacta gttattaata gtaatcaatt acggggtcat 8760

tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat ggcccgcctg 8820

gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa 8880

cgccaatagg gactttccat tgacgtcaat gggtggagta tttacggtaa actgcccact 8940

tggcagtaca tcaagtgtat c 8961

<210> 1452

<211> 100

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 1452

uucguaucug uaaaaccaag guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 1453

<211> 100

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 1453

uuaccuguac cauaaccagg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 1454

<211> 100

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 1454

cucuuaccug uaccauaacc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 1455

<211> 100

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 1455

ugcaccucug gggaggaccu guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 1456

<211> 100

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 1456

ccuaccuguc accaggacca guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 1457

<211> 100

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 1457

caccuaccua agaaccaucc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 1458

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1458

acucacccag cauccccagc 20

<210> 1459

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1459

uggaaaucug ggggucagaa 20

<210> 1460

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1460

guuacccacc uaagcagguc 20

<210> 1461

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1461

cugccagcgg cugaacugug 20

<210> 1462

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1462

ccucccacug cuuggagcuc 20

<210> 1463

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1463

ucugccagcg gcugaacugu 20

<210> 1464

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1464

gucugccagc ggcugaacug 20

<210> 1465

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1465

gaagugccag ggccagcugg 20

<210> 1466

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1466

ccaugugcca uccguccuug 20

<210> 1467

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1467

uuugcagcca gagcuggggc 20

<210> 1468

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1468

gguucugcaa ugagacacuc 20

<210> 1469

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1469

gagcuaggag aggagagagc 20

<210> 1470

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1470

ucacuuaccu gagcaaagga 20

<210> 1471

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1471

cucacuuacc ugagcaaagg 20

<210> 1472

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1472

cugcagcugg uggcacaguc 20

<210> 1473

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1473

cggccagcac ugugccggcg 20

<210> 1474

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1474

aucuuccauu ggauuggcaa 20

<210> 1475

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1475

gaucuuccau uggauuggca 20

<210> 1476

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1476

aaccugagag gggaagcaau 20

<210> 1477

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1477

aaaccugaga ggggaagcaa 20

<210> 1478

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1478

gugcagagcc gucugguggg 20

<210> 1479

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1479

cucccaccgc agcgagcucc 20

<210> 1480

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1480

cucccaaagu ucguggcacu 20

<210> 1481

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1481

aggugcagag ccgucuggug 20

<210> 1482

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1482

ggugcagagc cgucuggugg 20

<210> 1483

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1483

ucucccaaag uucguggcac 20

<210> 1484

<211> 20

<212> RNA

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1484

ugucccagug ccacgaacuu 20

<210> 1485

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1485

aaggugcaga gccgucuggu 20

<210> 1486

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1486

uccuaucgag ugcuggacgc 20

<210> 1487

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1487

caaggugcag agccgucugg 20

<210> 1488

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1488

ggugcgccag ggggcucagu 20

<210> 1489

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1489

gggcugccca cugagccccc 20

<210> 1490

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1490

aggugcgcca gggggcucag 20

<210> 1491

<211> 20

<212> RNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1491

ggccaggauc caaaccccgc 20

<210> 1492

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1492

cgccagugga uuggcccaac 20

<210> 1493

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1493

aagaagcagc gccaguggau 20

<210> 1494

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1494

aaagacacug ggcagauggu 20

<210> 1495

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1495

gcuuaccugg auaagcugac 20

<210> 1496

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1496

uuccagagcu ggggaaagaa 20

<210> 1497

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> CDS

<222> (2)..(19)

<400> 1497

g gag gct atc cag cgt gag t 20
Glu Ala Ile Gln Arg Glu

1 5

<210> 1498

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 1498

Glu Ala Ile Gln Arg Glu
1 5

<210> 1499

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1499

cttggtttta cagatacgaa 20

<210> 1500

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1500

ccacacccaa aaggccacac 20

<210> 1501

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 1501

Leu Gly Phe Thr Asp Thr Asn
1 5

<210> 1502

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> CDS

<222> (1)..(21)

<400> 1502

cca gac agg ccc tgg aac ccc 21

Pro Asp Arg Pro Trp Asn Pro

1 5

<210> 1503

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 1503

Pro Asp Arg Pro Trp Asn Pro

1 5

<210> 1504

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1504

caaccggcag cgagggaacg 20

<210> 1505

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1505

gcagcaccca gggcaggacc 20

<210> 1506

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1506

cagcacccag ggcaggacca 20

<210> 1507

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1507

aggcccaggc ggcagggccc 20

<210> 1508

<211> 20

<212> DNA

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1508

cagcaaccag acggacaagc 20

<210> 1509

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 1509

cacccagaaa acaacacagc 20

<210> 1510

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1510

gagccagcca cagggccccc 20

<210> 1511

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1511

acaggcccag gacacagagc 20

<210> 1512

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 1512

gggaaacaag agaccagagc 20